DEVELOPING THE FOUR-PARENT MAIZE POPULATION: PHENOTYPIC AND GENOTYPIC CHARACTERIZATION

A Dissertation

by

ADAM LYLE MAHAN

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Seth C. Murray
Committee Members,	Patricia Klein
	William L. Rooney
	Amir Ibrahim
	Devin Nichols
Head of Department,	David D. Baltensperger

May 2015

Major Subject: Plant Breeding

ABSTRACT

In order to gain a better understanding of the effects and amount of recombination that can be available to future plant breeding endeavors, a unique mapping population was created. A di-hybrid maize (*Zea mays L.*) cross was used in this study to create inter-crossed recombinant inbred lines (IRILs). The four parents include two with blue aleurone, one with high oleic acid, and a highly aflatoxin resistant publicly available inbred, Tx772. The 1291 line population consisted of six sub-populations with various levels of crossing and inter-crossing which afforded the opportunity to examine an array of mapping population designs for phenotypic diversity, as well as genetic mapping effectiveness. An especially important outcome of this study for future maize genetic studies was to determine whether it is more prudent to increase the size of mapping populations, to increase the number of parents, or to add additional cycles of intermating. The added increased effective recombination in several of the sub-populations produced greater diversity of phenotypes and an increase in quantitative trait locus/ loci (QTL) accuracy and resolution in some cases. Phenotypic variation and genetic mapping resolution were to a greater extent affected by population size.

Significant variation was observed for days to anthesis and silk, plant and ear height, leaf rolling, as well as cob and kernel color. Questions and outcomes that this population addressed concerning practical plant breeding applications included the added difficulty that various mapping population designs create, and that a direct relationship between increased phenotypic and genotypic variation and additional

parents or intermating cycles did not seem to exist. Instead, variation was specific for the population and target trait in question. From a molecular and genetic mapping standpoint, the 4way population without intermating detected the gene for cob color, while the only other two groups that detected it were the 4way3sib, as well as all 4way subpopulations combined; both which had the largest population sizes, excluding the entire population which contained bi-parental populations. Because only one of the four parents, B73*Olc1*, expressed red cob, this observation may prove useful for future studies where a target trait is under represented in available germplasm. This indicates that larger population sizes, more so than exotic population designs may lead to improved results. Due to the incorporation of diverse mating designs, this population is a unique resource for future research including additional phenotypic analysis, genetic map construction, QTL linkage mapping, and analysis of recombination rate.

DEDICATION

To my wife Jennifer for showing me love, support and understanding while I follow my dreams. Also to my parents for providing the childhood necessary to mold me into the man I have become.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Seth C. Murray, for providing me the opportunity to obtain my graduate education and for the many times he pushed me to learn new things and grow as a person both professionally and personally. Thanks also to my committee members, Dr. Patricia Klein, Dr. William Rooney, and Dr. Amir Ibrahim for advising me throughout my research and providing answers to the many questions that arose. If it wasn't for the use of Dr. Klein's laboratory resources, her firsthand knowledge, as well as the countless hours she spent teaching me, as well as analyzing data for my project, I have no doubt that my dissertation, as well as my PhD education would not have become the success that it is today.

As a graduate student support group, I was lucky to start my graduate career sharing office space with Gerald De La Fuente and Dr. Ivan Barrero-Farfan. The three of us spent a lot of time on each other's projects and I thank them for their help and camaraderie. Also, thanks to the current graduate students in the corn lab: Steven Anderson, Schuyler Smith, Justine Christman, Yuanyuan Chen, Zoran Ilievski, Nancy Wahl, and Rupa Kanchi for all they helped me with.

I also received a lot of support from many fulltime staff members. The support I received from Jacob Pekar, Jeff Savage, and Jim Wilborn during planting, field maintenance, and seed processing were crucial to managing the large amount of field space that my project required year in and year out. Thank you to Natalie Patterson for

teaching me the laboratory techniques necessary for the molecular work I conducted, and for helping me through every step of plant DNA preparation.

I have been very fortunate to be surrounded by many great undergraduate student workers. Although they are often required to perform remedial tasks that they may not always find exciting, they show up to work every day and do what is asked of them. Thank you to Ryan Mchugh, David Rooney, Travis Rooney, Davis Michalewicz, and Amee Bumguardner for putting up with me as your "boss" when you were helping me with my research. Special thanks also to Dr. Allen Wright for donating seed for B73*Olc1*.

Of course, none of my graduate research or education would be possible without financial contributions. Thank you to the Monsanto Fellowship in Plant Breeding, the Texas A&M University Whole Systems Genomics Initiative Traineeship, and the Tom Slick Fellowship for funding my tuition and large portions of my stipend. Thank you to the Texas Corn Producer's Board, Texas AgriLife Research, and the Texas AgriLife Monocot Improvement Award for funding my research.

Lastly, thanks to my parents and extended family that I am blessed to be a part of for their thoughts and prayers that has given me the strength and courage to pursue graduate education.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

Page

1. INTRODUCTION

Crosses are conducted by breeders to create variation and to attempt to combine the best alleles from two (or more) parents from which the superior progeny are selected. In hybrid crops such as maize, a traditional population design scheme includes two parents, creating what is known as a bi-parental population. Bi-parental populations require fewer crosses resulting in less labor and time for population creation, and allow for straight-forward analysis of progeny. Where the bi-parental populations are weakest (limited to variation between two parents, limited recombination), multi-parent populations and/or intermating methods are strongest. The advanced intercross (Darvasi and Soller, 1995) and the extension of this idea to multi-parent designs (Mott et al., 2000) in mice helped to eliminate the narrow focus of bi-parental populations and provide the opportunity to determine the genetic control of traits with multiple alleles/genes. Multi-parent advanced generation intercross (MAGIC) populations have been developed for increased resolution and eventual cloning of QTLs for quantitative traits in multiple plant species (Cavanagh et al, 2008; Huang et al., 2012; Bandillo et al., 2013; Balasubramanian et al., 2009). Increasing the number of founding parents increases number of traits that can be observed in the subsequent progeny, while also increasing recombination and decreasing linkage disequilibrium within subsequent progeny. Recombination can be increased by intermating before inbreeding, shown in the creation of the intermated B73 x Mo17 (IBM) population with a resulting increase in genetic map distance of 3.86 fold (Lee et al., 2002).

Recombination of alleles is one of, if not the limiting factor for the development of superior progeny. Effective meiotic recombination events (EMREs) are allelic recombination events within progeny that can be detected by phenotypic and/or genotypic methods. Next-generation sequencing approaches can increase the number of markers and individuals that are able to be screened efficiently, thereby taking advantage of the additional recombination available from multi-parent and intermated population designs. While early genotyping technologies such as restriction fragment length polymorphisms or simple sequence repeat markers were often cost prohibitive at more than $1 per marker per individual (Bernardo and Yu, 2007), genotyping-by-sequencing produces millions, if not billions of nucleotides per sequencing run (Shendure and Ji, 2008), reducing the overall cost per data point to conservatively $0.0004 per marker per individual for this study.

It has been shown with the MAGIC and IBM populations that increasing the number of parents and/or the generations of intermating in a breeding program will increase effective recombination in the resulting progeny. What has not been directly evaluated is the comparison within an interconnected population of the value of increased parent number and/or the addition of intermating prior to inbreeding as it relates to field and molecular plant breeding.

To that end, the primary objectives of this dissertation included: 1) developing a unique multi-parent population consisting of subpopulations representing varying generations of mating using a four-parent cross in maize; 2) collecting phenotypes for agronomic and seed color traits; 3) adapting a digital genotyping method from sorghum

to prepare maize DNA libraries and collect SNPs across the maize genome via genotyping-by-sequencing; 4) determining major gene control of blue aleurone in maize kernels; and 5) analyzing the subpopulations for differences in genetic mapping accuracy and resolution for kernel and cob color using association mapping methodology.

2. LITERATURE REVIEW

Parent selection is of the utmost importance when developing a mapping population. The quantity of polymorphisms possible in any mapping population is determined by how polymorphic the parents are to one another (Anderson et al., 1993). Parents chosen for greater allelic diversity are generally observed to produce more diverse progeny than that of more closely related parents (Hung et al., 2012). Parent selection is also critical in determining which traits can be measured and mapped for QTL, which is why target traits and parents polymorphic for said traits must be identified before population creation. If the chosen parents are too diverse for non-target traits, variation attributable to these additional traits may distort the results for target traits. For example, flowering time variation often distorts many traits of interest when plants of wide ranging maturities are compared. For the purpose of QTL mapping, it has been common to make a cross between two parents, and self for several generations until near homozygosity, producing bi-parental recombinant inbred lines (RILs).

2.1 QTL Mapping Population Design and Consequences

RILs are useful scientific tools for many reasons including but not limited to; they are immortalized and thus need only be genotyped once, multiple plants can be phenotyped from each individual to reduce error from individual, environmental, and measurement variability and finally, additional effective recombination breakpoints allow for higher mapping resolution (Broman, 2004). These populations have successfully identified many QTL linked to traits of interest (Korte and Farlow, 2013;

Wurschum, 2012). Unfortunately, many of these mapping populations identify large effect, qualitative genes that can often be just as easily phenotyped or are already under selection; or identify very large segments of the chromosome which make marker assisted selection and gene cloning challenging. Two of the most important factors in determining how close to the causative gene or non-gene element a QTL can be detected are the recombination rate within individuals of the mapping population and the size of the population. A method of increasing detectable / effective recombination is to intercross related progeny before inbreeding, creating intermated recombinant inbred lines (Balasubramanian et al., 2009). The intermated B73 x Mo17 (IBM) population highlighted this phenomenon with a 3.86-fold increase in genetic map distance through the intermating of two of the most historically important maize parents B73 and Mo17 (Lee et al., 2002). Other examples of mapping populations that performed intermating along with multi-parent designs include the MAGIC populations in wheat (Huang et al., 2012; Mackay et al., 2014; Verbyla et al., 2014), and the collaborative cross in mice (Collaborative Cross Consortium, 2012). In terms of genetic mapping, avoiding intentional selection and choosing random plants to intermate is crucial since random mating following a one-locus model can lead to a 50% increase in genetic variation within BC_1 populations in maize (Arbelbide and Bernardo, 2004).

