PHENOTYPIC RESPONSE TO SELECTION FOR PROTEIN IN MAIZE KERNELS

BY

NICHOLAS HELLER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Stephen Moose, Chair
Professor Frederic Kolb
Professor Lisa Ainsworth
Assistant Professor Alex Lipka

ABSTRACT

Long-term, divergent selection for protein concentration in maize grain has been conducted at the University of Illinois since 1896. Grain protein concentration in maize and other cereal crops is strongly associated with productivity, nutritional quality, and processing characteristics. Although grain protein concentration responds to selection in maize, the genetic basis for that response is largely unknown. The proteins that responded most to selection were the alpha-zeins, a class of the abundant seed storage proteins that are lacking in several essential amino acids. Thus, efforts to increase the nutritional quality of maize grain have focused on decreasing the zein proteins. The genomic organization of zein genes, however, limits genetic analysis and the effectiveness of selection for variation in zein accumulation. Developing a method to directly monitor changes in alpha zein accumulation while concurrently maintaining the accumulation of other seed proteins could enable more effective selection for improvements in yield, nutrition, and sustainability.

The Illinois Long-Term Selection Experiment (ILTSE) was remarkably successful in changing the protein concentration in the grain. The Illinois High Protein (IHP) population is the only original population still undergoing selection. IHP has now undergone 116 cycles of selection, and represents the biological extreme for protein concentration in maize kernels. Although IHP does not appear to be responding to selection, the variance for grain nitrogen concentration has not decreased. Response to selection has also ceased in Illinois Reverse High Protein (IRHP), but the phenotypic variance in this population was low. Two new IRHP populations were started from IHP cycle 100, and despite no further response to forward selection in IHP, response to reverse selection is documented here. IRHP2 has shown a more dramatic response to selection than has IRHP3, but a genotyping analysis shows patterns that might suggest an unintentional cross with the original IRHP. IHP continues to be grown to serve as a control for two new reverse selection populations, the first replicated aspect of this study since its inception.

The biological extreme for low protein was achieved by the Illinois Low Protein (ILP) population 50 years ago, and this feat has since been replicated by the Illinois Reverse High

Protein (IRHP) population despite initial selection for 47 years for higher protein concentration. Selection has ceased in both of these populations. The Illinois Reverse Low Protein (IRLP) population continues to slowly respond to selection at a rate similar to what IHP did. In addition to tracking the phenotypic response to selection, recent advances in technology allowed for genotyping a representative subset of the ILTSE populations genome-wide. Analyses revealed relatedness and genetic structuring within the protein populations which mirrored the protein concentration phenotype.

A transgenic reporter line was utilized to visualize the accumulation of a specific alpha-zein. The *red fluorescent protein* (*RFP*) gene was attached to the 3' end of the *Floury2* (*FL2*) gene with its native promoter, allowing a quick, nondestructive assessment of the accumulation of that alpha-zein in a maize kernel. This FL2-RFP reporter line has been used to document variation due to genetic, epigenetic, and environmental sources. A first experiment with a set of hybrids grown in paired high and low nitrogen environments showed that FL2-RFP responded to available nitrogen supply in a similar manner to protein concentration. The relationships between FL2-RFP accumulation and six other important agronomic traits was investigated in a population of hybrids that varied greatly for alpha-zein accumulation. The FL2-RFP phenotype showed the most variation among the measured traits, but surprisingly was not strongly correlated with either grain protein concentration or other agronomic traits. These results suggest that selection for changes in FL2-RFP accumulation can be achieved without affecting other aspects of hybrid performance.

Finally, observations from prior studies with protein selection in the Illinois Long-Term Selection Experiment and other work raise the possibility that epigenetic factors may contribute to phenotypic variation in alpha-zein accumulation. This hypothesis was directly addressed through the creation and analysis of a population where the FL2-RFP reporter was introgressed into the B73 inbred line and crossed to a B73 near-isogenic line harboring the *mediator of paramutation* (*mop1*) mutation. The *mop1* gene encodes the RNA-dependent RNA polymerase required for RNA-dependent DNA methylation, and loss of *mop1* was predicted to alter known epigenetic marks. This population of "epigenetic near-isogenic lines" (epiNILs) produced a high frequency of phenotypic variants with reduced fitness that were not observed in the control population, including significant variation in FL2-RFP intensity. Selection was largely effective in stabilizing these phenotypic variants. High-density genotyping of 90 epi-NILs revealed all

were indeed like B73 except for two genomic regions around *mop1*and *FL2-RFP*. Hybrids were created with these epiNILs to establish a heritability estimate for the FL2-RFP trait in this population. A strong narrow-sense heritability estimate of 0.84 was obtained for the FL2-RFP phenotype. Largely, the FL2-RFP phenotype explained most of the variation present when all three red, green, and blue channels were averaged, but this composite phenotype was not highly correlated with grain protein concentration; instead, a contrast between the green and red channels correlated more strongly with the total protein concentration. The FL2-RFP reporter could be used in breeding programs to increase the number of individuals which are monitored for alpha-zein accumulation. The insights into the makeup of the FL2-RFP phenotype allow for selection indices to be built and utilized differently based on the goals of that breeding program.

ACKNOWLEDGEMENTS

I must first and gratefully thank my family. From the beginning, my parents have always shown me love, encouragement, and willingness to supply anything needed to feed my intellectual curiosity. I thank my dad for his guidance and wisdom – seeing my love for numbers and plants, he encouraged me to follow my heart into a course of study that wasn't law. My grandfather, Harlan Heller, was also particularly instrumental in developing me into the man I am today, and also himself a lawyer, taught me a lot of about gardening and supported my passion for plants. In a way, I have been breeding plants my whole life. My parents let me have part of the garden when I was five, and let me continue to tear up larger and larger sections of the yard to grow all kinds of plants from around the world (tobacco, peanuts, banana trees, pineapple, avocado, cotton, etc.), try my hand at breeding (like cabbage plants that could survive the winter, and breeding the gray stripe out of the Mammoth Gray Stripe sunflower), and planting probably all but five of the one-hundred trees now in their yard. My brother not-so-affectionately referred to this as my tree farm, as he hated mowing around all the small twigs that I was nurturing. While at Illinois Wesleyan, I brought home several five-gallon buckets of acorns, each from a different favorite tree on the quad and started my own nursery, much to my brother's chagrin. I'm sure they were rolling their eyes, but my parents facilitated my every whim.

Also while at Illinois Wesleyan I was further inspired by the wonders of the biological world. From wading in creeks to seine for aquatic life to studying forest health in Illinois and in Costa Rica and even spotting pit vipers, many ecology and ornithology trips with Dr. Harper stand at the front of my memories. It was genetics, however, that I knew I wanted to pursue further. Dr. Dey first started me on this exciting journey – after taking Plant and Fungal Diversity and then Microbiology with him, I also became a TA for him in both of those courses. Preparing all of the specimens, hearing Dr. Dey's stories for a second time, and just watching his enthusiasm for the subject really was a turning point for me. I saw that it was okay to be excited about fungi or fun corn mutants or teaching – interacting with students, helping my peers with a commonly-perceived 'difficult course', and seeing Dr. Dey's passion for helping his students learn developed my love for teaching. Although very ill near the end of his life, Dr. Dey was

there every day sharing his expertise. Finally, I need to acknowledge my advisor, Dr. Bollivar, who helped me through losing a mentor and scientific role model. Dr. Bollivar also helped me prepare for graduate study in a department that has relatively few (student) plant enthusiasts. His genetics course and recommendation to spend a summer with Dr. Johal at Purdue ultimately decided my future path.

I need to acknowledge Dr. Johal for his guidance during that summer. He has an incredible collection of natural variation in maize and this cemented to me even further that this is what I wanted to do. He was always demonstrating best scientific practices, yet emphasized the importance of including an artistic touch, and I have tried to model this into my approach to breeding. He also reinforced and celebrated my habits of careful observation. Finally, I must thank Dr. Johal for his selflessness – he not only recommended me to Dr. Moose, but wrote a letter for me to join someone else's lab at a rival school.

Dr. Moose and I shared a common bond immediately – beyond our excitement for maize genetics, we had both played golf at a small, academically-rigorous university. He offered to not only be my graduate advisor, but also my career coach, and I could not be more thankful for his support from the very beginning. We have talked nitrogen and corn on the golf course, and we have talked golf in the corn field. There isn't a field on the South Farms that I haven't found a golf ball in. His ability to make long hours more enjoyable, whether it was a conversation about the latest in science or the latest in sports late at night while still in the lab, this helped to reset the clock and is definitely something I will take with me. As a mentor, he was caring and encouraging even when I was at my worst. Dr. Moose is a big part of my professional achievements and I cannot thank him enough for his support, but also for letting me fail and allowing me to learn from these failures to be an even better scientist because of it.

I am very thankful to the Department of Crop Sciences for putting together a great group of people to train young scientists. I want to acknowledge: Dr. Bullock and Dr. Anderson (in CITL) for their guidance in teaching and developing my teaching skillset even further; Dr. Kolb for his many plant breeding insights and discussions of not only weather, side projects, and fishing but also discussion about my teaching philosophy and careers in academia; Dr. Bollero, Dr. Butts, and Dr. Lipka for their thoughts in teaching and help with statistics when I got stuck; Dr. Ainsworth for her scientific guidance and reassurance in my first semester and beyond and fun discussions of running; and Dr. Studer for an awesome class and help with anything I've

asked for. Further, the supporting cast has always gone above and beyond with any request I've ever had – Megan Ward, Dianne Carson, Patty Sarver, Tracey Malkovich, Amber Adams Ruedi, Greg Parks, and Chinedu Onyejekwe.

I gratefully acknowledge the Illinois Corn Marketing Board Fellowship for funding the entirety of my graduate education, and Dr. Moose for providing materials and funding needed to complete lab and field work for this project. I would like to thank all members, former and present, of the Moose Lab for their camaraderie and the creation and continuation of the germplasm that I have had the privilege to work with. Special thanks to all of those who have helped me with phenotyping the Floury2-RFP kernels: Margaret Byrne, Parrish Smith, Michelle Wagner, and Sam Shaffer. That is a lot of dusty hours in a dark dungeon, and I am forever thankful for your help. Also special thanks again to Sam Shaffer for helping with field phenotyping and harvest, and also to my mother, who upon hearing of my long hours volunteered herself and spent several days harvesting and wrapping RFP ears in pollinating bags with me in the field. She also spent several days analyzing sentence structure and catching silly typos in this finished work.

I would like to acknowledge all of those people who let me destroy part of their yard to put in a small isolation plot, monitoring those plots, and sending me pictures and updates for two summers: Dr. Studer, Dr. Taylor Lovell, John and Betty Telford, Kent and Lori Heller, Marc Mills, Allison Krill, Harlan Heller, and Paul Daily.

I would also like to thank the people in the surrounding community for welcoming this crazy young man into their lives. Thank you to my neighbors who have tolerated my non-conventional yard (and chickens) and asked me about the many plants I have grown in my yard. Sarah Becker has been especially encouraging of my ideas and has watched my home and pets while I was at Maize Genetics Conferences or spending two weeks at a molecular biology workshop. And a huge thank you to the Crazy Runners and Second Wind Running Club. Running is a great stress reliever for me, and having other people to run with regardless of the time of day or distance to be covered definitely helped with my sanity. And thank you for listening to me droning on about my corn adventures and running by my field five times per run to keep the birds from eating all the freshly-germinating seedlings.

Finally, I would like to acknowledge Michele, whom I met while running through a corn field out by Sidney and have since started a family with. She has supported me in every way

possible and she picked up my slack at home when I have spent long hours in the field or at the lab. She has brought dinner to the field so that I could keep setting up the hundreds of pollinations for the next day, spent 30-hour weekends with me shelling corn, and put up with boxes of corn everywhere in our tiny house while I sorted the kernels before I could take pictures of them. I could not have accomplished this without you and look forward to the many adventures we have ahead of us.

This thesis is dedicated to my son, Harlan Heller. I hope to teach you to appreciate the wonders of the natural world and to notice all its intricate, beautiful details. You have brought your mother and I so much joy, and you have inspired and motivated me immensely.

TABLE OF CONTENTS

CHAPTER 1

RESPONSE TO SELECTION IN THE ILLINOIS LONG-TERM SELECTION EXPERIMENT

ABSTRACT

Recurrent, divergent selection in the Illinois Long Term Selection Experiment (ILTSE) has produced the known phenotypic extremes for grain protein concentration, known as the Illinois High Protein (IHP) and Illinois Low Protein (ILP) populations. The ILTSE was initiated in 1896 by Cyril Hopkins to determine if kernel protein and oil concentrations could be altered by selective breeding (Hopkins, 1899). The answer to this became apparent quickly and the experiment continued, but now with a new goal in mind – determine the limits of artificial selection on these aforementioned traits (Dudley and Lambert, 2004). Many limits have been established, and understanding the genetic basis for the response to selection has become the next goal. As of 2017, 116 cycles of recurrent selection has been done for IHP, the only active original population. Many reverse selection populations have been initiated to assess the remaining variability including two new populations started from IHP cycle 100. Because the original population of Burr's White cannot be obtained, disentangling the impacts of drift and selection on genetic variation in the original populations would be difficult; thus, the focus is on these reverse selection populations. Individuals from all active populations were genotyped genome-wide to allow investigation of genetic responses to long-term selection relative to each other population. Analyses revealed relatedness and genetic structuring within the protein populations which mirrored the protein concentration phenotype. Finally, one additional new population has been started to select on higher protein content in the kernels per plant.

INTRODUCTION

The Illinois Long-Term Selection Experiment was started with a practical goal in mind when Dean Davenport suggested to a young chemist Cyril Hopkins in 1896 that he should test if the chemical composition in maize (*Zea mays*) kernels could be changed. Increasing protein and oil was stated to be desired for food and feed purposes; low protein and oil would be high in carbohydrates and of potential interest to millers. Only three years into the experiment, Hopkins

(1899) reported that protein and oil concentrations could be altered by selective breeding and were already among the extremes in known varieties of the time. Further, he systemically confirmed his ability to classify high versus low protein and high versus low oil kernels by visually inspecting characteristics of the kernel – protein was largely discriminated by the starchy section of the endosperm being enlarged or not, and oil was largely discriminated by the size of the germ. Although current methods allow researchers to phenotype the kernels that will be selected and planted for the next generation, methods of the time required destructive sampling which Hopkins (1899) noted as a distinct disadvantage compared to selection experiments in sugar beets, for example. The other advantage in sugar beets was not only could a researcher phenotype a plant and still use that plant in sexual reproduction, one could phenotype the population before flowering and mate only the best individuals based on their superior phenotype. In corn, it was the product of the reproduction process that was then measured. The breeding and selection method used was modeled after animal experiments of the time, just as humans had been doing for centuries, selecting for desired qualities (without knowing the inheritance of those particular traits). This was all a year before botanists DeVries, Correns, and von Tschermak rediscovered Gregor Mendel's work on the laws of inheritance.

Introducing genetics into this experiment has enabled many questions and offers a unique system to utilize current technologies available (Moose et al., 2004). Genetics, combined with very careful documentation and observation, led Edward Murray East to notice that all ears from the Illinois High Protein (IHP) population traced back to one original ear. In contrast to characteristics like pea color, shape, and position of the flower, maize kernel protein and oil concentration were quantitative traits. Also dissimilar from Mendel's work, each population (IHP, ILP, IHO, ILO) was planted in isolation and allowed to open-pollinate – not knowing the parents in a cross disallowed the genetic ratios to be determined. East started doing controlled crosses and inbreeding at Illinois, but it was after doing controlled crosses at Yale and Harvard that he tried to offer interpretations of continuous traits in corn based on Medelian principles (1908, 1910). He proposed that the same phenotype could be achieved via different chemistries (as in seed color) and that even within one common pathway, many factors could be segregating independently making it difficult to track enough plants to make sense of the inheritance. The controlled crosses that were started in Illinois became the source of two inbred lines used in the first commercial double-cross maize hybrid grown in 1918 (Jones, 1922).

There were many avenues that this experiment, referred to here as the Illinois Long-Term Selection Experiment (ILTSE), underwent genetic diversity reductions. Certainly undocumented, inadvertent self-pollinations occurred within each population as it was planted with (half, or potentially full) siblings in each row and allowed to open pollinate, increasing the expected frequency of homozygous loci. Further, by cycle 17, each of the four original populations, Illinois High Protein, Illinois Low Protein (ILP), Illinois High Oil (IHO), and Illinois Low Oil (ILO), could be traced back to their own founding ear (Winter, 1929). From these four original Burr's White mother plants there were undoubtedly many pollen-donors, but this still represents a strong bottleneck and founder effect in these populations. Despite these apparent reductions in genetic diversity, the results from this experiment demonstrate a profound change in phenotype over time. The other strong force that is in play in this population is genetic drift. Each population was selected in the indicated direction for the indicated trait as its name suggests, but each population was also in reproductive isolation. Through continuous, divergent selection, the phenotypes of each of these populations changed quite quickly. By the 7^{th} and 8^{th} year of the experiment, there was only one overlapping data point each year. The last time there was ever *any* overlap between IHP and ILP in percent protein was cycle 26 (and between cycles 9 and 26, the phenotypes for 9 of those years had no overlap) (Figure 1.1, 1.2). The experiment continued, but now with a new goal in mind – determine the limits of artificial selection on these aforementioned traits (Dudley and Lambert, 2004).

The traits of interest in the ILTSE, although both quantitative traits, are different in many ways. East (1899) noted that the oil trait was much more straightforward than was protein concentration. The oil in a corn kernel is almost entirely contained in the germ portion of the seed, thus simply by judging its relative size the oil concentration could be estimated with surprising accuracy. From numerous QTL studies done with the crosses of IHOxILO and IHPxILP, Dudley (2007) recounts all the estimated numbers of significant factors for these phenotypes. It was estimated that there were probably around 50 genomic regions involved in the determination of the oil concentration phenotype. Further, Dudley (2007) estimated that up to 50 genomic regions were responsible for protein concentration, down from an earlier estimate of 122 using classical quantitative genetic procedures (Dudley 1977). Regardless of the specific number of expected associations, the large number of small-effect associations shows the complexity in a system with unparalleled variation for kernel protein accumulation. Some of the

other factors in play are starch concentration (negatively correlated with oil, and protein, concentration), seed size (negatively correlated with oil, and protein, concentration), and phytate concentration (negatively correlated with oil concentration, but positively correlated with protein concentration; Dudley and Lambert, 2004; Raboy et al., 1989). With all of these pathways having an influence on the phenotype, it is maybe less surprising that so many genetic factors are thought to be influencing the phenotype. Furthermore, grain protein concentration is a broad, and cumulative trait encompassing all proteins plus any other nitrogen-containing species from both the germ and the endosperm parts of the maize kernel.

So, what are the limits of selection for maize kernel protein and oil? Dudley and Lambert (2004) report that ILP and ILO seemed to have reached a limit to selection – ILO had concentrations of oil so low that the method used to measure that phenotype could no longer detect that species of chemical and it suffered from poor germination, and ILP had not changed phenotype in 30 years and also suffered from poor germination. It is thought that these may have established physiological lower limits in maize. Although Dudley and Lambert speculated that IHO and ISHO (Illinois Switchback High Oil, another high oil population established after selecting for lower oil concentrations in IHO for five years) had not reached a limit to selection, the interest in high oil was not sufficient to warrant its continuation. The 108^{th} cycle of selection for IHO was the most recent, grown in 2010. The oil concentration in the grain in that population had not increased in the eight years since Dudley and Lambert had hypothesized its unreached upper limit (206 g kg^{-1} in 2010 compared to 218 g kg^{-1} in 2000). At several times in its history, there was great interest in IHO – germplasm from the IHO population has even been used by several companies in commercial hybrids (Lambert et al., 2004). However, of the original four populations, only IHP continues today.

As of 2017, 116 cycles of recurrent selection have moved the average IHP grain protein concentration from 11% to around 30%. After 100 years, Dudley and Lambert concluded that no further increase has been seen in the most recent cycles of selection for higher protein in IHP; however, its continuation is warranted for several reasons: phenotypic variability was still prominent, and genetic variance had not decreased in the past 30 years, and viability problems had not been documented in IHP which was different than experiments in other species (Dudley and Lambert, 2004). With IHP at the high end of the spectrum for this trait, the ILTSE has

produced the known phenotypic extremes for grain protein concentration. In maize, the majority of the endosperm storage proteins is made up of a group of proteins called zeins and this is no exception in IHP. Zeins are alcohol-soluble, proline- and glutamine-rich prolamins, and it appears that the phenotypic response to selection in protein concentration is largely attributable to these zein molecules (Figure 1.1b). Zein proteins are composed of several groups, the main one being the alpha-zeins (Song and Messing, 2002). In fact, the alpha-zeins constitute almost 50% of the total protein in the kernel (Huang et al., 2004), but the regulation of these zein genes remains largely unknown.

Further reason for the continuation of IHP is seen in the success of several reverse selection populations. After 48 cycles of forward selection in IHP, Illinois Reverse High Protein (IRHP) was started to see if sufficient variation remained to get a phenotypic response to reversing the direction of selection despite the fact that by cycle 17 IHP could be traced back to one original ear (Winter, 1929). The same was done with ILP, which started Illinois Reverse Low Protein (IRLP) (Leng, 1962; Woodworth, et al., 1952). Both IRHP and IRLP have shown a response to selection (Figure 1.1a). These populations indicate that genetic variance was clearly not exhausted by cycle 48. New reverse selection populations have been started from the 100th cycle of IHP to test if after 100 cycles of selection the grain protein phenotype still responds to selection.

MATERIALS AND METHODS

Germplasm and Selection

The plant materials used in this study are the result of many years of selection prior to the author's involvement. Started in 1896, 163 ears (2 bushels) of "Burr's White" were measured for kernel composition. Seeds from the 24 ears with the highest protein concentration, 24 ears with the highest oil concentration, 12 ears with the lowest protein concentration, and 12 ears with the lowest oil concentration were the founders of this experiment. Dudley and Lambert (2004) provide an overview of the many breeding schemes used to arrive at the materials that were used in the current study. Briefly, the current method of maintaining and selecting populations follows this scheme:

The twelve (12) most-desirable ears of each strain (highest protein concentration for IHP, lowest protein concentration for IRHP3, highest protein content for IHPC, etc.) are selected and ranked from most desirable to 12th best. The odd ranks become group A and the even ranks become group B. Each selected ear gets catalogued and kept separate. From all the ears in group A, 25 seeds from each ear are taken out and bulked. This mixed sample is then separated out into 10 sets of 15 seeds. The same is done for group B. These are then planted in the field as alternating rows, ABABABA... Visual inspections are made to ensure interpopulation crossed did not happen in the previous summer – hybrid plants are rogued or marked for exclusion from the crossing scheme. Pollinations are made by hand, using each plant only once either as a male or female. Pollen donors are usually detasseled after a pollen sample is collected and used. Pollination continues among the 300 potential plants until 100 pollinations (with 200 parents) have been made or no pollen or ears remain. At harvest, at least 60 well-filled, disease-free ears receive a unique number identifier and all ears harvested from a given population are photographed (Supplemental Figures A.1 – A.14 show the ears harvested from the two most recent years). The grain composition phenotypes are then measured after shelling each ear individually. The data are analyzed for extreme outliers, the 12 ears with the most extreme values in the direction of selection are chosen and kept separate. The remaining ears are bulked as an additional archive in addition to the 12 selected ears, which again, are kept as separate individuals. This is done for each population being continued (IHP, IRHP2, IRHP3, IRLP, IRLP2, and IHPC). There are a couple of minor modifications to this general scheme for IHPC. The end plants of each row and all plants in the outer-most rows (rows 1 and 20 if planted in a block) are used ONLY as male plants. Because content is calculated as percent nitrogen in the kernels multiplied by how much mass is in kernels, and end plants receive more light and experience less competition, the goal is to have female plants that *have* neighbors and like neighbors. If a plant produces more than one ear per plant, the second can be open-pollinated, harvested, and analyzed but not used for seed. Because grain mass is a product of both kernel number and kernel weight, which are impacted by incomplete pollination, pollinations are delayed until three to four days after initial silk emergence. In addition to the kernel composition traits, the estimate of grain moisture from the NIR scan is used to adjust the grain mass to a dry-matter basis. This further demands good, full pollinations done when almost all of the silks have initiated growth (which is less problematic with a population where one is not trying to make any

self-pollinations). Without any restrictions on which plants are pollinated, IHP grain mass per plant has been recorded every year, too, as a control so a comparison between nitrogen content in IHPC and IHP can be made for demonstrative purposes.

IHP, IRHP, IRLP, Illinois Reverse Low Protein 2 (IRLP2), Illinois Reverse High Protein 2 (IRHP2), Illinois Reverse High Protein 3 (IRHP3), Illinois Switchback High Protein (ISHP), and Illinois High Protein Content (IHPC) are all white-seeded populations with white cobs and enormous tassels with copious amounts of pollen produced. Many other features are quite variable between the populations, for example IHP is generally characterized as having thick, dark green leaves borne from a very thick stalk with decreasingly elongated nodes especially above the ear node; whereas, ILP has narrow, yellow-green leaves and a tiny root from a much less robust stalk which maintains node elongation and achieves heights much taller than IHP. Inbreds were created to represent many of the populations in the ILTSE. These are grown and self-pollinated to serve as controls of fixed genotype whose phenotype can be compared across years and used to help normalize the environmental contributions to phenotype (Uribelarrea et al., 2009; Lucas et al., 2013). Inbreds for some populations are still being developed; in those instances, the S_x plants are grown next to representatives of the population from which they came. The population control plants are sib-mated and the inbreds-to-be are self-pollinated. The self-pollinated ears that are most similar to their respective founding population each year are selected for the next generation – the current author started developing inbreds from the oils populations via the pedigree method after some inbreeding of the bulk population.

Field Trials and Phenotyping

The plant materials were grown in Urbana, IL in years 2013-2017, although not all measures were taken each year. Modifications were made to common agronomic practices used in central Illinois for corn production. These materials are hand-planted 15 seeds in a plot that is 3.66 m long in 76.2-cm-spaced rows. Supplemental nitrogen was added in most years with the goal of creating a homogeneous experimental space. These plots were also rotated in soybean-corn every other year. The exception to the last two conditions is in 2017, these materials were grown in a field that had been corn the previous year and did not receive supplemental nitrogen to facilitate adjacent experiments. All phenotyped plants were hand-pollinated. Kernel

composition traits (grain oil concentration, grain nitrogen (~protein) concentration, and grain starch concentration) were all determined with the Perten NIR spectrophotometer (DA 7200) based on a custom calibration to handle the extremes in maize grain composition.

Genotyping Individuals from the ILTSE

In 2015, 24 plants from each population growing in the field were marked as sampled plants. Each of these plants had the tip of their youngest leaf removed, DNA extracted from that tissue sample, and the samples were plated and sent to Cornell for genotyping-by-sequencing (GBS) following the protocol from (Elshire et al., 2011). Each of these genotyped plants was then self-pollinated and kept separate. Because these plants are not inbred materials, a higher-depth sequencing was performed and unimputed genetic data was utilized in downstream analyses. A basic clustering procedure was performed in TASSEL (Bradbury et al., 2007) on the whole dataset, then the sites were filtered to less than 10% missing data and a minor allele frequency boundary set at 0.05, and the same basic clustering procedure was performed on this truncated dataset of 23,751 genetic markers. The two clustering procedures gave remarkably similar outputs (data not shown) confirming that the filtering step did not bias the interpretation of the data.

To get an initial indication of genetic structure within the ILTSE materials, the filtered data set was analyzed using the Bayesian clustering algorithm implemented in STRUCTURE (v2.3.4; Pritchard et al., 2000; Falush et al., 2003). This analysis included all loci obtained from the filtered dataset from TASSEL. A VCF containing these loci was converted to a STRUCTURE file in PGDspider (Lischer and Excoffier, 2012) per the specifications required to implement the clustering analysis in STRUCTURE, chiefly a single SNP per loci. The analysis in STRUCTURE was completed with the following specifications: admixture model, correlated allele frequencies, and no prior population information. Each clustering analysis utilized 10,000 burn-in steps and 20,000 replicates of one through ten (1-10) genetic groups (K), which was repeated 3 times for each possible K value.

Statistical Analysis

Statistical analyses were performed using the statistical analysis software, SAS (9.3 and 9.4). Proc Reg was used to assess slope of response over years. For IHP, there were somewhat influential data points (identified with Cooke's D scores) especially since the switch to the Perten. IRHP2 and IRHP3 did not have these influential points except for in their second year and the few lower points in the last four years. The test of variances was done with the Brown-Forsythe where IRLP was not included because it was flooded out in 2015. The independent samples t-test was performed with Proc TTEST and not adjusted because there were equal variances between IRHP 2015 and ISHP 2016 ($F=1.02$, $p=0.95$); however, the variances between ILO2 and ILO3 were different ($F=4.82$, $p=0.005$) and thus the Satterthwaite adjusted t-value and its p-value is reported – this difference in variance is driven by one distinct value in ILO2.

IHP variances were calculated and ranked from largest-to-smallest. Of the 116 years that IHP has been grown, the variance for 2012 is 29th largest, 2013 is 3rd, 2014 is 7th, 2015 is 10th, 2016 is 24th, and 2017 is 37th largest. In fact, there are only three years prior to 1975 with variances larger than the 2012 year (1939 at 16th, 1938 at 23rd, and 1957 at 28th), highlighting the argument that the sustained higher variance may suggest that IHP has not hit a limit quite the same way ILP did. Additionally, with excellent germination rate, IHP will be continued.

Finally, for the analysis of all populations grown in 2015 and 2017, an ANOVA model was fit with the seed nitrogen concentration being described by seed source (genotype). Assuming a comparison-wise Type I error rate of 0.05, the p-values of the contrasts should be compared to the Scheff*é*-adjusted experiment-wise Type I error rate. For comparisons within these trials, contrast statements were used to provide F-values for the comparison and the p-values were adjusted as mentioned. Though this research was not specifically interested in all pairwise comparisons, a Tukey adjustment could be made for a concise way of showing all the means and sample sizes (IRHP23c0 is the same thing as IHPc100) which does not come to a different conclusion on any occasion from the previous adjustment.

RESULTS AND DISCUSSION

Illinois High Protein and Selection Reversal after One-hundred Years of Forward Selection

Illinois High Protein (IHP) has produced the biological extreme for maize kernel protein concentration. Although Dudley and Lambert suggested that further increases in protein may not be achievable (2004), selection in IHP has been continued for many reasons. After the 2017 field season, Illinois High Protein (IHP) has been undergoing forward selection for higher grain protein for 116 cycles. Over the past 38 years (1978-2017), the slope is not different from zero ($\hat{\beta}_1 = 0.004$, F=0.38, p=0.54), but the variance has not decreased as was seen in ILP when it had reached a plateau (Figure 1.2). In fact, the variance was as high in 2013 – 2017 as it has ever been despite lower phenotypic values for kernel nitrogen concentration than in recent years. This variability remaining has not been accompanied by a lack of viability of the seeds, as was also the case for ILP. In the field, the visual phenotypes have remained consistent – thick, wide, dark-green leaves, internodes that don't elongate (especially near the top of the plant), and small ear shoots with tiny, compact seeds.

Over those past 40 years when the slope for seed nitrogen (percent per year) has been zero, several different machines have been used to measure the nitrogen in the seed. In the most recent cycles of the experiment an NIR spectrophotometer has been used which enables all the seeds from an ear to be included in the phenotyping and then those same seeds planted for the next generation. Earlier methods of phenotyping required a subset of seeds from an ear be ground into powder either for wet chemistry analysis or earlier versions of the NIR method (Dudley and Lambert, 2004). Despite the seed nitrogen percentage not increasing in recent years, this population continues to be the highest protein concentration materials and serves as a great comparison for other current experiments. For example, two reverse populations were started from IHP cycle 100 to test if selection could influence the tremendous variability still seen in the protein concentration phenotype. It is necessary to continue selecting for higher protein from this same base population to be able to compare the lower protein concentrations of the reverse populations to IHP such that the interpretation of change in phenotype of the reverses is more than just favorable field conditions and changes in phenotyping instruments. A distinct advantage of plant systems is that seed can be archived and then several cycles of an experiment

grown at the same time, in the same field, and phenotyped with the same instrument. This has been done several times in the past, and was done again in 2015 and 2017. In 2015, the selected cycles to be grown were chiefly chosen for which samples the lab needed to get DNA samples for: the oldest IHP samples that are available (from 1964), the founding cycle of the newest reverse high protein populations (from 2001), and the current cycle of the population for continued selection. The seeds produced from the 2015 IHP population had a higher percent protein than 1964 ($F=44.6$, $p<0.0001$) and lower protein than the 2001 IHP representatives ($F=18.9$, $p<0.0001$). Although we were unable to quantify this caveat, by chance the plants from 1964 and 2001 were grown in an area of the field which was not nearly as effected by early season floods as where the 2015 plants were growing (these were all grown together as per the methods of continuing selection). Finally, IHP1 is an inbred that was derived from IHP cycle 90 and has been grown many of the past years to provide a frame of reference for seed nitrogen concentrations (Figure 1.3).

IHP1 is also showing this same general trend of slightly less protein percent over the past 13 years (where phenotypes were recorded in 11 of those years) with the protein concentration falling below 250 g kg^{-1} in 2017. In most years, the inbred is in the lower quartile of the population values. It was noticed while creating the inbreds to represent the populations that the process of inbreeding moved the phenotypes away from the extreme (data not shown, Moose Lab). This is seen in the IHP inbred, ILP inbred, as well as the inbreds to represent the extremes for oil concentration – at several stages during the inbred line development process, individuals were chosen that best represent the population first for the phenotype that it is named after then for all other phenotypes as well. To minimize spatial variation, IHP1 has been grown near the IHP cycle in the recent past. This also means that IHP1 experienced two waves of substantial flooding in 2015, where the ground was completely underwater and then saturated for a whole week early in June, and again in late June (Supplemental Figure A.15). Maize grain nitrogen concentration increases when nitrogen fertilizer levels are increased in the field (Uribelarrea et al., 2009), and although soil tests were not performed in July of 2015, there was visual evidence of limited available nitrogen to the plants in the soil during the 2015 field season.

IHP1 had a low protein concentration again in 2017 (Figure 1.3). To accommodate another experiment, in 2017 the field site chosen had not grown soybeans the previous year as is

typical; instead, corn had been grown in this experimental field the previous year. Additionally, no supplemental nitrogen fertilizer was applied to the field in 2017. While this was different than the most recent years, nitrogen fertilizer was not applied to the ILTSE its first 52 years (Dudley 2007); however, not applying nitrogen fertilizer does decrease the potential to have high protein concentration phenotypes (Uribelarrea et al., 2009). These IHP1 plants might have been more adversely affected by stresses than the population plants due to their inbred status, but other populations of IHP also showed lower-than-expected protein concentrations in 2017 (Figure 1.4). The oldest seeds available in the experiment were retrieved from archive, as were a sample from each 5-year increment in the experiment. The exception is the current cycle and IHP1 which were created in the previous growing season. Fifteen seeds were planted for each of the archived populations, and it is not altogether surprising that in a field setting the older populations failed to germinate and establish stands like the more recent populations (number of ears measured from each population, Figure 1.4b). By cycle 65, nitrogen fertilizer had begun to be utilized as a part of standard procedure, so all seeds sown in 2017 were produced in fertilized soil, but this does not explain the significant difference between all materials and the current cycle of IHP.

After 100 years of forward selection for higher protein, Illinois Reverse High Protein 2 (IRHP2) and Illinois Reverse High Protein 3 (IRHP3) were started from the IHP population with the goal of decreasing seed nitrogen. In the first 14 cycles of reverse selection, IRHP2 has seen a slope of -1.28 ($F=1520.22$, $p<0.0001$) compared to -0.29 of IHP over the same time ($F=129.73$, $p<0.0001$) (Figure 1.5). Meanwhile in the first 15 cycles of reverse selection, IRHP3 has seen a slope of -0.68 ($F=508.95$, $p<0.0001$), again compared to -0.29 of IHP (Figure 1.6).

These two populations (IRHP2 and IRHP3) are the only replicated trials in the ILTSE in the sense that they were started from the same materials and selected in the same manner. There is a difference, though, in how they were grown – for the first three years, IRHP2 was grown in the low-nitrogen nursery with no supplemental nitrogen fertilizer applied while IRHP3 was grown in the high-nitrogen nursery. From the fourth year on, they were both grown in the same conditions. IRHP2 has undergone one less cycle than has IRHP3 due to a planting error in 2007 which was discovered too late to plant the correct seeds that summer and were kept until the next year. Visually, the two strains still look the same, and like IHP. Additionally, a hybrid made between IRHP2 cycle 11 and IRHP3 cycle 12 showed no visual heterosis – these plants were not

remarkably taller than either parent and the ears did not look any different from IRHP2 or IRHP3 – grown in 2016, the IRHP2 plants had a range of heights from 170-215cm, IRHP3 180-205cm, and their inter-population cross 200-210cm).

Additionally, the founders of IRHP2 and IRHP3 were grown in 2015. This snapshot allowed us to see that both IRHP2 and IRHP3 have lower seed nitrogen concentration than their founding IHP plants ($F=784.6$ and $F=62.4$, respectively, both $p<0.0001$). This comparison is more robust than comparing to historical numbers because they were grown in the same year and the phenotype was measured on the same instrument. Additionally from that 2015 plot, IRHP2 and IRHP3 had lower seed nitrogen than the IHP cycle 114 population ($F=1042.6$ and $F=23.6$, respectively, both $p<0.0001$). Since 2015, the phenotypic values for IRHP2 and IRHP3 have continued to decline whereas IHP has remained fairly steady (Figure 1.5, 1.6). Finally, IRHP2 has responded to selection more quickly than IRHP3 has, with IRHP2 having significantly lower seed nitrogen concentration than IRHP3 in 2015 ($F=753.0$, $p<0.0001$). The response to selection seen in IRHP2 resembles that seen in IRHP in its first fifteen cycles (Figure 1.1).