Other methods for increasing QTL mapping resolution and phenotypic variation are increasing the number of parents beyond traditional bi-parental population designs (as with the MAGIC populations) and by increasing the number of individuals in the population (Lu et al., 2010). The number of parents can be increased by a multiple-line

cross (Yan et al., 2008) where three or more inbreds are mated and selfed out to RILs to form multiple populations or by di-hybrid mating to create four or eight-way crosses (Xu, 1996). Larger mapping population sizes increase the probability of observing novel cross-over events in the progeny, and the unique phenotypic and genotypic differences that may exist within the population on the tails of a normal distribution. Because of this, large population sizes of 500 or more individuals are required to map small, quantitative QTL (Wurschum, 2012). However, to date, direct comparisons of multiple mating designs and population sizes have not been made and few, if any, populations of 500 individuals have been investigated for this purpose.

2.2 Blue Aleurone

Genetic diversity in maize includes an array of kernel colors (red, blue, and purple), which commercially are minimally used specialty corns compared to yellow and white maize. Anthocyanins, flavonoids, phenolic acids, etc. are antioxidant phytochemicals produced as secondary metabolites that often condition these kernel colors (Del Pozo-Insfran et al., 2007). Antioxidants have been linked to anti-cancer (Hyun and Chung, 2004; Zhao et al., 2004) and other anti-inflammatory health benefits (Tsuda et al., 2003). Blue maize is used to produce blue maize chips and tortillas and this appears to be a growing industry (Betran et al., 2000; Lopez-Martinez et al., 2009). Because blue pigment is primarily found in the aleurone layer, it can easily be dominated by competing pigments such as yellow endosperm to create a grayish/green tortilla chip that is aesthetically unpleasing to consumers (Salinas-Moreno et al., 2007).

Anthocyanins (red and purple), carotenoids (yellow), and white pigment have been genetically characterized as controlled by dominant alleles for red aleurone (*pr1*), colored aleurone (*c1*), colored (*r1*), and yellow endosperm (*Y1*) which is dominant over white (*y1*). The genetics behind blue and dark-purple maize is comparatively poorly understood and no genes have conclusively been reported (Ford, 2000). Additional factors controlling color expression within the anthocyanin pathway include *a1*, *a2*, *bz1*, *bz2*, *c2*, and *vp1* (Piazza et al., 2002; Selinger & Chandler, 1999; Hattori et al., 1992; Styles and Ceska, 1972). Eight factors are enzyme-related genes (*a1, a2, bz1, bz2, c2, chi*, *pr*, and *whp*) which are used to catalyze the production and transport of anthocyanin as well as five genes responsible for regulation (*b, c1, pl, r, and vp1*) which control expression of anthocyanins in specific plant organs (Hanson et al., 1996). Specifically the *c1* gene is responsible for the actual production of pigmentation in the seed tissue while *r1* regulates which tissues will express pigmentation (Petroni et al., 2000). Although the *c1* gene refers to a single gene, the *r1* gene region is a complex of multiple *r1* alleles which result in specific pigmentation patterns. These structural and regulatory genes have only been investigated as major gene mutants and it is unclear what, if any, of these genes can be used as selection criteria toward improving commercial blue maize production in fixed blue aleurone breeding lines. Blue maize developed by Texas AgriLife Research is a rare example of inbred breeding efforts in place for the improvement of temperate and subtropical colored maize and provides an excellent opportunity to further study blue color by incorporating blue aleurone inbreds with high yielding adapted yellow maize into a mapping population. Both blue parents in this

population (Tx903 and Tx906) were also diallel parents for A. Mahan's M.S. work (Mahan et al., 2013).

2.3 Recombination

Recombination contributes to increased diversity through novel combinations of alleles that produce diverse gametes that differ in DNA sequence. This provides for natural selection by segregating advantageous and deleterious genes independently which can be selected upon in the progeny (Liu et al., 2012). Genetic recombination is crucial for creating opportunities to select for useful diversity within a germplasm pool but also has knowledge gaps that have limited the use of molecular techniques in plant breeding. Of particular interest to plant breeders, geneticists, and this study are effective meiotic recombination events (EMREs) which result in swapping of non-homologous genetic material between individuals which can be observed in progeny. The number of EMREs is always less than the total genetic recombination events observed and expected because recombination of homogenous chromosomal regions does not result in a change in phenotype or genotype in the resulting progeny. In other words, although recombination events almost certainly occur when inbred lines are selfed, there are no resulting detectable EMREs. The study of recombination and its effect on QTL mapping is also important in quantifying the genetic diversity, as well as novel allelic combinations that may lead to improved and selectable phenotypes related to desirable traits. An increased ability to select from novel allelic combinations through mating designs that provide the opportunity for increased recombination will lead to a more efficient use of breeder resources.

2.4 Genetic Characteristics of Maize

The sequence differences found between parents of a mapping population, known as allelic variation (Springer and Stupar, 2007), are crucial for QTL mapping to observe recombination and map traits. Maize was domesticated from its wild ancestor teosinte (*Zea mays* subsp. parviglumis) ~10,000 years ago (Hufford et al., 2012). Allelic variation and genetic diversity has developed from the variation that survived selection pressures from this domestication. Approximately 85% of the sequenced B73 maize genome is composed of transposable element groups, dispersed non-uniformly across the genome (Schnable et al., 2009). This widespread transposon activity as well as gene duplication contributes directly to variation among gene rich regions across maize lines, resulting in an estimated 10,000 non-shared (hemizygous) gene rich regions between maize inbred lines B73 and Mo17 (Springer and Stupar, 2007). This hemi-zygosity, also termed non-colinearity (Fu and Dooner et al., 2002) creates diversity but can also lead to sequence and genetic map inconsistency across lines.

Analysis of randomly selected sequences of B73 and Mo17 showed insertion-deletion polymorphisms every 309 bp and single nucleotide polymorphisms (SNPs) every 79 bp (Vroh Bi et al., 2005). Across all maize, there was once thought to be a polymorphism every 100 bp between randomly chosen inbreds (Tenaillon et al., 2001; Ching et al., 2002). That number has since decreased to 44 bp with improved genotyping technology (Gore et al., 2009).

Among the many facets of the maize genome is its highly replicated nature, which includes a duplication of a paleopolyploid ancestor approximately 70 million

years ago (Paterson et al., 2004) as well as a duplication 5 to 12 million years ago (Blanc and Wolfe, 2004; Swigonova et al., 2004), which separated maize from its close relative *Sorghum bicolor* (Paterson et al., 2009). Due in large part to these duplication events, an estimated one third of all maize genes are present multiple times in the genome (Gaut, 2001). Regions with high sequence similarity due to genome duplication can result in markers that will be mapped to multiple locations when compared with the B73 reference genome and are therefore generally discarded.

2.5 Next-Generation Sequencing

The B73 maize genome is predicted to contain 32,000 genes (Schnable et al., 2009) across a 2300 Mbp genome and many studies have characterized the vast number of polymorphisms present across different sets of germplasm. With the second installment of the maize haplotype map which genotyped 103 lines across domesticated and predomesticated germplasm, 55 million SNPs were identified (Chia et al., 2012). Several million polymorphisms (SNPs, INDELs, etc.) have been identified in another study among 27 diverse maize inbreds (Gore et al., 2009). Between 2,815 maize inbred accessions from the United States National Plant Germplasm System, 681,257 SNPs were discovered (Romay et al., 2013). These findings were all made possible with next generation sequencing technologies (NGS).

The success of plant breeding and the sequenced genomes of many species have led to a high demand for sequence data that spans the genome at a low cost per data point. Early genotyping marker technologies such as restriction fragment length polymorphism or simple sequence repeat markers cost more than $1 per marker per

individual (Bernardo and Yu, 2007). Technological advances have occurred to develop DNA marker and genotyping technology referred to as NGS which has led to methodology such as genotyping-by-sequencing (GBS) with related procedures termed RAD-seq and digital genotyping. GBS leverages parallel sequencing and imaging to produce several 100 million, if not several 100 billion nucleotides per sequencing run (Shendure and Ji, 2008). Current technologies that facilitate GBS include the Roche 454 FLX Titanium (Thudi et al., 2012), Illumina MiSeq and HiSeq2500 (Bentley et al., 2008), and the Ion Torrent PGM (Rothberg et al., 2011). The large quantity of SNP markers available across the genomes of many species now detectable with NGS include sorghum (Morris et al., 2013), wheat (Poland et al., 2012), and maize (Elshire et al., 2011). This has led to the use of GBS to identify and exploit them for high resolution mapping. The detection of SNPs using GBS is done by digesting high quality DNA with select restriction enzymes. Methylation sensitive enzymes are often used to target gene rich regions of the genome due to the relationship between highly methylated DNA and non-coding regions (Yisraeli and Szyf, 1984; Schwartz and Dennis, 1986). Focus on the gene rich regions is especially important in a crop that has a highly repetitive, highly methylated genome such as maize. Protocols for template DNA preparation for GBS pipelines follow similar steps with subtle differences in restriction enzymes used, amount of DNA digested, number of samples in a GBS lane, purification methods, and PCR settings (Elshire et al., 2011; Morishige et al., 2013). Once DNA is digested, restriction-site associated DNA (RAD) barcodes (Baird et al., 2008; Lewis et al., 2007; Miller et al., 2007) can be applied which serve as identification tags for individual DNA

samples to facilitate the pooling of multiple samples for GBS. Restriction sites are leveraged to create short lengths of sequence which are read by next-generation sequencing instruments like the Illumina HiSeq2500, which generates short reads ranging from 50 to 300 bp which are later used to discover SNPs in downstream sequence analysis.

2.6 Genetic Map Creation

Genetic maps are a set of markers, with known distances between them, serving as benchmarks along partial or entire chromosomes (Cheema and Dicks, 2009). A critical statistic in determining if markers within a genetic map are linked is the logarithm of odds (LOD) score. The LOD score (Haldane and Smith, 1947; Morton, 1955) compares the amount of recombination between two markers under the assumption that there is linkage (< 50% recombination rate) against the same amount of recombination between two markers under independent segregation (~50% recombination rate). Linkage map construction consists of three parts: grouping markers by linkage, ordering of the markers within each linkage group, and spacing the markers accurately so that linkage block length can be determined (Cheema and Dicks, 2009). Programs such as MAPMAKER (Lander et al., 1987), MadMapper (Kozik and Michelmore, 2006), and MST_{MAP} (Wu et al., 2008) have often been used to group markers in a bi-parental population. To order markers, it is important to determine the quality of the marker order, and determine if and how one marker order is better than another (Cheema and Dicks, 2009). JoinMap is a program which can group markers and also has options for multiple statistical methods in determining marker order (Stam,

1993). In addition to these programs which are used to individually deal with marker grouping and ordering, Map Manager QTX (Manly et al., 2001) and THREad Mapper (Cheema et al., 2008) work through both steps simultaneously, which may or may not improve genetic map quality. Additional resources within the environment of R include the statistical packages R/qtl (Broman et al., 2003) and R/mpMap (Huang and George, 2011) which serve to handle mapping populations with greater than two parents.

2.7 Linkage Mapping Analysis

Traditional bi-parental mapping populations have been used for the development and application of most QTL analysis statistical software and computing programs and these applications are straightforward. The simplest method of QTL mapping uses a linear regression between a single marker and a phenotypic trait (single marker analysis). The main deficiencies of this method are that many markers may exhibit some level of linkage to a QTL, and the linked markers may not be allelic to the QTL, so QTL position and effect cannot be fully determined (Kearsey and Farquhar, 1998). Methods with increased complexity such as interval (Lander and Botstein, 1989) and composite interval mapping (Zeng, 1994) have been developed to use the power of the genetic map and to take advantage of the fact that two linked markers should share similar results. Intervals composed of adjacent pairs of markers along each chromosome are analyzed and the likelihood of a QTL being associated to a particular interval is determined with a LOD score. For increased accuracy of QTL effects, composite interval mapping can be used to combine multiple regression and interval mapping methods (Zeng, 1994).

The statistical methods for analyzing traditional bi-parental populations include simple least squares (Haley, 1999), weighted least squares (Han and Xu, 2008), maximum likelihood (Zhang and Xu, 2005), and Bayesian methods (Xu et al., 2009; Yandell et al., 2007).

While bi-parental populations segregate at two alleles per locus, di-hybrid or multiline cross populations have the potential to segregate for multiple alleles per locus, increasing the resolution to detect desired trait QTL (Liu and Zeng, 2000). However, there are a number of issues that complex mating designs present over bi-parental populations such as irregular inbreeding due to intermating, multiple generations of mating and inbreeding, accounting for the possibility of multiple alleles and additional segregation distortion (Yi and Xu, 2002). When the parents of a complex mating design are not closely related, traditional bi-parental cross mapping may be used directly with mixed results, due to common lines sharing a percentage of common alleles. This genetic structure is not accounted for when applying bi-parental analyses methods in multi-parent mapping (Liu and Zeng, 2000). Therefore, for multiple cross and multiple parent populations, other statistical methods have been developed which include extending the composite interval mapping method from crosses of two inbred lines (Liu and Zeng, 2000) and leveraging the R statistical package HAPPY to look at the structure of the genomic sequence of each line as a mosaic of the parents, or founders of the population (Kover et al., 2009).

An inability to know which parent is responsible for each allele in multiple parent crosses directly effects the ability to create genetic maps and map QTL. Computer

software has been developed to attempt to handle multiple parents (Huang and George, 2011; Hu et al., 2012).

When performing linkage mapping, segregation distortion can be observed when allele frequencies at a locus or region are statistically greater or less than expected and are the result of unintentional human (or natural) selection. Segregation distortion can also create spurious linkages and inaccurate mapping results due to an effect on marker order (Balasubramanian et al., 2009). This is why when creating mapping populations, especially those with extra generations of crossing such as intermating, it is important to randomly choose plants to cross, as well as randomly choose ears and kernels with which to advance the next generation.

2.8 Association Mapping

Association mapping is a statistical approach to mapping QTL that can be used for unstructured populations such as single linkage mapping populations and structured populations such as those assembled from diverse individuals. Compared with QTL linkage mapping, association mapping of diverse assembled populations has an advantage in that it explores historical recombination events, mutations, and all alleles lending itself to studying greater diversity (Yu and Buckler, 2006). This statistical approach is also relevant to multi-parent populations that may not fit traditional linkage methods which require the construction of a genetic map.

A well-developed study combining linkage populations and association analysis jointly was performed using the nested association mapping population in maize (NAM) (Yu et al., 2008; Buckler et al., 2009; McMullen et al., 2009). The NAM population

consists of 25 diverse inbred lines crossed to the B73 maize reference genome line, creating 25 families of ~200 lines. Analysis of the 4,699 RILs discovered ~136,000 recombination events (McMullen et al., 2009). For the NAM study, a 4-6 cM sliding window analysis, where a snapshot of sequence data is taken, was used to analyze recombination events and differences across the 25 populations. This analysis revealed hot and cold spots of recombination throughout the genome and it also revealed that recombination events were statistically different in different maize lines. However, no QTL for recombination rate could be detected. The vast numbers of markers that are available with next generation sequencing and genotyping technologies have also led researchers to use the sliding window approach to search dense marker regions for specific trait associations beyond single marker analysis. Graphical Assessment of Sliding P-values (Mathias et al., 2006) is a recent tool which can help to better visualize and understand the output from a sliding window analysis.

2.9 Using Association Analysis in Multi-Parent Mapping Populations

An important component of the association mapping approach is to control for structure present within germplasm. Breeding efforts have led to subpopulations within maize germplasm which differ for alleles and allele frequencies of certain traits, leading to unexpected linkage disequilibrium (LD): the decreased probability of alleles which are closer to one another of being separated by recombination (Yan et al., 2011). Population structure would not be expected in a 4-way cross population because multiple recombination events and independent assortment would have broken linkages in the

population that were within each of the four parents and allele frequencies should be relatively balanced within this type of population.

An increase in intermating and in the number of parental lines in population design decreases the size of linkage blocks between alleles and decreases the effect of LD. However, because multi-parent populations typically use four to eight parents, LD is expected to be much higher than reported in association studies assembled from diverse panels. Genetic drift, natural selection, and interbreeding of isolated populations generally result in rapid LD breakdown in diverse panels (Flint-Garcia et al., 2003, Gaut and Long, 2003; Yu and Buckler, 2006) and is a main factor to consider when analyzing populations with association mapping methods. One drawback to lower LD and/or increased recombination events is the need for a larger quantity of markers to obtain increased resolution for trait mapping (Morgante and Salamini, 2003).

Elite germplasm with a relatively narrow genetic base has a slower decline in LD (Ching et al., 2002). Similarly, in multiparent linkage mapping, the closer related the parents, the less diversity that will be available to map crossover events and the greater the LD within the population will likely appear. These findings are relevant to breeders using breeding pools that have undergone artificial selection bottlenecks and also suggest the ability of multi-parent populations in providing additional genetic variation and crossover events to decrease LD and by extension, increase genetic mapping resolution and phenotypic variation for selection.

3. CREATION AND PHENOTYPIC CHARACTERIZATION OF A MULTI-PARENT MAIZE POPULATION

To measure the effectiveness of multi-parent population design in maize, an unprecedented, four-parent maize (FPM) population was developed using a series of different mating designs. The FPM population incorporated up to three generations of intermating to allow for comparison of traditional bi-parental, multi-parent, and multi-parent intermated populations for phenotypic diversity. A total of 1,291 inbred lines were measured with at least one replication among 5,551 total plots across two inbred trials in College Station in 2013 and three in 2014. These trials were phenotyped for days to anthesis and silking, plant and ear height, leaf rolling, and cob and kernel color. Significant genetic variation was found for all traits analyzed, as well as substantial environmental variation for days to anthesis and days to silk. Overall, single replications of population entries performed well for agronomic trait analysis as indicated by low residual variation. Although flowering time means did not show a particular relationship to specific subpopulations, plant and ear heights in some of the 4-parent populations were significantly higher and more variable than the bi-parental populations, with some of this attributable to larger population size. The creation of this population and the analysis of these traits showcases the advantages of traditional and complex mating designs and reveals the potential to research these and other traits with this population resource.

3.1 Introduction

Populations developed for both genetic improvements as well as genetic characterization have primarily been centered on the use of bi-parental populations derived from two founding parents. Bi-parental populations are used due in part to their ease of development (fewer crosses needed), the speed of population creation, and the simplicity of comparing the progeny for improvement against the two elite parents. Major weaknesses of bi-parental populations are that the resulting progeny are limited to the variation between the two parents, and a limited opportunity for recombination which may lead to an insufficient amount of phenotypic variation from which the breeder can perform selection on.

Parent selection is of the utmost importance when developing a population. The quantity of polymorphisms is determined by how diverse the genomes of the parents are to one another (Anderson et al., 1993); however progeny can, and often do, show transgressive segregation beyond the genetic diversity of the parents. Parent selection is also critical in determining which traits segregate and their proportion of variation. For genetic mapping, target traits and parents segregating for these traits should be identified before population creation, although transgressive segregation will be valuable for some traits. More complex population designs including six to eight parents were discussed some time ago (Allard, 1960; Jensen, 1970) but have been little used in breeding and never used in mapping population design until recently. Before the advent of next generation sequencing technology, the majority of genetic studies used a small number

of individuals with relatively small numbers of markers which would have been poorly equipped to take advantage of a multi-parent mating design.