IRHP and IRLP

Illinois Reverse High Protein (IRHP) was started in 1947 from IHP and has been selected for lower seed Nitrogen for 64 cycles (Figure 1.7). After 12 cycles, there was no overlap with IHP, and by cycle 31 IRHP had nearly the same seed Nitrogen as Illinois Low Protein (ILP). That said, the past 32 cycles (cycles 32 to 64) have done much what ILP did once it hit about 5% seed Nitrogen – ILP had a slope of -0.001 in its last 32 cycles ($F=0.65$, $p=0.42$), and IRHP had a slope of 0.004 in its last 32 cycles ($F=2.61$, $p=0.11$). In 2015, representatives of ILP were grown from 1964 and 1990 (which were not different for grain protein, $F=1.0$, $p=0.31$), and IRHP was not different from the ILP representatives ($F=0.0$, $p=0.94$). In addition to the slope of IRHP being statistically zero, the variability has decreased – the variance in 2015 was different from all other cycles ($F=18.86$, $p<0.0001$). This can be seen in the field, too – the plants in this population are much more uniform than any of the others currently being grown. It is for these reasons that the author suggested to cease selection in IRHP for lower protein. Instead, selection was initiated for higher seed nitrogen concentration. A reverse selection experiment was started in ILP when it was discontinued, and the same was desired from IRHP. The difference here is

that IRHP came from IHP cycle 48 which had been selected for higher protein whereas ILP had only ever been selected for lower protein concentrations. To create this population, seeds from the IRHP population which were created in 2015 were pooled and grown in 2016 as the cycle 0 of Illinois Switchback High Protein (ISHP, named similarly to Illinois Switchback High Oil). With the caveat in mind that these plants were grown in two different years, the seed nitrogen concentration was similar in 2015 and 2016 ($t=0.18$, $p=0.8572$), which should be the case since there was no selection in 2015, but ISHP was selected for higher seed Nitrogen concentration after the 2016 cycle. Figure 1.7 shows a larger range of values spreading upward in 2017 despite the field not receiving nitrogen fertilizer, but it is too early in the experiment to make too many conclusions.

Illinois Low Protein (ILP) has not been grown since 1994. As discussed previously, lack of progress combined with poor viability were deciding factors when Dudley and Lambert stopped ILP after 93 cycles (2004). Illinois Reverse Low Protein (IRLP) was started in 1947 from ILP and has been selected for higher seed Nitrogen concentration for 66 cycles. In the field, IRLP is tall, has high ear placement, and lodges severely; but, IRLP is still making progress (slope = 0.18, $F=14001.99$, $p<0.0001$; Figure 1.7). The rate of progress is smaller in magnitude than IRHP, but is similar to IHP 1896-1977 where the slope was 0.18 ($F=29967.4$, $p<0.0001$). The ILP founders of the IRLP population had comparatively low phenotypic variation and low protein concentrations, but despite this, IRLP continues to increase in protein concentration and begs the question of whether IHP-like levels of seed Nitrogen concentration can be obtained from a population stemming from ILP cycle 48.

IRLP2

While the first reverse selection population from ILP has been increasing in protein concentration, the question was asked ‘could the same be seen after an additional 30 years of selection for low protein and maintenance of very small phenotypic variance?’ Concurrent with a cease in ILP, another reverse strain was started – Illinois Reverse Low Protein 2 (IRLP2). ILP had been at a similar level of seed nitrogen concentration from cycle 57 to 93. After 16 cycles of IRLP2, the slope was actually negative at -0.03 ($F=21.5$, $p<0.0001$), showing no progress toward

higher seed nitrogen concentration. Then, a dramatic response was seen starting in cycle 17 (Figure 1.8).

When the data for each individual ear were analyzed instead of just the mean of each year, a peculiar pattern was noticed (Figure 1.9) – nine of the measures made in 2008 far exceeded the previous range of values seen. While this observation is potentially possible, it seems highly unlikely. Because of the population size and breeding scheme, even a dominant mutation would not show a phenotype as seen in Figure 1.9 that quickly, especially when you consider the ILTSE breeding scheme (Moose et al., 2004). With the large number of populations in this experiment and how similar the names of the populations are, it is hypothesized that seeds from a different population were mislabeled or mixed into this population inadvertently. Alternatively, an inappropriate plant could have been used as a pollen parent and the resulting progeny grown in 2008 could have not been noticed as overly hybrid-like as these materials were being tended to by a new researcher at this time. Despite the error made, it was desired to make an effort to 're-start' the selection process from before this mixing event occurred.

To determine how far back in the archives to go for the seeds to resume selection in IRLP2, representatives from several cycles of the population were grown in 2016. In addition, the twelve selected ears from 2006 were also grown for crossing as a population for continuation if it passed the visual and kernel composition expectations of those materials. Just by visual examination of the kernels from these archived populations it is apparent that something drastic happened in this short time frame as the kernels went from chalky and opaque to translucent and increased greatly in size (Figure 1.10).

In addition to the striking differences in the kernels in the archives, the plants growing in the field in 2016 were also qualitatively distinct. The 2005 and 2006 plants looked like ILP and ILP-derived materials also growing in the field in 2016 – all had light, pale-green, narrow leaves and strikingly elongated internodes. The 2008 plants were segregating plants that looked like ILP and others that looked like hybrids, and the 2009 plants all looked like the more robust plants from the 2008 section. Pollinations were made in all sections and the seed nitrogen concentration was measured for the ears from these plants, showing very little overlap between 2006 and 2009 (Figure 1.11). The pollinations in the section that was planted from the 2006 selected ears was

done following the standard procedure for maintaining and selecting a population, thus, essentially re-growing the 2007 'cycle 17' in 2016. Selections were made and another cycle was grown in 2017 showing phenotypes similar to that seen in 2016 (Figure 1.12). While IRLP has shown a strong response to selection, the response was slower than any of the selections from high protein to low protein. This is also consistent with the oil experiment where reverse selections from high oil responded much more quickly than did reverse selections from low oil (Dudley and Lambert, 2004). The current population of IRLP2 seems to fit this observation and more cycles of selection will make this more clear.

IHPC

If we believe that IRLP2 was contaminated and was most likely a cross-contamination or mis-labeling within ILTSE-derived materials, it represents an inter-population mixing event. This assertion is based on the fact that IRLP2 still had tassels, silks, plant characteristics, white seeds, and white cobs like the rest of the ILTSE materials. After this mixing event, the IRLP2 population has been in isolation since, and in the Moose lab since 2009. Since IRLP2 cycle 23 was not going forward, the author used these materials to start a new population using a new selection criterion: selecting for higher protein content per plant. From a production standpoint, it could be imagined that getting as much nitrogen in corn kernels per area (or per plant, assuming equal spacing as is done for all populations in the ILTSE) would be desirable – this idea of having moderate protein concentrations but high yields was first discussed by Dudley et al. (1977), discussing ways to get the best production of protein per hectare. Illinois High Protein Content (IHPC) was started from the former IRLP2 which was in cycle 23 (which is 8 cycles after the mixing event). This 23rd cycle of IRLP2 was grown in 2015 and was considered cycle 0 of IHPC. The grain yield was measured in addition to seed quality traits, and those ears were selected on the total content of N in the seeds from each plant (the adjusted grain yield multiplied by the seed nitrogen concentration). The grain mass per plant has not been measured in the other populations in the ILTSE for many years, but the author made these measurements on all the populations grown in 2015 (data not shown). As suspected, the ears from the IHP population confirmed that individuals from that population did not put the most nitrogen into seeds when compared to other populations in the ILTSE. It made sense to do the suggested selection procedure in the former IRLP2 population because the top 30 plants for nitrogen content in the

seed all belonged to this population. Even in just the first three years' worth of data, there are some very interesting combinations of yield and seed nitrogen concentration that lead to some plants getting – or not getting – selected, as was suggested by Dudley and Lambert (1977). The yield and protein concentration were both lower in 2017, likely due to less fertile soils. More cycles of selection will be able to show if selection for this trait can be successful. For comparison, the protein content per plant was also calculated for IHP in 2015, 2016, and 2017. The selected plants from IHPC are higher than those of IHP (Figure 1.13). One additional restriction is placed on IHPC in regard to which plants can be utilized. In all ILTSE materials, each plant can only be used once – either as a male or female. This is also true in IHPC; but, plants at the end of any row and all plants in the first and last row (outside rows) of the IHPC block in the field can also NOT be used as females, but may only be pollen donors or not used at all. (IHP did not have this restriction, but was shown merely for comparison.)

Inbred Lines to Represent the Oil Selection Populations

The ILTSE was started to select for altered protein and oil concentrations. The discussion thus far has focused on the protein populations. This is largely because the populations that were being selected for different oil concentrations have not been selected since 2010. Concurrent with the discontinuation of these populations, inbred line development was started to have representative lines for the oil populations. During this process, a small subset of each oil population was grown to compare phenotypically to those plants undergoing inbreeding. Inbred lines had already been started before 2010 for Illinois High Oil (IHO) and Illinois Low Oil (ILO). IHO1, IHO2, IHO3, and IHO4 were grown in 2014 along with a pooled sample of a cycle of IHO which the inbred lines were to represent, and IHO2 and IHO3 were grown in 2015 along with a pooled sample of a cycle of IHO which the inbred lines were to represent. The other inbred lines did not represent the cycle as well as IHO2 did (data not shown). All original IHOx inbred lines still have seed, but additional selections were made in 2015 and 2016 in IHO2 to better represent the cycle and have a more stable phenotype. ILO2 and ILO3 represented the ILO cycle data for oil concentration, but ILO2 was markedly different than ILO3 for protein (protein concentration means of 12.6 and 8.6, respectively; $t=7.14$, $p<0.0001$) and morphology in the field. IHO2 and ILO3 were grown to represent the oil populations in 2017 and are included in Figure 1.1 and Figure 1.14 as controls because they were started from the same Burr's White

population in 1896 and have not been selected for protein concentration – the measurements for ILO-derived materials in 2016 and 2017 were the first measures made and included in the experiment since 1988. In Figure 1.14, IHO and ILO are plotted as separate values whereas the average of IHO and ILO are plotted for control in Figure 1.1 with the exception of 1988-2015 being just an average of all IHO ears measured.

Similar to the protein experiment, three other populations were initiated to reverse the response to selection for oil concentration. Illinois Reverse High Oil (IRHO), Illinois Reverse Low Oil (IRLO), and Illinois Switchback High Oil (ISHO) were all last grown and pollinated as a population in 2010, but inbred lines to represent the IRHO and ISHO populations are all at the S7 generation. The S5 IRLO looked so markedly different from the population of IRLO, that I self-pollinated all plants from the cycle of IRLO that was grown in 2016 to restart the inbreeding process (Figure 1.15). IRHO and ISHO have several lines that are similar phenotypically with regard to both morphology and seed quality traits. The bulk method was used for the first three generations in all populations, subsequently the pedigree method has been used to track specific lineages and enable the use of family structure information in the selection of the best inbred line to represent its respective population.

Genetic Patterns in the ILTSE

With recent advances in technology, genotyping a representative subset of a population genome-wide has become possible. Of particular concern was choosing enough individuals from each population such that a sampling bias was avoided as much as possible. Genotyping-by-sequencing (GBS) was the method used to obtain genome-wide SNP information for comparing the populations within the ILTSE. All populations grown in 2015 were sampled, including all of the active protein populations and three oil populations: IHO, IRHO, and IRLO. In addition, B73 (a model inbred line, also the genotype that was sequenced and these GBS markers mapped on) was included as an outgroup. It is not surprising that the STRUCTURE results demonstrated a strong signal between the ILTSE and the outgroup B73. The strongest split in the clustering procedure showed a strong division between the populations selected for protein (IHP, IRHP, IRHP2, IRHP3, IRLP, IRLP2) and those selected for oil (IHO, IRHO, IRLO). Further grouping revealed the presence of fine-scale structuring within the protein populations. The analysis

provided K = 6, where there were 4 distinct groups within the protein populations (Figure 1.16). Although the STRUCTURE analysis was blind to phenotype, each of these four clusters seemed to correspond to the protein concentration phenotype. The magenta cluster identity is strongly IHP with high protein concentration. Large portions of identity for IRHP3 and IRHP2 also belong to this magenta cluster. IRHP was also founded from IHP, but after 64 cycles of selection for lower protein concentration, the SNPs from IRHP separate itself from IHP almost entirely. Some of the green cluster identity is also shared by members from IRHP2. Two testable hypotheses are that 1) the same regions/SNPs that changed from IHP to IRHP are also starting to shift in frequency in the IRHP2 population, and 2) that there has been an inadvertent introduction of IRHP (either by pollination or mishandling during seed processing) into IRHP2. Seed for the selected ears from each cycle of IRHP2 and IRHP3 have been archived and are available for genotyping to assess the populations for these possibilities.

The other population that was thought to have experienced an interpopulation mixing event was IRLP2. The data presented here are of that population sampled in 2015 before this realization was made. IRLP2 has a distinct cluster identity but also shares some of the magenta cluster membership potentially indicating some high-protein SNPs being introduced into this population. Even with the potential mixing event, this ILP-derived population had only undergone 23 cycles of selection for higher protein concentration, thus the hypothesis is that the light-blue cluster membership represents low-protein SNPs. IRLP defines the other distinct cluster and shares no membership with either cluster making up IRLP2 membership. If the hypothesis that ILP would have been the light-blue cluster is granted, IRLP is similar to IRHP in that none of its membership is like its founding population. As all ILTSE materials came from Burr's White, this is also the case for protein concentration selection versus selection for oil concentration. Finally, the protein concentration phenotype in IRLP is different from other populations with a high protein concentration in a distinct way, biologically. Plants from the IHP population hyper-accumulate asparagine, whereas IRLP has – thus-far – found a different way to increase protein concentration (Moose Laboratory, Figure 1.1). This finding further demonstrates the complex phenotype that is kernel protein concentration and illustrates different pathways being used to achieve changes – after all, it only took three years for Hopkins to note that other morphologies were changing to these plants that resulted in the measured phenotype (1899).

Conclusions and Further Directions in the ILTSE

The ILTSE has produced the biological extremes for protein concentrations and oil concentrations in the maize kernel, and after 116 cycles it is still being continued. Although three of the four original populations are no longer undergoing selection and the fourth, IHP, has not seen recent improvement, reverse selection experiments have been started and offer unique opportunities to explore genetic, metabolic, and evolutionary questions. The phenotype of protein concentration in the kernel is affected by interrelations with other kernel constituents – protein and starch show a strong negative correlation with one another (Dudley and Lambert, 2004). Because the selection has been on concentration, this interplay is exaggerated, forcing one species to decrease in percentage if the other increases. To obtain the higher protein phenotype, these aspects could have been selected on as well: nitrogen uptake ability, improved nitrogen storage and/or remobilization in the plant, and a reduction in the starch storage in the seed and smaller seed size were noted only three years into the experiment. It was further noticed that IRLP and IHP plants tend to flower later than some of the other ILTSE materials, and IRHP2 and IRHP3 started to flower earlier than IHP concurrent with their gradually lower protein concentrations. To get an initial indication of whether flowering time should be considered for this trait, pollinations made in 2017 were noted whether they occurred on or before July 26 (time1) or on or after July 28 (time 2). There was no interaction between genotype and pollination date and there was also no significant difference in protein concentration between the two times in any genotype indicating that there was not a clear bias toward later-flowering plants producing a higher concentration of protein (Figure 1.17).

The ILTSE was started with a very practical goal in mind – an animal scientist collaborated with a chemist to address the possibility of changing maize kernel composition to better the end product. This practical goal has been wildly successful: high-oil corn has been licensed numerous times (Lambert et al., 2004), one half of the first commercially-available double cross corn hybrid was derived from high protein and low protein (Jones, 1927), and low protein and low oil are indeed of interest to those in need of high starch maize inputs (Ramchandran et al., 2017). Although the efforts to achieve higher protein corn has been successful, the largest class of proteins that increased in response to this selection regime was the zeins which have only further accentuated the lacking nutritional quality of maize protein. This

unique germplasm resource can allow for future studies into the regulation of the zein genes, and also demonstrates the incredible power of artificial selection during the breeding process. A further advantage of plant systems such as this is the ability to save seeds and regenerate populations that a direct comparison can be made between materials during the past 50 years. Over this time, IHP has continued to have a large phenotypic variance while approaching a plateau. Careful precautions have been taken and will be taken when exploring the reverse experiments, but as IHP is at a biological extreme for this phenotype, this is one population where further progress or change toward higher protein concentration is very unlikely to be due to contamination.

TABLE AND FIGURES

Table 1.1. Current and older populations from the ILTSE were grown in 2015 to compare phenotype on the same instrument from plants grown in the same year. The Tukey-adjusted mean separation is presented for a concise representation of the interpopulation comparisons that could be made.

Tukey's Studentized Range (HSD) Test for protein

Note: This test controls the Type I experiment-wise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	364
Error Mean Square	4.904714
Critical Value of Studentized Range	4.58113
Minimum Significant Difference	2.0346

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	Genotype
A	29.1618	22	IRHP23c.0 (IHP2001)
B	26.7582	60	IHP2015
B	24.7950	60	IRHP3_2015
C	22.2331	13	IHP1964
D	17.5961	23	IRLP2015
E	13.7007	60	IRHP2_2015
F	9.9559	17	IRLP1968
F	9.5740	20	IRHP1966
G	6.8709	22	ILP1964
G	6.4772	60	IRHP2015
G	6.1556	18	ILP1991

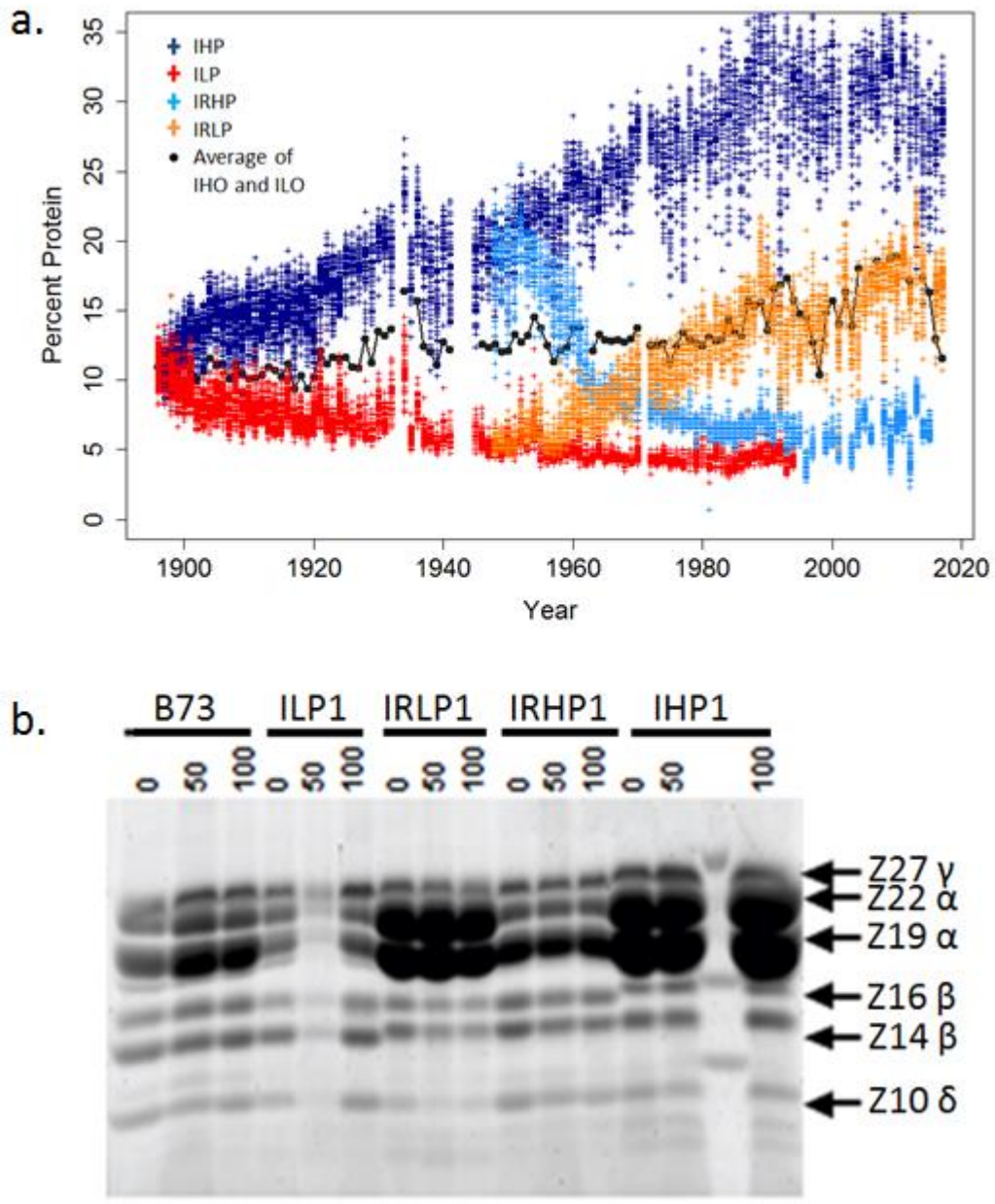


Figure 1.1. (a) The concentration of protein for each ear measured in IHP (navy), ILP (red), IRHP (light blue), and IRLP (orange) is plotted against the year its parent plant was grown. The black symbols represent the mean protein concentration of IHO and ILO that were grown that year. (b) SDS-PAGE analysis of zein proteins extracted from 16 DAP inbred maize kernels from the five genotypes labeled above the solid black lines from three nitrogen treatments in the field (0, 50, 100 kg ha^{-1}) (Moose, unpublished).

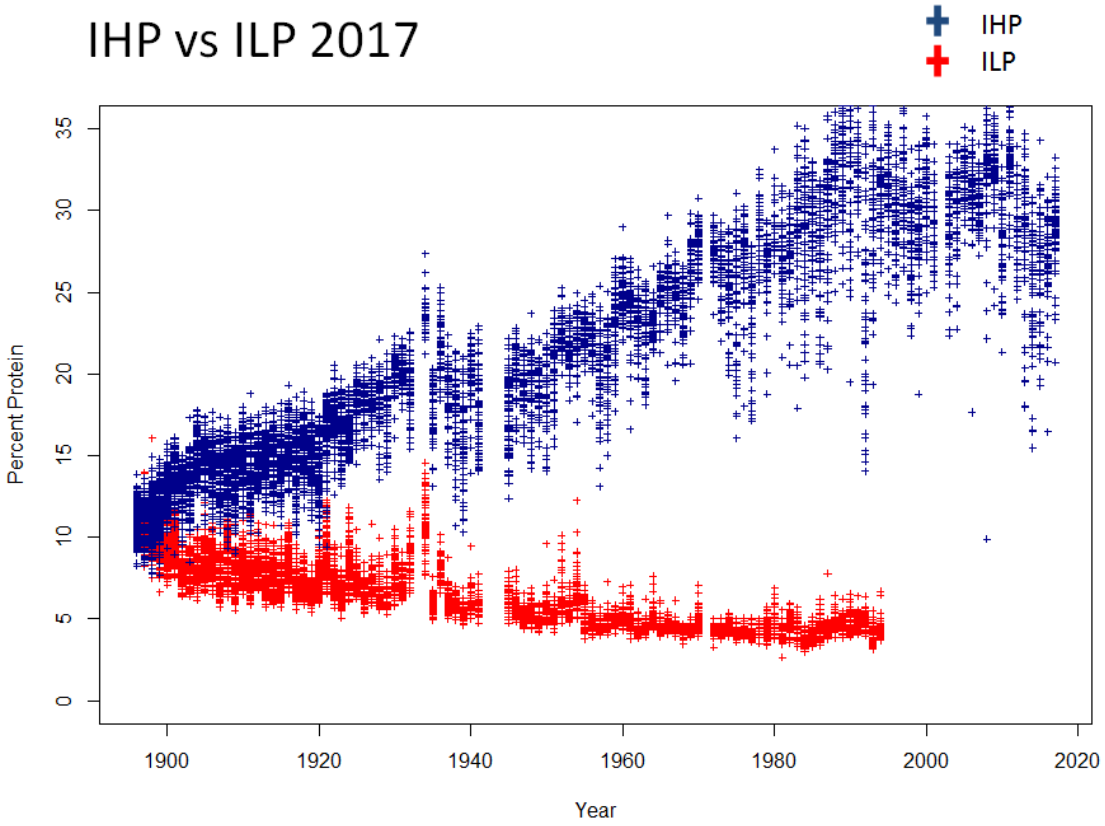


Figure 1.2. The concentration of protein for each ear measured in the experiment is plotted against the year its parent plant was grown. IHP is plotted in blue symbols and ILP is plotted in red symbols.

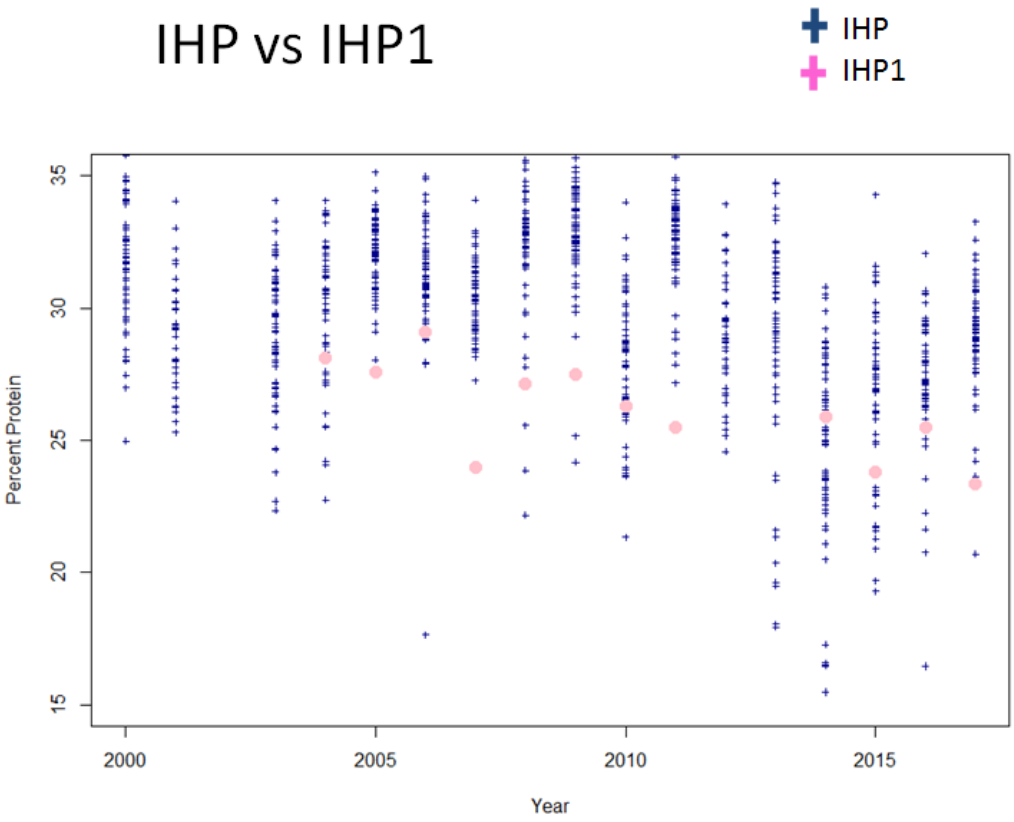
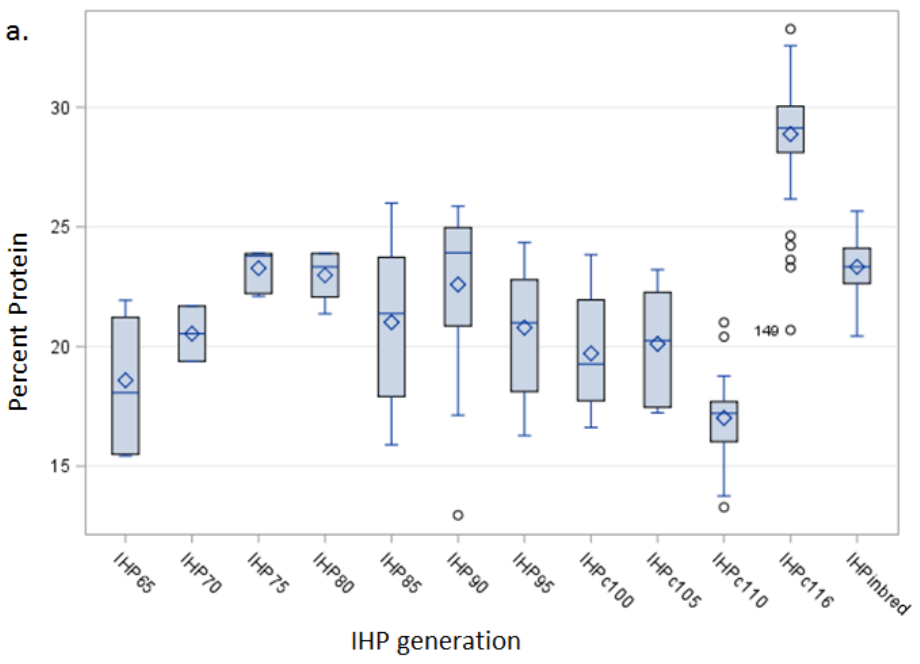


Figure 1.3. The concentration of protein for each ear measured in the experiment is plotted against the year its parent plant was grown. IHP is plotted in blue symbols and IHP1 is plotted in pink circles.



b.

Means with the same letter are not significantly different.

Tukey Grouping			Mean	N	Genotype
A			28.876	70	IHPc116
B			23.329	38	IHPinbred
B			23.277	6	IHP75
B			22.980	4	IHP80
B			22.592	14	IHP90
C	B		21.017	12	IHP85
C	B	D	20.790	9	IHP95
C	B	D	20.537	3	IHP70
C	B	D	20.110	6	IHPc105
C	B	D	19.715	12	IHPc100
C		D	18.591	7	IHP65
		D	17.018	13	IHPc110

Figure 1.4. (a) The protein concentration for each population of IHP is plotted as a box plot. (b) A Tukey separation shows all pair-wise comparisons between the different generations of IHP and the inbred materials.

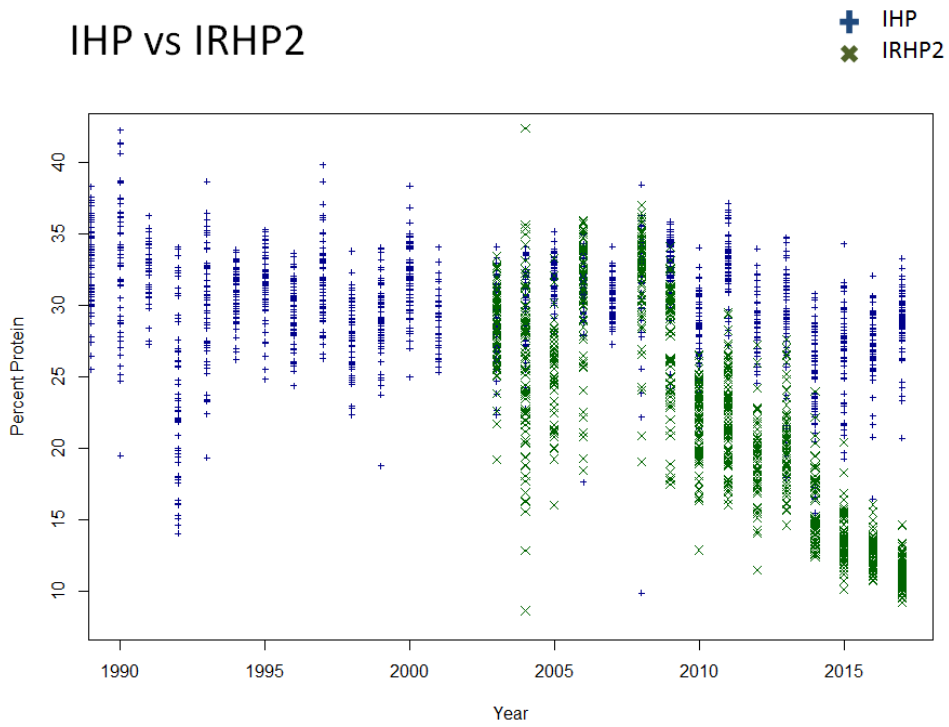


Figure 1.5. The concentration of protein for each ear measured in the experiment is plotted against the year its parent plant was grown. IHP is plotted in blue symbols and IRHP2 is plotted in green symbols. It was not documented why, but the IRHP2 ear with less than 10% protein in 2004 was not selected as a parent for the next cycle.

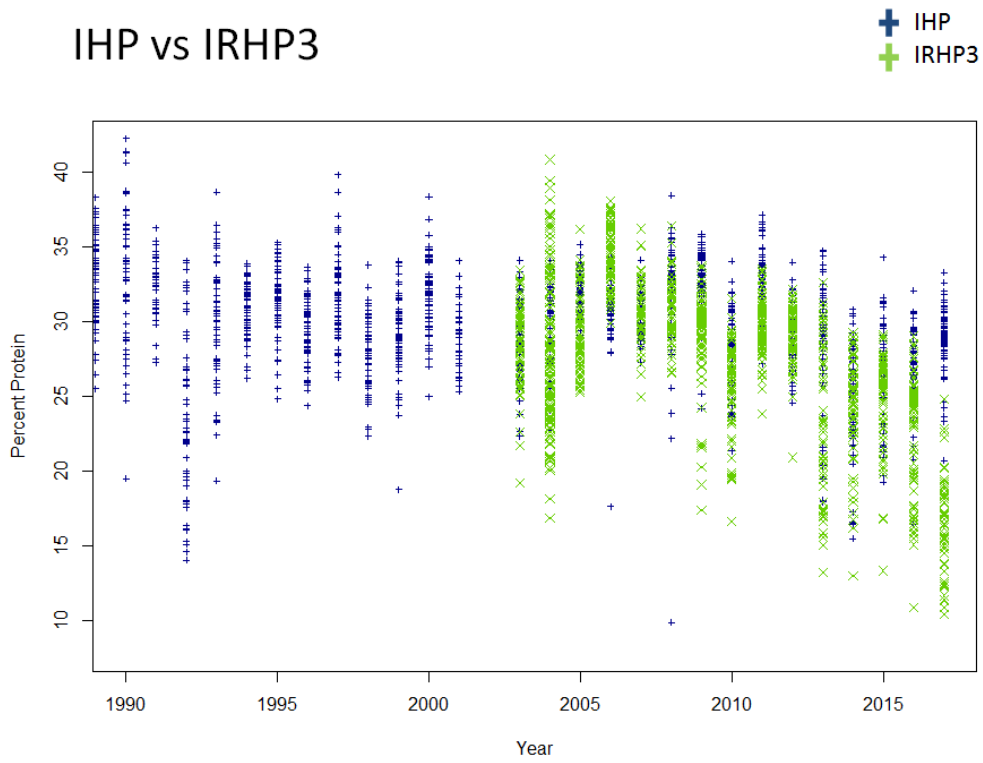


Figure 1.6. The concentration of protein for each ear measured in the experiment is plotted against the year its parent plant was grown. IHP is plotted in blue symbols and IRHP3 is plotted in lime-green symbols.

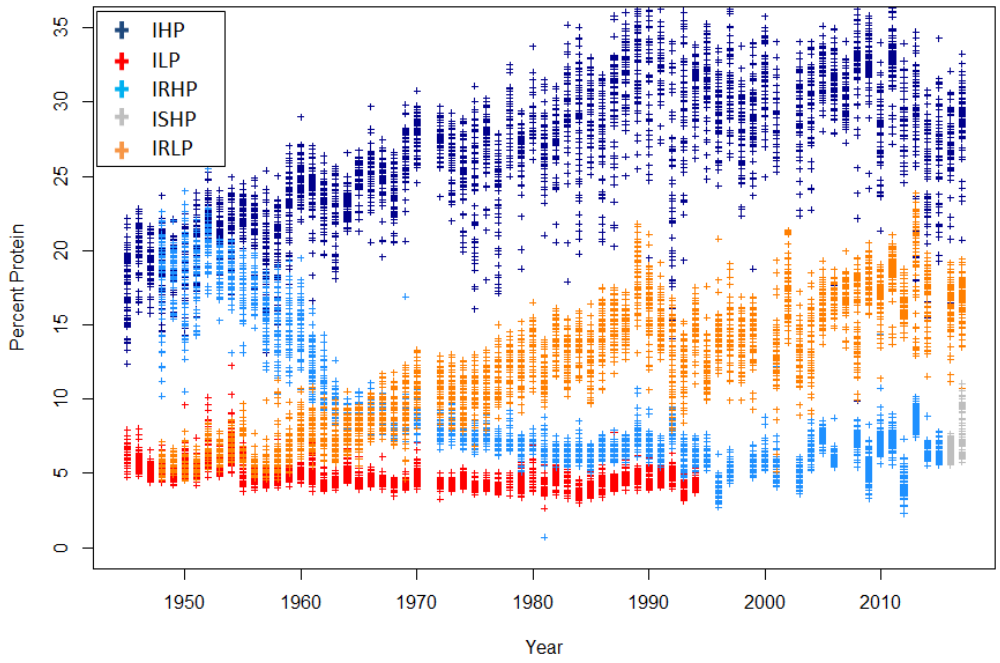


Figure 1.7. The concentration of protein for each ear measured in the experiment is plotted against the year its parent plant was grown. The key in the upper right-hand corner details which color symbols represent each population.

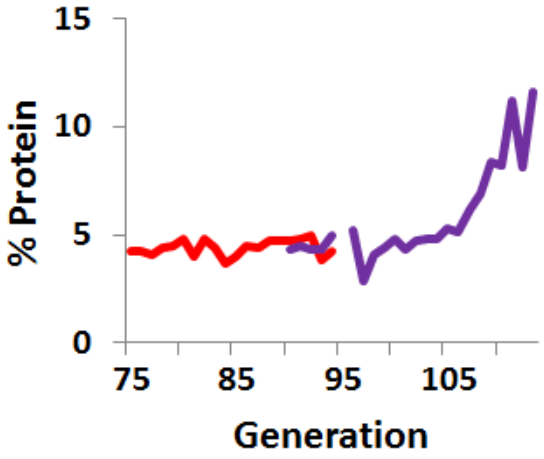


Figure 1.8. The concentration of protein for all ears measured in a year were averaged and are plotted against the generation of the ILTSE (which here corresponds to years 1975 to 2013) in which the parent plants were grown. ILP is plotted as a red line and IRLP2 is plotted as a purple line.

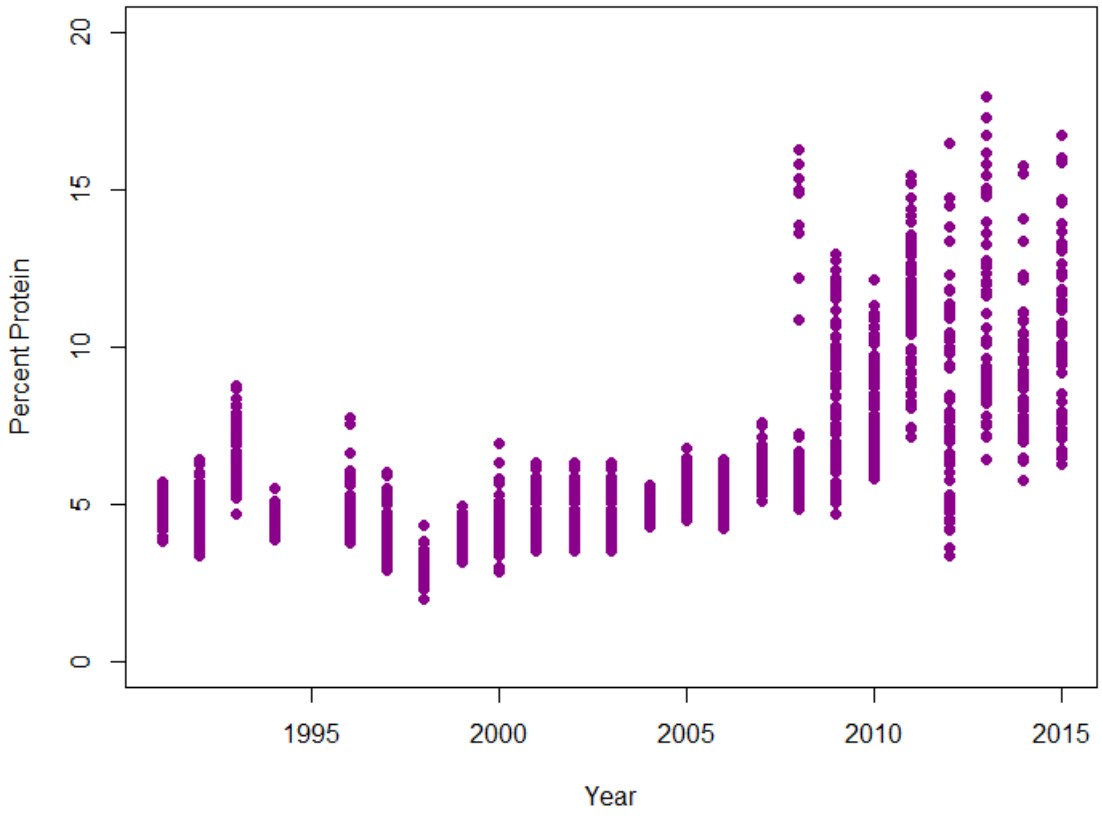


Figure 1.9. The concentration of protein for each ear measured in the experiment is plotted against the year its parent plant was grown. IRLP2 is plotted in purple symbols.

Figure 1.10. Seeds from three cycles of IRLP2 are shown, from left to right: 2005 bulk, non-selected bulk from 2008, and selected ears bulk from 2009.

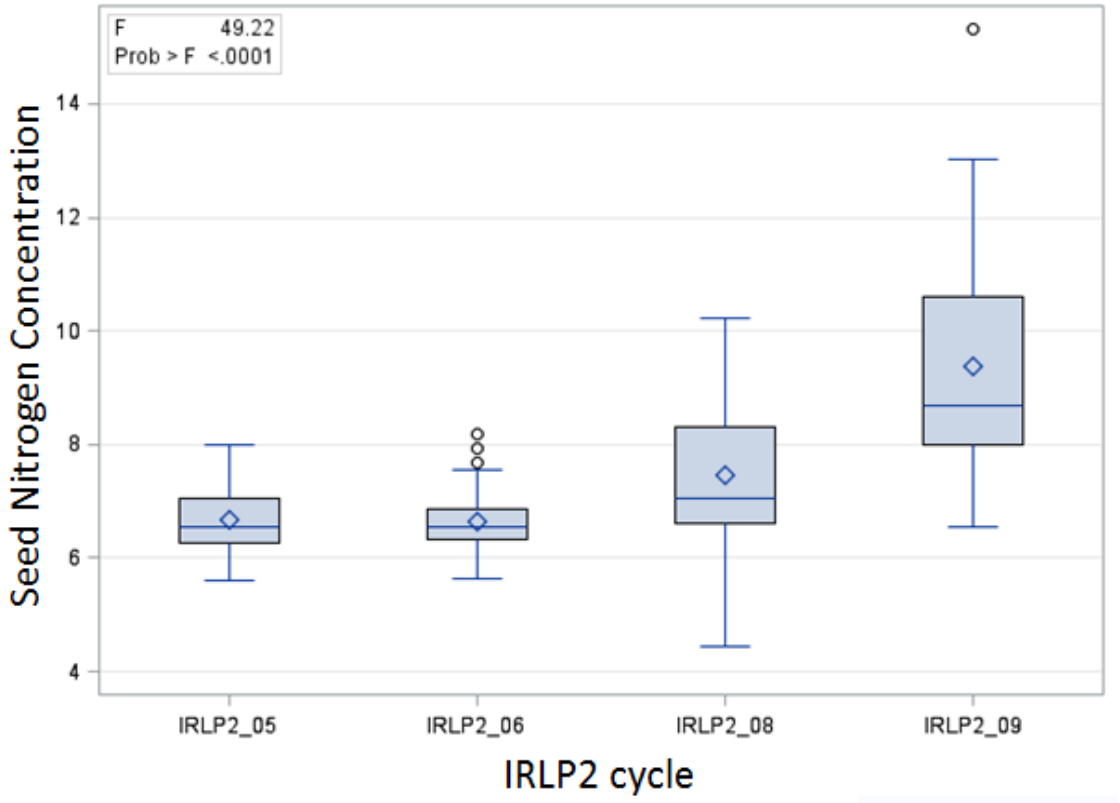


Figure 1.11. A box plot of the seed nitrogen concentration for each of the IRLP2 cycle and the F-test statistic for the one-way ANOVA model addressing the hypothesis that the seed nitrogen concentration was the same for all four cycles of IRLP2 is shown. This hypothesis is rejected.

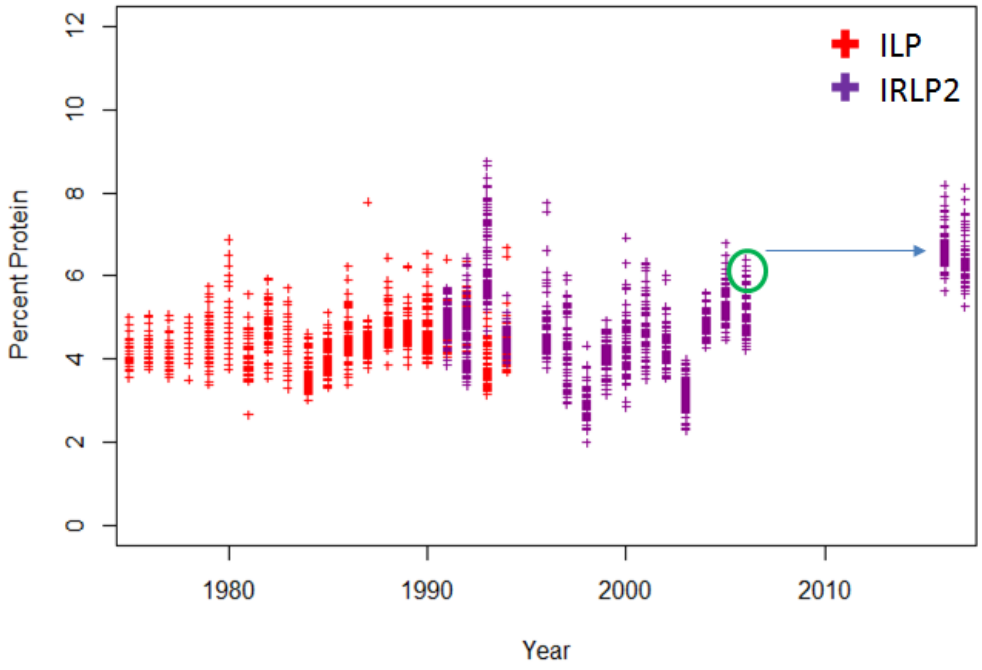


Figure 1.12. The change made to IRLP2 is shown where the concentration of protein for each ear measured in the experiment is plotted against the year its parent plant was grown. ILP is plotted in red symbols and IRLP2 is plotted in purple symbols.

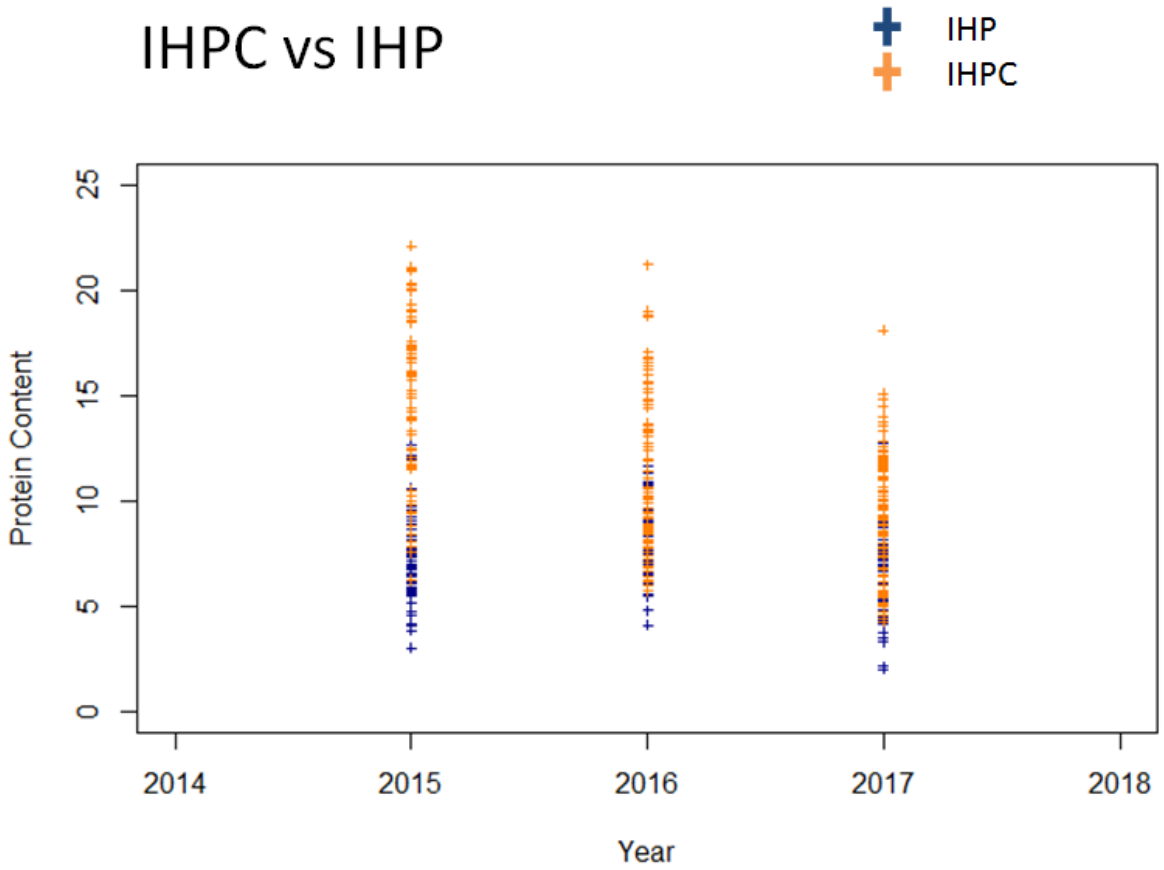


Figure 1.13. The content of protein for each ear measured in the experiment is plotted against the year its parent plant was grown. IHP is plotted in blue symbols and IHPC is plotted in orange symbols.

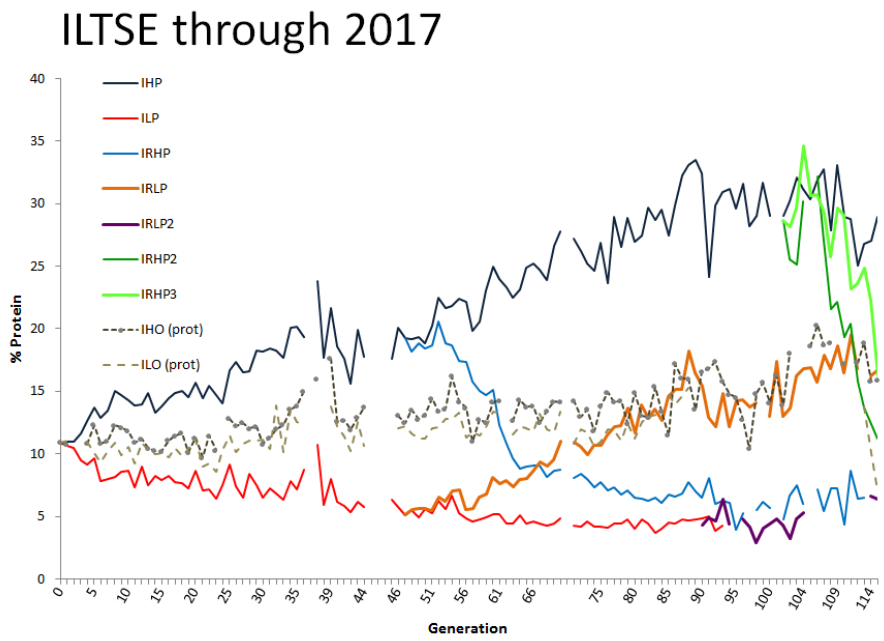


Figure 1.14. A line graph of the ILTSE displaying the mean of all ears from each year the experiment was grown, where the x-axis is generation of the experiment. The kernel protein concentration is also shown for the oil lines though they were not selected for protein concentration.

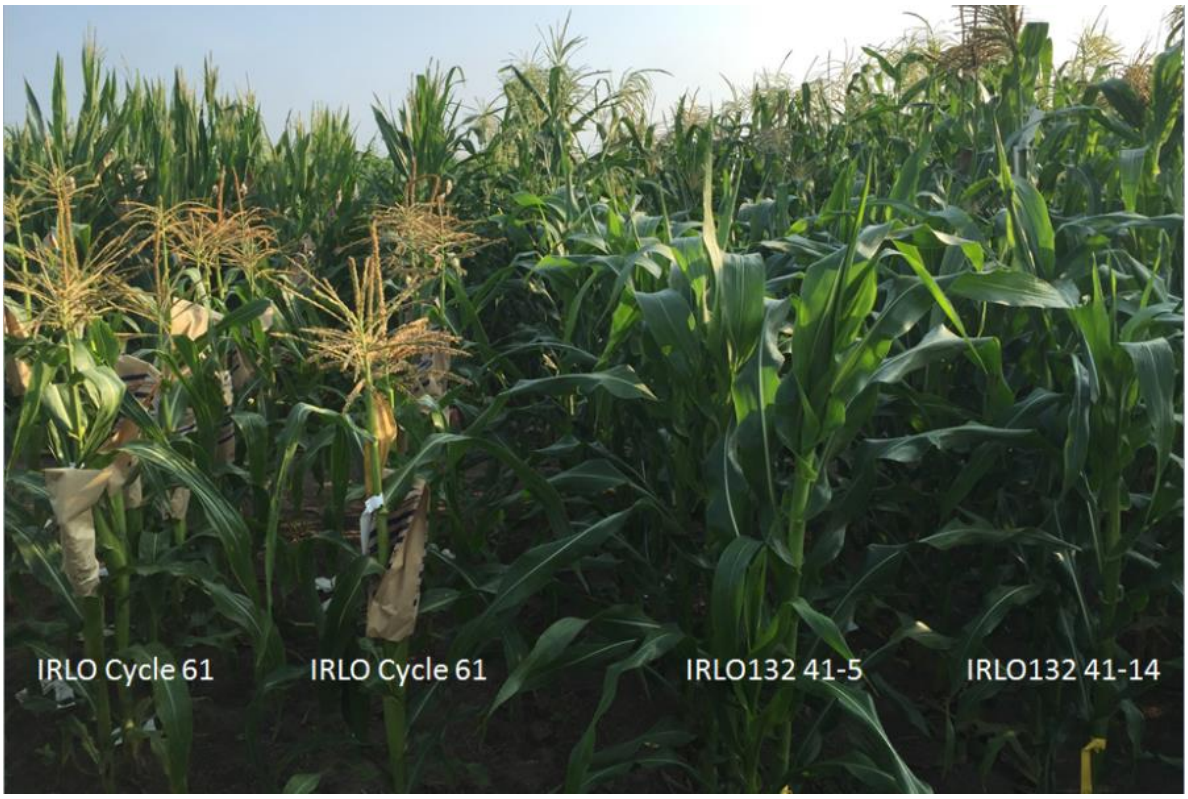
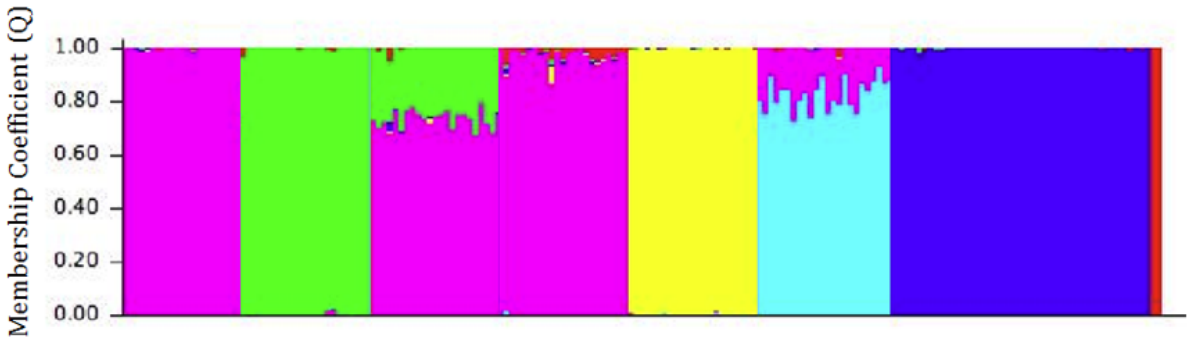


Figure 1.15. In the foreground, the left two rows are plants from IRLO Cycle 61 and the right two rows are S4 plants that were meant to represent IRLO Cycle 61. These IRLO Cycle 61 plants, grown in 2016, were self-pollinated to re-start the creation of a representative inbred for IRLO.



	IHP	IRHP	IRHP2	IRHP3	IRLP	IRLP2	IRHO	IRLO	IHO	B73
% Kernel N										
2015	27.3	6.4	13.7	24.0	-	10.6	11.7	13.5	18.8	
2017	28.8		11.4	16.2	16.7		11.7	10.6	15.9	10.0
% Kernel Oil										
2015	6.5	5.1	4.3	6.0	-	3.6	4.3	4.6	16.5	
2017	6.8	4.3	4.6	5.7	4.0		5.1	4.7	14.0	3.9

Figure 1.16. The analysis from STRUCTURE shows the probability of cluster identity for each individual. Because there was clear separation between the different populations, each population was labeled and the average phenotype for two different years (the year that the samples were taken, and the most recent year) are shown as well.

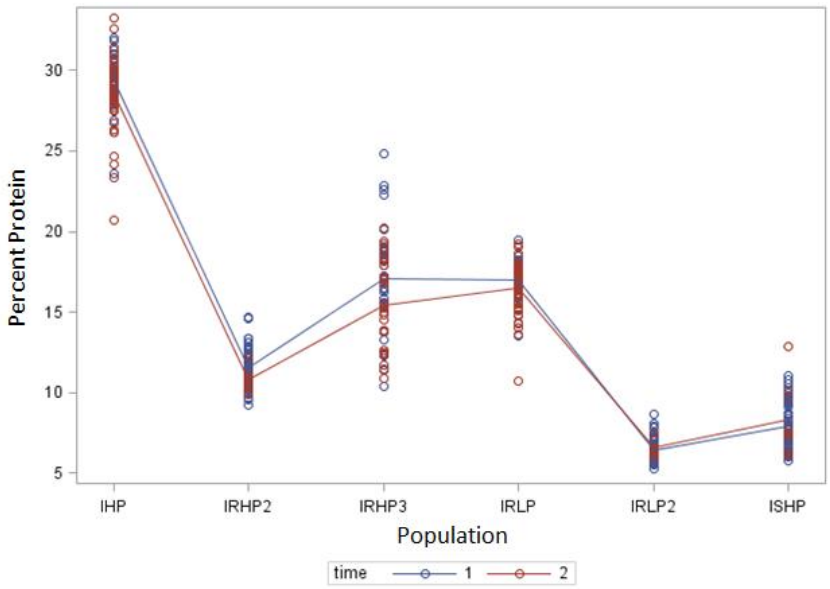


Figure 1.17. The protein concentration for each ear that was pollinated in 2017 is plotted by which population that plant belonged to. The ears that were pollinated earlier are shown in blue and those that were pollinated later are shown in red.

REFERENCES

Below, F.E., J.R. Seebauer, M. Uribelarrea, M.C. Schneerman, S.P. Moose. 2004. Physiological changes accompanying long-term selection for grain protein in maize. Plant Breed Rev. 24: 133–151.

Bhattramakki, D., M.M. Sachs, A.L. Kriz. 1996. Expression of genes encoding globulin and prolamin storage proteins in kernels of Illinois long term chemical selection strains. Crop Sci. 36: 1029–1036.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformstics. 23: 2633-2635.

Dudley, J.W. 2007. From means to QTL: the Illinois long-term selection experiment as a case study in quantitative genetics. Crop Sci. 47. 10.2135/cropsci2007.04.0003IPBS.

Dudley, J.W., D. Clark, T.R. Rocheford, J.R. LeDeaux. 2007. Genetic analysis of corn kernel chemical composition in the random mated 7 generation of the cross of generations 70 of IHP x ILP. Crop Sci. 47: 45–57.

Dudley, J.W., R.J. Lambert. 2004. 100 Generations of selection for oil and protein in corn. Plant Breeding Reviews: Long-Term Selection: Maize, Volume 24, Part 1, 79-110, (ed J. Janick), John Wiley & Sons, Inc., Oxford, UK. doi: 10.1002/9780470650240.ch5

Dudley, J.W., R.J. Lambert, I.A. delaRoche. 1977. Genetic analysis of crosses among corn strains divergently selected for percent oil and protein. Crop Science. 17: 111-117.

East, E.M. 1908. Inbreeding in corn. Rep. Conn. Agric. Exp. Stn. 419-428.

East, E.M. 1910. A Mendelian interpretation of variation that is apparently continuous. The Amer. Naturalist. 44: 65-82.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, S.E. Mitchell. 2011. A robust, simple genotyping-by sequencing (GBS) approach for high diversity species. PLoS One. 6: e19379.

Falush, D., M. Stephens, J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 164: 1567-1587.

Hopkins, C.G. 1899. Improvement in the chemical composition of the corn kernel. Illinois Agricultural Experiment Station Bulletin. 55: 205-240.

Huang, S., A.R. Whitney, Q. Zhou, K.P. Malloy, D.A. Voyles, J. Anthony, et al. 2004. Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. Journal of Agricultural and Food Chemistry. 2004: 1958-1964. DOI: 10.1021/jf0342223

Jones, D.F. 1922. The production of single and double first generation corn hybrids. J. Am. Soc. Agron. 14: 241-252.

Lambert, R.J., D.E. Alexander, I.J. Mehaya. 2004. Single kernel selection for increased grain oil in maize synthetics and high oil hybrid development. Plant Breeding Reviews. 24: 153-175.

Laurie, C.C., S.D. Chasalow, J.R. LeDeaux, R. McCarroll, D. Bush, B. Hauge, et al. 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics. 168: 2141-2155.

Leng, E.R. 1962. Selection reversal in strains of corn previously long-term selected for chemical composition. Crop Science. 2: 167-170.

Lischer, H.E.L. and L. Excoffier. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 28: 298-299.

Lucas, C.J., H. Zhao, M. Schneerman, S.P. Moose. 2013. Genomic changes in response to 110 cycles of selection for seed protein and oil concentration in maize. Seed Genomics, (ed W. Becraft), Wiley-Blackwell, Oxford, UK. doi: 10.1002/9781118525524.ch12

Moose, S.P., J.W. Dudley, T.R. Rocheford. 2004. Maize selection passes the century mark: a unique resource for 21st century genomics. Trends Plant Sci. 9: 358–364.

Pritchard, J.K., M. Stephens, P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 155: 945-959.

Raboy, V., F.E. Below, D.B. Dickinson. 1989. Alteration of maize kernel phytic acid levels by recurrent selection for protein and oil. J. of Heredity. 80: 311-315.

Ramchandran, D., S.P. Moose, K. Low, J. Arp, C.M. Parsons, V. Singh. 2017. Ethanol yields and elevated amino acids in distillers dried grains with solubles from maize grain with higher concentrations of essential amino acids. Industrial Crops and Products. 103: 244-250.

SAS Institute. 2013. The SAS system for Windows. Release 9.4. SAS Inst., Cary, NC.

Song, R., J. Messing. 2002. Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. Plant Physiology. 130: 1626-1635.

Uribelarrea, M., S.J. Crafts-Brandner, F.E. Below. 2009. Physiological N response of field-grown maize hybrids with divergent yield potential and grain protein concentration. Plant and Soil. 316:151.

Winter. 1929. The mean and variability as affected by continuous selection for composition in corn. *Journal of Agronomic Research*. 39: 451-476.

Woodworth, C.M., E.R. Leng, and R.W. Jugenheimer. 1952. Fifty generations of selection for oil and protein in corn. Agronomy Journal. 44, 60-65.

CHAPTER 2

RAPID, NONDESTRUCTIVE METHOD TO QUANTIFY ALPHA-ZEIN ACCUMULATION IN MAIZE KERNELS

ABSTRACT

Kernel composition traits in maize, especially protein accumulation, are of broad interest. The most abundant proteins in maize endosperm, the alpha-zeins, respond positively to soil nitrogen supply and exhibit significant genetic variation. Targeted reductions in alpha-zein accumulation improve the nutritional quality of maize grain and improve nitrogen utilization, but have required expensive and typically destructive phenotyping methods. We showed previously that a Floury2-RFP transgene (FL2-RFP) is a faithful reporter for genetic regulation and developmental timing of alpha-zein accumulation. This study developed a method to quickly document and quantify FL2-RFP accumulation using high-resolution images and image processing software. The FL2-RFP phenotyping system was used to evaluate a population of maize hybrids for zein accumulation, grain protein concentration, and the relationships with these kernel phenotypes to other agronomic traits. Results show that the FL2-RFP reporter phenotype effectively reports the strong maternal control of alpha-zein accumulation and, like grain protein concentration, responds to soil nitrogen supply. Unlike grain protein concentration, FL2-RFP is not negatively correlated with starch concentration in the kernel and FL2-RFP is more sensitive to heterosis, but such that FL2-RFP may be useful to select for low zein while maintaining total grain nitrogen concentration.

INTRODUCTION

Modulating grain protein accumulation in major crops is desirable, especially in maize (*Zea mays*). Maize grain is used in many ways, from industrial starch extraction and ethanol production where it is preferable for as much of the seed as possible to be starch, to those that benefit from a much higher concentration of protein as for human and animal nutrition. In

developing countries that rely heavily on maize for their dietary needs, Kwashiorkor is an illness that predominantly affects children as a result of extreme protein deficiencies (Kismul et al., 2017). Maize grain is relatively low in protein and the desire for higher-yielding varieties perpetuates this problem because there is an inverse relationship between grain yield and protein concentration in the grain (Uribelarrea et al., 2004). This mirrors the condition of Kwashiorkor – the total energy provided is sufficient, but the nitrogen concentration is not high enough in these food stuffs. Of the proteins in the maize kernel, the major contributor to total protein is the zeins, so selecting for increases or decreases in zein will be most effective (Lucas et al., 2013), but these proteins are not easy to directly monitor. Being able to easily assess the protein concentration, and the zein proteins specifically, could increase the selection efficiency for changes in zein accumulation and discoveries of basic regulatory mechanisms that could be applied to other cereals. Here, protein concentration and an efficient way to track a specific zein protein are discussed.

Commonly, the phenotype of seed nitrogen concentration is measured via Near-Infrared Reflectance (NIR). The value obtained from this method is accurate because the proportion of amide bonds, which are not present in other major kernel components, is being reported. Thus, NIR is used as a standard method for estimating grain protein concentration. In unique materials where other seed constituents with amide bonds are present, the seed nitrogen concentration is not exactly protein concentration. Other kernel composition traits have minimal effect on this NIR-derived phenotype, but because this is a concentration, all components measured add to 100%. For example, the case where two grain samples have the same amount of starch but different amounts of protein will result in the grain concentrations and starch concentrations to be different between the samples. Grain protein concentration is a quantitative, broad, and cumulative trait encompassing all proteins in the maize kernel.

There are also many proteins in the maize kernel, thus choosing to follow one protein that is also a major contributor to the phenotype would allow for a more targeted approach to dissecting this trait. The majority of the grain protein in maize is a group of endosperm storage proteins called zeins (Nelson, 1979). Zeins respond positively to nitrogen supply, but breeding for yield has suppressed this response to favor kernel growth with excess nitrogen. An ERA study from DeBruin et al (2017) and insights from the Illinois Low Protein (ILP) germplasm

suggest that this strategy of suppressing the zein response to nitrogen in favor or kernel growth has merit for future maize improvement. There are, however, many zein genes and pseudo-genes in the maize genome and there is high sequence similarity among them (Feng et al., 2009). Zein proteins are alcohol-soluble, proline- and glutamine-rich prolamins, and are composed of several groups, the main one being the alpha-zeins (Song and Messing, 2002). The alpha-zeins constitute almost 50% of the total protein in the kernel (Huang et al., 2004), but one alpha-zein gene in particular, *Floury2*, accounts for a large proportion of the alpha-zein expression in the inbred B73 (Song and Messing, 2003; Feng et al., 2009). In addition to B73, resources with extensive genetic variation for grain protein concentration will be used in this study.

The Illinois Long Term Selection Experiment (ILTSE) is a unique genetic germplasm resource developed at the University of Illinois and maintained by the Moose laboratory, where continuous divergent selection for grain protein concentration has been practiced at the same location since 1896 (Dudley and Lambert, 2004; Moose et al., 2004). The ILTSE was initiated to determine if kernel protein and oil concentrations could be altered by selective breeding (Hopkins, 1899) and has since produced the known phenotypic extremes for grain protein concentration, known as the Illinois High Protein (IHP) and Illinois Low Protein (ILP) populations. Furthermore, the amount of alpha-zeins produced by IHP and ILP are dramatically different when grain proteins were run on an SDS-PAGE gel (Bhattramakki et al., 1996; Below et al., 2004). Reverse selection experiments have also been conducted from IHP and ILP – zein proteins were also targets of the response to reverse selection. SDS-PAGE gels document zein and non-zein proteins of Illinois Reverse High Protein (IRHP) and Illinois Reverse Low Protein (IRLP) – it is the output of zein protein that has most changed in these lines (Lucas, 2014). The ILTSE materials provide an insight into the largely unknown regulation of zein genes.

The amount of the Floury2 protein could be determined via chemistry assays on several kernels from each plant of interest, but a precise, less-expensive, nondestructive, and higher-throughput alternative was sought after. Of the many fluorescent protein reporter lines created by the collaborative efforts of the Jackson laboratory, the Sylvester laboratory, and the Plant Transformation Center at Iowa State University (Mohanty et al., 2009), the *Floury2* gene and its promoter were used to create the reporter line of interest here. The reporter itself is a monomeric Red Fluorescent Protein (RFP) derived from the coral reef *Discosoma* genus (Campbell et al.,

2002). The *RFP* gene was fused to the C-terminus of the 22kD alpha-zein gene *Floury2*. So that the regulation of this fusion protein mirrored that of the *Floury2* gene, the transgenic insertion includes the intact alpha-zein gene, the promoter (approximately 2000 bp of flanking sequence), and about 1000 bp of 3' sequence. To identify a single transgenic event to use, several transgenic events were introgressed into inbred lines derived from the ILTSE (IHP1, ILP1, IRHP1, and IRLP1) and B73 for at least six backcrosses. The FL2-RFP phenotype in these transgenic lines was assessed for zein-like spatial and temporal expression and it was concluded that the *FL2-RFP* transgene is a valuable tool for looking at expression of a single alpha-zein gene, and only materials from one transgenic event were used in further experiments (Lucas et al., 2013). The zein-like localization is further confirmed by images of nuclei prepared from 14-DAP endosperm which show FL2-RFP in the endoplasmic reticulum in protein bodies just like zeins would be (Figure 2.1). The images of DAPI-stained DNA and red fluorescence from the RFP were collected using 3D deconvolution microscopy as separate wavelengths and pseudocolored for image overlay of both channels (DAPI and CY3, Bass et al., 2014). Likely due to the high expression of the relatively stable *Floury2* gene, and the fact that monomeric RFP does not require multimerization to emit fluorescence, the FL2-RFP phenotype can be visualized in white light (Wenck et al., 2003; Campbell et al., 2002). With this reporter, the FL2-RFP phenotype allows us to dissect the maize kernel composition trait by more easily tracking the expression of an individual alpha-zein gene. The methods described here to analyze the FL2-RFP phenotype were developed to allow accurate, inexpensive measures of this reporter phenotype.

The FL2-RFP reporter phenotype was incorporated into many experiments, especially those where there was variation for grain protein. While the FL2-RFP reporter phenotype is useful in many diverse maize genotypes, evidence that the alpha-zeins have changed drastically in the ILTSE (Lucas et al., 2013) has most motivated the development and optimization of another precise phenotype which should provide insights into the control of grain protein.

MATERIALS AND METHODS

Plant Materials

Yet another resource that has come out of the ILTSE is the Illinois Protein Strain Recombinant Inbred Lines (IPSRIs). This genetic resource was created by crossing individuals

from the IHP and ILP populations, both from their respective cycle 70, random-mating those offspring for seven generations (Dudley et al., 2007), and inbreeding for six generations. After two generations of self-pollinating there were 200 plants or families; thus, there is a smaller group of inbred lines that represent the whole IPSRI population (n=500). Both the full set of 500 inbred lines and the core subset of 138 inbred lines used here show a normal distribution and an identical, noticeably large, phenotypic range for seed nitrogen concentration (Lucas et al., 2013). To create hybrids, the IPSRI lines were pollinated by B73: FL2-RFP.

Kernel Microscopy

Three-dimensional images were collected on a DeltaVision deconvolution microscope using a 60X lens and 0.2 µm Z-step optical sections as in Bass et al. (2014). Datasets of DAPI and Red Fluorescence were captured for the entire nuclei and deblurred by deconvolution. Images were adjusted by linear scaling of intensity shown as single wavelength or pseudocolored overlay. DAPI stains all of the chromatin in the nucleus and the bright foci of RFP appear tethered to the isolated nuclei from a 14 days after pollination kernel.

FL2-RFP Phenotyping

A direct, non-destructive imaging method for the quantification of the FL2-RFP phenotype in whole, mature kernels is described here, modified from Lucas (2014). After pollination, the ear on the maize plant is kept out of the sun in a pollinating bag while physiological development progresses. At maturity, the ear is harvested and immediately wrapped in this bag to avoid photo-bleaching. The FL2-RFP phenotype is visualized in just white light, thus any number of image-capturing devices can be used to directly record and digitally preserve the phenotype, and the photographs can be analyzed with computer software. Photographs are taken in a room with no windows of the whole ears and kernels from those ears laid out at the same depth of focus with background lighting controlled for uniformity. Alternatively when a grain yield measure is also taken, the picture of the intact ear is less desired especially since the FL2-RFP phenotype is more quickly, easily, and more accurately documented from shelled kernels laid out embryo-side down. A ruler and color card are included in the photographs to allow color correction and size normalization among batches of photos if needed. After being photographed, the shelled kernels are kept in the dark as well. Although,

once the photographs are taken, the phenotype is preserved digitally which eliminates the opportunity for changes in the phenotype.

To analyze the photographs, a combination of AxioVision (Carl Zeiss, Obercohen, Germany) and Adobe Photoshop (Adobe systems, San Jose, CA) are used to identify the suitable areas for measurement. Here, 'suitable' is defined as free from glare from the camera flash or lighting, free from shadows from other kernels, free of disease on the kernel, and not on the dent portion of the kernel (Woo et al., 2001). The inbred materials grown 2010-2012 were phenotyped while the kernels were mostly still on the ear; however, with the hybrid ears, the FL2-RFP phenotype is exclusively taken from the shelled kernels laid out embryo-side-down. To look into how uniform the phenotype was among kernels on the same ear, Lucas (2014) used a bootstrapping technique to identify that 12 kernels (subsamples) per sample were sufficient – including more kernels only costs one the time to measure/select more kernels (Figure 2.2). The FL2-RFP phenotype is remarkably consistent among kernels on an ear but only within the same class of kernel pigmentation (e.g. yellow versus white endosperm, aleurone anthocyanins); the FL2-RFP phenotype is not robust to the scientist selecting kernels that have other pigmentation differences also segregating on the same ear. For this reason, only one type of kernel was chosen for measurements at a time – this is important to note when any materials with white kernels (like ILTSE materials) are used.