Additional generations of crossing in multi-parent populations have been implemented in *Arabidopsis* and wheat (Cavanagh et al., 2008; Huang et al., 2012; Kover et al., 2009). These populations are referred to as MAGIC populations. One major advantage to using multi-parent populations from a field breeding perspective is the greater number of traits that can be targeted. With this advantage in mind, and the proven success of MAGIC populations, this study takes the population design complexity a step further by incorporating generations of intermating prior to inbreeding. The IBM population highlighted this approach in maize resulting in a 3.86-fold increase in genetic map distance by intermating progeny of two of the most historic maize parents B73 and Mo17 for three generations before inbreeding (Lee et al., 2002). Herein, we discuss the creation of a population that melds the MAGIC and intermated designs together, while also allowing for comparisons among various mating designs. We hypothesized that a multi-parent population would have greater phenotypic variability than traditional bi-parental populations. Furthermore we hypothesize that additional generations of intermating would break linkages to expose additional combinations of alleles and thus greater phenotypic variation.

The objectives of this research were: 1) to develop a novel multi-parent population with subpopulations varying in mating design methodologies; 2) to collect phenotypic data on agronomic traits and cob and kernel color; and 3) to compare phenotypic variation present across subpopulations.

3.2 Materials and Methods

3.2.1 Four-Parent Maize Population Construction

The four inbred lines Tx903, Tx906, Tx772, and B73*Olc1* (Figure 1) were selected as founders for their unique phenotypic contributions.

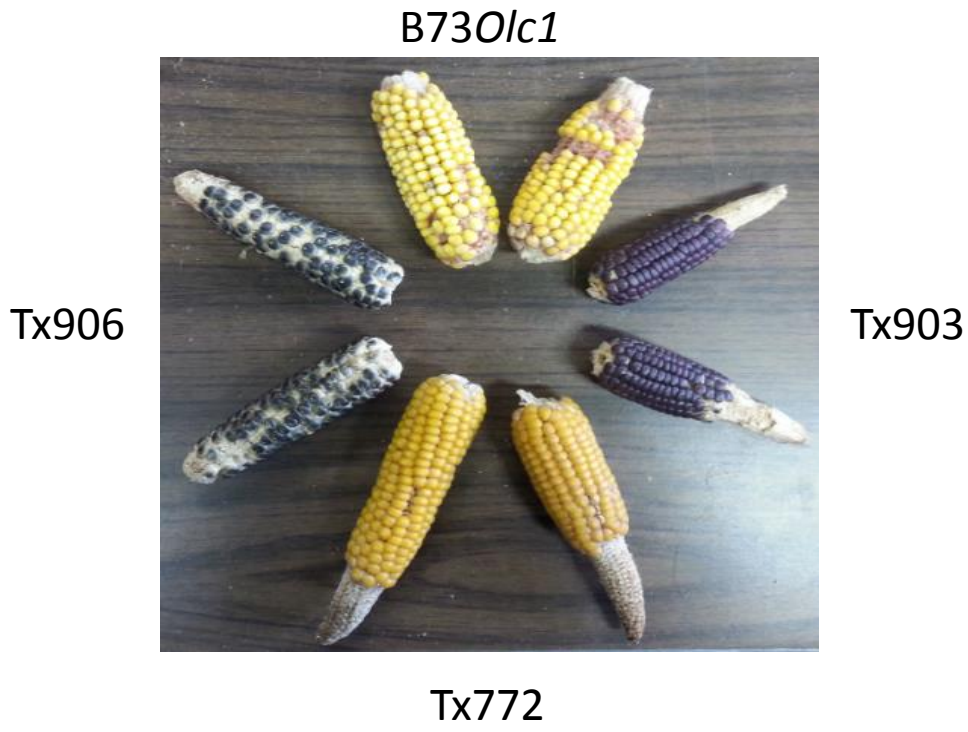

Figure 1. Examples of maize ears typical of a Texas environment from all four founding parents.

Parents Tx903 ((Lfy2361-B/Tx114 (B73w)-B Dark blue-B)Tx114/Lfy2304-B-B-B-1-3-B-B-B-3-B-B) and Tx906 (Ethiopia15-B-5-1-B-B2-B-1-B-B-1) are two unrelated

sources of the blue aleurone trait that is used in blue tortilla products and provided the opportunity to further analyze the genetic control of blue kernel color. Tx772 (Tx772-B-B-B-B-B-1) (Llorente et al., 2004) incorporated aflatoxin resistance into the population which helped keep the grain cleaner (less aflatoxin accumulation) but also provides the opportunity for future research into aflatoxin resistance. The fourth founding line, B73*Olc1*, is expected to be isogenic to the maize line from which the maize reference genome was drafted (Schnable et al., 2009), but also has an ethyl methanesulfonate (EMS)-induced mutation that nearly doubles the oleic acid content (Wright, 1995) providing another unique trait for further investigation.

The four parent maize (FPM) population was derived from a four way cross of the hybrids B73*Olc1*/Tx903 and Tx772/Tx906. The FPM population consists of six subpopulations derived from these hybrids, the subsequent di-hybrid and different generations of intermating before inbreeding (Figure 2) totaling 1291 phenotyped individuals.

The complete course of action taken to develop this population and divide it into subpopulations is detailed in Table 1. Briefly, the original bi-parental F_1 hybrids were produced in College Station, TX (CS) in 2009, and the di-hybrid was created in the subsequent Weslaco, TX (WE) 2009 winter nursery. The multi-parent individuals underwent three successive generations of intermating with the final intermating in CS 2011 producing 430 four-parent, three sibling intermated individuals. It was in CS 2011 that selfing began on all other subpopulations (4way2sib, 4way1sib, 4way0sib, B73*Olc1* x Tx903, and Tx772 x Tx906). In the WE 2011 winter nursery, all subpopulations

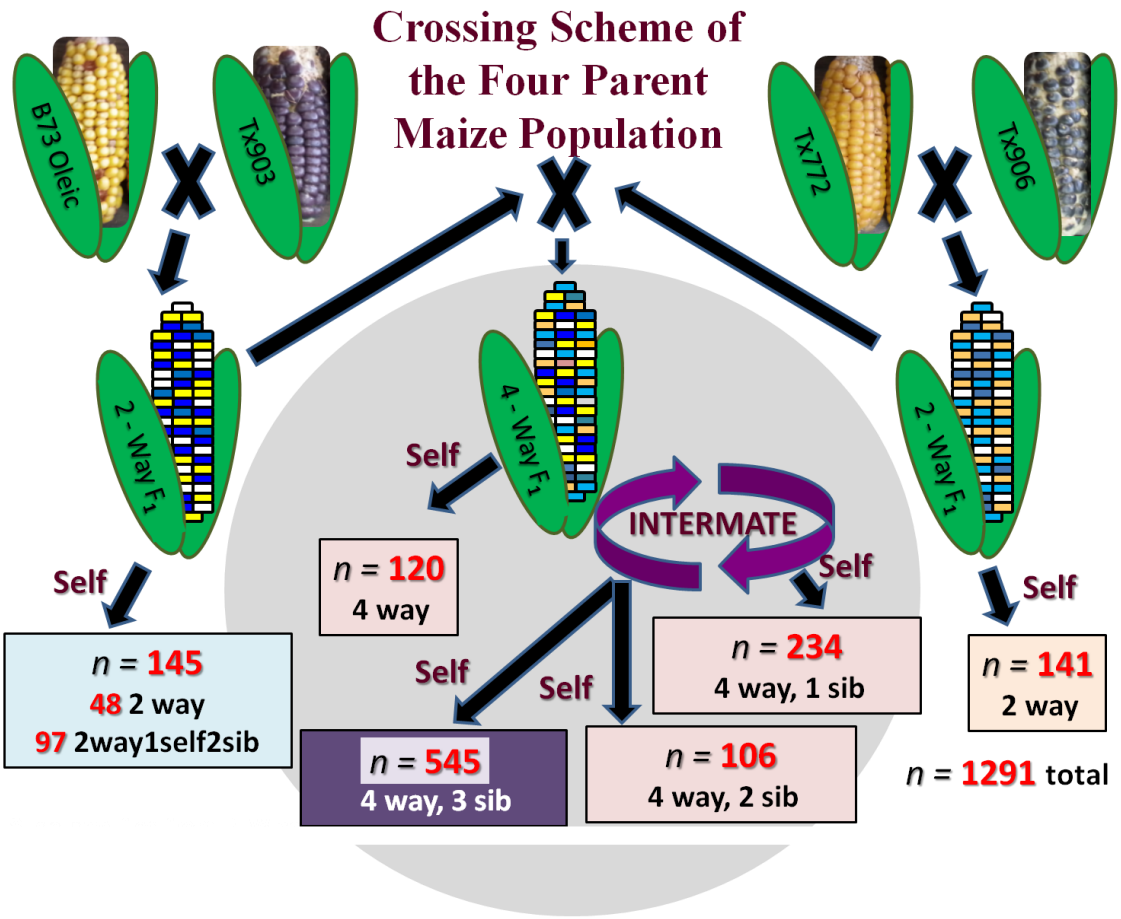

Figure 2. Mating design, creation of subpopulations, and number of phenotyped individuals.

were selfed, with some plots having two individuals randomly selected to increase the size of a respective subpopulation before inbreeding. This included selecting two ears at random from 147 plots of the 4way3sib population, from 21 plots of 4way2sib, from 84

Table 1. Sequence of events during the development of the four-parent maize population and resulting subpopulations.

Subpopulation	College Station 2009	Weslaco 2009	College Station 2010	Weslaco 2010	College Station 2011
4way3sib		Di-hybrid cross	Intermated, harvested 95 ears	Bulked 3 kernels from each CS-10 ear, planted 8 plots and intermated within and between plots, harvested 78 ears	Intermated between 79 plots, harvested 430 ears
4way2sib					114 selfed ears harvested
4way1sib					Selfed 89 plots, harvested 295 ears
4way					Selfed 6 plots, harvested 49 ears
B73*Olc1* x Tx903	hybrid cross				Selfed 3 plots, harvested 15 ears
Tx772 x Tx906	hybrid cross				Selfed 6 plots, harvested 40 ears
B73*Olc1* x Tx9031self2sib			1 plot selfed	1 plot intermated, 8 ears harvested	intermated 8 plots, harvested 41 ears
Total					984 individuals

Table 1. Continued.