When an area on a kernel is identified, AxioVision collects the color data split into the three channels: red, green, and blue (RGB). An 8-bit photograph could have a range of values for every pixel from 0 – 255, or 2^8 shades of gray from 0 being black to 255 being white. By collecting data in all three channels, each pixel gets a value in the red, green, and blue channel each ranging from 0 – 255. Even with a photograph on this scale, the precision achieved is far beyond what the human eye can detect. Supplemental Figure A.16 shows a diagram of the RGB model and how colors are built using combinations of these three channels using this 0 – 255 scale. However, the FL2-RFP reporter system allows for the photograph to be taken using a 48-bit camera. Thus, the range for each of these channels is 0 – 6,535 (2^{16} shades) instead of 0 – 255. All materials reported here use the high-resolution photographs to collect data on the photographs of the kernels. Wild type maize kernels are yellow and maize kernels with the *FL2-RFP* transgene usually range in color from yellow to red. In figure 2.3, that gross movement is

along the green axis, but there is also movement within the RGB cube necessitating the collection of all three channels and the use of multivariate techniques in analysis of the data. There are many advantages of having three numbers for this phenotype including the numerous RGB combinations used to detect subtle differences in this phenotype.

Another advantage of this reporter system is the robustness of the FL2-RFP phenotype. The source of the *FL2-RFP* transgene can be introduced by pollen to obtain the phenotype of the maternal plant. Utilization of these tools have provided a visualization of a well-documented observation of the inheritance of grain protein concentration (Tsai et al., 1990). The maternal effect describes that the protein concentration in the grain is much more determined by the mother plant than by which genome is provided to the grain by the pollen. Lucas (unpublished data) has shown this with ILTSE materials, and since then many crosses, especially those crosses using IHP1:FL2-RFP and ILP1:FL2-RFP as pollen donors, illustrate this phenomenon – the same FL2-RFP phenotype is seen on a given inbred mother regardless of the pollen source, which is contrasted by the FL2-RFP phenotype when IHP1:FL2-RFP and ILP1:FL2-RFP are the mother plant (Figure 2.3). Even dosage effect from the double-fertilization process that happens in the endosperm does not create significant variability in grain protein (Reggiani et al., 1985), or the reporter FL2-RFP phenotype.

Other Phenotyping

In addition to the FL2-RFP phenotype, measurements were taken on other physiological and agronomic traits of the corn plants. The plant materials were grown in Urbana, IL in years 2010-2014, although not all measures were taken each year. The Illinois Protein Strain Recombinant Inbred (IPSRI) hybrid materials were only measured in 2014; the inbred IPSRI parents had been grown in several years, therefore the inbred phenotypes for inbred grain nitrogen concentration and FL2-RFP phenotype (in the three channels) are Best Linear Unbiased Predictors (BLUPs) calculated with SAS 9.4 Proc Mixed (SAS Institute) using genotype and years as random effects. Only slight modifications were made to common agronomic practices used in central Illinois for corn production, like the fact that post-emergence herbicide could not be used on the 19th century-derived IPSRI materials. Each genotype was grown in a single row plot 5.34m in length in 76.2-cm rows. In plots that had more seeds germinate than the desired

final stand, the inbreds were thinned to 18 plants per plot and the hybrids were thinned to 30 plants per plot. The non-IPSRI hybrid genotypes were grown in a nitrogen response nursery where nitrogen had been monitored for the past 10 years prior to the study. Each genotype was grown in paired high nitrogen and low nitrogen plots. The high nitrogen plots had nitrogen applied by sprinkling a pre-weighed amount of granulated ammonium sulfate at the base of the plants at growth stage V5. Hybrid plants were allowed to open-pollinate to facilitate the measurement of agronomic traits, whereas the recombinant inbred line plants had to be hand-pollinated to provide the FL2-RFP reporter gene from the B73 pollen parent. The total number of nodes, ear node number, and plant height were measured at maturity in the field. Total dry biomass produced was estimated from a plant mass at harvest from which a subsample was taken and moisture determined from the difference in the wet and dry subsample. Total ear mass (including the cob) and grain mass were measured after harvest with a scale and adjusted for any variability in moisture. Grain oil concentration, grain nitrogen (~protein) concentration, and grain starch concentration were all determined with the Perten NIR spectrophotometer (DA 7200) based on a custom calibration to handle the extremes in maize grain composition. This calibration was built using materials from the ILTSE to take advantage of the biological extremes for grain protein observed in this germplasm. Samples were measured on the Perten, and were also ground to a fine powder and nitrogen concentration was analyzed using a combustion technique (EA1112 N-protein analyzer; CE Elantech, Inc., Lakewood, NJ). The FL2-RFP phenotype was also recorded for these plants as values in the red, green, and blue channels as described above.

Statistical Analysis

Various multivariate approaches were implemented for data dimensionality reduction purposes on the 16 original variable and to explore relationships between variables. For the IPSRI hybrid materials, a principal components analysis was conducted using SAS 9.4 Proc Princomp (SAS Institute) to reduce the dimensionality of the dataset and explore how the variance could be best explained. Based on the eigenvalues from the correlation matrix, there were six principal components that individually explained more variance than did any one original variable; thus, those six were kept and cumulatively explain almost 78% of the variance in the original dataset (Supplemental Table A.1).

After reducing the 16 original variables to these six uncorrelated variables (**RFP**, Yield versus Protein, **IPSRI RFP**, **Hybrid Productiveness**, **Protein versus Starch**, and **Hybrid Oil**), a principal components score for each of the principal components was calculated for all the observations. These six variables are what were used as the traits in a genome-wide association study. Leaf samples from all inbreds were collected for DNA extraction in 2015, and the single nucleotide polymorphism (SNP) markers on these materials were obtained following a genotyping-by-sequencing (GBS) pipeline (Elshire et al., 2011). The Buckler Lab did the SNP calling and imputation based on their "production pipeline" – a platform built to utilize the many thousands of samples that they have done, giving them more power to accurately call SNPs to increase the number and diversity of markers that can be used. Genome Association and Prediction Integrated Tool (GAPIT; Lipka et al., 2012) is a publicly available tool used to perform the genome-wide association analysis. SNP markers with a minor allele frequency of less than 0.05 were not used in the analysis. GAPIT also corrected for population structure using up to four principal components, and reports false discovery rate adjusted p-values for the markers used in the analysis in addition to the raw p-values for association between marker and phenotype. For all other specifications, the default settings were used.

For the non-IPSRI hybrid materials, a MANOVA was run using SAS 9.4 Proc GLM. The variables included in the MANOVA were protein concentration of the hybrid grain, the red channel reading of the FL2-RFP phenotype as well as the green channel, and the ratio of the red channel to the green channel. Lucas (2014) reported that the ratio of the red channel to the green channel correlated with protein concentration. For those traits where a significant result was obtained in the MANOVA at alpha=0.05, a follow-up analysis of variance was run on that trait. Microsoft Excel (2007) was used to make the plot to understand the genotype-by-nitrogen availability interaction.

When modeling communality and uniqueness, SAS Proc Factor reported that there was at least one common factor (chi-square = 2265.6, $p < 0.0001$) and that more than four factors were needed (chi-square = 153.8, $p < 0.0001$). I failed to reject the null hypothesis that 5 factors are sufficient (chi-square = 41.0, $p > 0.10$). Kaiser's measure of sampling adequacy is 0.59, which is close to the suggested threshold of 0.6 (Bollero, personal communication), the residual correlations have mostly very small values, the root mean square off-diagonal residuals are less

than 0.05 (0.0197) and the partials are close (0.0577). Based on these diagnostics, the factor analysis is doing a good job of separating out communality and uniqueness.

For the canonical correlation analyses, SAS Proc Cancorr was utilized. For the first canonical correlation analysis, I used the red, green, and blue channels of the FL2-RFP phenotype as the v variables and the other hybrid phenotypes as the w variables. For the second canonical correlation analysis, the inbred protein concentration and stand were the v variables and all the hybrid phenotypes were the w variables.

RESULTS AND DISCUSSION

Response to Nitrogen of the FL2-RFP Reporter Phenotype

Maize grain nitrogen concentration increases when nitrogen fertilizer levels are increased in the field (Figure 2.4; Uribelarrea et al., 2009). Grain nitrogen concentration and the amount of the alpha-zein reporter, FL2-RFP, in maize kernels are two different phenotypes, but respond in a similar fashion to nitrogen application in the field. To illustrate this relationship, several different hybrids were grown in 2012 and 2013. Kernels from these ears were analyzed by NIR for grain quality traits and were photographed so that the FL2-RFP phenotype could be quantified. A MANOVA returned a significant result, so each individual trait was also used in an analysis of variance. There is a significant interaction between genotype and nitrogen availability in these materials (Red channel, Green channel, Red/Green, and grain nitrogen concentration $F=4.6, 8.8, 11.3, 5.1$, $p=0.0004$, and then <0.0001 for others, respectively); the main driver of this interaction being the fact that hybrids with ILP as a parent do not respond to nitrogen application with respect to these phenotypes (Figure 2.4). It is well-documented that grain protein is different for these genotypes and at differing nitrogen application levels (Uribelarrea et al., 2007), and the FL2-RFP reporter phenotype also shows these trends. Interactions with year, and the year effect itself, were not significant. Previous studies in the Moose laboratory have suggested that the nutritional status of the mother plant is the most important factor in seed nitrogen concentration in the grain. Because these plants were also being grown for other purposes, they were allowed to open-pollinate. The fact that these results are seen even with

open-pollinated ears is further support for this maternal effect theory. For the FL2-RFP trait, the pollen that was providing the *FL2-RFP* reporter gene could have been any number of genotypes, including IHP1:FL2-RFP, B73:FL2-RFP, or ILP1:FL2-RFP. The FL2-RFP reporter phenotype allows the maternal effect to be visualized (Figure 2.3). Further, despite FL2-RFP kernels potentially all having different pollen sources and genotypes, there was no visually detectable variation for the intensity of the FL2-RFP phenotype within a single ear.

The FL2-RFP Phenotype in a Population with Unparalleled Variation for Grain Nitrogen Concentration

Since corn production in the Midwest uses hybrids and because the inbred IPSRI plants have been described previously (Lucas, 2014), we wanted to see how this material with unparalleled variation for nitrogen utilization performs as a hybrid. The IPSRIs were crossed by the maize inbred line B73 with the introgressed FL2-RFP reporter transgene. How these materials respond as hybrids, describing the relationship between inbred protein concentration and hybrid phenotypes, and how this highlights the interplay between nitrogen concentration in the seed and how many seeds can be made were all questions of interest. The next question, and of particular interest in this study, is the novel phenotype of the alpha-zein reporter, FL2-RFP, and how that relates to the grain nitrogen concentration trait, as well as other agronomic traits.

A principal components analysis was conducted to determine if a smaller set of uncorrelated variables exists that will explain the relationships among the original variables, how many of these variables there are, and what they represent. Due to the variances for the variables having different magnitudes, the correlation matrix is used. The data were pruned such that there were no missing data points, giving N = 131. Looking at the Correlation Matrix (Supplemental Figure A.17), we can see very strong correlation between the FL2-RFP phenotypes as well as between total ear weight and total grain weight. The grain quality traits are also interrelated, each being negatively correlated with the others. The IPSRI protein concentration (inbred protein concentration) is positively correlated with its hybrid's protein concentration and negatively correlated with hybrid grain yield, hybrid starch concentration, and the green channel of the hybrid FL2-RFP phenotype. The magnitude of the correlation between grain nitrogen concentration and the FL2-RFP phenotype (intensity of alpha-zein-RFP) shows some

relationships, but the lack of a strong correlation is not surprising since they are different phenotypes.

The loadings for each variable within each eigenvector are shown and highlighted to illustrate the interpretation of the contrasts within each eigenvector (Figure 2.5). The first principal component is the **FL2-RFP phenotype** of the grain as those loadings are very high compared to anything else within that eigenvector. Principal component two is contributed to by many of the variables, but could be summarized as yield versus protein as this is most strongly total ear weight and total grain weight, supplemented slightly by the nodes and slightly contrasted by the inbred protein concentration. Principal component three is the **IPSRI FL2-RFP versus hybrid nodes** as this is mostly the green and blue channels of the inbred FL2-RFP phenotype and some hybrid grain protein contrasted by the total nodes, ear node, and height of the hybrid. Principal component four is the **Productiveness of the hybrid** as this is a contrast between hybrid grain and biomass production versus total number of nodes (and the inbred Green channel). Principal component five is the **Protein versus Starch concentration** in the hybrid grain as this is mostly the protein concentration contrasted by starch concentration. There is a well-described trade-off between these two components of the maize grain. Finally, principal component six is the **hybrid oil concentration**.

These inbred lines were created from crossing two populations that showed extreme grain protein concentration phenotypes. Thus, it makes intuitive sense that the hybrids would also vary for protein concentration and that trait is a contributor to many of the principal components. Further, although the loading is small in comparison to the FL2-RFP phenotype, hybrid and inbred grain protein work in the same direction as each other in the first principal component. It is interesting to see that the novel phenotype of quantifying FL2-RFP is the largest part of principal component one – this first principal component explains over 20% of the variability in the dataset. Finally, since we are talking about hybrid materials, it is not surprising to see hybrid performance (in terms of grain production) as the second principal component, especially as somewhat of a function of inbred protein concentration. The two parents used to create this population were not only at different extremes for protein concentration, but also for agronomic desirability – ILP produces a decent amount of grain and combines well (Dudley et al., 1977)

and IHP produces very few, tiny kernels, which are packed with protein, but the yield is far from desirable.

The principal components score for each of the six uncorrelated variables (**FL2-RFP**, Yield versus Protein, **IPSRI RFP**, **Hybrid Productiveness**, **Protein versus Starch**, and **Hybrid Oil**) were used as the traits in a genome-wide association analysis. The Manhattan plots (Supplemental Figure A.18) show the same result as sorting the list of associated markers by false-discovery rate (FDR) adjusted p-values. There is not a significant association between marker and phenotype that passes the 0.05 threshold for FDR-adjusted p-values, but future studies could investigate the genomic regions identified here. From visually scanning the plots, each trait has different regions that might pass as being signal peaking above the background noise. This is perhaps not a surprise for principal component scores that partition phenotypic variation into independent components. Developing a selection index on the principal component scores from this many variables may result in some competing genetic architecture for simultaneously increasing yield and decreasing FL2-RFP intensity. More straightforward techniques could be used to decrease alpha-zeins while maintaining agronomically-desirable traits.

The regulation of alpha-zeins is attributable to many factors, many of them not cis sequences (Miclaus et al., 2011). Further, Dudley (2007) estimated that up to 50 genomic regions were responsible for protein concentration, down from an earlier estimate of 122 using classical quantitative genetic procedures. Using the IPSRI materials, a major-effect genomic region probably should not have been detected. If a cis element was a major contributor to variation in grain protein concentration, 70 years of divergent selection (when the IPSRIs were made) would have likely already fixed these alleles and a lack of further progress would have been documented. This is not the case as after 100 years of selection, sufficient genetic variability is still present for progress to be observed in the ILTSE (Dudley and Lambert, 2004). Further, not getting a major effect from the genomic region of introgression of the transgenic marker makes a stronger argument for this phenotype reporting kernel biology and not an artifact of where it was integrated into the genome (for discussion on the potential genomic region of the transgenic marker, see Chapter 3).

Regardless of the specific number of expected associations, the relatively low statistical power for this experiment may preclude detecting small-effect associations. This system does, however, demonstrate the utility of the reporter FL2-RFP phenotype in explaining variance in a system with unparalleled variation for kernel protein accumulation. Principal components analysis tries to explain as much of the total variation as possible, but the eigenvalues could be inflated with error variance. Further, if there is interest in how the variance is partitioned between common variance and unique variance, a factor analysis can be run. Based on the diagnostics and chi-square significance tests described in the materials and methods, five factors are used and their rotated factor pattern is shown in Supplementary Table A.2. At first glance these resemble what was represented by the principal components, but the loadings for each of the factors is much more targeted to one or two variables. For example factor one is a combination of the three RGB channels and very little of any of the other phenotypes. The unique variance contributed by a variable provides insight into variation that none of the other traits measured have explained, and between the three channels of the FL2-RFP phenotype, very little variance was given to uniqueness. The only two variables contributing a large amount of unique variance were the hybrid grain oil concentration and the stover biomass; these traits contribute variation that is not explained by communality. The test of the relationship between the FL2-RFP phenotype and the other hybrid phenotypes concluded that there is no canonical correlations to describe that relationship (even when controlling for inbred protein concentration and stand, Wilks' Lambda = 0.82, $p > 0.399$). The IPSRI materials were developed to vary for protein concentration, and even in the hybrid, the FL2-RFP phenotype explains the largest amount of the variation present in the hybrids.

Finally, the relationship between the inbred protein concentration and the hybrid phenotypes can be described by 2 canonical correlations (Wilks' Lambda = 0.47, $p < 0.0001$; Canonical Correlation 1 = 0.70, approximate F = 6.74, $p < 0.0001$; Canonical Correlation 2 = 0.37, approximate F = 2.43, p=0.018). The first canonical correlation relates inbred protein concentration with a combination of hybrid traits. Lower inbred protein concentrations are correlated with more productive hybrids with low grain nitrogen and high value (less red appearance) of FL2-RFP in the green channel, but lower values in the red channel (Supplemental Figure A.19). The second canonical correlation relates progressively higher inbred protein concentrations with higher cob mass (the contrast between total ear weight and total grain

weight). The first canonical correlation, however, is strong and highlights a contrast between the green and red channels in the FL2-RFP phenotype that is not captured in a principal components analysis, where the first principal component was represented by an average of the three channels. The contrast between the red and green channel was also what Lucas (2014) found to be most correlated with protein concentration just in the IPSRI lines, and also highlights the finding that most of the variation for alpha-zeins is instead described by an average of the red, green, and blue channels. This insight into the makeup of the FL2-RFP phenotype can be utilized differently based on the goals of a breeding program.

This FL2-RFP reporter phenotype could easily be adopted in any breeding program. Because the phenotype is not influenced by the pollen parent, a pollen-donor line could be maintained in the homozygous state for the FL2-RFP marker and crossed onto any genotype of interest. The FL2-RFP pollen could even be mixed in with a desired pollination, the non-FL2-RFP seeds separated as genetically pure still and the FL2-RFP seeds kept for phenotyping. The data from the FL2-RFP phenotype is multivariate in nature, but plant breeders in general take notes on as many traits as possible that matter to the breeding goal. In these instances a selection index can be created to allow the breeder to select for several traits simultaneously using a single value (built from the summation of weighted values multiplied by each phenotype), and the genotypes ranked on this one value. Different from a principal component, this would let the breeder build their index based on their experience and end-goal rather than selecting on the dimension that explains most of the total variation, though the principal components vector could help to inform the creation of the selection index. Higher intensities of the FL2-RFP phenotype is indicative of more alpha-zein proteins which might be of interest for increasing the nitrogen content in the grain; however, for nutritional breeding goals, even with higher nitrogen concentrations, maize is low in some essential amino acids, still necessitating supplemental nutrition. A selection index could be created instead to enable the selection for genotypes with high yield and high seed nitrogen concentration, but low FL2-RFP intensity. Because the correlation between inbred and hybrid was not overwhelmingly strong, the selection would have to be done primarily in the hybrid as is done with current commercial breeding programs.

For applications like ethanol production, the low FL2-RFP intensity extreme could be explored as well, but in this case the seed nitrogen concentration could be low as well. Higher

nitrogen concentration in the grain is not needed in ethanol production or beer-brewing, thus extremely low FL2-RFP intensities could be used as a proxy for higher starch and higher yield potential. This has a potential to be even more efficient than just selecting for the highest grain yield and lowest grain protein because the FL2-RFP phenotype is more targeted than total grain protein which is a massively cumulative trait dependent on all proteins in the kernel and other kernel constituents. The FL2-RFP phenotype, instead, allows a direct route for selecting for or against a specific kernel storage protein that is a main contributor to total kernel protein, but is not significantly correlated with that phenotype allowing an opportunity for taking advantage of possible independent regulation. In conclusion, the alpha-zein RFP reporter system and this inexpensive method could be used widely to quickly quantify kernel biology in many systems and breeding programs.

FIGURES

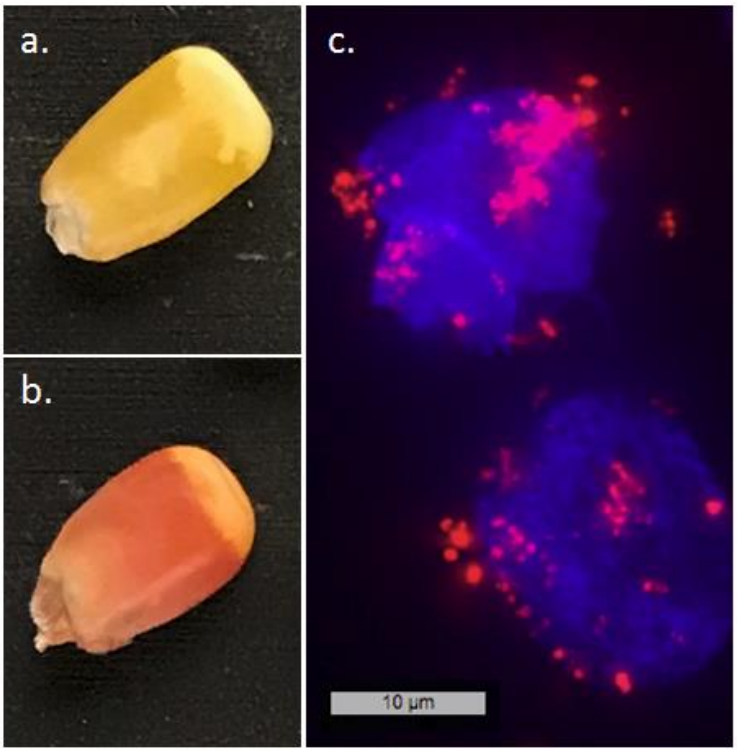


Figure 2.1. From the same ear, (a) a wild-type kernel is compared to (b) a FL2-RFP kernel showing that the FL2-RFP accumulation in maize kernels can be seen in white light. (c) Three-dimensional deconvolution microscopy of 14-days-after-pollination nuclei shows all of the chromatin in the nucleus stained by DAPI and the FL2-RFP fusion protein accumulating as expected of an alpha-zein in protein bodies on the rough endoplasmic reticulum associated with the nuclear envelope. A scale bar (10 microns) is shown for reference.

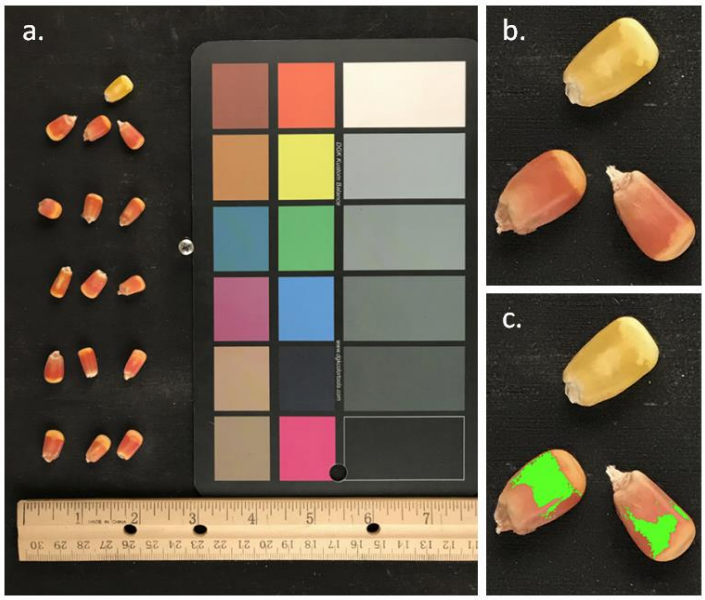


Figure 2.2. (a) The whole image for this hybrid sample. A yellow kernel is included as a control in this photograph. To get precise data, the photograph is expanded such that a single kernel fills the computer screen, though (b) shows an example slightly less zoomed than this; (c) shows the areas that were selected as suitable for data collection (green shading) from these example kernels in this sample.

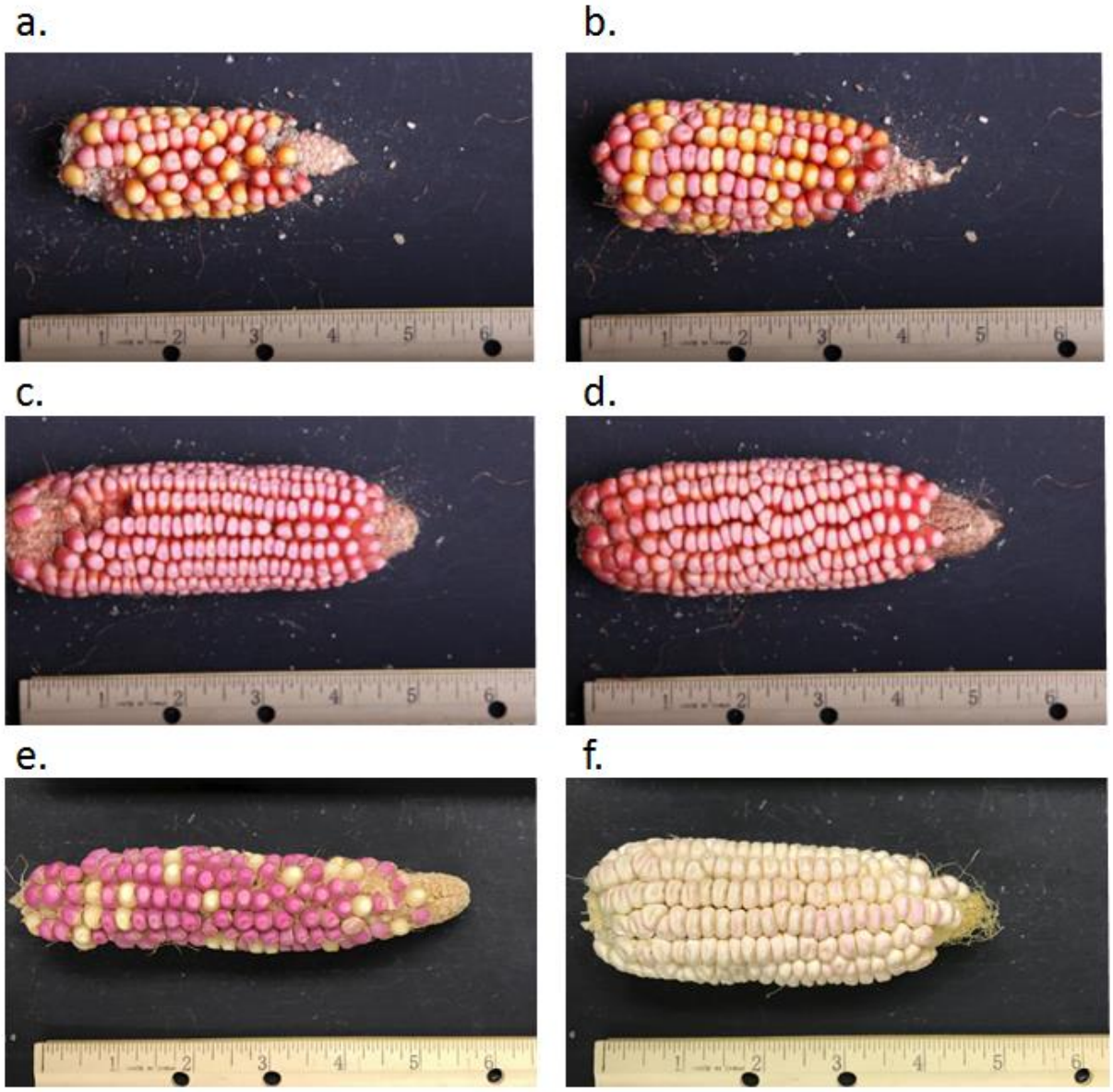


Figure 2.3. The FL2-RFP transgene responds to the maternal effect much in the way that grain protein does – (a) and (b) show ears from the same inbred line (B73) but that were pollenated by (a) IHP:FL2-RFP or (b) ILP:FL2-RFP. (c) and (d) also show ears from the same inbred line (B73: FL2-RFP) that were pollenated by (c) IHP:FL2-RFP or (d) ILP:FL2-RFP. This is contrasted by ears from (e) IHP:FL2-RFP/0 and (f) ILP:FL2-RFP/FL2-RFP which visually show the disparity between these two genotypes as mother plants for the expression and accumulation of FL2-RFP content, which is consistent with the alpha-zein content in these genotypes.

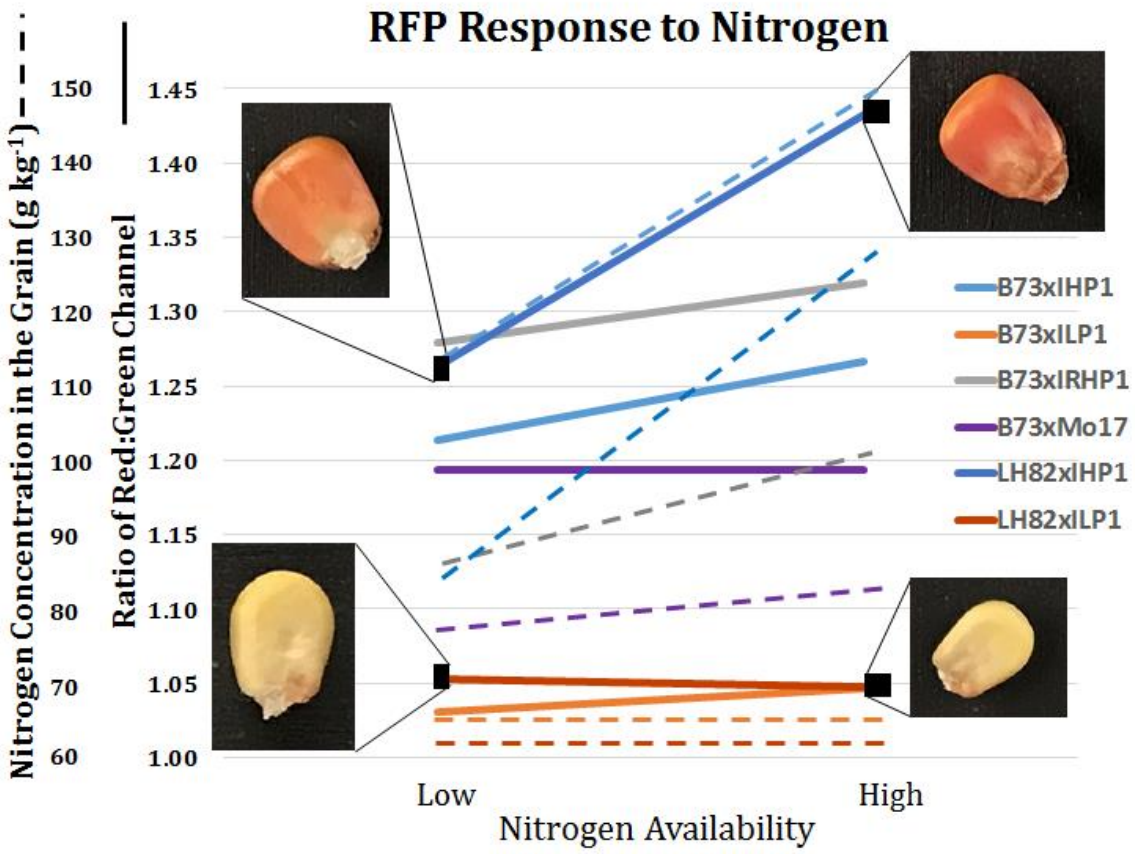


Figure 2.4. The ratio of the red channel to the green channel (solid lines) and the nitrogen concentration in the kernels (dashed lines) are plotted against the nitrogen availability in the field for six hybrids. When IHP1 is a parent (blue lines), there is a response to nitrogen by the FL2-RFP phenotype; however, when ILP1 is a parent (orange lines), there is no response to nitrogen by the FL2-RFP phenotype. Representative kernels are shown for these cases mentioned.

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
Hyrbid Grain Oil (%)	0.15387	-0.0991	0.16705	-0.2252	0.17725	0.76576
Hybrid Grain Protein (%)	-0.262	-0.1893	-0.2854	0.1957	0.41556	-0.2963
Hybrid Grain Starch (%)	0.12011	0.33723	0.11443	0.04148	-0.5121	-0.2343
Hybrid Total Ear Mass (g)	0.21998	0.32676	-0.0926	-0.3928	0.37171	-0.1783
Hybrid Total Grain Mass (g)	0.2173	0.33752	-0.095	-0.3892	0.36395	-0.1738
Hybrid RFP - Red Channel	0.448	-0.2478	0.09463	0.17211	0.10358	-0.1142
Hybrid RFP - Green Channel	0.47268	-0.2348	0.0756	0.14883	0.07765	-0.0772
Hybrid RFP - Blue Channel	0.4556	-0.244	0.05741	0.19219	0.08529	-0.1039
Hybrid Stover Biomass (g)	0.05998	-0.0319	0.05609	-0.3297	-0.17	0.13906
Hybrid Plant Height (cm)	0.03768	0.20539	0.32665	0.04853	-0.0643	-0.0348
Hybrid Total Plant Nodes	-0.0965	0.27861	0.41453	0.32052	0.27554	0.05091
Hybrid Ear Node	-0.1159	0.28315	0.37482	0.31768	0.29694	0.06068
IPSRI Inbred Grain Protein (%)	-0.2726	-0.2748	-0.0048	-0.0092	0.19082	0.10283
IPSRI Inbred RFP - Red	0.19749	0.2453	-0.1858	0.25082	-0.054	0.19615
IPSRI Inbred RFP - Green	0.15729	0.26083	-0.4361	0.30384	-0.0254	0.22874
IPSRI Inbred RFP - Blue	0.01296	0.19793	-0.4417	0.21162	0.04335	0.22562

Figure 2.5. The loadings of each eigenvector from the principal components analysis is shown to illustrate the interpretation of each principal component. Each column is independent of all other columns where the scale is from dark red with strongly negative loadings and dark green for strongly positive loadings.

REFERENCES

Bass, H.W., E.E. Wear, T. Lee, G.G. Hoffman, H.K. Gumber, G.C. Allen, et al. 2014. A maize root tip system to study DNA replication programmes in somatic and endocycling nuclei during plant development. Journal of Exp Bot. 65: 2747-2756.

Below, F.E., J.R. Seebauer, M. Uribelarrea, M.C. Schneerman, S.P. Moose. 2004. Physiological changes accompanying long-term selection for grain protein in maize. Plant Breed Rev. 24: 133–151.

Bhattramakki, D., M.M. Sachs, A.L. Kriz. 1996. Expression of genes encoding globulin and prolamin storage proteins in kernels of Illinois long term chemical selection strains. Crop Sci. 36: 1029–1036.

Campbell, R.E., O. Tour, A.E. Palmer, P.A. Steinbach, G.S. Baird, D.A. Zacharias, et al. 2002. A monomeric red fluorescent protein. Proc Natl Acad Sci U S A. 99: 7877–7882.

DeBruin, J.L., J.R. Schussler, H. Mo, M. Cooper. 2017. Grain yield and nitrogen accumulation in maize hybrids released during 1934 to 2013 in the US Midwest. Crop Science. 57: 1431-1446.

Dudley, J.W. 2007. From means to QTL: the Illinois long-term selection experiment as a case study in quantitative genetics. Crop Sci. 47. 10.2135/cropsci2007.04.0003IPBS.

Dudley, J.W., D. Clark, T.R. Rocheford, J.R. LeDeaux. 2007. Genetic analysis of corn kernel chemical composition in the random mated 7 generation of the cross of generations 70 of IHP x ILP. Crop Sci. 47: 45–57.

Dudley, J.W., R.J. Lambert. 2004. 100 Generations of selection for oil and protein in corn. Plant Breeding Reviews: Long-Term Selection: Maize, Volume 24, Part 1, 79-110, (ed J. Janick), John Wiley & Sons, Inc., Oxford, UK. doi: 10.1002/9780470650240.ch5

Dudley, J.W., R.J. Lambert, I.A. delaRoche. 1977. Genetic analysis of crosses among corn strains divergently selected for percent oil and protein. Crop Science. 17: 111-117.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, S.E. Mitchell. 2011. A robust, simple genotyping-by sequencing (GBS) approach for high diversity species. PLoS One. 6: e19379.

Feng, L.N., J. Zhu, G. Wang, Y.P. Tang, H.J. Chen, W.B. Jin, et al. 2009. Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize alpha-zein super gene family. Plant Mol Biol. 69: 649–659.

Hopkins, C.G. 1899. Improvement in the chemical composition of the corn kernel. Illinois Agricultural Experiment Station Bulletin. 55: 205-240.

Huang, S., A.R. Whitney, Q. Zhou, K.P. Malloy, D.A. Voyles, J. Anthony, et al. 2004. Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. Journal of Agricultural and Food Chemistry. 2004: 1958-1964. DOI: 10.1021/jf0342223

Kismul H., M.A. Mapatano, J.P. Banea. 2017. Diet and Kwashiorkor in the Democratic Republic of Congo. In: Preedy V., Patel V. (eds) Handbook of Famine, Starvation, and Nutrient Deprivation. Springer, Cham. 1-19. DOI: https://doi.org/10.1007/978-3-319-40007-5_38-1

Laurie, C.C., S.D. Chasalow, J.R. LeDeaux, R. McCarroll, D. Bush, B. Hauge, et al. 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics. 168: 2141-2155.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, et al. 2012. GAPIT: genome association and prediction integrated tool, Bioinformatics. 28: 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Lucas, C.J., H. Zhao, M. Schneerman, S.P. Moose. 2013. Genomic changes in response to 110 cycles of selection for seed protein and oil concentration in maize. Seed Genomics, (ed W. Becraft), Wiley-Blackwell, Oxford, UK. doi: 10.1002/9781118525524.ch12

Lucas, C.J. 2014. Genetic analysis of grain protein concentration and related traits in the Illinois protein strain recombinant inbred population of maize. Ph.D. diss., University of Illinois, Urbana-Champaign.

Miclaus, M., J.H. Xu, J. Messing. 2011. Differential gene expression and epiregulation of alpha zein gene copies in maize haplotypes. PLoS Genet. 7: e1002131.

Mohanty, A., A. Luo, S. DeBlasio, X.Y. Ling, Y. Yang, D.E. Tuthill, et al. 2009. Advancing cell biology and functional genomics in maize using fluorescent protein-tagged lines. Plant Physiol. 149: 601–605.

Moose, S.P., J.W. Dudley, T.R. Rocheford. 2004. Maize selection passes the century mark: a unique resource for 21st century genomics. Trends Plant Sci. 9: 358–364.

Nelson, O.E. 1979. Genetic control of polysaccharide and storage protein synthesis in the endosperms of barley, maize, and sorghum. Adv Cereal Sci Technol. 3: 41–71.

Reggiani, R., C. Soave, N. Di Fonzo, E. Gentinetta, F. Salamini. 1985. Factors affecting starch and protein content in developing endosperms of high and low protein strains of maize. Genet Agric. 39: 221–232.

SAS Institute. 2013. The SAS system for Windows. Release 9.4. SAS Inst., Cary, NC.

Song, R., J. Messing. 2002. Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. Plant Physiology. 130: 1626-1635.

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

Tsai, C.L., I. Dweikat, C.Y. Tsai. 1990. Effects of source supply and sink demand on the carbon and nitrogen ratio in maize kernels. Maydica. 35: 391–397.

Uribelarrea, M., S.P. Moose, F.E. Below. 2007. Divergent selection for grain protein affects nitrogen use efficiency in maize hybrids. Field Crops Res. 100: 82–90.

Uribelarrea, M., F. E. Below, S. P. Moose. 2004. Grain composition and productivity of maize hybrids derived from the Illinois protein strains in response to variable nitrogen supply. Crop Sci. 44:1593-1600. doi:10.2135/cropsci2004.1593

Uribelarrea, M., S.J. Crafts-Brandner, F.E. Below. 2009. Physiological N response of field-grown maize hybrids with divergent yield potential and grain protein concentration. Plant and Soil. 316, 151.

Wenck, A., C. Pugieux, M. Turner, M. Dunn, C. Stacy, A. Tiozzo, et al. 2003. Reef-coral proteins as visual, non-destructive reporters for plant transformation. Plant Cell Rep. 22: 244–251.

Woo, Y.M., D.W.N. Hu, B.A. Larkins, R. Jung. 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.

Wu, Y., J. Messing. 2012. RNA interference can rebalance the nitrogen sink of maize seeds without losing hard endosperm. PLOS ONE. 7: https://doi.org/10.1371/journal.pone.0032850

CHAPTER 3

RESPONSE TO SELECTION ON INDUCED VARIATION IN FL2-RFP IN THE INBRED LINE, B73 AND ITS HYBRIDS

ABSTRACT

Many kinds of variation exist, and all are concurrently subjected to selection in a breeding program. In an effort to isolate variation created by the removal of 24-nt sRNAs from maize plants, the mutant *mediation of paramutation1* (*mop1*) was utilized. In homozygous *mop1* plants, the maize genome and its gametes are exposed to a situation where 24-nt sRNAs are not being adequately produced. Then, a wild type allele is reintroduced. We dubbed this treatment 'epimutagenesis' because it produced epigenetic variant lines (epiNILs) with stochastic, pleiotropic phenotypes. To track a known gene, the Floury2-Red Fluorescent Protein (FL2-RFP) reporter phenotype was used, which provides a quick, nondestructive assessment of the accumulation of that alpha-zein in a maize kernel. Variation was documented in the FL2-RFP phenotype of the epiNILs beyond what is seen in the native B73 background. Of the many lines produced, selection for more and less intense FL2-RFP expression identified a few epiNILs to be sampled for genetic comparison to B73 to test the easiest hypothesis of contamination in their creation. Except for two genomic regions around *mop1*and *FL2-RFP*, the epiNILs were largely B73-like. Hybrids were created with these epiNILs to establish a heritability estimate for the FL2-RFP trait in this population. Despite a strong narrow-sense heritability estimate of 0.84 for the FL2-RFP phenotype, the epiNILs did not cluster together when all measured traits of the hybrids were used. Heritable variation independent of DNA sequence change provides another way plant breeders can regulate traits of interest.

INTRODUCTION

It is hard to imagine the phenotypic response to selection by the ILTSE purely in the sense of genetic variation, especially noting that all ears by cycle 17 traced back to one original ear (Winter, 1929). Without specifically intending to target a specific class of protein, much of the selection effected the alpha zeins in the Illinois Long Term Selection Experiment (ILTSE). The majority of the grain protein in maize is this larger family of endosperm storage proteins

called zeins (Nelson, 1979). There are, however, many zein genes and pseudo-genes in the maize genome and there is high sequence similarity among them (Feng et al., 2009). Zein proteins are alcohol-soluble, proline- and glutamine-rich prolamins, and are composed of several groups, the main one being the alpha-zeins (Song and Messing, 2002). The alpha-zeins constitute almost 50% of the total protein in the kernel (Huang et al., 2004), but one alpha-zein gene in particular, *Floury2*, accounts for a large proportion of the alpha-zein expression in the inbred B73 (Song and Messing, 2003; Feng et al., 2009). The *Floury2* gene and its promoter were used to create a reporter line (Mohanty et al., 2009). The reporter gene is the *red fluorescent protein* (*RFP*) which is fused to the tissue-specific promoter for *FLOURY2* (*FL2*) and the *FL2* gene, which encodes an abundant alpha-zein seed storage protein. The *FL2-RFP* transgene is an excellent reporter gene for three reasons: 1) the phenotype (red seed) can be quantitatively measured through imaging techniques; 2) the alpha-zeins are responsive to nitrogen supply; and 3) alpha-zein genes are known to be sensitive to epigenetic regulation (Miclaus et al., 2011).

Even with a targeted reporter gene, teasing apart effects of genetic selection, drift, and epigenetic variation over the past 115+ years especially without seed further back than cycle 65, is a dauntingly complex experiment. Thus, a different system will be utilized to look specifically into the potential epigenetic effect in selection. Advances in sequence data have been a boon for basic biology and also helps elucidate this complimentary theme. This new sequence information has led to amazing advancements; however, the heritability of many phenotypes is not fully explained by genomic DNA sequence, especially for complex traits. For these traits, genome-wide association studies (GWAS) explain only a minority of estimated heritability, which has been described as the "missing heritability problem". The principles of expression and heredity are currently based on knowledge of variation in protein coding genes, and more recently regulatory RNAs. There is growing recognition that epigenetic variation, often driven by non-protein-coding elements, contributes to the inheritance of complex traits (Eichten et al., 2011).

Transposable elements contribute to epigenetic variation by remodeling chromatin, often via small RNAs (sRNAs) and DNA methylation. The discovery of sRNA is somewhat recent and has added another layer to decipher in the transcriptomic analysis (Hamilton and Baulcombe, 1999; Fire et al., 1998). Small RNAs are short, non-coding, regulatory RNAs with a length between 20 and 30 nucleotides (nt) (Li et al., 2012). MicroRNAs (miRNAs) and small

interfering RNAs (siRNAs) are two major types in plants. MicroRNAs are usually 20 – 22 nt small RNAs that target specific messenger RNAs (mRNAs) by sequence pairing for directed transcriptional cleavage or translational repression (Chen, 2010). Post-transcriptional regulation is accomplished by the RNA-induced silencing complex (RISC complex), which directs the complementary binding of the mature miRNA to mRNA transcripts. Known miRNAs play essential roles in plant growth, regulating development, and responses to stress (Lauter et al., 2005; Zhang et al., 2009). Short inhibitory RNAs (siRNAs), on the other hand, are 21-24nt small RNAs that can be generated from transposable elements and tandem repeats in the organism's genome (Meister, 2013), or from exogenous genes (RNA viruses and transgenes; Hamilton et al., 2002).

The great majority of endogenous maize siRNAs are 24-nt siRNAs (Li et al., 2012; Barber et al., 2012) that act mainly to repress transposon activity by DNA methylation, thereby maintaining genome integrity (Lisch, 2012; McCue and Slotkin, 2012). Specifically, DNA methylation is a chromatin modification that changes the accessibility of transcription machinery to segments of DNA and leads to the suppression of their respective expression (Eichten et al., 2013; Napoli et al., 1990). Thus, there can be a change in gene expression – and a change in phenotype – without a change in DNA sequence! Many of our analyses and technologies take advantage of our ability to make the genome two-dimensional – DNA sequence, one letter after another, in a particular order. In a cell, though, the genetic material occurs in three-dimensions, associated with proteins, the chromatin tightly packaged into nucleosomes. This extra dimension may offer some information to help fill in the "missing heritability" problem, too.

To test and explore the epigenetic regulation and underlying genetic architecture of 51 maize inbreds lines or diverse genotypes, Eichten et al. (2013) used a genome-wide methylated DNA immunoprecipitated-chip (meDIP-chip) analysis to discover and locate nearly 2000 common differentially methylated regions (DMRs) and almost as many rare DMRs distributed throughout the genome. Maybe not surprisingly, there is much interplay among all mechanisms in the cell – genetic and epigenetic effects do not occur entirely independently. About half of the DMRs in this study were found to be associated with DNA sequence variation. As it is relatively inexpensive to collect dense single nucleotide polymorphism (SNP) datasets compared to DNA methylation profiles, it is important to understand what can be inferred about DNA methylation

variation from the genotype information. For the other half that are not closely associated with SNP variation, the importance of this third dimension highlights the positioning of proteins associated with the DNA which have a substantial regulatory function. Each nucleosome consists of a histone protein octamer as a core around which 146 bp of DNA is wrapped (Luger et al., 1997). Between each nucleosome is a short segment of exposed DNA called linker DNA. The availability of the DNA to function as binding sites *in vivo* for transcription factors and transcription machinery is dependent on the positioning of these nucleosomes. Open or accessible segments of the genome are most commonly targeted by regulatory elements, and have been characterized by nuclease hypersensitivity *in vivo* (Gross and Garrard, 1988). Repeat-associated siRNAs (rasiRNAs) and trans-acting siRNAs (tasiRNAs) have also been characterized, and have functions similar to the rest of the small RNAs: mediating post-transcriptional regulation, RNA-directed DNA methylation, and chromatin remodeling (Smith and Weigel, 2012; Chen, 2010). In maize, the effect of sRNAs is shown to be different in inbreds versus the effect seen in hybrids (Barber et al., 2012).

As recent as the discovery of sRNAs is, hybrid vigor has been documented for quite some time and yet there is no unifying theme for the mechanism for this vigor. Charles Darwin observed and tested this observation over 150 years ago, describing the trends of inbreeding depression and heterosis (1876). The phenomenon of hybrid vigor was tested in maize and applied to plant breeding strategies starting near the beginning of the 20th century by George Shull (1908). He and Edward East (1908) described inbreeding depression in maize, but this reduced fitness was a small obstacle to maneuver for the superior growth and grain yield of the resulting hybrid plants. Their work was the beginning of how corn would be grown, and significant increases in yields have been seen since hybrids have been implemented in corn production. Because of the importance of inbreds and hybrids in the commercial production of corn, the current study includes both inbred and hybrid plants.

Though not unique in this feature, the *mediator of paramutation* (*mop1*) mutation manifests very differently with regard to the phenotype of the inbred and hybrid. The *mop1-1* mutation has been described previously and a brief discussion is included here. As the name might suggest, the *mop1-1* mutant was first described as being required for the phenomenon called paramutation (Dorweiler et al., 2000). Paramutation is an epigenetic interaction between

two alleles that heritably changes the other allele, and plants homozygous for *mop1-1* inhibit the establishment of paramutation at several loci (2000). These homozygous *mop1-1* plants were also noted to have stochastic, pleiotropic phenotypes (and usually developmental in nature) relative to wild-type siblings, specifically noted were stunting and delayed flowering (2000). This recessive mutation results in a dysfunctional RNA-dependent RNA Polymerase 2 (RDR2) which the *Mop1* gene encodes (Alleman et al., 2006). In the biogenesis of 24-nt sRNAs, *Mop1* builds a second strand on the precursor RNA molecule (Figure 3.1), and thus without a functional *Mop1* allele there is a global decrease of 24-nt sRNAs (Nobuta et al., 2008; Jia et al., 2009). These molecules are associated with maintaining the silencing of transposons through the RNA-directed DNA methylation pathway which targets chromatin (Alleman et al., 2006). Labonne et al., (2013) showed there are changes in nucleosome position at transcriptional start sites of specific genes in *mop1-1* plants. With the DNA methylation activity lost in *mop1* plants, RNA-directed DNA methylation which usually enforces the boundaries between heterochromatin and euchromatin may allow open chromatin to spread to adjacent transposons (Li et al., 2015). It is through this mechanism by which we propose epialleles could be created.

Aware of the speculated regulatory functions of sRNAs, Barber et al. (2012) set out to test the effect of removing 24-nt siRNAs globally in maize inbreds and hybrids. To do this, the *mop1-1* loss-of-function allele was introgressed into the inbred lines B73 and Mo17. B73 and Mo17 are both widely studied as inbred models and their hybrid had tremendous commercial success in the 1970s. The loss-of-function *mop1* led to a global decrease of 24-nt siRNAs, but no change in the abundant 21- and 22-nt siRNAs in the inbreds or the hybrids shown by a Northern Blot Analysis. This had drastic consequences for the inbreds – the mutant plants are stunted, have less biomass, and the cob weight is half that of the wild type inbred plants. When the *mop1* mutant inbreds are crossed (B73: *mop1/mop1* x Mo17: *mop1/mop1*) to produce a hybrid, however, the mutant hybrid mirrors its wild type counterpart – the hybrids are phenotypically less affected by the missing 24-nt sRNAs (2012).

This surprising observation raises the possibility that maize breeders have selected for variation in 24-nt sRNAs that contribute to improved performance of inbred lines, and that the same variation in 24-nt sRNAs perhaps offers no advantage in hybrids. Inbred maize is after all an artificial, man-made scenario – maize is an out-crossing species and would only self-pollinate

as a last resort to eke out a few seeds. Over the course of the last 100 years, maize inbreds have been created that can produce as much grain (on a land-area basis) as early hybrids did. The B73 inbred is one of these usually prolific maize inbred lines, but when harboring the *mop1* mutation in homozygous condition, a myriad of developmental misregulations appear (Dorweiler et al., 2000). These surprising observations and the previous studies of *mop1* prompted the question of whether this pathway could be utilized to create heritable variation. The first step to address this question was to develop a population of epigenetic variants, a set of epiNILs (Cortijo et al., 2014). This study will address the question of how variation created by a loss of regulatory molecules, formerly coded for by *Mop1*, may contribute to three economically important traits in maize: composition of the grain, nutrient utilization, and hybrid vigor.

MATERIALS AND METHODS

Plant Materials

The population of plants in this study was created using the B73 inbred background, which has a publically-available genome sequence. As documented by Barber et al. (2012), the *mop1* allele was received from Jane Dorweiler. The *mop1* allele was then backcrossed into the B73 background; during this process, other mutations (*b1*, etc.) were segregated out such that the stock used in this study was B73: *mop1*.

In addition to harboring the *mop1* mutation, the B73 inbred line has been crossed to a near-isogenic B73 inbred line containing a reporter gene for a specific alpha-zein grain protein. The *Floury2* (*FL2*) gene and its promoter were used to create a fluorescent fusion protein construct with a monomeric red fluorescent protein (RFP) which did not require multimerization to emit fluorescence (Mohanty et al., 2009). The resulting transgenic reporter line shows a phenotype that could be visualized in white light and was backcrossed into B73. The B73: FL2-RFP line that was used in this study had undergone at least seven backcrosses.

Epimutagenesis – Creating Phenotypic Variation in a Genetically Uniform Population

The epiNIL population was created to address the research hypothesis that we can create and select for variation that is not due to DNA sequence. To generate epigenetic variants, the maize mutant *mediator of paramutation1* (*mop1*) was utilized. As described above, the

hypothesis is that *without* the wildtype level of 24-nt sRNAs to maintain transposon silencing, normally-silenced sections of the genome can be transcribed and these molecules change the three-dimensional configuration of the genome. When a wildtype allele of *Mop1* is then re-introduced, these new conformations of the genome will be enforced and stably inherited in their new state. In addition to initially harboring the *mop1* mutation, the B73 inbred line containing a reporter gene for a prominent alpha-zein, the same class of protein that showed response to selection in the ILTSE, was introduced to allow for tracking a very specific phenotype.

Originally, the idea was to generate homozygous 'epi-alleles' using the *mop1*/*mop1* background, but the homozygous *mop1* plants have markedly reduced fitness, so much so that they are practically sterile (Moose, Unpublished data). Different inbred backgrounds respond slightly differently, but at an advanced inbred generation, the homozygous *mop1* plants make only about 10-15 seeds on a good ear (Barber et al., 2012). Also, these plants have a reduced tassel, if any tassel at all, so even use as a male parent is limited. Thus, the combination of hoping to get a functional/fertile tassel that actually has anthers *and* an earshoot – at all, but hopefully one – with silks *and* a plant big enough to support filling a kernel is a rare combination in homozygous *mop1* plants. It is for these reasons that one homozygous *mop1* plant was crossed to B73 in 2011 to allow a bigger population to be grown (Figure 3.2). The whole population is derived from this one B73 ear with one pollen donor homozygous for *mop1*.

These resulting seeds were grown in 2012 and were pollinated by B73 with the FL2-RFP transgene introgressed into it; the B73:FL2-RFP plants have very stable RFP expression and have continued to be grown with the treatment plants to establish the range of phenotypic values without selection and to illustrate the variation from the environment (Figure 3.2). The B73: *mop1; FL2-RFP* and B73: *FL2-RFP* inbred populations were grown in a uniformly low-nitrogen nursery in all years to accentuate any variation for nitrogen utilization. Of the ears created in 2012, 20 FL2-RFP seeds from each of 100 ears were planted in rows in the field in 2013. All plants in this population were genotyped to determine if they had the mop1 mutation or not (see Genotyping), and all plants were self-pollinated (where possible). Ten (10) seeds from each of the 234 ears that had been identified as having a *mop1* allele in 2013 were planted in one row each in summer 2014 – there was no selection on whatever variation there was in the FL2-RFP phenotype except maybe to ensure the whole spectrum observed was planted. Every individual in

this *mop1*-exposed group was genotyped for the presence of the mutant allele. Additionally, 10 seeds from each of the 164 ears that did not show positive signs of the *mop1* allele from 2013 were grown as a control population of B73 that had been treated the same as those that had the *mop1* allele. DNA from all plants in a row were pooled to confirm the 2013 result showing a lack of a mutant allele. Finally, 2 rows were dedicated to the B73: FL2-RFP ear that had been used to generate the epiNILs. From this population in 2014, at least two self-pollinations were attempted in every row. From these ears produced, the ears with the lowest FL2-RFP intensity, the ears with the highest FL2-RFP intensity, and a subset from the middle were selected from the *mop1*-exposed group. A random sample of ears from the control group were selected, and all of them had moderate FL2-RFP intensity.

The selected ears were grown in their own row in 2015. All of these 90 rows, plus two samples of B73, were sampled for genotyping-by-sequencing (GBS). At least one ear in each row was self-pollinated. From the *mop1*-exposed group, fifteen rows of the 'lowest FL2-RFP intensity' (555, 556, 557, 575, 576, 581, 597, 600, 601, 602, 618, 622, 626, 627, 629), fifteen rows of the 'highest FL2-RFP intensity' (565, 566, 567, 569, 570, 572, 573, 585, 588, 589, 592, 593, 606, 610, 613), and four rows of the 'moderate FL2-RFP intensity' (630, 632, 640, 641, 648) were also chosen as parents to create hybrids. Four rows of the control B73: FL2-RFP were also included in the hybrid creation (657, 659, 660, 661). Hybrids were made with inbreds PH207, A632, Mo17, PHZ51, PHG39, LH82, PHJ31, IHP1, ILP1, PHJ33, NC350, LH1, and PHG84. Some hybrid combinations were attempted to be made again in 2016. These selfed ears and the hybrids created were grown in 2016 and 2017. There was no intent to establish which genotype created the best hybrid with B73 – numerous other studies have exhaustively described these crosses. Such that non-heterotic hybrids were not immediately adjacent to heterotic hybrids in the field, the hybrids were blocked by tester parent; thus, no further discussion on the individual tester parents is included.

Genotyping Individuals from the *mop1* Population Development, PCR, and primers

The *mop1-1* loss-of-function allele has been previously described (Alleman et al., 2006). The mutant allele was backcrossed for seven generations into the B73 inbred background and then was self-pollinated for four generations. Heterozygous *mop1-1* B73 mutant plants were

differentiated using the following primers that assay for the presence of the Mutator insertion in exon 4 of Mop1: WT allele, Mop1 forward, TTCGACGAGTTCCTGGACGC, Mop1 reverse, GGGTGGTAGGTCACGTGGTA, expected amplicon size of 290 bp; mutant allele, mop1Mu forward, GCGCCCTGATGACCTACTAC, mop1Mu reverse, TGCGTCTCCAAAACAGAGAA, expected amplicon size of 170 bp (Barber et al., 2012). The creation of the epiNIL population was completed in a monitored low-nitrogen nursery in Urbana, IL (Figure 3.2). All plants were genotyped in 2013, and based on the presence or absence of a mutant *mop1* allele, a different genotyping scheme was utilized in summer 2014. In the rows where the parent was identified to have had a mutant *mop1* allele, each plant was genotyped; whereas, in the rows where no mutant *mop1* allele was previously detected, plant tissue from all ten plants in the row was pooled and tested for the presence of a mutant *mop1* allele. In 2015 and 2016, tissue from all plants in a row was pooled.

Genotyping-By-Sequencing

To test genetic relationships of individuals in a population, a genome-wide molecular marker data set obtained by a technique called Genotyping-By-Sequencing (GBS) was used. GBS is a high-throughput, low-cost technology that provides a large amount of genotypic data across the genome. The whole genome can be sampled by a complexity reduction technique using restriction enzymes and Illumina Sequencing. This breadth of coverage is an advantage over previous techniques like array-based genotyping, which is subject to the ascertainment bias. A major disadvantage of GBS is the large amount of missing data (which is more troublesome for some intended analyses than for others) and uncalled single nucleotide polymorphisms (SNPs) that cannot be used (Elshire et al., 2011). For the epiNIL experiment though, the GBS data are used to compare versions of B73 to the reference B73 (the control) and checking introgression regions, so missing data between individuals and heterozygous regions should not pose problems for the analysis. Fresh tissue was harvested from experimental plants growing in the field in 2015, DNA was extracted, and the samples were plated with other materials from the Moose Lab and sent to Cornell for GBS following the protocol from Elshire et al. (2011). The Buckler Lab did the SNP calling and imputation based on their "production pipeline" – a platform built to utilize the many thousands of samples that they have done, giving them more

power to accurately call SNPs to increase the number and diversity of markers that can be used (Glaubitz et al., 2014).

Field Trials and Phenotyping

The plant materials were grown in Urbana, IL in years 2011-2017. The inbred materials were hand-planted in 2011-2015. The number of seeds planted and the length of the row changed each year, but the rows were always spaced 76.2 cm apart with intra-row spacing between 15 and 25 cm. The inbred materials were machine planted in 2016 and 2017 in 76.2-cm-spaced rows with intra-row spacing of 20 cm. Germination rate and plot stand were monitored and corrected as needed to maintain a homogeneous experimental space. All plots in the experiment were at least two rows of hybrid border from an irrigation alley or the edge of the field. Irrigation was only applied as needed in 2012 and on 10 June 2017. Supplemental nitrogen was only added in 2015 and 2016 with the goal of creating a homogeneous experimental space for adjacent experiments; in all other years, the plants were grown in a low-nitrogen environment. These plots were also rotated in soybean-corn every other year. The exception to this is in 2017, these materials were grown in a field that had been corn the previous year and did not receive supplemental nitrogen. All phenotyped inbred plants were hand-pollinated. Hybrid materials were machine planted in 2016 and 2017 in 76.2-cm-spaced, 5.33 m rows with intra-row spacing of 10 cm (seeded population of 40,000 plants/acre). Due to anticipated variable viability in the hybrid seed in 2016, a twin row of non-experimental fill hybrid corn was planted two-inches (5 cm) directly to the East of the experimental plot. At the V3 growth stage, all experimental plots were thinned to 34 plants (34,000 plants/acre). If that plot had more than 34 seeds germinate, all plants in the accompanying twin row were removed. For those plots with less than 34 seeds germinating and surviving bird attacks, the twin row fill plants were kept where needed to establish 34 plants per plot between the experimental plants and the fill plants combined. This allowed the within plot competition to be equal among all plots as well as no plot being adjacent to an empty plot. In 2017, any plot with less than 34 plants at growth stage V1 were re-sown with experimental plants where gaps in the rows occurred; all other plots were thinned to 34 plants. All hybrid materials were allowed to open-pollinate to ensure the best, full pollination possible. For an indication of yield, five ears were randomly selected from the middle of each plot at harvest. The ears were weighed on a laboratory scale both before shelling the kernels and

after, the difference inferred to be the weight of the cob. The kernels from these same ears were used in kernel composition and FL2-RFP phenotyping. Kernel composition traits (grain oil concentration, grain nitrogen (~protein) concentration, and grain starch concentration) were all determined with the Perten NIR spectrophotometer (DA 7200) based on a custom calibration to handle the extremes in maize grain composition. Plant height was done in the field directly after pollination, usually early August, with a large ruler and was measured to the tallest point of the plant, usually the top of the tassel. High-wind storms occurred in both 2016 and 2017 at the beginning of September just before harvest, so lodging notes were made on these plots as well. The percentage of plants experiencing each type of lodging within each plot was recorded. The types of lodging were described as root lodging (where the roots were visible due to the stalk being blown sideways), economic or below-ear lodging (where the stalk snapped below the ear node), and cosmetic lodging (where any stalk breakage above the ear node resulted in the tassel not being displayed upright).

FL2-RFP Quantification

The FL2-RFP phenotype is a visual phenotype seen even in just white light, thus this phenotype can be recorded with any number of image-capturing devices and analyzed with computer software. An additional advantage of this phenotype is that it is a non-destructive sampling technique. Photographs taken in 2013, 2014, and 2015 are of the whole ears, and kernels from those ears are laid out at the same depth of focus. These materials were inbreds and the area of the kernel proximal to the cap could be seen and reliable data could be obtained from the side of the kernels while still on the ear. In a few cases where this was not the case, a few kernels were shelled and placed in the same depth of focus next to the ear. In 2016 and 2017, all ears were shelled and photographed as described in Chapter 2. To analyze the photographs, a combination of AxioVision (Carl Zeiss, Obercohen, Germany) and Adobe Photoshop (Adobe systems, San Jose, CA) are used to identify the suitable areas for measurement. Here, 'suitable' is defined as free from glare from the camera flash or lighting, free from shadows from other kernels, free of disease on the kernel, and not on the dent portion of the kernel (Woo et al., 2001). Especially with hybrid ears, the FL2-RFP phenotype is exclusively taken from the shelled kernels laid out embryo-side-down to avoid any measure being taken from the cap of the kernel. If 12-to-15 suitable areas were not identified, data collection was stopped with fewer subsamples

measured rather than collecting data from unsuitable areas to get a larger number of subsamples. All inbred kernels in this experiment were consistent for yellow color; the hybrid kernels, however, had occasional white kernels. Only yellow kernels with the FL2-RFP reporter were measured in this experiment.

Statistical Analysis

To perform the principal components of the RGB data for the FL2-RFP phenotype, SAS Proc Princomp was used on the correlation matrix. Eigenvalues less than one were not considered, and principal component scores were calculated for each individual enabling each individual to have one variable describing most of the total variation in the dataset. Correlation matrices were also produced with Proc Princomp and color was added to aide in visualization in Microsoft Excel. SAS Proc Corr was used to describe the linear correlation between the principal component scores for each individual and the visual rating given to those ears. SAS Proc Univariate was used to conduct the tests of normality, and the Shapiro-Wilks test is reported. SAS Proc GLM was used to conduct the Brown-Forsythe non-parametric test of the variances as well as provide visual representations of the data to supplement what was obtained by SAS Proc Boxplot. To compare medians, SAS Proc npar1way was used, and the Wilcoxon test of two independent groups is reported, and was not different from any other comparable test like the Kruskal-Wallis, but the latter is generally used with more than two groups. Tests between the means of two groups were conducted using SAS Proc Ttest and Satterthwaite-adjusted statistics are reported where the variances between the two groups were unequal. To obtain standardized values for a variable, SAS Proc Standard was used.

Custom scripts written and executed in RStudio (CRAN, R Development Core Team) were used to plot values with differing symbols and color in x- and y-space. To analyze the GBS data, TASSEL (Bradbury et al., 2007) was used to visualize the data and save the dataset as a hapmap file. Microsoft Notepad was used to eliminate all parts of the text file before the genetic data and headers. RStudio was used to prune the dataset to only those sites that had homozygous SNP calls for both of the B73 control samples. The genome was then divided into 50 SNP windows with a 25 SNP overlap, and each epiNIL sample was compared to the B73 samples. The percent of SNPs in each window that did not match the B73 samples were saved. The size of

the symbol plotted was increased proportionally for how many of the epiNIL samples also had that percent of non-B73 SNPs in a given window.

These 37 epiNILs were designated as one group and the 14 tester inbreds were designated as the other group in the factorial mating design. This design allows the decomposition of the genotypic effect of which hybrid into general combining ability and specific combining ability. Another goal of this design is to establish estimates for the variance components, enabling the estimation of the heritability of measured traits. Some caution is taken to ensure a representative group of tester inbred parents are used for those traits of interest – since selection in the epiNILs was for the intensity of FL2-RFP, the known extremes were included (IHP1 and ILP1) as well as many other inbreds in between. These genotypes must also be considered random effects, which is indeed the case – the epiNILs were a subset of the possible lines that could have been used and the tester inbreds are representatives of Midwestern germplasm from the last 50 years. The model used was:

$$y_{ijkl} = \mu + Y_i + GCA_j + YG_{ij} + GCA_k + YG_{ik} + SCA_{jk} + YS_{ijk} + \varepsilon_{ijkl}$$

where y_{ijkl} is the observed phenotypes of the hybrid,

μ is the grand population mean,

Y_i is the random effect of i^{th} year, NID(0, σ_Y^2)

GCA_j is the random effect of the general combining ability of the j^{th} inbred parent, NID(0, σ_{GCA}^2)

YG_{ij} is the random interaction between year and the general combining ability of the first parent, NID(0, σ_{YGCA}^2)

GCA_k is the random effect of the general combining ability of the k^{th} inbred parent, NID(0, σ_{GCA}^2)

YG_{ik} is the random interaction between year and the general combining ability of the second parent, NID(0, σ_{YGCA}^2)

SCA_{jk} is the random effect of the specific combining ability of the j^{th} and k^{th} inbred parents, NID(0, σ_{SCA}^2)

YS_{ijk} is the random interaction between year and the specific combining ability of the parents used in a hybrid cross, NID(0, σ^2_{YS})

ε_{ijkl} is the random error term, NID(0, σ^2_e)

This model was analyzed in SAS Proc Mixed. The residual (restricted) maximum likelihood method was used to be able to obtain the most realistic estimates of variance components. Solutions were obtained to enable the calculation of BLUPs for each hybrid for the three channels of the FL2-RFP phenotype. The BLUPs for each hybrid were then used in a principal components analysis in SAS Proc Princomp as has been done for these highly correlated channels. Explaining 91% of the total variance present, one principal component was kept, and principal component scores were calculated for each hybrid plot grown from the loadings on the three channels (red = 0.566, green = 0.603, blue = 0.562). The principal component scores for the hybrids were plotted against the inbred intensity of FL2-RFP (green channel average of the two years the inbreds and hybrids were grown) in RStudio plot command.

Now that each plot has one value, a univariate analysis was completed again using the model described directly above for this one new variable (principal component one scores) which explains most of the variation present. Variance components were obtained from this analysis and were used to estimate the narrow-sense heritability (h^2) and broad-sense heritability (H^2) for the composite trait of FL2-RFP with the following equations:

$$h^2 = \frac{\sigma^2_{GCAepi} + \sigma^2_{GCAtester}}{\sigma^2_{GCAepi} + \sigma^2_{GCAtester} + \sigma^2_{SCA} + \frac{2\sigma^2_{YG}}{years} + \frac{\sigma^2_{YS}}{years} + \frac{\sigma^2_e}{years * reps}}$$

$$H^2 = \frac{\sigma^2_{GCAepi} + \sigma^2_{GCAtester} + \sigma^2_{SCA}}{\sigma^2_{GCAepi} + \sigma^2_{GCAtester} + \sigma^2_{SCA} + \frac{2\sigma^2_{YG}}{years} + \frac{\sigma^2_{YS}}{years} + \frac{\sigma^2_e}{years * reps}}$$

where the number of years this study was conducted was two. The number of reps was unequally distributed allowing all genotypes to be grown at least once and those with more seeds to increase the size of the error term. Although 240 hybrids were replicated three times, another 200

hybrids were only grown in one plot. A few other hybrids were grown in two reps, or four or more reps, so a conservative estimate is calculated using just two as the number of replications.

The narrow-sense heritability was then used to assess the response to selection. The following equation was used:

$$R = k_p \sqrt{\sigma_p^2} \; h^2$$

where R is the response to selection, k_p is the standardized selection differential, and h^2 is the narrow-sense heritability of the FL2-RFP trait in this population. The hypothetical proportion of individuals selected was based on a normal distribution and calculated in Microsoft Excel; the plot was also made in Microsoft Excel.

To look at the relation of the individual epiNILs, all data collected on the hybrids was clustered using Ward's Method in SAS Proc Cluster. The residuals from a MANOVA model with just the year, tester parent, and their interaction were clustered the same way. Finally, a MANOVA model with year, tester parent, epiNIL parent, and all interactions was used. All terms in the model were deemed significant, and the individual tests of each term were also inspected on all variables.

RESULTS AND DISCUSSION

Characterization of the epiNILs Population

The epiNIL population was created to address the research hypothesis that we can generate and select for variation that is not due to DNA sequence. The first part of that is to assess whether reductions in 24-nt sRNAs conditioned by *mop1* leads to increased phenotypic variation in an otherwise homogeneous inbred genetic background. A homozygous *mop1* plant produced enough pollen in 2011 to pollinate a wild type B73 plant. The resulting plants from this ear were all heterozygous for *mop1* and pollinated by B73: FL2-RFP in 2012. Thus in 2013, the plants were either heterozygous for the *mop1* allele (*mop1* exposed) or wild type, serving as controls that went through the same backcrossing scheme but were not exposed to the mutant *mop1* allele. In 2013, 1277 ears were created by self-pollination and were harvested. All of these

ears were inspected for phenotypes and given a visual rating of the FL2-RFP intensity. A subsample of those ears were chosen at random to be planted in the following year; some of these (n=98) were analyzed with Axiovision to explore the FL2-RFP phenotype further. The red, blue, and green channels were highly correlated, so principal components analysis was used. The first principal component (λ=2.77) was found to explain 92.5% of the total variation in this dataset and was built essentially as an average of the three channels (red = 0.565, green = 0.586, blue = 0.581), thus this one variable will be used instead of the raw data from all three color channels. The principal component score of each individual was significantly linearly correlated to the visual score for those ears (ρ = -0.59, $p < 0.0001$). This is a strong correlation especially considering that the researcher only felt comfortable reliably discriminating these shades of red visually into 8 distinct groups whereas the computer software can place each channel into a range of 6,536 shades of color, which is many more combinations of color and this creates an essentially continuous distribution. With only 8 distinct groups, it is expected that the distribution of residuals are not normally distributed (Shapiro-Wilk W = 0.94, $p < 0.0001$), and the variances of the two groups (those individuals that were identified to be heterozygous for the *mop1* mutant allele and those that were wild type) were not statistically different (Brown-Forsythe F = 0.33, $p > 0.56$). Despite both groups having a median of 3.0 for the visual rating, they were deemed to have come from different populations (Wilcoxon, $p < 0.0001$; Figure 3.3). With the visual ratings, 3.0 was meant to represent what the phenotype of what the control B73: FL2-RFP usually is. In this population in 2013, the *mop1* heterozygotes tended to deviate slightly more than the control population, and many more had a weaker FL2-RFP intensity than was seen in the control population.

In the summer of 2014, 10 seeds from each of the 234 ears that had been identified as having a *mop1* allele were planted in one row each – there was no selection on whatever variation there was in the FL2-RFP phenotype. Additionally, 10 seeds from each of 164 ears that did not show positive signs of the *mop1* allele were grown as a control population of B73 that had been treated the same as those that had the *mop1* allele. The null hypothesis was that the two populations would not be different because, in essence, they were both B73 crossed by B73.