	Weslaco 2011	**College Station 2012**	**Weslaco 2012**	**College Station 2013**
4way3sib	Planted 428 plots, selfed, harvested 575 ears.	Planted 575 plots, selfed, harvested 575 ears.	Planted 575 plots, selfed, harvested 545 ears.	Planted 545 nursery plots, selfed, harvested 527 ears. Planted 502 yield trial plots.
4way2sib	Planted 112 plots, selfed, harvested 173 ears.	Planted 133 plots, selfed, harvested 133 ears.	Planted 133 plots, selfed, harvested 106 ears.	Planted 106 nursery plots, selfed, harvested 101 ears. Planted 89 yield trial plots.
4way1sib	Planted 295 plots, selfed, harvested 253 ears.	Planted 253 plots, selfed, harvested 253 ears.	Planted 253 plots, selfed, harvested 235 ears	Planted 235 nursery plots, selfed, harvested 228 ears. Planted 205 yield trial plots.
4wayF6	Planted 49 plots, selfed, harvested 133 ears.	Planted 133 plots, selfed, harvested 133 ears.	Planted 133 plots, selfed, harvested 120 ears.	Planted 120 nursery plots, selfed, harvested 116 ears. Planted 110 yield trial plots.
B73*Olc1* x Tx903-F5	Planted 15 plots, selfed, harvested 55 ears.	Planted 55 plots, selfed, harvested 53 ears.	Planted 53 plots, selfed, harvested 52 ears.	Planted 52 nursery plots, selfed, harvested 49 ears. Planted 49 yield trial plots.
Tx772 x Tx906-F5	Planted 39 plots, selfed, harvested 140 ears.	Planted 140 plots, selfed, harvested 140 ears.	Planted 140 plots, selfed, harvested 137 ears.	Planted 137 nursery plots, selfed, harvested 134 ears. Planted 124 yield trial plots.
B73*Olc1* x Tx9031self2sib	Planted 41 plots, selfed, 103 ears.	Planted 103 plots, selfed, harvested 103 ears.	Planted 103 plots, selfed, harvested 97 ears.	Planted 97 nursery plots, selfed, harvested 91 ears. Planted 86 yield trial plots.
Total	1432 individuals	1430 individuals	1291 individuals	1246 individuals

plots of 4way0sib, from 40 plots of B73*Olc1* x Tx903, from 101 plots of Tx772 x Tx906, and from 62 plots of the B73*Olc1* x Tx903-1self-2sib population. For each generation of intermating and successive generation of selfing, one random ear was selected from each plot (representing one individual) and planted ear-to-row. All individuals were planted and selfed during the CS 2012 and 2013 summer nurseries, as well as the WE 2012 and 2013 winter nurseries.

3.2.2 Experimental Design of Nursery and Observation Trials

Phenotypes were measured within five separate trials of the population each planted in an augmented design. Table 2 includes information regarding the size of each trial. The number of plots per check is indicative of the number of blocks in each environment. The CS 2013 early and late nursery plantings were affected by weather and flowered at similar times. Significant differences were not found for flowering time and thus these two plantings were combined for analysis, leading to five environments.

Although the five trials differed in entry size, the fields were evenly blocked to help account for field variation with three check inbreds replicated twice within each block for a total of six check inbred plots per block. In CS 2013, the three checks were LH195 (Holden's Foundation Seeds, 1991), PHV63 (Pioneer Hi-Bred International, 1988), and LH82 (Holden's Foundation Seeds, 1985). In CS 2014, LH195 and PHV63 were used as checks, but LH82 was switched for PB80 (DeKalb-Pfizer Genetics, 1988). By design, entries were not replicated within a trial but replicated checks and multiple planting dates across two years were used in the statistical analysis.

Table 2. Total number of entry, parent, and check plots for six environments in College Station 2013 and 2014.

Trial	Entries	Parent plots	Check plots	Total Plots
CS 2013 nursery early planted	646	4 (1 per parent)	54 (18 per check)	704
CS 2013 nursery late planted	646	4 (1 per parent)	54 (18 per check)	704
CS 2013 observation trials late planted	1290	8 (2 per parent)	108 (36 per check)	1406
CS 2014 nursery	654	16 (4 per parent)	48 (16 per check)	718
CS 2014 observation trials early late planted	1173	16 (4 per parent)	78 (26 per check)	1267
CS 2014 observation trials late planted	1142	16 (4 per parent)	72 (24 per check)	1230
Total	5551	64	414	6029

Beginning in CS 2013, an early nursery planting of the FPM population was planted on March 4th, followed by a delayed nursery planting on March 13th. A second observation trial was planted a week later. For the CS 2014 growing season, three additional trials were conducted. An observation trial as well as a smaller nursery trial was planted on March 14th, and another observation trial was planted on March 28th. The nursery trials were single row plots 3.05 m in length and row spacing of 0.76 m. The observation trials were single row plots 6.40 m in length and row spacing of 0.76 m.

3.2.3 Phenotypic Measurements

Nursery trials were used to inbreed the population to RILs. Flowering time was taken in the 2013 nursery but not in the 2014 nursery, in addition to plant and ear height which were taken in both 2013 and 2014 nurseries. Once nursery plots were hand

harvested, Mendelian traits for cob color and kernel color were taken during seed cleaning and processing. Observation trials were not harvested, but flowering time and plant and ear height notes were taken. Plant height was measured in the field from the base of the plant to the tip of the tassel and ear height was measured from the base of the plant to the top ear node. Flowering time was measured by the number of days from planting to when 50% of the plants were either shedding pollen (days to anthesis) or silk (days to silk). Anthesis-silking interval was calculated by subtracting days to anthesis from days to silk. The amount of leaf rolling that was observed across the population was noted on a scale of 0-3 based on severity of leaf rolling; 0- no observable leaf rolling, 1- slight leaf rolling, 2- moderate leaf rolling, 3- excessive leaf rolling. The leaf rolling data was taken by the same person (ALM) across multiple measurement replications in the CS 2014 early and late observation trials twice in the morning and twice in the evening on separate days to account for subjective measurement error and time of day since multiple hours elapsed between the observation in the first plot and the last plot.

Self-pollinated ears from the nursery were shelled using a hand sheller, and a single, random ear was chosen to be planted each subsequent growing season. During shelling, cob color was recorded (red or white) as well as kernel color (yellow, blue, or white). At a later date, endosperm color (yellow or white) was confirmed on individuals with blue aleurone by cross sectioning several kernels to fully observe the endosperm. To determine if there were any significant variations in the intensity of yellow endosperm and blue aleurone that could not be detected by simply labeling the

individuals as yellow, white, or blue, a Chroma Meter CR-410 (Konica Minolta, Ramsey, NJ) was also used on kernels from the CS 2013 summer nursery. A colorimeter reports a set of three measurements to represent the color of each sample including "L" (lightness) which measures how white or black the sample is, "a" (red-green), and "b" (blue-yellow) (Jha, 2010). Hue ($\tan^{-1}(b/a)$) and chroma ($\sqrt{(a^2 + b^2)}$) are additional values that are calculated from the computer generated values of lightness, red-green, and blue-yellow. Due to a lack of sufficient seed for all individuals, a reduced set of 949 genotypes were analyzed with the Chroma Meter.

3.2.4 Statistical Analysis

Analysis of variance was performed using PROC MIXED in SAS 9.4 (SAS Institute, Cary, NC) with an all random model to produce variance component estimates so that percent of total variation could be determined from significant sources of variation. This was modeled as the phenotypic observation y_{ijk} on individual i in block j of trial k:

$$Y_{ijk} = \mu + e_k + g_i + (b/e)_{jk} + (g*e)_{ik} + \varepsilon_{ijk}$$

where μ is the grand mean; e_k is the random effect of environment k; g_i is the random effect of individual i; $(b/e)_{jk}$ is the random effect of block j nested in environment k; $(g*e)_{ik}$ is the random effect of the interaction between individual i and environment k; and ε_{ijk} is the random residual effect for individual i in the block j of environment k. In a two step analysis, the best linear unbiased predictors (BLUPs) adjusted by the random model were used for a means analysis using Duncan's multiple range at the $p < .05$ level to compare days to anthesis and silk, anthesis-silking interval, and plant and ear height

within environments by subpopulation. Parents and checks were compared separately for these traits using least-squares means from a fixed model. Heritability was calculated for agronomic traits by dividing the total genetic variation by the sum of the total genetic variation, genotype by environment variation, and the residual variation.

3.3 Results and Discussion

The four founder parents were first crossed in the CS 2009 nursery to create the founding hybrids. The initial crosses (B73*Olc1* x Tx903 and Tx772 x Tx906) were made so that each blue aleurone, white endosperm parent (Tx903 and Tx906) would be mated to a clear aleurone, yellow endosperm parent. The progeny of B73*Olc1* x Tx903 showed dramatically lower hybrid vigor, likely due to being closely related, with Tx903 having 50% parentage from Tx114, a white conversion of B73. As a result of B73*Olc1*'s lack of Texas adaptation, lower hybrid vigor, and fewer resulting F_2 seed, a smaller number of progeny were available after several generations of selfing (52 F_5 RILS). Because of the smaller size of the B73*Olc1* x Tx903 population at the F_2 stage, an additional subpopulation was created by intermating these F_2 progeny twice, before selfing to RILs; this added an additional 97 progeny derived from the two parent cross.

One of the main goals of this population was to maintain a representative sampling of all alleles, minimize segregation distortion, and minimize genetic drift by maintaining a large population size and restricting intentional selection of any kind. Therefore, one ear was randomly selected from a segregating plot of selfed ears, although ears with substantial ear rot or pre-harvest sprouting were selected against because of viability concerns. Between the 1st and 4th generation of selfing the initial

1432 lines, a total of 141 lines (9.8%) were lost. Low vigor, low pollen or silk shed, and other poor agronomics were observed in some lines during the inbreeding process; a result of random selection. These issues were likely exacerbated in what can be an extreme growing environment in Texas for maize any given summer. Asynchrony of male and female flowering (anthesis-silking interval; ASI) was also observed in many of the lines which has been well documented in the tropics as highly correlated with a lack of adaption and low yield (as reviewed in De la Fuente et al., 2014). In order to successfully pollinate with sufficient kernel set, many plots had to be "cut back" which required trimming the end of the developing ear shoot to allow the silk to emerge at the same time as the pollen was shed, or prior to tassel burnout. This added a significant amount of effort to an already labor intensive task of managing such a large population and resulted in fewer plants being pollinated. It was also difficult to train workers to consistently look at ear shoots that either were not yet silking, or were on plants whose tassels were fully developed but not yet flowering, which was observed to be the best time to cut-back ear shoots. For these reasons, which ultimately affected the population survival and breadth of research use, it is recommended that additional selection beyond germination and viability concerns should be conducted, and maintaining a large population size should alleviate some concern for reducing overall population variation.

3.3.1 Phenotyping Trials

Because of the field space required to grow out 1000+ plots, nursery plots (where controlled inbreeding and seed saving was conducted) were planted at half the row length of the observational plots. Along with saving field space, the shorter plots allowed

for easier access to plots for manual pollination and required covering less ground. Because the lines were not replicated, a set of three check lines were replicated in blocks to quantify field variance and adjust for field spatial variation. An effort was made to block all five of the environments with replicated check plots. It has been shown that single replicate trials are the most efficient use of program resources (Moehring et al., 2014) and this is especially true in linkage mapping populations where the different combinations of alleles are of most interest and are replicated among the individual lines.

After looking at the results from 2013, it was determined that there were very few individuals that flowered early enough to justify the use of a check line that flowers as extremely early as LH82. LH82 was replaced with a more vigorous, moderate maturity inbred in PB80. The number of entries in each of the observation trials fluctuated to some extent (Table 2). This was solely a product of short seed supply which may have prevented multiple plantings of some individuals or inbreeding depression leading to loss of the line completely. The total number of check plots fluctuated based on overall trial size, and the number of parent plots increased in 2014 due to preferences of the research team to have increased observations of the parent lines within each environment. The CS 2013 nursery, given its size (1292 entry plots), was split into an early and delayed planting to distribute labor but the two plantings flowered at nearly the same time due to late germination and growth from the cold weather early in the growing season. Treating these two plantings as separate environments resulted in no significant differences ($p<0.05$). Because of this, the early and late CS 2013 nursery

plantings were combined for phenotypic analysis of agronomic traits. Phenotypic observations of the 5551 experimental plots allowed a better understanding of the variation represented by this population as a whole and any differences in variation present among the population designs, represented here as subpopulations. To determine if environments could be combined into one analysis, the homogeneity of error variances was calculated for each location and each trait using both Hartley's and Bartlett's tests. Both tests indicated homogeneity of error variances for each trait, confirming that environments could be combined.

3.3.2 Agronomic Traits

Flowering time and height traits were helpful phenotypes to provide a basic characterization of experimental germplasm, and observe differences caused by environments and field variation in these highly heritable traits. Across all five environments (Table 3) field variation (block) was significant, but the proportion of the total variation coming from this effect was low. For both days to 50% anthesis and days to 50% silking, the environment accounted for ~75% of the variation. This is likely a result of different heat units accumulated in each environment to reach flowering. Plant height and ear height had substantially less variation from environment than flowering time but an increased residual. It is suspected that more phenotyping error was introduced from multiple people assisting in taking the height notes and inconsistent measuring points, especially for ear height; in contrast, a single person phenotyped days to anthesis and days to silk.

Table 3. The percentage of observed variation and significance of each variance component explained across five environments. The percentage of observed variation was calculated by dividing each variance component by the total of all variance components plus error variance.

Traits	Environment	Entry	Entry x Environment	Block	Residual	Heritability
Days to silk, 50%	74.6**	12.3**	3.1**	4.8**	5.2	0.60
Days to anthesis, 50%	79.7**	11.3**	1.6**	2.4**	5.0	0.63
Anthesis-silking interval	22.6**	25.2**	24.5**	6.6**	21.1	0.36
Plant height, cm	10.3**	59.5**	14.6**	4.0**	11.6	0.69
Ear height, cm	23.7**	43.4**	11.6**	4.2**	17.1	0.60

** Significant at the 0.01 probability level.

Anthesis-silking interval (ASI) was calculated because many individuals were observed to exhibit asynchrony when having to cut back ear shoots for manual pollination. Significant variation was found for ASI, as well as the largest amount of variation for the interaction of each experimental line with the environment (GxE). This larger GxE variation resulted in a much lower heritability value for ASI, while the heritability estimates for the other four traits were similar. The relatively low levels of residual variation for all traits were encouraging given single replications among experimental entries can often lead to unaccounted sources of variation.

Including block and using the replicated check plots as fixed effects resulted in an improved analysis. Accounting for field variation is always important but often carries more importance for yield trials in Texas where furrow irrigation is used as it was

here; water may be distributed unevenly down rows, or unevenly from the front to the back of the field across ranges. Because all agronomic traits had significant genetic variation, significant differences between parents, checks, and subpopulations could be determined.

Comparisons were first made between all four parents and all four check inbreds (Table 4). Several observations of the parent's per se performance help to make more sense of the phenotypes within the subpopulations and the population as a whole. For flowering time, the bi-parental populations matched a late and early flowering parent with one another. B73*Olc1* and Tx906 were significantly later in flowering than Tx903 and Tx772. When the parent and check lines were ranked by days to anthesis, the check lines were dispersed evenly between the parents. The check lines were chosen intentionally to flower at different times to ensure that at least one check line would be flowering when experimental lines were flowering as a point of comparison. A problem with this population concerning flowering time was the widespread asynchrony displayed by a large portion of the progeny. The ASI of Tx903, Tx906, and Tx772 were significantly longer than B73*Olc1* and the check lines. B73*Olc1* was also significantly the tallest, and Tx903 was significantly the shortest among the four parents, while Tx906 and Tx772 were not significantly different from one another.

Means analysis was performed using BLUPs to determine differences between subpopulations and significant differences were found for all traits (Table 5).

Table 4. Least-squares mean and minimum and maximum values for each parent and check inbred line for flowering time and height traits. Significant differences between parents are reported by the letters after the mean as calculated by Duncan's multiple range at $p < .05$.

Parent lines	**B73*Olc1***		**Tx906**		**Tx903**		**Tx772**	
Traits	Mean ± C.I.	Min. and max.	Mean ± C.I.	Min. and max.	Mean ± C.I.	Min. and max.	Mean ± C.I.	Min. and max.
Days to anthesis, 50%	83.3 ± 0.9a	72 and 91	82.6 ± 0.9ab	70 and 93	80.1 ± 0.9c	70 and 89	79.0 ± 0.9c	69 and 88
Days to silk, 50%	84.8 ± 0.9a	72 and 93	85.7 ± 0.9a	73 and 96	83.7 ± 0.9ab	73 and 92	82.8 ± 0.9b	72 and 90
Plant height, cm	187.5 ± 4.1a	152.4 and 215.9	162.6 ± 4.1c	142.2 and 190.5	146.6 ± 3.6e	129.5 and 167.6	156.0 ± 4.3cd	142.2 and 175.3
Anthesis-silking interval	1.6 ± 0.6b	0 and 7	3.1 ± 0.6a	0 and 7	3.6 ± 0.6a	2 and 7	3.8 ± 0.6a	1 and 7
Ear height, cm	72.1 ± 3.6a	53.34 and 93.98	46.0 ± 3.6b	27.94 and 58.42	50.3 ± 3.6b	25.4 and 68.58	46.5 ± 3.6b	30.48 and 63.5
Check lines	**LH195**		**LH82**		**PB80**		**PHV63**	
Traits	Mean ± C.I.	Min. and max.	Mean ± C.I.	Min. and max.	Mean ± C.I.	Min. and max.	Mean ± C.I.	Min. and max.
Days to anthesis, 50%	83.6 ± 0.3a	71 and 93	77.0 ± 0.4d	75 and 87	79.7 ± 0.4c	67 and 85	81.7 ± 0.3b	69 and 92
Days to silk, 50%	84.5 ± 0.3a	71 and 94	77.9 ± 0.4d	75 and 87	80.5 ± 0.5c	68 and 86	82.9 ± 0.3b	70 and 93
Plant height, cm	160.0 ± 1.3c	129.5 and 180.3	129.3 ± 2.0f	106.7 and 147.3	176.3 ± 2.0b	139.7 and 228.6	150.1 ± 1.3de	116.8 and 182.9
Anthesis-silking interval	1.0 ± 0.2b	-1 and 5	0.9 ± 0.3b	-1 and 6	0.7 ± 0.3b	0 and 4	1.2 ± 0.2b	-1 and 6
Ear height, cm	46.2 ± 1.3b	25.4 and 71.1	34.5 ± 1.8c	22.9 and 61.0	68.1 ± 1.8a	43.2 and 88.9	45.7 ± 1.3b	25.4 and 66.0

Table 5. Means from individual BLUP estimates, standard deviation, and minimum and maximum values for each subpopulation for flowering time and height traits. Significant differences between parents are reported by the letters after the mean as calculated by Duncan's multiple range at $p < .05$.