The question that prompted the study was: can variation be created using the system described here? From a very qualitative measure of visual phenotype frequency in one group

versus the other, the answer was yes, epimutagenesis created variation. Supplemental Figures A.20–A.33 document some of the many visual phenotypes seen in this *mop1*-exposed population. As encouraging (and surprising) as it was to see frequencies and diversity of phenotypic variation reminiscent of chemical or radiation mutagenesis, many of the unexpected phenotypes (chlorophyll, dwarfing, hormone signaling, *etc*.) would be very difficult to characterize further by genetics, as many have not been observed previously and their genomic locations are unknown. In contrast, the FL2-RFP transgene is the known target (unexpected) variation in RFP intensity. Due to the diversity of variant phenotypes observed, a quantitative comparison of variant frequencies between the *mop1*-exposed population and the control population was not feasible for each of the numerous phenotypes seen. Thus in 2014, several phenotypes were measured on all plants: height, stem diameter, and FL2-RFP intensity (Table 3.1). In 2014, all harvested ears were the result of a hand-pollination, were inspected for phenotypes, and were given a visual rating of the FL2-RFP intensity. Of the 996 controlled pollinations attempted, nearly a quarter of these (236) were so moldy that they were left in the field, and many of the others had some degree of mold on them (Supplemental Figure A.27). The kernels from 524 ears were analyzed with Axiovision. The red, blue, and green channels were highly correlated, so principal components analysis was used. One principal component (λ=2.52) was found to explain 84.0% of the total variation in this dataset and was built essentially as an average of the three channels (red = 0.564, green = 0.604, blue = 0.564), thus this one variable will be used instead of the raw data from all three color channels. The principal component score of each individual was significantly linearly correlated to the visual score for those ears (ρ = -0.40, $p < 0.0001$). With only 8 distinct groups, it is expected that the distribution of residuals are not normally distributed (Shapiro-Wilk W = 0.94, $p < 0.0001$), and in fact most of the variables were not normally distributed, thus the means are only included for completeness.

The mean and median of the measured traits were comparable between the epiNILs and the B73 control population, but for each phenotype the variance was significantly larger for the epiNIL population. Off-type phenotypes were seen with a higher frequency in the epiNIL population than the B73 control which contributed to the variance being higher for that population. Most of these distributions were not normally distributed. For example Supplemental Figure A.26 shows a plot in the epiNIL population that was much shorter than the rest of the B73 plants grown which certainly skewed the distribution of plant heights. The visual ratings of the

FL2-RFP phenotype help to illustrate the phenotype in a comprehensible way – higher numbers are given to the darker red kernels, but with the human eye, only a small number of discrete classes could be reliably scored. Although potentially more difficult to comprehend, the FL2-RFP phenotype in three channels provides essentially a continuous variable as the computer software can establish many thousands more shades of red than the human eye. The negative correlation of the visual rating with the principal components score is based in the makeup of colors in the RGB model – less on the green axis results in a darker red color, as well as less on all axes get the color closer to black. The principal components score then moves from almost yellow (i.e. the least red) as the highest value to the darkest red being the lowest value. The epiNILs population had a long tail in the distribution for the principal components scores in the high direction (Figure 3.4), and this more variable population (Table 3.1) had more of these individuals where the FL2-RFP intensity was much lower than is seen in the control population; however, this variation for more intense FL2-RFP expression is not seen. This prompted me to investigate the heritability and response to further selection of these extreme FL2-RFP phenotypes.

The self-pollinated ears from 2014 were grown for three more years to track the FL2-RFP phenotype. From the ears in 2014, the 29 ears with the most intense FL2-RFP phenotype, the 46 ears with the least intense FL2-RFP phenotype, 27 ears with a medium FL2-RFP intensity, and 5 ears from the B73: FL2-RFP rows were all selected to be grown in 2015. As in 2014, the ears from the epiNILs had an unusually high incidence of mold – 64% of pollinations made in 2015 were left in the field due to the ear being completely consumed by mold (Supplemental Figure A.25). From this group grown in 2015, 14 of the "high" FL2-RFP, 15 of the "low" FL2-RFP, 4 "control" FL2-RFP, and 4 B73: FL2-RFP were grown again in 2016 and the self-pollinated ears were grown in 2017, also. The RGB channels of the FL2-RFP phenotype correlate with each other strongly within a year across all genotypes, but the correlation across years and across genotypes is weaker (Figure 3.5). A principal components analysis is a common way to represent highly correlated variables in a more succinct way by reducing the number of variables in the analysis. Previous principal components analyses with the RGB data have reported one principal component explaining between 80 and 93 percent of the variation in the dataset (Figure 2.3, Figure 2.4), this first principal component being essentially an average of all three channels in all cases. Because the goal was to track the computer-analyzed FL2-RFP phenotype over years,

each year was kept separate and the average of the standardized data are plotted in each year. Figure 3.6 shows that the average of the three channels does not completely separate out the selected groups from the unselected B73: FL2-RFP. The variation does stretch in the appropriate direction for the yellow symbols (less FL2-RFP intensity) and dark red symbols (more FL2-RFP intensity) despite the selection not being on this averaged standardized value – the selection was for just the absolute least red and the most red ears. To explore the variation in B73 itself, 37 B73 plants were pollinated by the 37 epiNILs that had shown divergence for FL2-RFP intensity in 2016. These B73 ears were also analyzed with AxioVision software and were shown to not be different from the average of all 37 self-pollinated epiNILs grown that year ($t=0.06$, $p>0.95$). The Satterthwaite adjustment was used for the comparison of the means in SAS proc ttest because the variance of the B73 ears – despite being pollinated by father plants that, as mothers, produced very different FL2-RFP intensities – was significantly lower than the variance seen in the epiNILs (28.6 and 156, respectively; $F = 5.45$, $p < 0.0001$). This low variance in B73 demonstrates how consistent the phenotype is despite never having selected for that trait, specifically. As was also demonstrated in Figure 2.3 and 2.4, the pollen source does not significantly affect the FL2-RFP intensity that will be determined by the mother plant bearing the kernels, and even just one copy of the *FL2-RFP* transgene in the endosperm produces a mean intensity not significantly different from similar plants with three copies of the transgene (most of the epiNILs are homozygous for the *FL2-RFP* transgene). Although a formal experiment has not been conducted to test this, kernels on an ear known to be segregating for kernels with different numbers of copies of the transgene could not visually be distinguished. Even the same ear pollinated with a mix of Illinois High Protein and Illinois Low Protein pollen does not segregate for intensity of FL2-RFP – in fact, the only time segregation in FL2-RFP intensity is seen within an ear is when another mutant is also segregating on that ear as is seen in *o2* (Lucas et al., 2013). Finally, it is apparent that these materials created by removing 24-nt sRNAs temporarily and then re-introducing them have significantly more variability than non-treated B73 plants.

Based on what is known about the *mop1* mutation, the working hypothesis is that *Mop1* maintained transposon silencing via DNA methylation. Being homozygous for *mop1* (having two dysfunctional alleles) drastically reduces the 24-nt sRNAs that are made (Barber et al., 2012) which cannot maintain the silencing of the transposons. Releasing this repression allows

chromatin to rearrange and new transposon patterns to express, and then crossing back in a functional allele of *Mop1* maintains this new pattern. I postulate that this is a function of half of the gametes not having a functional *Mop1* allele – epigenetic regulation is particularly important in the gametic tissues which have a flourish of sRNA production and DNA methylation. While paternally imprinted genes are more common, maternally imprinted genes are known (Liu et al., 2015) and may be especially important for endosperm development and protein concentration.

Although our experiment was designed to promote phenotypic variation via an epigenetic approach, other factors could contribute to the variation observed for FL2-RFP (and other phenotypes). The easiest explanation that should be explored is the hypothesis of impure pollinations. Describing the scenario in this experiment – where homozygous *mop1* plants have poor fertility and might also have some sort of aversion to self-pollinating and therefore, the resulting seeds might be a result of cross-pollinations – sounds eerily similar to the story of the *Arabidopsis* mutant *hothead*. With the case of *hothead*, Lolle *et al.* (2005) described a scenario where epigenetic mechanisms repaired a mutant version of the gene, but Peng et al. (2006) showed that in absolute isolation, this phenomenon never occurred, making a strong case for outcrossing (in a predominantly self-pollinating plant) or seed contamination as the cause for the presence of the wild-type allele in the filial generation. As the *Arabidopsis* plants used in Lolle et al. were in plastic sleeves, my pollinations only saw the inside of a carefully maneuvered pollination bag, but there are other ways to get outside pollen on maize silks (for example, corn rootworm beetles have a strong affinity for silks on late-flowering maize plants and can sometimes get inside the shoot-protective bag, or any exposure to air that is saturated with pollen could result in an outcross). Another potential possibility is that these phenotypes are due to residual genetic variability, including linkage drag around either the *mop1* or the *FL2-RFP* loci.

To test for the presence of foreign genetic material and genetic relationships of individuals in a population, a genome-wide molecular marker data set was obtained using genotyping-by-sequencing (GBS) to assess genetic variation. GBS provides a large amount of genotypic data across the genome which was used to compare different epiNILs, or versions of B73, to one another and to the reference B73 (the control), checking introgression regions and determining how much genotypic variation is present in this population. The null hypothesis here is that B73 was sequenced 92 times. Of particular interest is the size of the introgression

regions around *mop1* and *FL2-RFP*. The population is all derived from one plant, but those regions could have changed between individuals with subsequent meiosis events during generation advancement and selection. When considering which plants to genotype, the trackable phenotype of FL2-RFP was the primary concern - some of the highest FL2-RFP intensity, some of the lowest FL2-RFP intensity, some from the middle, the control B73:FL2-RFP, and B73 were chosen for genotyping. A summary of the genetic analysis is shown in Figures 3.7, 3.8, and 3.9 where deviations from the x-axis indicates non-B73 SNPs and the size of the dots indicate how many of the epiNILs deviate at that location in the genome. The *mop1* mutation has previously been mapped to the short arm of chromosome 2, and an introgression block is seen here (Figure 3.7). The *FL2-RFP* transgene was also not originally in the B73 background, and the only other place in the genome that also looks like this is on the short arm of chromosome 3 (Figure 3.8). The hypothesis is that this is the location of the *FL2-RFP* transgene in the genome. The FL2-RFP phenotype is visual and easily scored on very large numbers of individuals, which may explain why its introgression region is much smaller than the one surrounding the *mop1* location because *mop1* was followed with a molecular assay (Methods) performed on smaller populations. Other small differences between the epiNILs and the B73 control are seen on the other eight chromosomes (Figure 3.9). There are no other regions where all of the SNPs within a window are non-B73-like and no regions where all of the epiNILs move away from the B73 control.

Phenotypes of Hybrids Created from the epiNILs

Through selection on the FL2-RFP phenotype, epiNIL inbred lines were created that deviated from the phenotype commonly seen in the B73: FL2-RFP stock. The question then was: is this variation transferrable to hybrid plants? Siblings from the same 37 epiNILs that were grown in 2013-2017 were used to create hybrids, which included the highest and lowest FL2-RFP intensity in 2014 that were selected and grown in 2015, as well as some from the middle of the FL2-RFP intensity spectrum and B73:FL2-RFP. The epiNILs were crossed by a suite of other inbred lines: PH207, A632, PHG39, PHJ31, LH1, Mo17, PHZ51, LH82, PHJ33, PHG84, NC350, IHP1, and ILP1.There were several reasons for choosing several of these other inbreds. A preliminary clustering of a collection of maize inbred lines representing historical diversity was done based on their profiles of transposon-associated small RNAs (Barber, 2013). Inbreds

were chosen to span the degree of diversity relative to B73. Based on active transposable elements producing sRNAs, PHG39, PHG84, and PHZ51 were the most divergent from B73, the second most diverged, and third most diverged, respectively (of the genotypes which had been sampled for sRNA sequencing, data not shown). NC350, ILP1, LH1, and Mo17 were most like B73 for this sRNA-producing trait (most alike to less alike). IHP1 was also included to complement ILP1 in ensuring a representative distribution including the biological extremes for the main trait of interest, the intensity of FL2-RFP. These genotypes also established a strong genetic effect for high and low grain protein. The other inbreds listed above were included to ensure all three heterotic groups were represented and also to include more recent germplasm. Additionally, the B73 ears that had been pollinated by the epiNILs in 2016 were grown such that B73 could be considered a tester parent if desired – these backcross 'hybrids' did not display common signs of heterosis.

The hybrids were grown in 2016 and 2017. The FL2-RFP phenotype of these hybrid plants was of most interest, and thus this phenotype was analyzed first. A BLUP was calculated for each channel of the FL2-RFP phenotype such that each combination of inbreds grown (each hybrid) had a best estimate of its FL2-RFP phenotype. These BLUPs were then used in a principal component analysis and one principal component was identified which explained 91% of the variation present. The principal component scores of the hybrids' FL2-RFP composite trait were plotted against the epiNIL inbred intensity of the FL2-RFP phenotype in Figure 3.10. The hybrids made with ILP1 (the "+" symbols) show, by far, the least intense FL2-RFP phenotype regardless of which epiNIL parent it was crossed with. The effect of ILP1 is glaringly obvious, but the effect of tester parent is also apparent by the fairly uniform spread of values for each epiNIL hybrid. As was seen in Figure 3.6, the intensity of the FL2-RFP phenotype in the epiNIL inbreds shows a slight relaxation toward the middle of the spectrum from where they were selected.

The loadings from the first principal component of the hybrid FL2-RFP BLUPs were used to calculate a score for each plot in the experiment, also. This response variable was used in a univariate analysis to obtain estimates of the variance components (Table 3.2). The narrow-sense heritability (h^2) of the FL2-RFP composite trait was found to be 0.84, and the broad-sense heritability (H^2) was 0.89. This seems to indicate that the FL2-RFP phenotype can be strongly

described by the additive genetic variance. The gain in heritability observed by adding in the specific combining ability highlights the contribution of dominance variance in this trait, albeit small. The additive variance is a measure of the variation in the effects that are transmitted from one generation to the next; thus, it plays a key role in predicting the change in the population mean due to selection (Figure 3.11). With estimates of h^2 and the selection differential now established, the response to selection was plotted over a range of selection intensities. Not surprisingly, this shows that the more stringent the selection index is the stronger the response to selection. Additionally, the magnitude of response is quite substantial, which is largely a function of the high heritability estimates.

There were more traits measured on these hybrid materials than just the FL2-RFP phenotype. When all sixteen of the measured variables were analyzed with a principal components analysis, three principal components explained 67.3% of the variation in the dataset. The first principal component was largely a function of how heterotic the hybrid was, evidenced by the eigenvector loadings strongly favoring grain and total ear mass and height and slightly contrasted by all three of the FL2-RFP channels. The linear combination describing principal component two was largely fiber, the green channel, and the red channel of the FL2-RFP phenotype contrasted by density and slightly contrasted by protein. Principal component three also included the FL2-RFP phenotype as the blue and green channels and protein were contrasted by starch concentration. The principal components analysis was blind to the treatment variables, but mirrors what was found with a MANOVA analysis – the effect of tester parent had a much larger F-value ($F=43.09$, $p<0.0001$) than did the epiNIL parent ($F=3.19$, $p<0.0001$). The FL2-RFP phenotype was what was selected on in the epiNILs, so seeing it make up principal component two and three shows that variation was present, and the ANOVA of each trait (except for the three lodging traits) supported the conclusion that not all epiNIL parents were equal. Each epiNIL did not responded to hybridization with each tester parent in a uniform way, however. The residuals were obtained from a MANOVA model that accounted for the variation due to year, the tester parent, and their interaction. The question then was would the residuals cluster by epiNIL or at least by the direction of selection for the epiNILs, but there was no clear pattern by which the individuals were assigned to a cluster in the context of treatment groups. Although the epiNILs were created by selection on only the FL2-RFP phenotype, variation for other traits measured was present but did not help to describe patterns in the epiNILs themselves.

In addition to phenotypes that arose in the epiNIL inbreds, there was a phenotype that had not been seen in any of the epiNIL inbred lines, but showed up in several of the epiNIL hybrids. This phenotype manifests as yellow streaks in the adult leaves, but nothing in the juvenile leaves. This phenotype showed up in seven different hybrids in 2016 and in both reps of two of those (out of 271 hybrids that were grown, 234 had a second replicate planted). In most plots, two or three out of 40 plants showed this phenotype, but in one row, half of the plants had this phenotype (Figure 3.12). These off-type plants were not different in any other way from their siblings despite not having as much chlorophyll-containing area. Two of these hybrids had enough seed to be planted again in 2017, and the yellow streaks phenotype was seen in a couple plants in the expected rows again.

Heritable changes in gene expression have resulted in a wide array of variation. In addition to developmental off-types, variation was documented in the FL2-RFP phenotype, a known, quantitative trait. It is enticing and exciting to think about improving germplasm and making gains toward breeding goals without changing the DNA sequence, especially with the recent pressure from the media (in the US as well as abroad) creating an unfavorable situation for genetically modified crops. The understanding of epigenetic mechanisms and the underlying genomic landscape may lead many breeding programs to more aggressively tap into this source of previously unassessed heritable variation. The method of unleashing the epigenetic variation should be less harsh and easier to implement than the *mop1* system described here, but just as the accumulation of alpha-zeins were altered in the B73 inbred line, a leading inbred line in commercial production may possess untapped variation to make it even better.

TABLES AND FIGURES

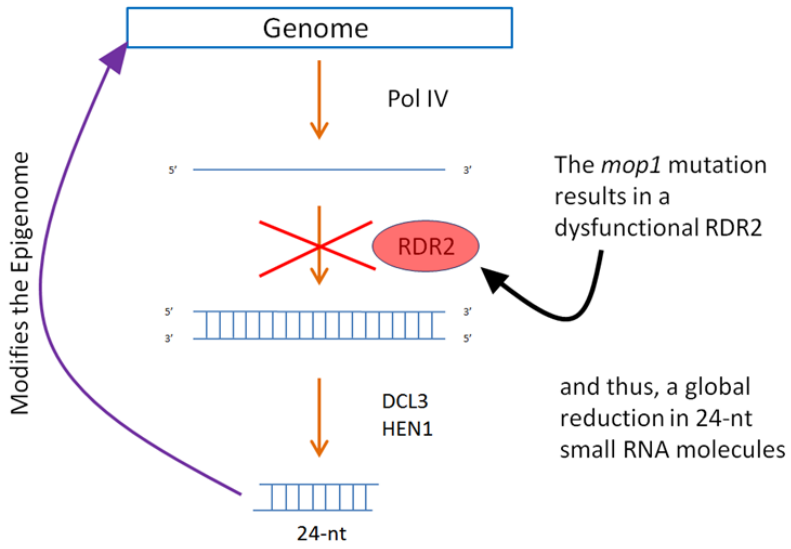


Figure 3.1. The maize mutant *mop1* was utilized to create epigenetic variants for the experiment. RDR2 is the *Arabidopsis* homolog of *Mop1*, which plays a fundamental role in the processing of sRNA molecules. Pol IV stands for RNA Polymerase IV. DCL3 stands for dicer-like3, and HEN1 stands for HUA1 Enhancer1; these molecules are involved in the resulting 24-nt sRNA product and its stability.

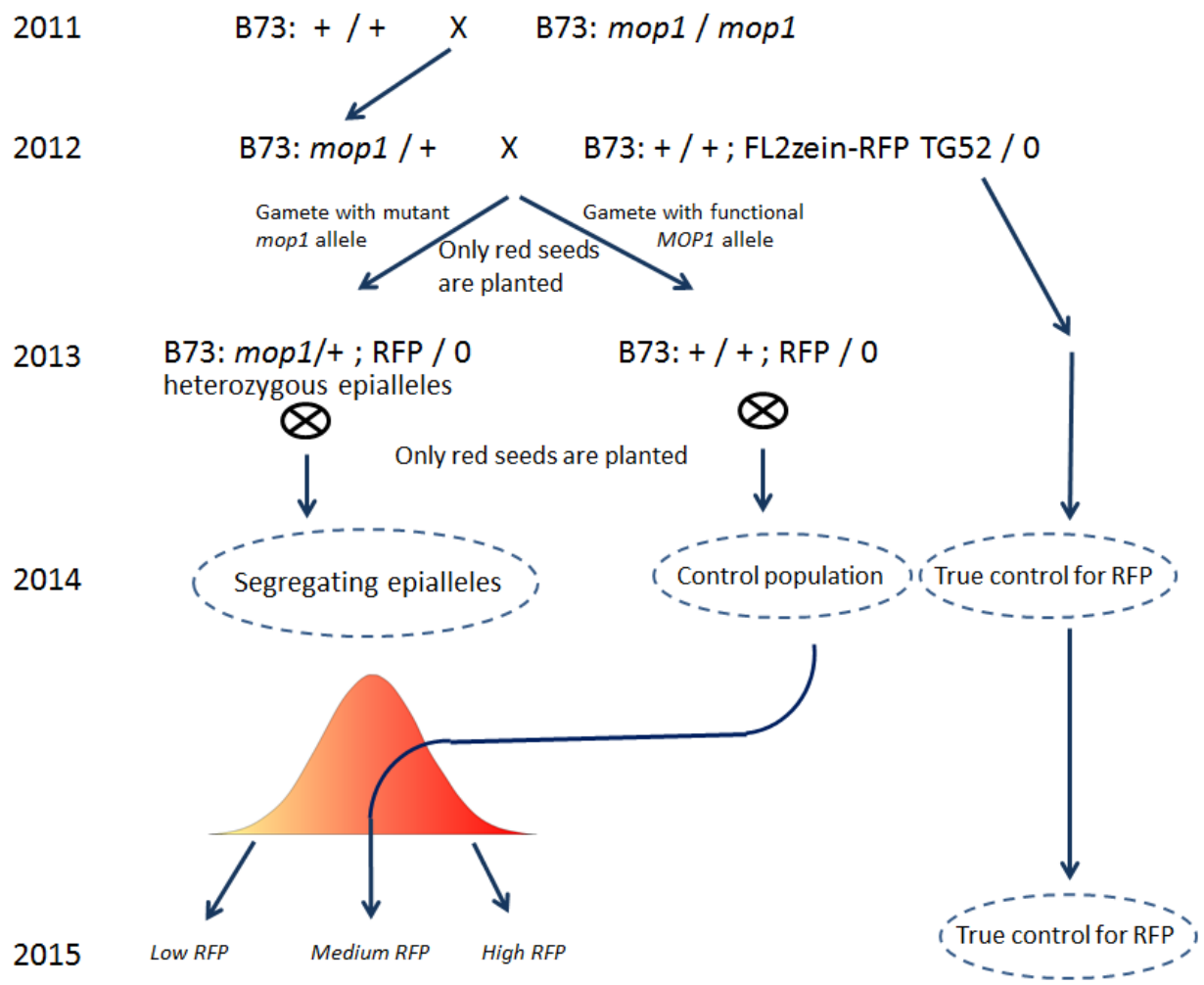


Figure 3.2. The breeding scheme used in the creation of the B73 epiNIL population. The epi-alleles most likely are being generated in the B73:*mop1* parent, as well as in the F1. Importantly, the FL2-RFP reporter was not directly exposed to the *mop1* homozygous status and so phenotypic variation is more likely to reflect *trans*-acting rather than *cis*-acting regulation.

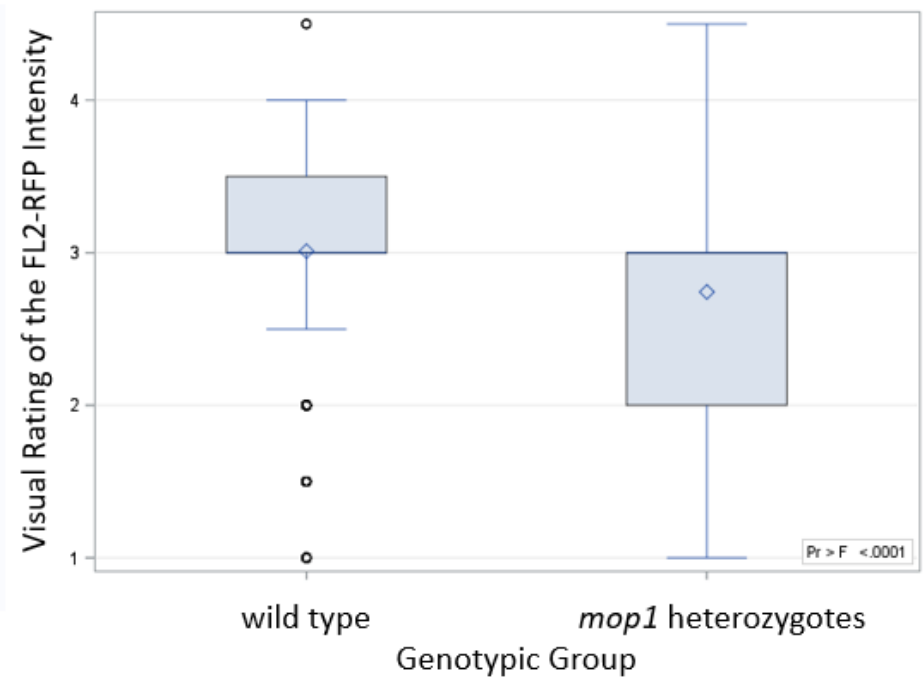


Figure 3.3. The distribution of the visual ratings of the FL2-RFP intensity is shown for the two groups of individuals segregating in the field in 2013 (*mop1* heterozygotes n = 411, B73 Control n = 739). The median and the first quartile are both 3.0 for the wild type plants. The median and third quartile are both 3.0 for the *mop1* heterozygote ears.

Table 3.1. Increased variation in the epiNIL population grown in 2014. All plants grown were measured for height and stem diameter (epiNIL n = 2172, B73 Control n = 1554), the visual FL2-RFP score, ranging from 1 – 5, is reported for all controlled pollinations made (epiNIL n = 468, B73 Control n = 270), and was used for pilot study data for the FL2-RFP phenotype. A high-definition picture was taken of each ear produced and the FL2-RFP intensity was measured using AxioVision software allowing a detailed analysis of FL2-RFP accumulation in the kernels. The principal components score for the FL2-RFP phenotype is from as many ears as could be measured (epiNILs n = 319, B73 Control n = 205). Significance is reported here with the following symbols: ^{t} represents $p<0.05$, * $p<0.01$, ** $p<0.001$, *** $p<0.0001$ The means are reported for completeness. Most populations were not normally distributed, thus the test of the medians used SAS Proc npar1way and the Brown-Forsythe test was used for variances.

	Group	Mean	Median	Variance
FL2-RFP (visual score 1-5)	EpiNILs	2.81	3.00	0.43***
	Control	2.87	3.00	0.23
FL2-RFP (PC1 Score from RGB)	EpiNILs	0.26***	– 0.086***	3.14***
	Control	– 0.45	– 0.50	1.07
Height (cm)	EpiNILs	207.7***	215.0	990.6***
	Control	212.5	214.0	262.5
Stem Diameter (mm)	EpiNILs	16.6	17.0^{t}	7.45***
	Control	16.6	17.0	4.87
Germ. Rate (% germination)	EpiNILs	92.8	100.0	174.0
	Controls	93.1	100.0	153.5

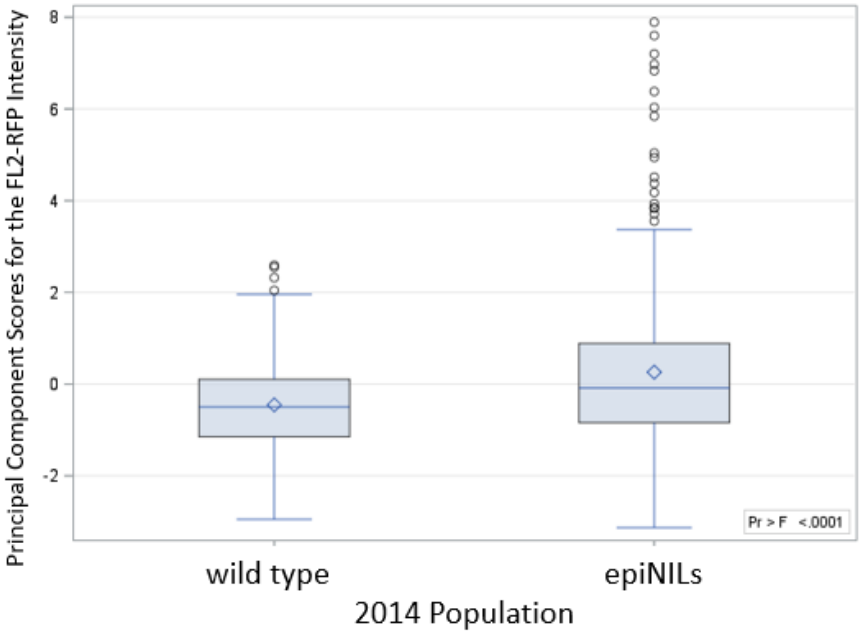


Figure 3.4. The distribution of the principal component scores of the FL2-RFP intensity is shown for the two groups of individuals grown in the field in 2014 (epiNILs n = 319, B73 Control n = 205). The epiNILs population shows a long tail for the less intense FL2-RFP phenotype.

	Blue Channel 2017	Green Channel 2017	Red Channel 2017	Blue Channel 2016	Green Channel 2016	Red Channel 2016	Blue Channel 2015	Green Channel 2015	Red Channel 2015	Blue Channel 2014	Green Channel 2014	Red Channel 2014	Blue Channel 2013	Green Channel 2013	Red Channel 2013
Blue Channel 2017	1	0.87602	0.90495	0.44375	0.52368	0.31611	0.35801	0.32802	0.23896	-0.02233	0.24512	0.14433	0.13364	0.0739	0.19287
Green Channel 2017	0.87602	1	0.89167	0.36321	0.5587	0.26579	0.34016	0.43528	0.23853	0.20228	0.57687	0.37355	0.27727	0.21781	0.46563
Red Channel 2017	0.90495	0.89167	1	0.38826	0.51907	0.25084	0.26203	0.30287	0.05457	0.07856	0.34847	0.13997	0.30263	0.20371	0.43526
Blue Channel 2016	0.44375	0.36321	0.38826	1	0.90562	0.83497	0.31283	0.22917	0.27	0.10101	0.03389	-0.08477	0.20609	0.13217	0.25889
Green Channel 2016	0.52368	0.5587	0.51907	0.90562	1	0.83564	0.30046	0.2851	0.22925	0.39301	0.4314	0.09231	0.52858	0.44454	0.59246
Red Channel 2016	0.31611	0.26579	0.25084	0.83497	0.83564	1	0.34267	0.29665	0.14964	0.26453	0.1068	-0.08135	0.2648	0.17874	0.15492
Blue Channel 2015	0.35801	0.34016	0.26203	0.31283	0.30046	0.34267	1	0.84078	0.5904	0.19288	0.08231	-0.16655	0.27826	0.28116	0.24134
Green Channel 2015	0.32802	0.43528	0.30287	0.22917	0.2851	0.29665	0.84078	1	0.66564	0.17176	0.19915	-0.10782	0.29255	0.31671	0.2957
Red Channel 2015	0.23896	0.23853	0.05457	0.27	0.22925	0.14964	0.5904	0.66564	1	-0.04215	-0.0625	-0.13878	-0.10008	0.00281	-0.07444
Blue Channel 2014	-0.02233	0.20228	0.07856	0.10101	0.39301	0.26453	0.19288	0.17176	-0.04215	1	0.72814	0.41596	0.37154	0.26578	0.23503
Green Channel 2014	0.24512	0.57687	0.34847	0.03389	0.4314	0.1068	0.08231	0.19915	-0.0625	0.72814	1	0.70728	0.6124	0.50335	0.62966
Red Channel 2014	0.14433	0.37355	0.13997	-0.08477	0.09231	-0.08135	-0.16655	-0.10782	-0.13878	0.41596	0.70728	1	0.08745	-0.01543	0.08027
Blue Channel 2013	0.13364	0.27727	0.30263	0.20609	0.52858	0.2648	0.27826	0.29255	-0.10008	0.37154	0.6124	0.08745	1	0.98182	0.93646
Green Channel 2013	0.0739	0.21781	0.20371	0.13217	0.44454	0.17874	0.28116	0.31671	0.00281	0.26578	0.50335	-0.01543	0.98182	1	0.9154
Red Channel 2013	0.19287	0.46563	0.43526	0.25889	0.59246	0.15492	0.24134	0.2957	-0.07444	0.23503	0.62966	0.08027	0.93646	0.9154	1

Figure 3.5. For the 37 epiNIL families that were grown each year from 2013 to 2017, the correlations between each channel in the RGB data are colored from negative correlations being progressively darker red and more positive correlations being progressively darker green.

Figure 3.6. The standardized data were averaged across the three channels to allow a visualization of the one dimension that explains the most variation in the dataset and this intensity of the FL2-RFP phenotype is plotted along the y-axis where higher numbers refer to weaker red kernels on an ear. Each epiNIL line is given its own combination of color and symbol and is plotted for which year it was grown. The yellow symbols are families that were selected for a less intense FL2-RFP, the dark red symbols for families selected for a more intense FL2-RFP phenotype, and the red symbols were for the families that had a moderate FL2-RFP intensity.

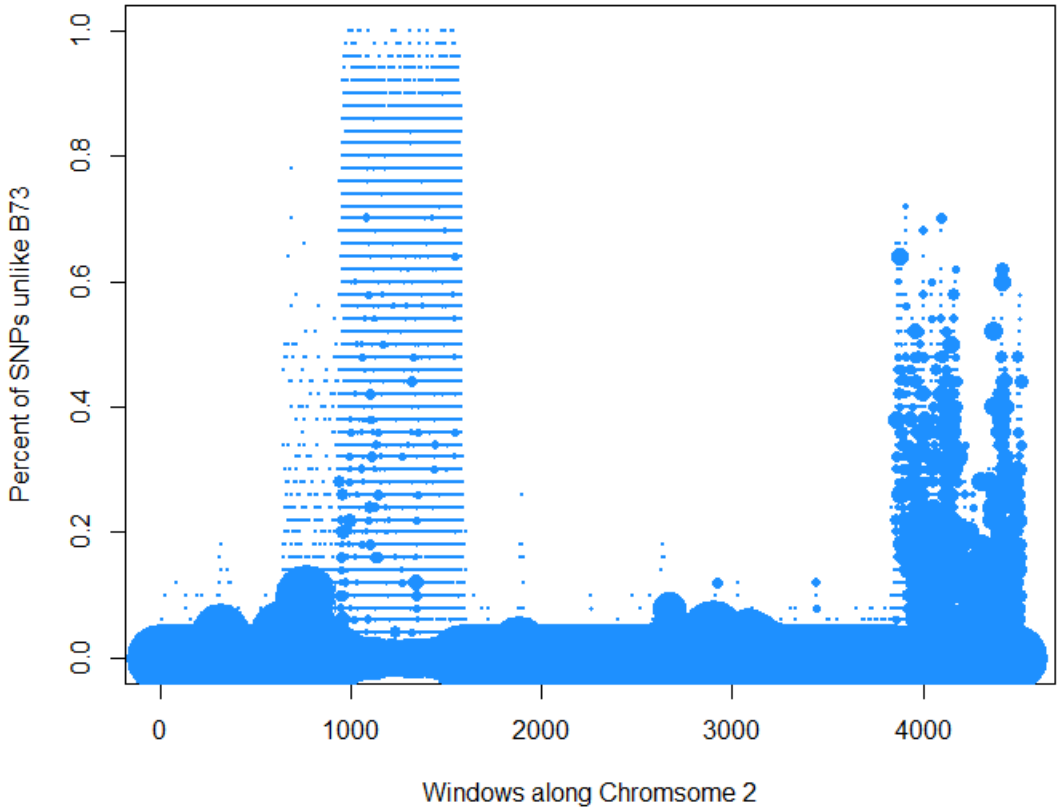


Figure 3.7. Molecular markers were used to assess the uniformity of the epiNILs genetically. The position and size of the symbols in the plot refer to where along chromosome 2 the 50-SNP-window was and the number of epiNILs that had the same number of non-B73-like SNPs in that window dictated the size.

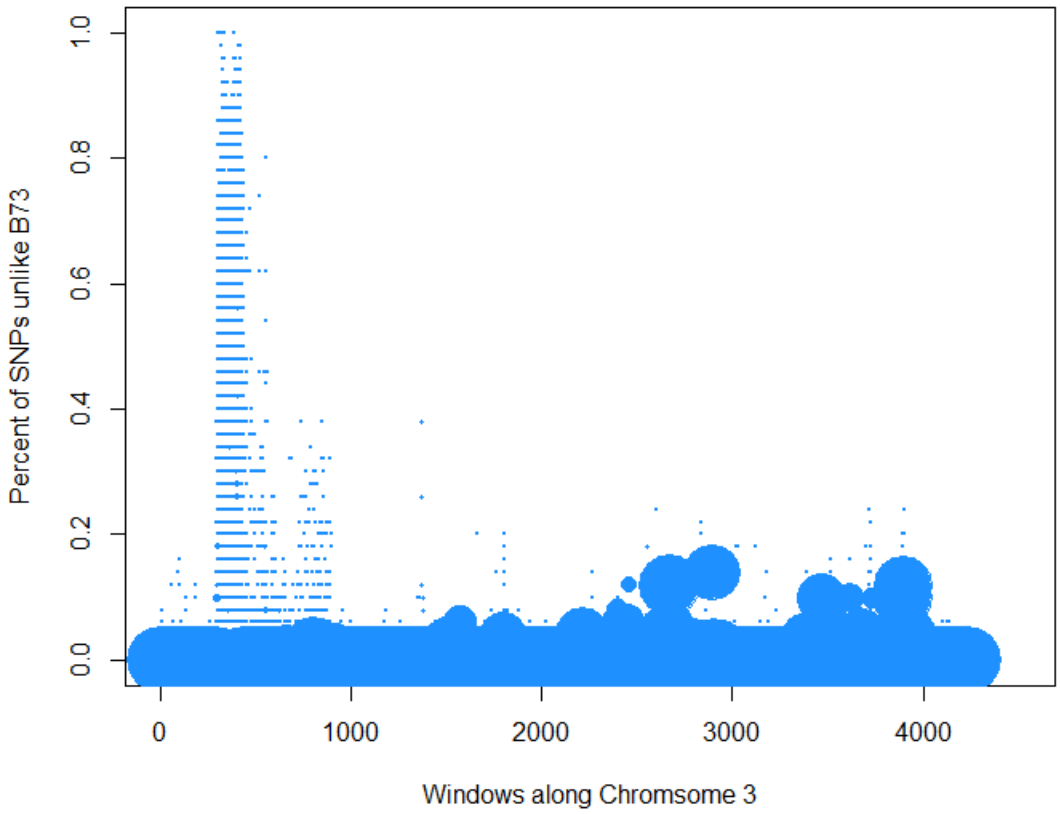


Figure 3.8. Molecular markers were used to assess the uniformity of the epiNILs genetically. The position and size of the symbols in the plot refer to where along chromosome 3 the 50-SNP-window was and the number of epiNILs that had the same number of non-B73-like SNPs in that window dictated the size.