Traits	B73*Olc1* x Tx903		Tx772 x 906		4way0sib	
	Mean ± SD	Min. and max.	Mean ± SD	Min. and max.	Mean ± SD	Min. and max.
Days to anthesis, 50%	80.6 ± 1.8a	75.9 and 84.3	81.4 ± 1.9ab	77.3 and 86.6	80.9 ± 2.1bcd	75.8 and 86.2
Days to silk, 50%	83.7 ± 1.8bc	77.7 and 88.4	84.3 ± 2.0a	79.4 and 89.7	83.7 ± 2.2bc	78.0 and 89.9
Anthesis-silking interval	2.96 ± 0.7a	1.1 and 4.8	2.92 ± 0.9ab	1.2 and 5.4	2.76 ± 0.8bc	1.4 and 5.3
Plant height, cm	160.8 ± 14.7cd	117.3 and 197.9	157.4 ± 13.5d	120.9 and 196.0	164.8 ± 16.6ab	128.7 and 209.3
Ear height, cm	55.5 ± 10.4ab	38.4 and 82.0	47.3 ± 8.1c	30.0 and 70.6	55.3 ± 9.7ab	33.8 and 85.7
Traits	**4way1sib**		**4way2sib**		**4way3sib**	
	Mean ± SD	Min. and max.	Mean ± SD	Min. and max.	Mean ± SD	Min. and max.
Days to anthesis, 50%	80.7 ± 2.3cd	74.1 and 87.8	81.6 ± 2.3a	76.2 and 87.9	81.2 ± 2.3abc	73.6 and 87.0
Days to silk, 50%	83.4 ± 2.4c	75.9 and 89.9	84.3 ± 2.3a	78.1 and 90.4	84.0 ± 2.4ab	75.1 and 91.4
Anthesis-silking interval	2.74 ± 0.9bc	0.3 and 6.3	2.69 ± 0.8c	1.2 and 4.7	2.80 ± 0.9abc	0.5 and 7.6
Plant height, cm	162.3 ± 16.1bc	110.3 and 206.1	161.7 ± 15.4bc	119.9 and 193.4	166.7 ± 16.7a	115.7 and 220.4
Ear height, cm	54.9 ± 10.6b	31.1 and 91.9	54.9 ± 9.9b	31.1 and 80.2	57.5 ± 10.0a	27.8 and 89.5

The B73*Olc1* x Tx903 bi-parental population was significantly earlier flowering than the Tx772 x Tx906 bi-parental population, which was not surprising based on the inbred line performance. No recognizable trend could be found for flowering time or ASI in relation to the number of parents or generations of intermating. In contrast, for plant height, the bi-parental populations were numerically the shortest, the Tx772 bi-parental was significantly different than all of the 4-parent subpopulations while the B73*olc1* bi-parental population was statistically different only from the 4way3sib population. The observed trend of increased plant height with additional parents and/or intermating was partially due to the wider variation available across all four parents, compared to the more limited variation sampled in a bi-parental design. This observation was also likely due in some part to population size, given that the 4way3sib population (largest) had the highest mean plant height.

Conversely, the 4way0sib population had fewer individuals than both bi-parental populations but was also significantly taller than both. It is clear from the results of plant height that multi-parent crosses increased phenotypic variation in this population. The subpopulation ranks for ear height were nearly identical across all populations, except that the Tx772 x Tx906 population was significantly the shortest. The B73*olc1* x Tx903 bi-parental population having a higher ear height mean was not surprising based on the comparison of parent and check inbred lines where a significantly taller ear height was observed for B73*Olc1*.

3.3.3 Leaf Rolling

Due to the size and complexity of the population and the amount of labor required to create, maintain, genotype (Chapter 4) and analyze it phenotypically and genetically, data on a limited number of phenotypic traits were taken. An interesting and unexpected phenotype that was observed frequently across plots while taking flowering time and height notes was a phenomenon commonly known as leaf rolling. Tx772 and derivative lines have been observed to be especially prone to this phenotype and were the most likely cause for leaf rolling is segregating in the overall population. However, leaf rolling was also observed in individuals from the B73*Olc1* x Tx903 bi-parental subpopulation. Although rarely reported in maize, leaf rolling has been heavily studied in upland rice (non-flooded) production as a mechanism of drought avoidance that can lead to a reduction in transpirational water loss (Price et al., 1997). Many plants in this population experienced leaf rolling on hot, sunny days even after being irrigated and in standing water. By visually scoring leaf roll on a scale of 0-3 for leaf rolling, sufficient segregation and genetic variation was found within this population to explore this trait further (Table 6).

Table 6. The percentage of variation and significance each component explains for leaf rolling phenotype in the CS 2014 early and late plantings combined.

Leaf Roll	Environment	Entry	Entry x environment	Block	Technical measurement replication	Residual
CS 2014 combined	14.6*	18.3**	28.0**	3.7**	2.4**	33.1

* Significant at the 0.05 probability level.
** Significant at the 0.01 probability level.

With the combined analysis of leaf rolling observations from the CS 2014 early and late observation trials a substantial portion of variation came from GxE. Although there was some significant block variation, a larger residual error than other measured traits indicated sources of unaccounted variation which could include the actual method for recording the leaf roll phenotype. Although efforts were made to observe leaf rolling quickly and efficiently, a ~1200 plot trial still required approximately three hours to cover. Because the degree of leaf rolling can change based on the environment (temperature, sunny or cloudy) it would likely be better to plant smaller trials for the purpose of leaf roll observation so that less time elapses between the first and last plot recorded. Interestingly, the difference in measurements between morning and night were a significant, but not substantial source of variation as might have been expected given the significant GxE interaction.

3.3.4 Kernel and Cob Color

A primary objective beyond comparing the effectiveness of population designs for phenotypic discovery was to investigate further the behavior of the blue aleurone trait

in maize. The simple qualitative measure of color was taken during seed shelling and processing of nursery plots. Phenotypes were taken for kernel and cob color beginning with seed harvested from the WE 2011 winter nursery, and continued each subsequent generation to monitor residual segregation and as an additional check for errors in seed planting and processing. The distribution of cob, endosperm, and aleurone color phenotypes recorded from CS 2013 are presented in Table 7.

Table 7. Segregation of cob, endosperm, and aleurone color within each sub-population.

	B73*Olc1* x Tx903	**Tx772 x Tx906**	**4way 0sib**	**4way 1sib**	**4way 2sib**	**4way 3sib**	**Total**
			Cob color				
Red cob	43%	0%	31%	40%	34%	31%	31%
White cob	54%	100%	65%	58%	60%	66%	66%
Segregating cob	3%	0%	4%	2%	6%	3%	3%
			Endosperm color				
Yellow endosperm	52%	32%	39%	35%	44%	50%	44%
White endosperm	43%	60%	54%	61%	47%	44%	50%
Segregating endosperm	5%	8%	7%	4%	9%	6%	6%
			Aleurone color				
Blue aleurone	25%	24%	21%	26%	27%	23%	24%
Clear aleurone	71%	71%	71%	68%	66%	69%	69%
Segregating aleurone	4%	5%	8%	6%	7%	8%	7%

In addition to a simple visual characterization of kernel color, a colorimeter was used to quantify kernel color so that a more specific and quantitative value could be used in subsequent genetic analysis (Chapter 4). A total of 949 individuals were scanned with

the colorimeter from CS 2013 nursery seed, and significant variation was found for colorimeter values L, a, and b, as well as the values of hue and chroma (Table 8). Visual appearance of kernel color was used to divide the individuals into groups for analysis.

Table 8. The percentage of variation and significance that visual kernel color (yellow, white, blue) and cob glume color (red, white) had on colorimeter color characterization values.

Colorimeter values	Kernel color	Cob glume color	Residual
Lightness (black/white)	62.9**	10.8**	26.2
a (red/green)	39.9**	34.8**	25.3
b (blue/yellow)	88.3**		11.7
Hue ($\tan^{-1}(b/a)$)	8.5**	54.4**	37
Chroma ($\sqrt{(a^2 + b^2)}$)	88.2**	0.2**	11.6

* Significant at the 0.05 probability level.
** Significant at the 0.01 probability level.

Because the population segregated for red or white cob color, some individuals had kernels with red glumes still attached after seed processing. This was significant for the "a" and to a lesser extent the "L" values. Since "a" is a measure of red/green and "L" is a measure of black/white, it was not surprising that the addition of red glumes in some kernel samples added red to the value, as well as darkened the sample. The value of hue, which is calculated with both "L" and "a" was especially affected by those samples having red glumes. For studies in maize where hue has more importance, it would be best to subset those individuals with excessive red glumes so that they do not affect the colorimeter values.

The colorimeter was also able to effectively separate individuals grouped according to visual kernel color (Table 10). This data shows the trend of lightness values to be highest among white kernels, and red-green and blue-yellow values to be highest among yellow kernels. The ability to separate kernel colors and in some cases even those individuals that were segregating, in addition to detecting differences in glume color showcase the precision of the colorimeter.

Table 9. Least-squares mean for kernel color measured as colorimeter values L, a, b, hue, and chroma. Significant differences between parents are reported by the letters after the mean at $p < .05$.

	Lightness		**Red/**		**Blue/**				
Color[†]	**(L)**	**Color**	**Green (a)**	**Color**	**Yellow (b)**	**Color**	**Hue**	**Color**	**Chroma**
W	65.8a	**Y**	6.7a	**Y**	31.4a	**B**	81.2a	**Y**	32.2a
WY	64.9ab	**WY**	6.3a	**WY**	30.8a	**BW**	80.4ab	**WY**	31.5a
Y	64.2b	**W**	4.8b	**W**	27.0b	**W**	79.8ab	**W**	27.5b
BY	58c	**BY**	4.2bc	**BY**	21.4c	**BY**	79.5abc	**BY**	21.9c
BW	57.8c	**BW**	3.4cd	**BW**	20.0c	**WY**	78.1bc	**BW**	20.4c
B	54d	**B**	2.7d	**B**	16.0d	**Y**	77.6c	**B**	16.4d

[†]W, white endosperm; WY, segregating white or yellow endosperm; Y, yellow endosperm; BY, blue aleurone with yellow endosperm; BW, blue aleurone with white endosperm; B, blue aleurone.

3.4 Conclusion

The described population is unlike any in *Zea mays* or other species that has been created to our knowledge; constructed so that direct comparisons can be made between population mating designs. This population has demonstrated a large range of phenotypic variation in the few currently observed traits and provides the opportunity for an exhaustive list for future research. It was demonstrated that it is important to create large populations with practicality in mind, as well as maintain sound research practices such as balancing random selection of individuals to advance, which is at odds with breeding goals of selecting only superior individuals and in maintaining a population that suffers from inbreeding depression.