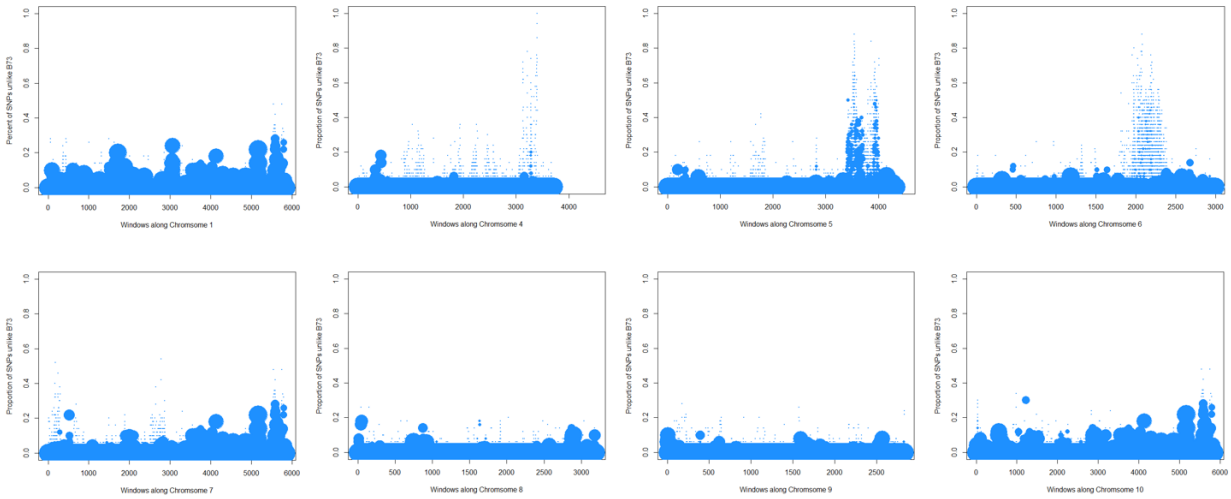


Figure 3.9. Molecular markers were used to assess the uniformity of the epiNILs genetically. The position and size of the symbols in the plot refer to where along the other eight chromosomes the 50-SNP-windows were and the number of epiNILs that had the same number of non-B73-like SNPs in that window dictated the size.

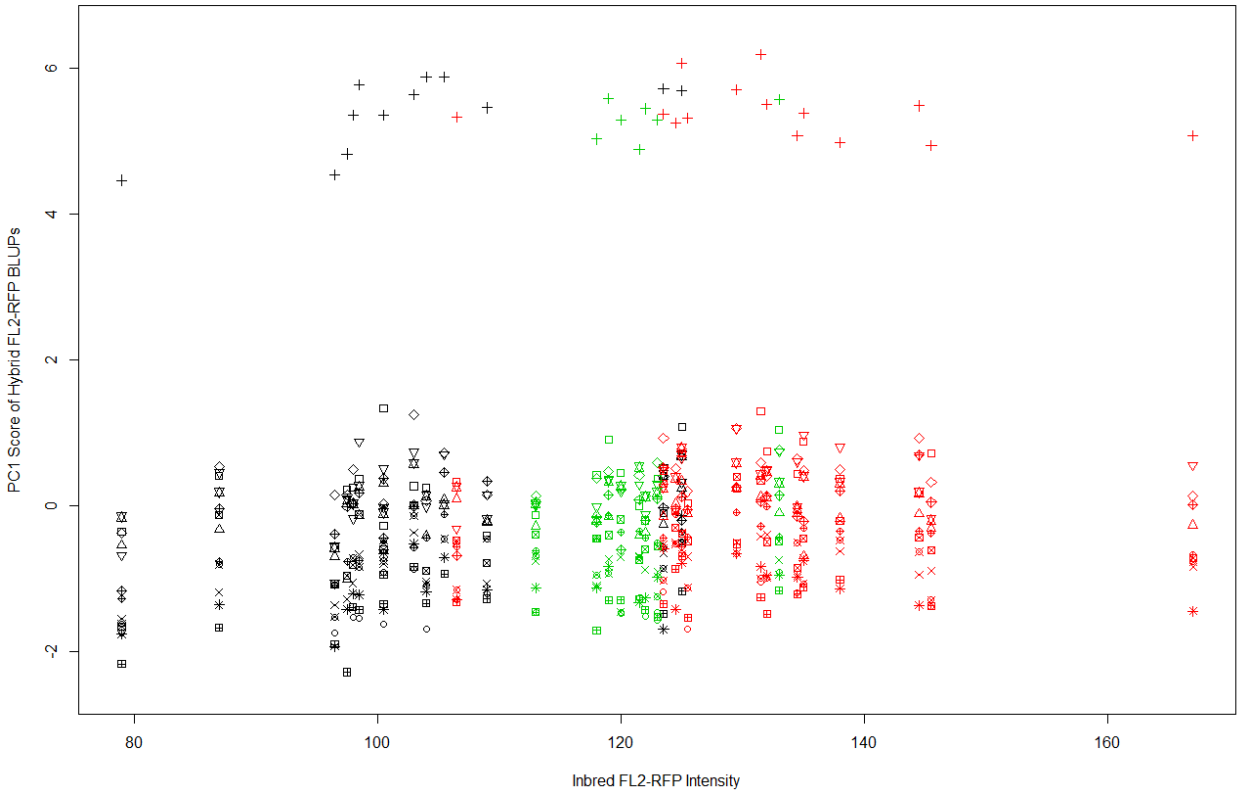


Figure 3.10. The first principal component scores of the hybrid FL2-RFP BLUPs are plotted against the epiNIL parent FL2-RFP intensity used in that hybrid (represented by the value of the green channel). Thus, all hybrids made with a given epiNIL are at the same x-coordinate. Each tester parent is represented by a unique symbol, and the colors refer to the direction of selection that epiNIL parent experienced, black is the darkest red epiNILs, red is the lightest red epiNILs, and green is the epiNIL controls.

Table 3.2. Estimates of the variance components.

Variance Component	Estimate of the Variance Component
σ^2_{GCAepi}	66.45
$\sigma^2_{GCAtester}$	382.6
σ^2_{SCA}	30.30
σ^2_{YG}	38.88
σ^2_{YS}	13.12
σ^2_e	125.9

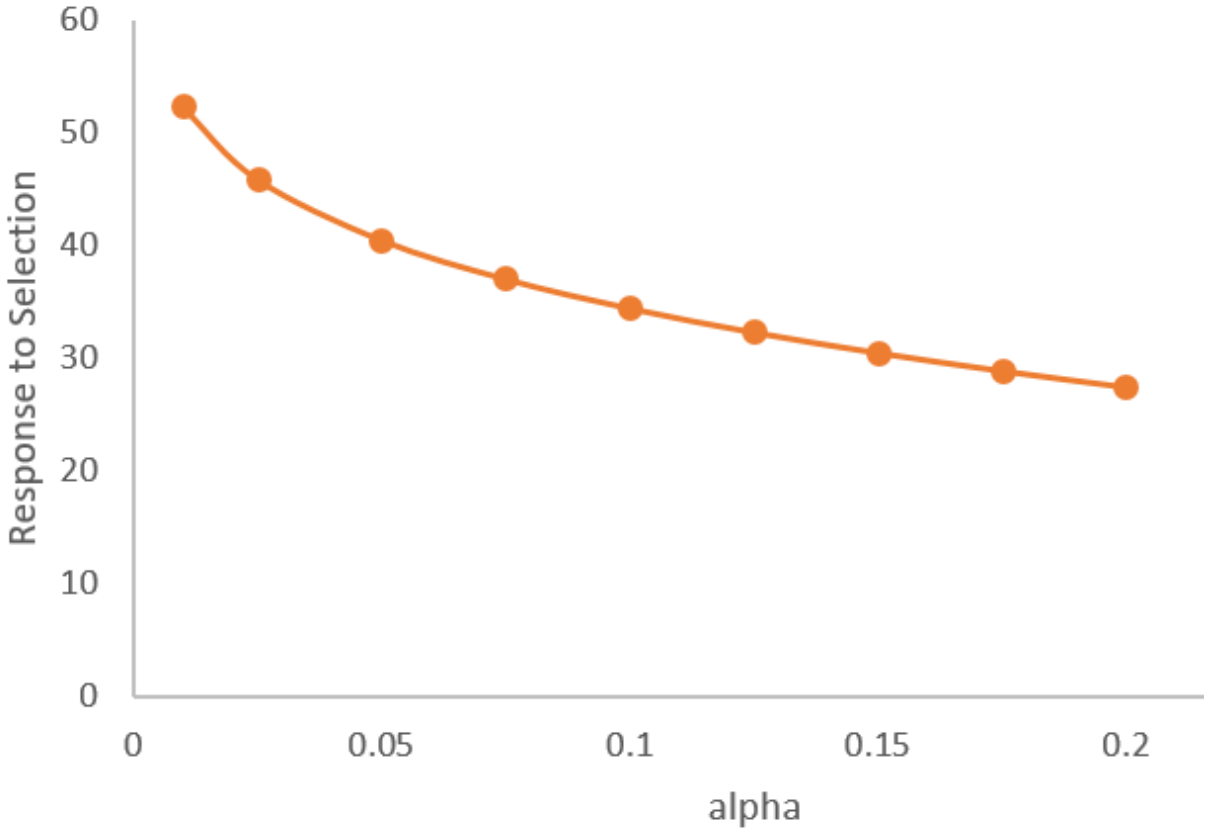


Figure 3.11. The response to selection is plotted against the alpha value used to determine the proportion of individuals that would be selected for the FL2-RFP trait in this population.

Figure 3.12. Three photographs showing the phenotype that was seen for the first time in the epiNIL hybrids (epiNIL 555, 566, 567, 571, 573, 589, 597 crossed by NC350, Mo17 (both reps), PHJ31 (both reps), PHZ51, IHP1, NC350, and PHJ31, respectively). a) Shows the phenotype on younger plants (23June), while b) and c) show different hybrids at a later stage (11July).

REFERENCES

Alleman, M., L. Sidorenko, K. McGinnis, V. Seshadri, J.E. Dorweiler, J. White, et al. 2006. An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature*. 442, 295-298.

Barber, W.T., W. Zhang, H. Win, K.K. Varala, J.E. Dorweiler, M.E. Hudon, et al. 2012. Repeat Associated Small RNAs Vary among Parents and Following Hybridization in Maize. Proceedings of the National Academy of Science, U S A. 109, 10444-10449.

Barber, W.T. 2013. The contribution of repeat associated small RNAs to genetic variation, hybrid vigor, and inbreeding depression in maize. Ph.D. diss., University of Illinois, Urbana-Champaign.

Chen, Z. J. 2010. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Science*. 15, 57-71.

Cortijo, S., R. Wardenaar, M. Colome-Tatche, A. Gilly, M. Etcheverry, K. Labadie, et al. 2014. Mapping the epigenetic basis of complex traits. Science. 343, 1145-1148.

Darwin, C.R. 1876. The effects of cross- and self-fertilisation in the vegetable kingdom. London: John Murray.

Dorweiler, J.E, C.C. Carey, K.M. Kubo, J.B. Hollock, J.L. Kermicle, V.L. Chandler. 2000. *mediator of paramutation1* is required for establishment and maintenance of paramutation at multiple maize loci. The Plant Cell. 12, 2101-2118.

East, E.M. 1908. Inbreeding in corn. Rep. Conn. Agric. Exp. Stn. 419-428.

Eichten, S.R., R.A. Swanson-Wagner, J.C. Schnable, A.J. Waters, P.J. Hermanson, S. Liu, et al. 2011. Heritable epigenetic variation among maize inbreds. PLOS Genetics. 7: 11. e1002372. https://doi.org/10.1371/journal.pgen.1002372

Eichten, S.R., R. Briskine, J. Song, Q. Li, R. Swanson-Wagner, P.J. Hermanson, et al. 2013. Epigenetic and genetic influences on DNA methylation variation in maize populations. The Plant Cell. 25, 2783-2797.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, S.E. Mitchell. 2011. A robust, simple genotyping-by sequencing (GBS) approach for high diversity species. PLoS One. 6: e19379.

Feng, L.N., J. Zhu, G. Wang, Y.P. Tang, H.J. Chen, W.B. Jin, et al. 2009. Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize alpha-zein super gene family. Plant Mol Biol. 69: 649–659.

Fire, A., S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, C.C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in *C. elegans*. Nature. 391, 806-811.

Gross, D.S., W.T. Garrard. 1988. Nuclease hypersensitive sites in chromatin. Annual Reviews of Biochemistry. 57, 159-197.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One. 9, e90346. https://doi.org/10.1371/journal.pone.0090346

Hamilton, A., O. Voinnet, L. Chappell, D. Baulcombe. 2002. Two classes of short interfering RNA in RNA silencing. EMBO Journal. 21, 4671-4679.

Hamilton, A., D. Baulcombe. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science. 286, 950-952.

Huang, S., A.R. Whitney, Q. Zhou, K.P. Malloy, D.A. Voyles, J. Anthony, et al. 2004. Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. Journal of Agricultural and Food Chemistry. 2004: 1958-1964. DOI: 10.1021/jf0342223

Jia, Y., D.R. Lisch, K. Ohtsu, M.J. Scanlon, D. Nettleton, P.S. Schnable. 2009. Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLOS Genetics. 5, e1000737.

Labonne, J.D.J., J.E. Dorweiler, K.M. McGinnis. 2013. Changes in nucleosome position at transcriptional start sites of specific genes in *Zea mays mediator of paramutation1* mutants. Epigenetics. 8, 398-408.

Lauter, N., A. Kampani, S. Carlson, M. Goebel, S.P. Moose. 2005. microRNA172 down-regulates *glossy15* to promote vegetative phase change in maize. Proc Natl Acad Sci USA. 102: 9412–9417.

Li, Q., K. Varala, S.P. Moose, M.E. Hudson. 2012. The inheritance pattern of 24-nt siRNA clusters in Arabidopsis hybrids is influenced by proximity to transposable elements. PLoS One. 7, e47043.

Li, Q., J.I. Gent, G. Zynda, J. Song, I. Makarevitch, C.D. Hirsch. 2015. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. PNAS, USA. 112, 14728-14733.

Lisch, D.R. 2012. Regulation of transposable elements in maize. Current Opinions in Plant Biology. 15, 511-516.

Liu, C., J. Wang, X. Mei, X. Deng, T. Yu, X. Liu, et al. 2015. Characterization of the imprinting and expression patterns of ZAG2 in maize endosperm and embryo. The Crop Journal. 3, 74-79.

Lolle, S.J., J.L. Victor, J.M. Young, R.E. Pruitt. 2005. Genome-wide non-mendelian inheritance of extra-genetic information in *Arabidopsis*. Nature. 434, 505-509.

Lucas, C.J., H. Zhao, M. Schneerman, S.P. Moose. 2013. Genomic changes in response to 110 cycles of selection for seed protein and oil concentration in maize. Seed Genomics, (ed W. Becraft), Wiley-Blackwell, Oxford, UK. doi: 10.1002/9781118525524.ch12

Luger, K., A.W. Mader, R.K. Richmond, D.F. Sargent, T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Angstrom resolution. Nature. 389. 251-260.

McCue, A.D., R.K. Slotkin. 2012. Transposable element small RNAs as regulators of gene expression. Trends in Genetics. 28, 616-623.

Miclaus, M., J.H. Xu, J. Messing. 2011. Differential gene expression and epiregulation of alpha zein gene copies in maize haplotypes. PLoS Genet. 7: e1002131.

Meister, G. 2013. Argonaute proteins: functional insights and emerging roles. Nature Reviews Genetics. 14, 447-459.

Mohanty, A., A. Luo, S. DeBlasio, X.Y. Ling, Y. Yang, D.E. Tuthill, et al. 2009. Advancing cell biology and functional genomics in maize using fluorescent protein-tagged lines. Plant Physiol. 149: 601–605.

Nelson, O.E. 1979. Genetic control of polysaccharide and storage protein synthesis in the endosperms of barley, maize, and sorghum. Adv Cereal Sci Technol. 3: 41–71.

Nobuta, K., C. Lu, R. Shrivastava, M. Pillay, E. DePaoli, M. Accerbi, et al. 2008. Distinct size distribution of endogenous siRNAs in maize: evidence from deep sequencing in the *mop1-1* mutant. PNAS, USA. 105, 14958-14963.

Peng, P., S.W. Chan, G.A. Shah, S.E. Jacobsen. 2006. Plant genetics: increased outcrossing in hothead mutant. Nature. 443, doi: 10.1038/nature05251

Shull, G.H. 1908. The composition of a field of maize. Am. Breeders Assoc. Rep. 4: 296-301.

Smith, L.M., D. Weigel. 2012. On epigenetics and epistasis: hybrids and their non-additive interactions. EMBO Journal. 31, 249-250.

Song, R., J. Messing. 2002. Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. Plant Physiology. 130: 1626-1635.

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

Winter. 1929. The mean and variability as affected by continuous selection for composition in corn. Journal of Agronomic Research. 39: 451-476.

Woo, Y.M., D.W.N. Hu, B.A. Larkins, R. Jung. 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.

Zhang, L., J.M. Chia, S. Kumari, J.C. Stein, Z. Liu, A. Narechania, et al. 2009. A genome-wide characterization of microRNA genes in maize. Plos Genetics. 5, e1000716.

APPENDIX A

SUPPLEMENTAL FIGURES AND TABLES

Figure A.1. The ears harvested from the IHP population grown 2017. Superimposed on the ears are their unique ID numbers (small black numbers) and their protein concentrations as measured by NIR (larger numbers below – those ears selected for advancement appear in red).

Figure A.2. The ears harvested from the ISHP population grown 2017. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

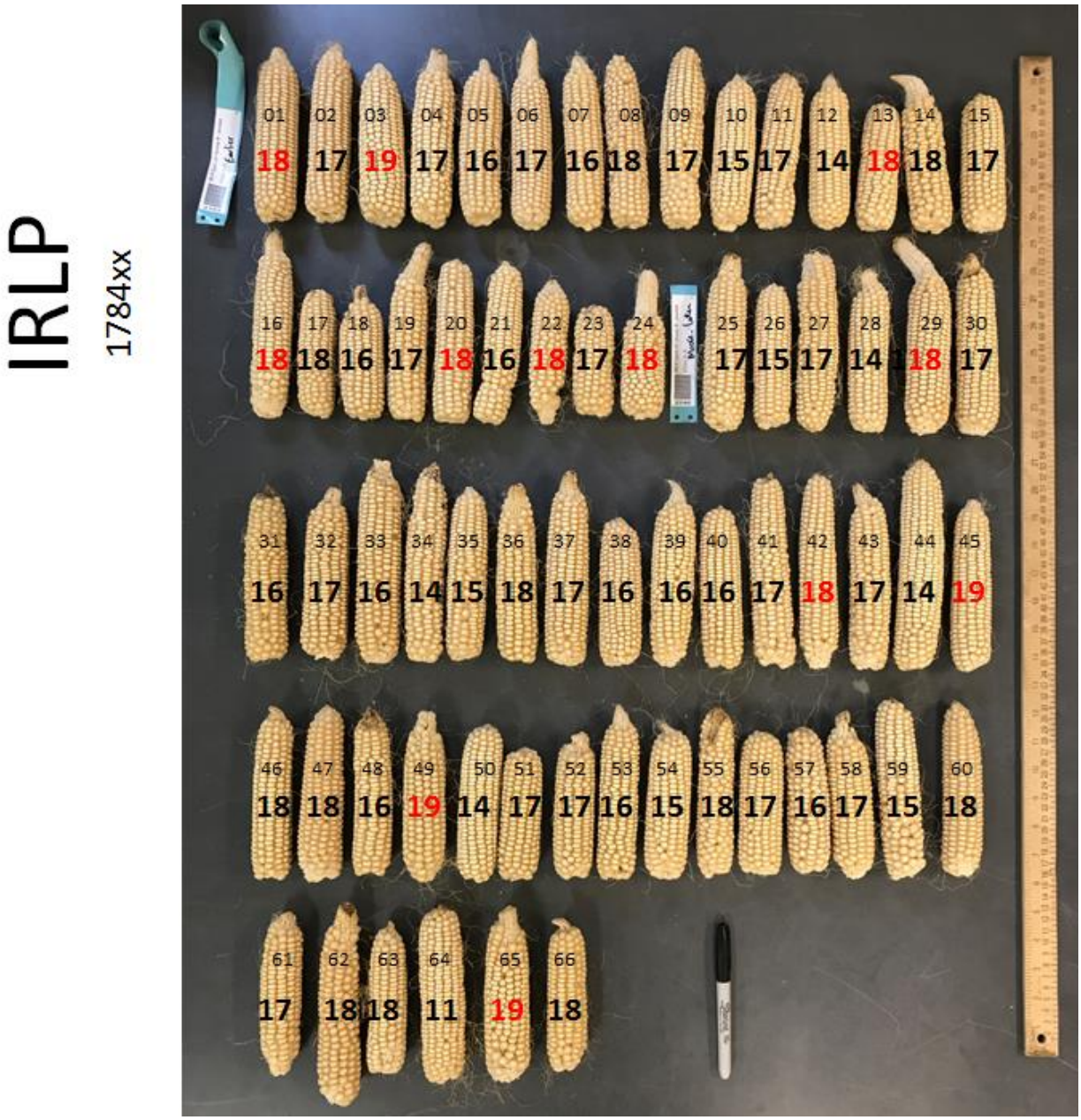


Figure A.3. The ears harvested from the IRLP population grown 2017. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

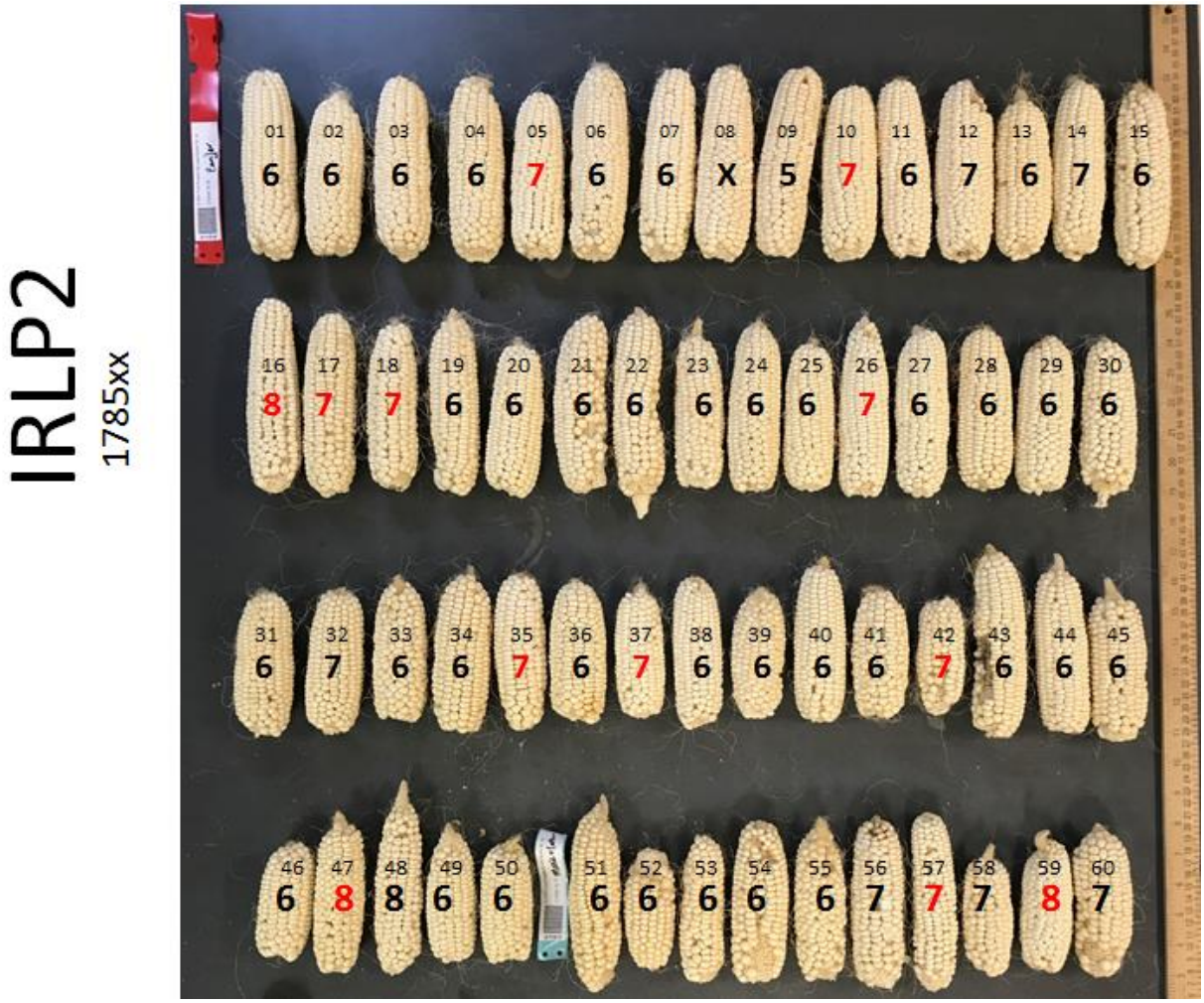


Figure A.4. The ears harvested from the IRLP2 population grown 2017. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

178601-178602 tassel tip. 178603 3 ears in one sheath.

IRHP2

1786xx

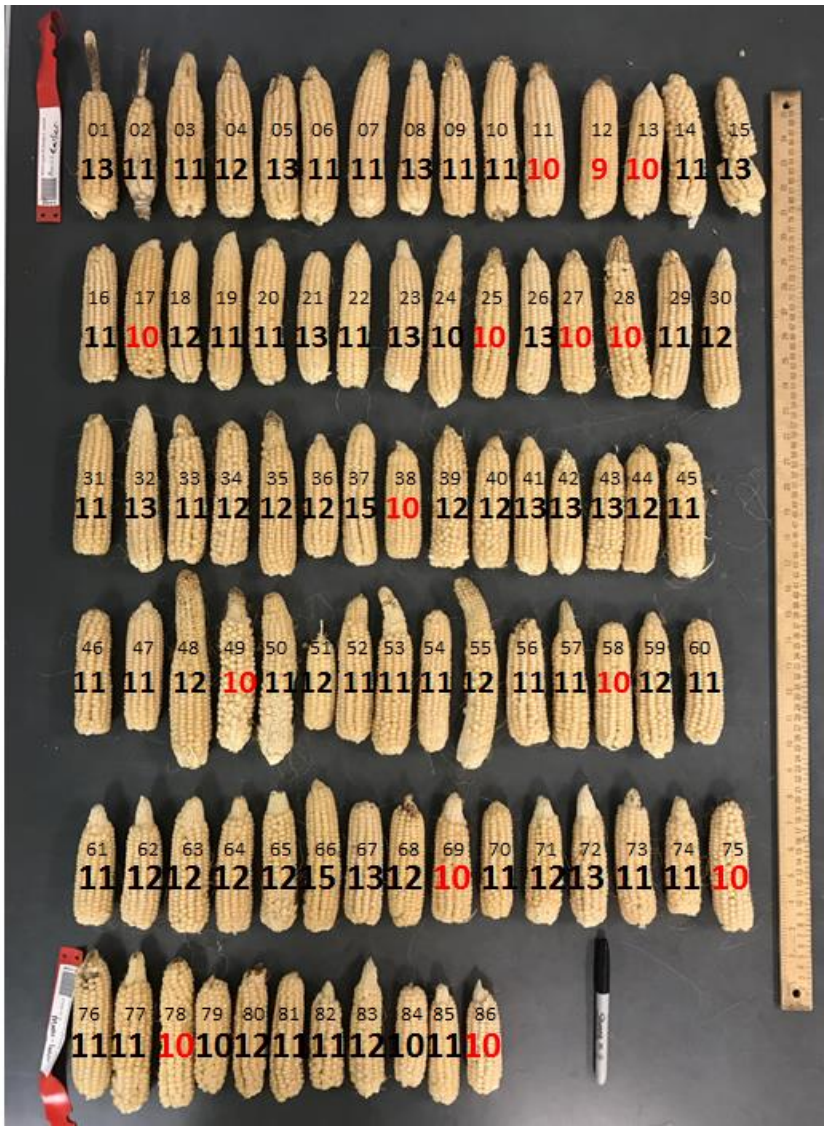


Figure A.5. The ears harvested from the IRHP2 population grown 2017. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

IRHP3

1787xx

Figure A.6. The ears harvested from the IRHP3 population grown 2017. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

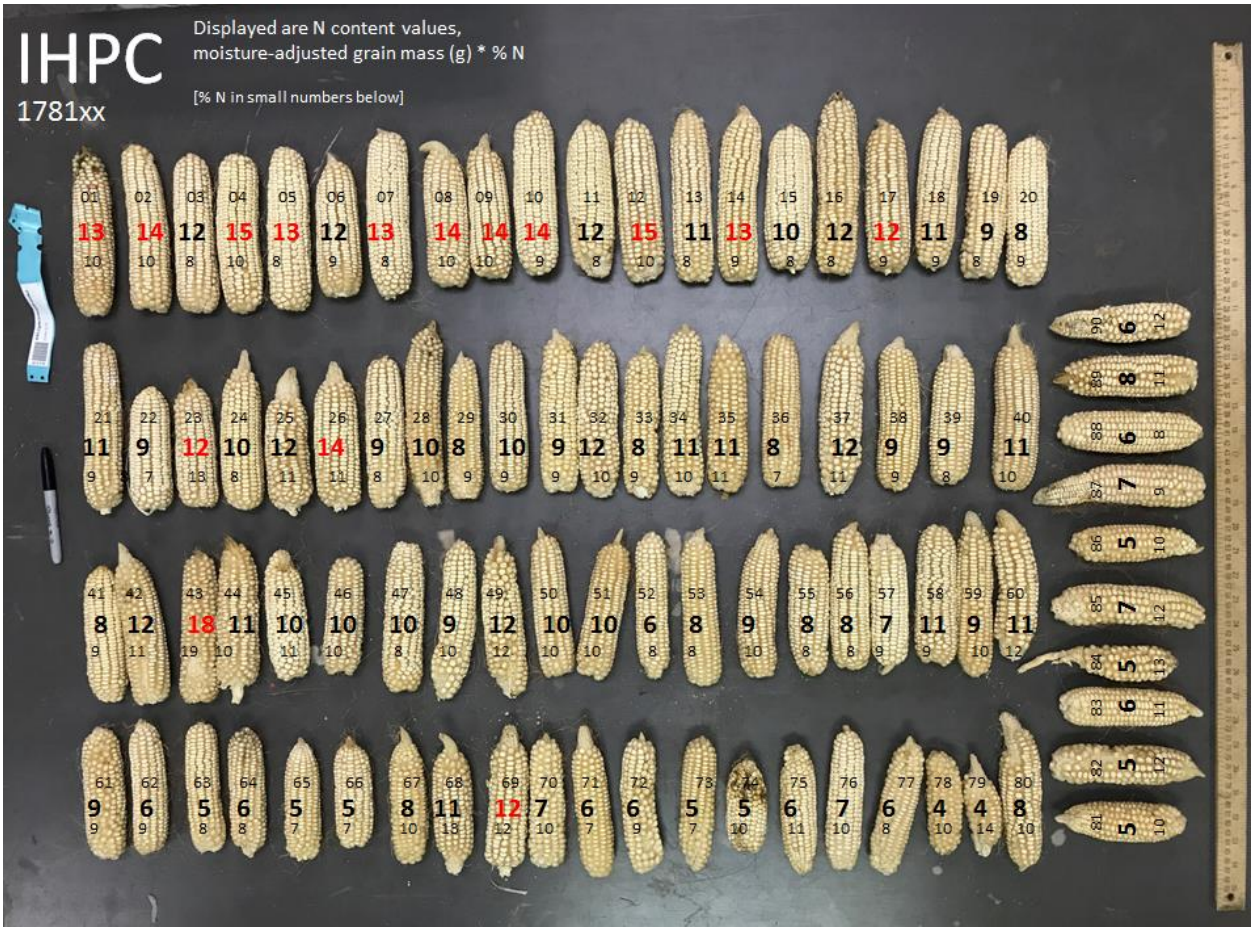


Figure A.7. The ears harvested from the IHPC population grown 2017. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

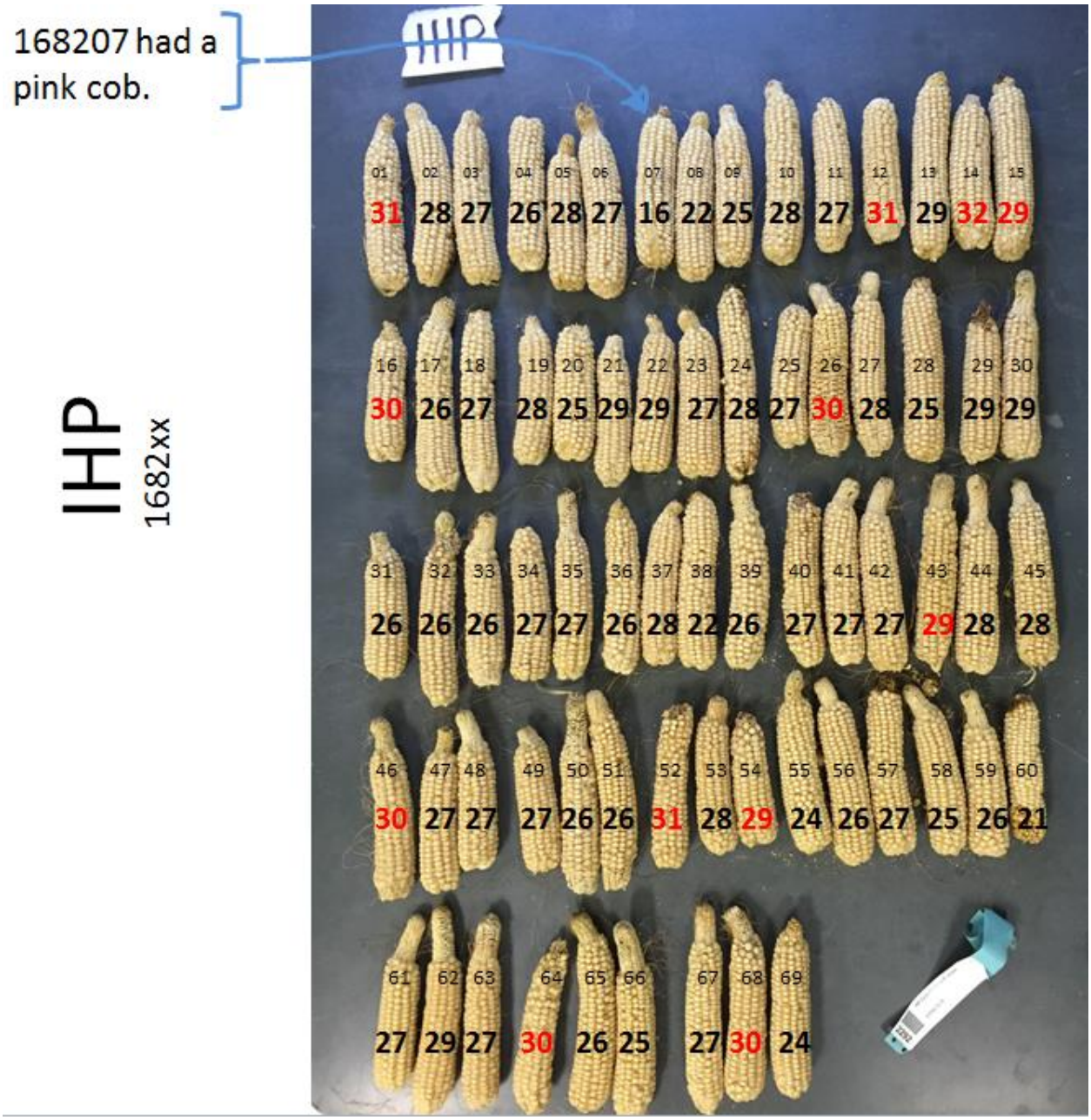


Figure A.8. The ears harvested from the IHP population grown 2016. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

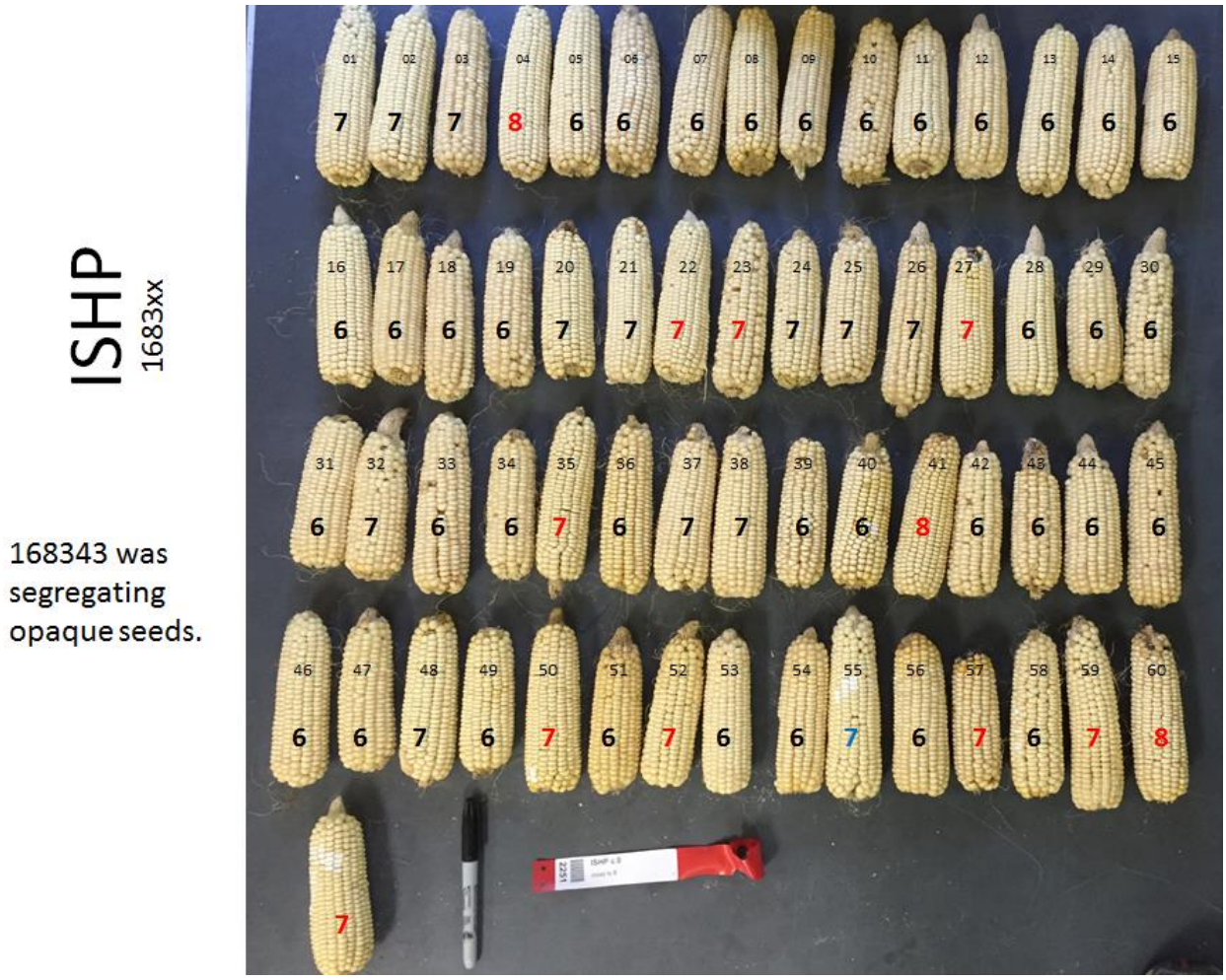


Figure A.9. The ears harvested from the ISHP population grown 2016. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

IRLP

1684xx

Figure A.10. The ears harvested from the IRLP population grown 2016. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

Figure A.11. The ears harvested from the IRLP2 population grown 2016. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

Figure A.12. The ears harvested from the IRHP2 population grown 2016. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

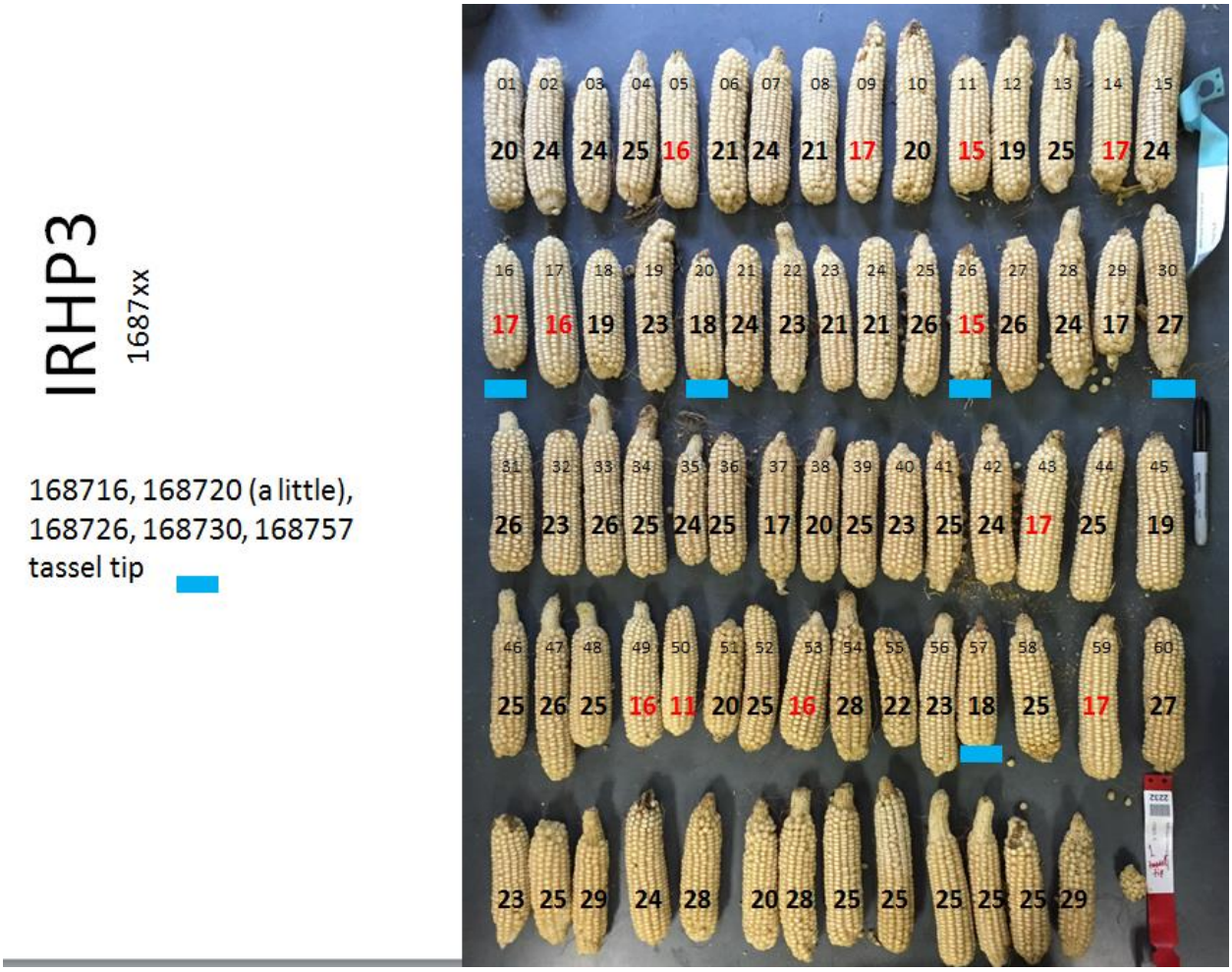


Figure A.13. The ears harvested from the IRHP3 population grown 2016. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

IHPC

1681xx

Displayed are N content values, moisture-adjusted grain mass (g) * % N

[% N in small numbers below]

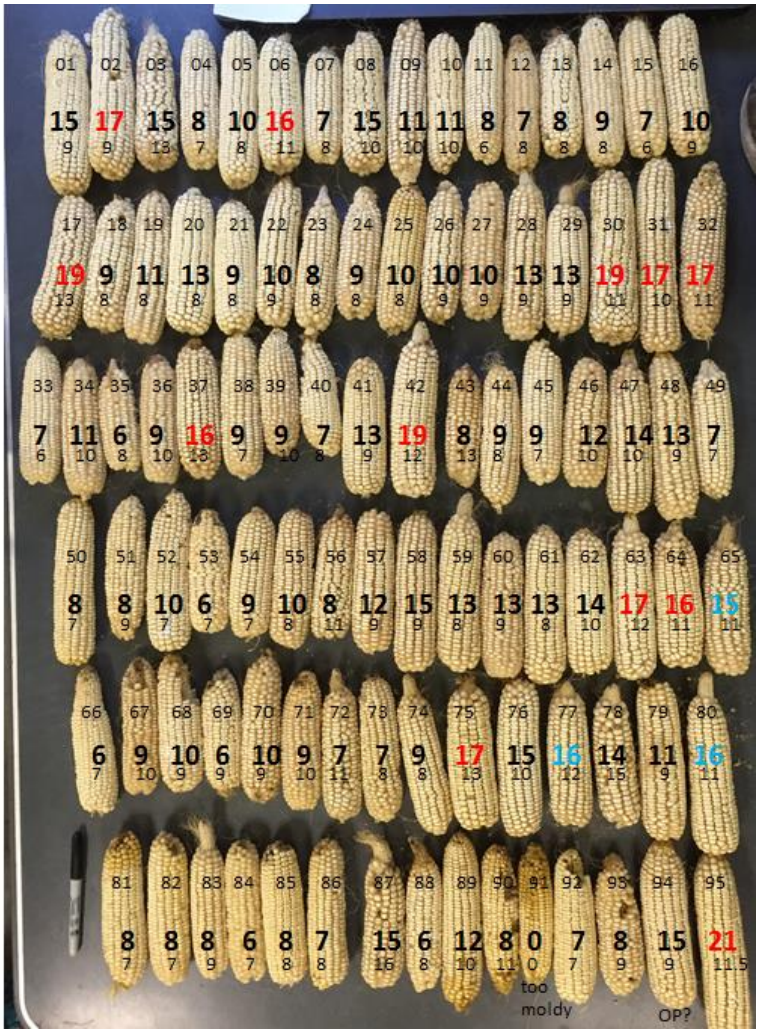


Figure A.14. The ears harvested from the IHPC population grown 2016. Superimposed on the ears are their unique ID numbers and their protein concentrations as measured by NIR.

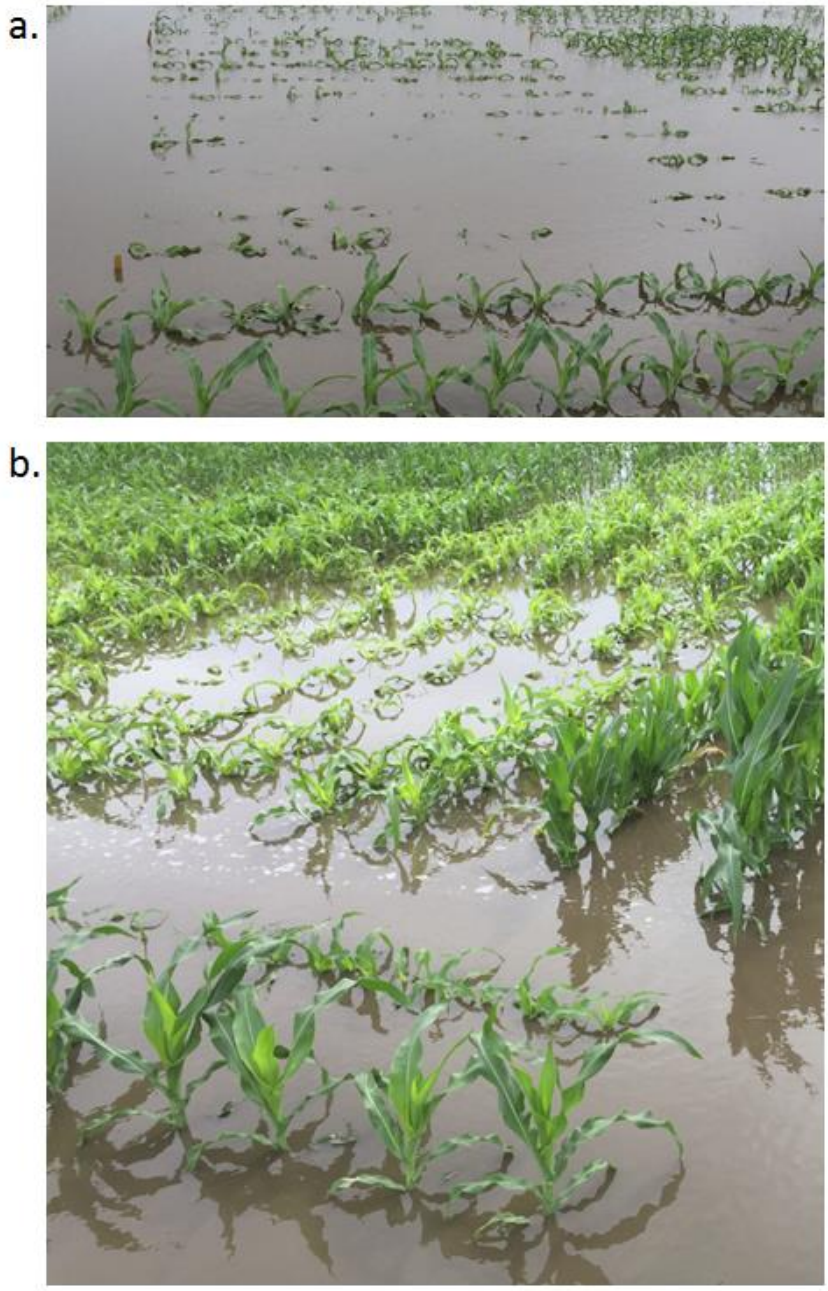


Figure A.15. Nitrogen availability in the field was low in 2015 largely due to a pair of gully washers that flooded the field plots. In the immediate foreground in both images is IRLP, which was damaged so severely that it was not phenotyped in 2015. (a) Looking North, several populations of the ILTSE are seen snorkeling early in the growing season on a beautiful day following a large rain. (b) Looking northeast, the same plot as in (a) can be seen again standing in deep water later in June. The 12-inch orange row stake in (a) is completely submerged in (b).

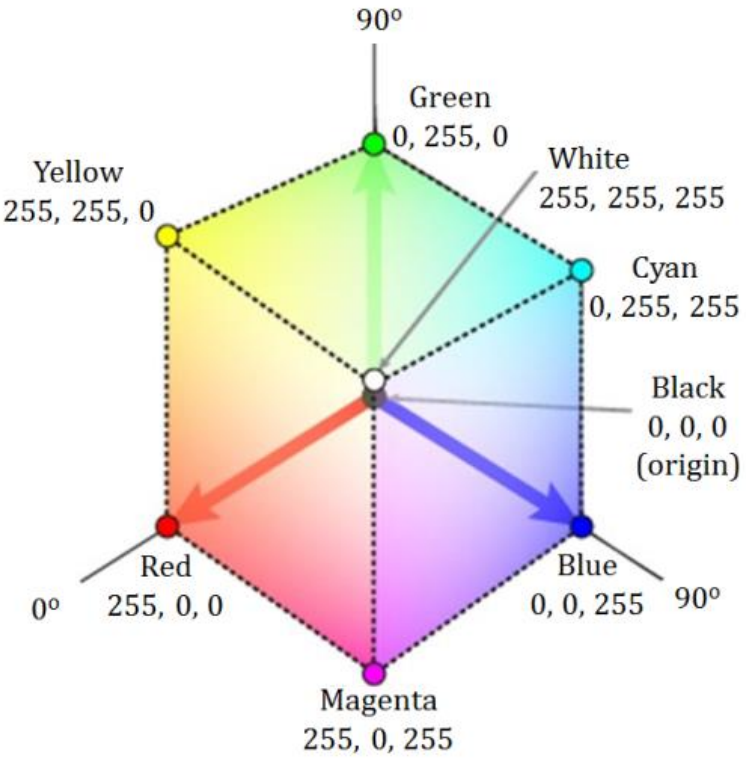


Figure A.16. The RGB color model mapped in a cube allows for visualization of how colors translate into three numbers and where the experimental corn kernels may fall. Modified from: social.technet.microsoft.com

Table A.1. The 16 eigenvalues and the proportion of variance that each explains from the principal components analysis of the correlation matrix.

	Eigenvalue	Difference between Eigenvalues	Proportion of Total Variance Explained	Cumulative Variance Explained
1	3.24994135	0.52088787	0.2031	0.2031
2	2.72905348	0.68229233	0.1706	0.3737
3	2.04676115	0.14258459	0.1279	0.5016
4	1.90417656	0.49134493	0.119	0.6206
5	1.41283164	0.33222103	0.0883	0.7089
6	1.08061061	0.16304144	0.0675	0.7765
7	0.91756917	0.04126567	0.0573	0.8338
8	0.8763035	0.16309315	0.0548	0.8886
9	0.71321035	0.19367473	0.0446	0.9332
10	0.51953562	0.32977982	0.0325	0.9656
11	0.1897558	0.07179056	0.0119	0.9775
12	0.11796525	0.02172042	0.0074	0.9849
13	0.09624483	0.01083785	0.006	0.9909
14	0.08540697	0.02685544	0.0053	0.9962
15	0.05855154	0.05646935	0.0037	0.9999
16	0.00208219		0.0001	1

Pearson Correlation Coefficients, N = 131

	Hyrbid Grain Oil (%)	Hybrid Grain Protein (%)	Hybrid Grain Starch (%)	Hybrid Total Ear Mass (g)	Hybrid Total Grain Mass (g)	Hybrid RFP - Red Channel	Hybrid RFP - Green Channel	Hybrid RFP - Blue Channel	Hybrid Stover Biomass (g)	Hybrid Plant Height (cm)	Hybrid Total Plant Nodes	Hybrid Ear Node	IPSRI Inbred Grain Protein (%)	IPSRI Inbred RFP - Red	IPSRI Inbred RFP - Green	IPSRI Inbred RFP - Blue
Hyrbid Grain Oil (%)	1	-0.4037	-0.3025	0.10771	0.10132	0.16219	0.20922	0.15882	0.10434	-0.0001	-0.0322	-0.0474	0.02623	-0.023	-0.0916	-0.1241
Hybrid Grain Protein (%)	-0.4037	1	-0.5971	-0.1927	-0.1984	-0.1454	-0.2048	-0.1555	-0.20215	-0.2869	-0.0346	0.00571	0.3609	-0.1125	0.01111	0.13254
Hybrid Grain Starch (%)	-0.3025	-0.5971	1	0.13967	0.15158	-0.0376	-0.0327	-0.0176	0.04076	0.18524	0.15893	0.14049	-0.3303	0.19292	0.19224	0.06546
Hybrid Total Ear Mass (g)	0.10771	-0.1927	0.13967	1	0.99745	0.0326	0.05728	0.02545	0.10309	0.08036	-0.0055	0.00066	-0.2933	0.13661	0.14805	0.0938
Hybrid Total Grain Mass (g)	0.10132	-0.1984	0.15158	0.99745	1	0.0216	0.04526	0.01409	0.10804	0.08098	-0.0004	0.01249	-0.3046	0.14808	0.15563	0.10723
Hybrid RFP - Red Channel	0.16219	-0.1454	-0.0376	0.0326	0.0216	1	0.90536	0.91127	0.04455	-9E-05	-0.0993	-0.1337	-0.1329	0.12819	0.04217	-0.112
Hybrid RFP - Green Channel	0.20922	-0.2048	-0.0327	0.05728	0.04526	0.90536	1	0.92624	0.02522	-0.0258	-0.1305	-0.1735	-0.2329	0.1146	0.08576	-0.0995
Hybrid RFP - Blue Channel	0.15882	-0.1555	-0.0176	0.02545	0.01409	0.91127	0.92624	1	-0.0128	-0.0193	-0.1267	-0.1615	-0.1466	0.12728	0.08469	-0.0421
Hybrid Stover Biomass (g)	0.10434	-0.2022	0.04076	0.10309	0.10804	0.04455	0.02522	-0.0128	1	-0.0541	-0.133	-0.1261	-0.0001	-0.0461	-0.169	-0.0705
Hybrid Plant Height (cm)	-0.0001	-0.2869	0.18524	0.08036	0.08098	-9E-05	-0.0258	-0.0193	-0.05407	1	0.31842	0.22354	-0.17	-0.0209	-0.0555	-0.035
Hybrid Total Plant Nodes	-0.0322	-0.0346	0.15893	-0.0055	-0.0004	-0.0993	-0.1305	-0.1267	-0.13297	0.31842	1	0.87463	-0.0435	0.0765	-0.0274	-0.0338
Hybrid Ear Node	-0.0474	0.00571	0.14049	0.00066	0.01249	-0.1337	-0.1735	-0.1615	-0.12613	0.22354	0.87463	1	-0.02	0.1204	-0.0098	-0.0263
IPSRI Inbred Grain Protein (%)	0.02623	0.3609	-0.3303	-0.2933	-0.3046	-0.1329	-0.2329	-0.1466	-0.00014	-0.17	-0.0435	-0.02	1	-0.3673	-0.2754	-0.0148
IPSRI Inbred RFP - Red	-0.023	-0.1125	0.19292	0.13661	0.14808	0.12819	0.1146	0.12728	-0.04609	-0.0209	0.0765	0.1204	-0.3673	1	0.54627	0.17483
IPSRI Inbred RFP - Green	-0.0916	0.01111	0.19224	0.14805	0.15563	0.04217	0.08576	0.08469	-0.16903	-0.0555	-0.0274	-0.0098	-0.2754	0.54627	1	0.68837
IPSRI Inbred RFP - Blue	-0.1241	0.13254	0.06546	0.0938	0.10723	-0.112	-0.0995	-0.0421	-0.07049	-0.035	-0.0338	-0.0263	-0.0148	0.17483	0.68837	1

Figure A.17. The correlation matrix for the 16 variables in this study. The cells of the table are colored to aid in quick identification of larger correlations, negative being darker red and positive being progressively darker green.

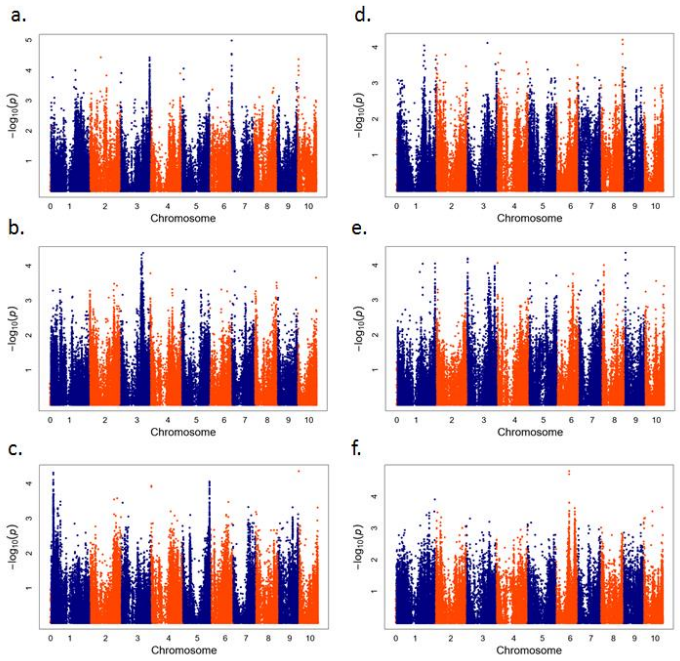


Figure A.18. The GAPIT analysis of all six traits produced six Manhattan plots of the negative logarithm of the probability of association for principal component scores for all individuals. Analyzed SNPs are plotted on the x-axis ordered by chromosomal position. The trait analyzed in each plot is (a) scores for principal component one, (b) scores for principal component two, (c) scores for principal component three, (d) scores for principal component four, (e) scores for principal component five, (f) scores for principal component six.

	Factor1	Factor2	Factor3	Factor4	Factor5
Oil	0.13739	0.07789	-0.2965	-0.7185	-0.0383
Prt	-0.1103	-0.0815	-0.5776	0.75691	0.27327
Stc	-0.0327	0.06591	0.98803	0.01961	-0.1342
tEW	-0.004	0.98974	0.06177	-0.0718	-0.1069
tGW	-0.0155	0.98578	0.07258	-0.0688	-0.1134
Red	0.94176	0.03658	-0.0117	-0.0565	-0.0188
Gee	0.95166	0.04919	-0.0205	-0.1024	-0.1213
Blu	0.96291	0.02777	0.00836	-0.0497	-0.0303
stand	0.11488	0.48079	-0.0229	-0.245	-0.0185
Svr	-0.0019	0.08144	0.04639	-0.2236	0.05873
inb	-0.1256	-0.1941	-0.1786	-0.0266	0.956

Table A.2. The loadings for each variable within each factor shown and highlighted to show the contrasts within each factor to make sense of what each factor is portraying.

a.

Canonical Correlation Analysis

Standardized Canonical Coefficients for the VAR Variables

	inbred1	inbred2
Inbred protein	-0.6339	0.7820
Hybrid stand	0.7040	0.7195

Standardized Canonical Coefficients for the WITH Variables

	hybrid1	hybrid2
Hybrid Oil	-0.3162	0.3542
Hybrid Protein	-0.6361	0.3381
Hybrid Starch	-0.2627	-0.3684
Hybrid Total Ear	1.1078	6.1697
Hybrid Grain Mass	-0.4222	-5.7997
Hybrid Red Channel	-0.5347	-0.2074
Hybrid Green Channel	1.0894	-0.5397
Hybrid Blue Channel	-0.3316	0.5227
Hybrid Stover Mass	-0.0464	0.4179

b.

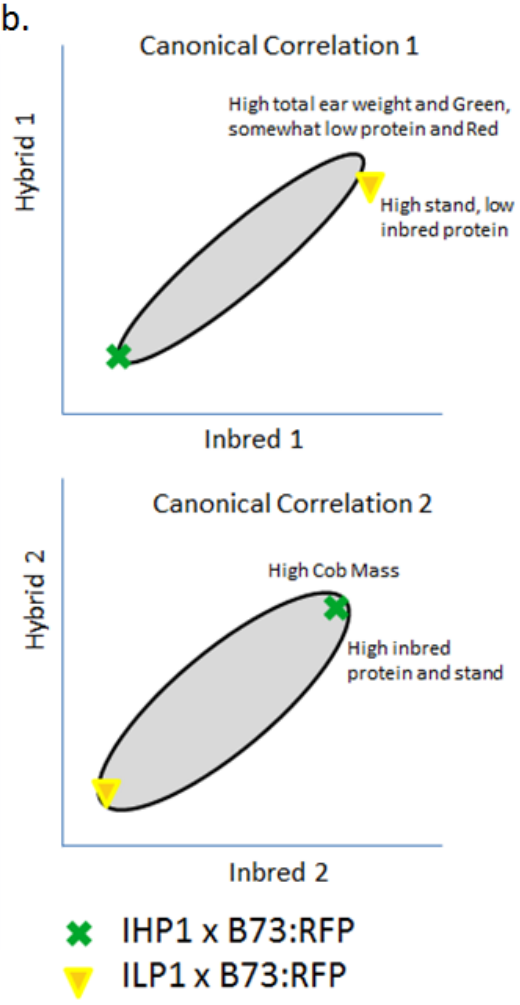


Figure A.19. The two canonical correlations that describe the relationship between the v variable inbred protein and the w variables of the hybrid phenotypes while controlling for hybrid stand are shown. (a) The standardized canonical coefficients are shown for each group to aid in seeing how an individual would be placed along each axis, while (b) shows a schematic of how these relationships look and where the high protein control hybrid IHP1 x B73: FL2-RFP (green symbol) and ILP1 x B73: FL2-RFP (yellow symbol) would be on those plots.

Figure A.20. From left-to-right, representative phenotype of this lineage grown in 2014, 2015, and 2016. An agronomically attractive dwarf was first seen in 2014, where 2 out of the 10 plants showed this phenotype and were self-pollinated. (Both plants were only 120cm tall. A third plant in that row was a dwarf, but was far from being agronomically desirable as it didn't even make an ear or seeds.) When the good-looking dwarfs were grown in 2015, both rows were 100% phenotypically the stout dwarf (10/10 and 12/12). Several of these were self-pollinated and a couple hybrids were made (PHJ31 as a pollen parent, and A632 as a mother). The self-pollinated (S2) good-looking dwarfs were grown in fourteen rows in 2016, and the range in plant height for the 190 plants was 90-115cm; and the hybrids were also uniform, but much taller. The PHJ31 hybrid was 285cm and the A632 hybrid was 250cm, both of these strikingly similar to their wild-type B73 hybrid, not showing any signs of a dwarfing. Genetic stocks were grown in 2016 to compare to the epidwarfs phenotypically and none of the genetic stocks looked like the epiDwarfs. Crosses were made between these materials and were grown in 2017 showing that this dwarf phenotype is not due to any of the genetic stocks maintained in the Moose Lab.

Figure A.21. Yellow pin stripes (yps) is what I have called a phenotype that is easily seen in the field with only the viens of the leaves being green and all the tissue between being yellow which looks like the phenotype of the maize mutant yellow-stripe (ys1 and ys3). Yps first appeared in the epiNILs in 2014 where four plants in a row of 10 had this phenotype; a) A representative plant in 2014. The yps plants did not make ears, and tassels came out long after any silks were receptive, but four wild-type siblings were self-pollinated. d) The resulting seed were grown in 2015 where three rows showed no yps phenotypes, but one row had 3 of 15 yps plants, the others being indistinguishable from B73. Self-pollinations of the rows that did not segregate yps in 2015 showed no yps in 2016; however, self-pollinations of wild-type siblings of yps-segregating rows segregated yps in four out of five rows in 2016 (3/13, 0/15, 2/11, 1/14, and 5/14, but the last row had variation for the intensity of the yps phenotype, b) and c). This plant grew over five feet tall in 2016, had a full tassel, and made an ear shoot early enough to attempt a pollination. Self-pollinations of the other yps plants were again attempted, but were unsuccessful as the yps plants do not generally make ears or seeds. e) In fact, the yps often don't make male gametes either – the tassels are often barren branches without anthers. A few anthers clung to some yps tassels in 2015 and these were carried to the only receptive silks left in the field, a delayed planting of PHJ31. f) In 2015 one PHJ31 plant was pollinated by a yellow pin stripe plant, and the resulting seeds were grown in 2016 and showed wild-type phenotype. Despite the lack of vigor in these yps plants (height for yps vs wild-type siblings: 78cm vs 219cm in 2014 and 161cm vs 222cm in 2016, the only yps plants in 2015 are shown in d) but were not measured), hybrids made with these yps plants are quite robust (growing to a height of 305cm and bearing large reproductive structures). These plants were backcrossed by pollen from a yps plant in 2016 and planted in 2017.

Figure A.22. In 2017, the seeds from two backcrosses with yellow pin stripes (yps) hybrids were each planted in their own row. In both rows, approximately half of the plants were yps (18/37 and 17/35) suggesting that this trait is acting as a recessive mutant. Also seen in this photo is the progeny of the plant in Figure A.21 b) and c) – this row was 100% yps plants (26/26). The ear that showed this phenotype in 2014 was planted again in 2017 and had 2 yps plants out of 30. The S1 that was grown in 2015 was grown again in 2017 and had 11 yps plants out of 39. Four other rows were planted with siblings of the plant shown in b) and c) and did not show a consistent trend except that they all segregated yps plants (19/40, 12/39, 4/23, 8/35). A wild type sibling of the plant in Figure A.21 b) and c) was pollinated by a yps plant and that row in 2017 had zero yps plants out of 26 plants.

Figure A.23. Seedling mutants have also been seen. Albino seedlings were seen in 10 rows in 2014 (4/10, 4/10, 1/10, 5/10, 3/10, 2/10, 2/10, 3/10, 2/10, 3/10) and yellow seedlings that were lethal and occurred in 1 row (3/10). Non-lethal yellow seedlings were seen in many rows, but most grew out of the early phenotype. In 2015, I planted the siblings of the lethal seedling phenotypes that had been self-pollinated and eight of the nine rows segregated lethal seedling phenotypes.

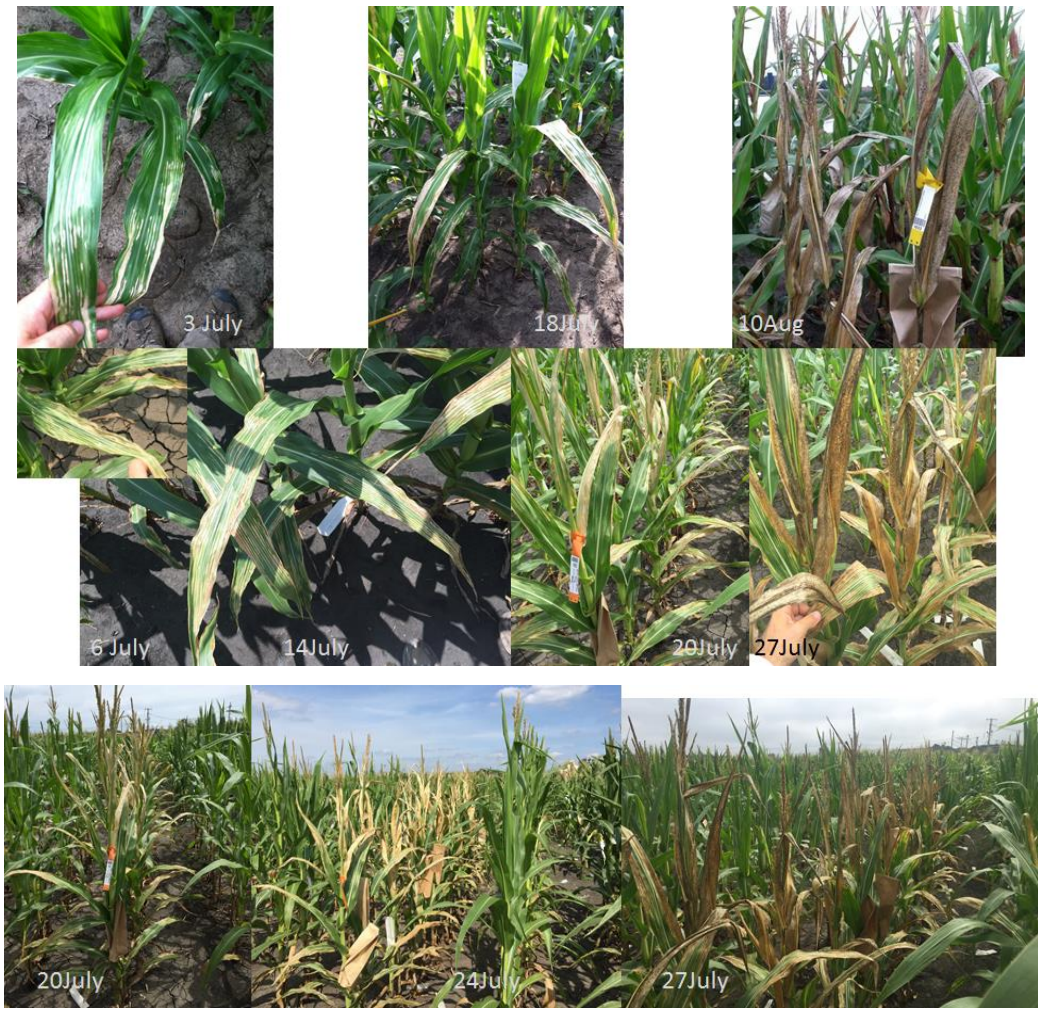


Figure A.24. An early senescence phenotype in the B73 epiNILs was also found. A few epiNILs have appeared completely B73-like all the way until flowering, but then turn brown at flowering. These plants have made no seeds as they blast right at anthesis. First seen in 2014 (top set of photographs), non-blasted half-siblings (that had been self-pollinated) showed the phenotype again in 2015. The second and third rows of photographs are from 2015 where the plants showed a spectrum of severity of burning – those blasting completely made no seeds, but there were a few plants that kept a little greenness and moisture. The rest were crispy before grain-fill could be initiated. The top three rows of photos show the progression of the blasting mostly on a plant or row; whereas, the third set shows that row in comparison to the wild-type B73 on either side of it for comparison.

Figure A.25. The blasting phenotype stocks were grown again in 2017. The photo was taken on 05 July 2017 showing that the blasting occurred long before anthesis likely due to hot and dry weather in June and no fertilizer being applied. If this is a pathogen susceptibility, growing in a field that had corn the previous year might have compounded the early symptoms. The table and dead plant tissue in the lower left is debris from tissue sampling the week before.

Figure A.26. A phenotype which I have called ‘dwarf sideways’ is a phenotype which I have yet to come across studying mutants of maize. This phenotype is also one that has been heritable – all 10 plants in the row in 2014 had the phenotype, though one was slightly taller than the others. When grown in 2015 and 2016, all offspring from dwarf-sideways self-pollinated plants have a myriad of adherent, knotted, dwarf, and loss-of-sense-of-which-way-is-up phenotypes. a) Young plants showing only slight indications of anything going wrong. After looking mostly normal for the first three-to-five leaves, the internodes of these plants never elongate. b) By 42 DAP, these plants are drastically shorter than their wild-type B73 counterparts (shown on the left side). The leaves begin to stick together (a slight adherent phenotype), and the apical meristem no longer knows which way is upward. c) In fact, when I was trying to pollinate these plants, it was common that their tassel was pointing toward the ground.

Figure A.27. The epiNILs also showed an incredible susceptibility to mold and an overall decreased resistance to pathogens. Approximately 25% of the pollinations attempted in these materials in 2014 and 64% of pollinations attempted in 2015 failed to create viable seed, with most having the ‘mummified’ ear appearance. Some could be shucked as normal, but all seeds were moldy and non-viable (left), but some had the ear sheath completely stuck to the ear (right). Some of the shuck is peeled back on a few for documentation purposes.

Figure A.28. A wide variety of tassel mutant phenotypes were seen in 2014.

Figure A.29. Instead of tassels bearing anthers, some plants resembled the phenotype of the tasselseed mutant.

Figure A.30. Many plants showed developmental misregulation in 2014. From the upper left, clockwise: this plant grew from two apical meristems after not developing a main stem; this plant grew an ear from its apical meristem; as a seed was germinating, the apical meristem started to develop two plants; two plants started growing from the same seed; plant with particularly stiff leaves rubbed the soil smooth; a harvested ear had developed two tips.

Figure A.31. Adherent phenotypes were seen from early stages resembling buggywhipping even into maturity where a plant would end its chances of being able to reproduce. Some plants were stunted because of this adherence or it was coupled with a dwarfing phenotype.

Figure A.32. One plant grown in 2015 had an unusually low ear placement and many more leaves above the ear node than is typical of B73.

Figure A.33. Leaf morphologies were varied in this population, some plants having much more narrow leaves than B73 (upper left and lower right). An additional variety of dwarfing phenotype is seen in the upper right. The lower left photograph shows another plant that produced two leaves but no main stem, and then two stems were initiated.