Significant variation was found for all agronomic traits measured. All traits could effectively separate subpopulations with the most interesting differences between subpopulations coming from plant height. The effect of four parents on increasing variation, as well as population size, was observed for plant height; surprisingly, increased population size resulted in a greater range of plant and ear height variation. The unique leaf rolling trait was found to be heavily segregating in this population which will need to be validated further with an emphasis to lower residual variation by observing leaf rolling in smaller trials and limiting the amount of time it takes to get through a set of plots. Furthermore it will be interesting to see if this is transferable to hybrids and if it correlates with yield under stress. The cob and endosperm color distribution within the population was as expected since one parent has red cob color (B73*Olc1*) and two parents have yellow endosperm. The blue aleurone is distributed as

if it is under the control of two major epistatic genes. This conclusion is reinforced by genetic analysis as described in Chapter 4.

4. MULTI-PARENT POPULATION MATING DESIGN EFFECTS ON GENETIC MAPPING RESOLUTION FOR MAJOR COLOR GENES IN MAIZE

Central questions surrounding QTL mapping and quantitative genetics is how accurate and reliable is the commonly used bi-parental methodology and can it be further improved? Why have so many QTL been reported but not cloned or resolved to genes? The first mapping results from a novel multi-parent, intermated population derived from four maize inbred lines are presented. The present study should allow for a better understanding of the practical effects of population design and size with regard to the resolution and accuracy of genetic mapping using Mendelian traits, expected to be easily mapped. To date there has been no known empirical comparison of multi-parent and/or intermating design methods with standard mating designs for mapping resolution or accuracy. The maize population developed herein adopts a series of designs with both of these strategies combined to create an unprecedented 1,207 individual linkage mapping population with 107,308 genetic markers allowing empirical comparisons of effects on genetic mapping resolution and accuracy. Using association mapping methods we identified two epistatic candidate loci co-localizing to known mutants for control of blue aleurone, believed to be under complex genetic control. These results suggest that these two genes largely control blue aleurone expression in these improved breeding lines. However, in both smaller sets of four way individuals and in bi-parental crosses, one or both genes were not always detected or were detected in the wrong location. Mendelian genes for yellow endosperm and red cob color showed similar but less dramatic losses

from reducing population size. Mapping results of red cob suggest that an uncommon trait that is only represented in one or a limited number of inbred lines may be more easily mapped with increased resolution if multiple parents are used, as long as population size is large (~500 individuals), even if the additional parents are not themselves expressing the target trait. The multi-parent populations with four founders and intermating resulted in less enhancement of locus resolution than anticipated. The effect of increased population size was the most beneficial modification. Although the 4-parent subpopulation with three generations of intermating was often the most successful in accurately detecting the target gene with enhanced resolution, it was difficult to determine if this was due to it being the largest subpopulation, or the additional intermating. There was a multi-parent effect on improving the mapping resolution in some cases but not all. This new population provides a design that can be used to better understand limitations in genetic mapping accuracy and resolution. From the results it is concluded that larger linkage mapping populations, such as this one, should be used for detection of traits with even simple inheritance. Furthermore it was found that the current software and approaches for genetic linkage map construction are unsuitable for large multi-parent populations with large marker sets.

4.1 Introduction

The progression of genotyping technologies has allowed increasingly greater marker density to be used in genetic mapping studies. As a result, mapping population design can begin to be refocused towards increasing recombination and capturing higher

levels of diversity than what can typically be observed in a traditional bi-parental mapping population.

Crosses are conducted by geneticists and breeders to create variation and to attempt to combine the best alleles from two (or more) parents from which the superior progeny are selected and for which recombination is believed to be a limiting factor (Darvasi and Soller, 1995). For row crops such as maize that have been under heavy plant breeding selection for decades, breeding is usually based on phenotypic selection and assisted by genotypic selection off of a previously validated and commonly used set of markers tailored to a specific set of germplasm. Genetic mapping becomes especially important in assisting the plant breeding pipeline when novel genes for desirable traits are discovered in order to create new genetic markers, especially when phenotyping proves difficult or costly.

Parent selection is critical in determining which traits can be measured and mapped as QTL; target traits and parents segregating for these traits must be identified before population creation (Anderson et al., 1993). If the chosen parents are too diverse for non-target traits, this may mask, distort, or confound the subsequent detection results. For the purpose of QTL mapping, it has been common to make a cross between two parents, and self for several generations until near homozygosity, producing bi-parental RILs. These progeny are useful because they are immortalized and thus need only be genotyped once, multiple plants can be phenotyped from each individual to reduce error from individual, environmental, and measurement variability, and additional recombination breakpoints allow for a higher mapping resolution than with an F_2

population (Broman, 2004). Some mapping populations have successfully identified QTL linked to traits of interest and, with substantial additional work, the causal gene that conditioned the QTL were subsequently cloned, validating the QTL mapping approach (Doebley et al., 1997; El-Din El-Assal et al., 2001; Kroymann et al., 2003; Mouchel et al., 2004; Werner et al., 2005). A primary reason for most QTL studies failing to result in cloned genes responsible for the trait is due to the low mapping resolution that bi-parental populations provide and the limited recombination events they produce.

In an effort to improve genetic mapping resolution, the advanced intercross was designed as an extension of traditional RILs (Darvasi and Soller, 1995) but still had the issue of a narrow genetic base (two parents) for each cross, limiting the analysis of multiple traits. As a result of research moving away from an individual Mendelian trait focus and moving towards applications determining the genetic control of traits under the control of multiple genes (yield, food quality, drought tolerance, etc.), the need for populations with the opportunity for a wider genetic and phenotypic base must be investigated. With an increase in recombination, comes the need for an increase in marker coverage. The next-generation sequencing technology movement has made way for genotyping-by-sequencing (GBS), which uses parallel sequencing to produce up to several 100 billion nucleotides per sequencing run. These short-read sequences (~50-300 bp) produced by platforms such as the Illumina HiSeq2500 (Bentley et al., 2008) are used to discover single nucleotide polymorphism (SNP) markers which are highly prevalent across the maize genome. A study of 2,815 inbred accessions from the United States National Plant Germplasm System, representing a large portion of maize genetic

diversity, recently discovered 681,257 SNPs (Romay et al., 2013), showing the effectiveness of GBS to produce SNP marker coverage across the maize genome.

The multi-parent mating design was first implemented to produce heterogeneous mouse populations (Mott et al., 2000). For the purpose of QTL mapping in plants, MAGIC populations have been developed in wheat (Cavanagh et al., 2008; Huang et al., 2012), rice (Bandillo et al., 2013), and *Arabidopsis* (Balasubramanian et al., 2009). In addition to increasing the opportunity for recombination by increasing the number of founding parents of a population, intermating the F_1 hybrids prior to inbreeding can further increase recombination.

It is widely known and demonstrated that increasing the generations of intermating and/or number of parents will increase the opportunity to observe recombinants and more closely detect QTL in the resulting population's progeny (Flint and Mott, 2001). A famous example of this is the intermated B73 x Mo17 (IBM) population in maize where intermating increased the genetic map distance by 3.86-fold (Lee et al., 2002). Examples of using large numbers of individuals can be found in fine-mapping QTL studies, such as the 3,742 individuals necessary to map a fruit weight QTL in tomato to a 150-kb interval (Alpert and Tanksley, 1996), as well as the 7,000 individual population that detected a QTL responsible for sugar content of tomato to a 484-bp interval (Fridman et al., 2000). A multiple parent population with various levels of intermating for comparison against related, bi-parental populations has not previously been developed and provides a unique resource to continue to examine the utility of multi-parent populations.

A multi-parent mating population with four parent lines which incorporates different levels of intermating was established to compare and contrast mating designs ranging from a traditional bi-parental design, to a four parent design with three generations of intermating. The ability to affect accuracy and resolution of genetic mapping was characterized by analyzing cloned genes for the Mendelian traits of yellow endosperm and red cob color. The gold standard for detection was the use of the entire unprecedented 1,207 individual mapping population with comparisons made within and between sub-populations. Specifically the objectives of this study were to 1) identify SNPs for genetic mapping across the maize genome using GBS technology; 2) leverage major Mendelian genes to validate sequence data and compare QTL detection accuracy and resolution of mating designs and population sizes; 3) identify genes controlling blue aleurone in this population; and 4) explore detectable quantitative variation for kernel color.

4.2 Materials and Methods

4.2.1 Population Development

The four founding parents include Texas adapted inbred lines with blue aleurone and white endosperm (Tx903 and Tx906), an aflatoxin resistant inbred line with yellow endosperm (Tx772) (Llorente et al., 2004), and an oleic acid mutant, B73*Olc1* (Wright, 1995) donated by Allen Wright. The subsequent steps leading to the creation of the population are discussed in greater detail in chapter 2 of this dissertation.

4.2.2 DNA Isolation and Quantification

Two separate and distinct DNA isolation protocols were conducted during the genotyping of this population. Initially, tissue was sampled from field grown, F_3 and F_4 plants from the 2011 Weslaco winter nursery at ~V8 growth stage. Samples taken from the youngest leaf tissue 3 cm in length were bulked from 10 plants per plot, and immediately stored in an ice chest and transported to College Station and placed in -80^0C storage prior to DNA isolation.

The first DNA isolation protocol followed an in-house, phenol-chloroform method. A detailed description of the protocol can be found in the Appendix. The extraction buffer consisted of Sorbitol, Tris, and EDTA, and the lysis buffer consisted of Tris, EDTA, EDTA, and CTAB. Briefly, tissue samples were arrayed in 96-well plate format. Eight to ten thin strips of leaf tissue 0.6 cm in length were placed inside 1.2ml tubes containing a steel rod 0.6 cm in length. During this process, care was taken to sterilize scissors and tweezers for each individual sample. Fresh buffer was added to each tube and the tissue was homogenized using a Geno/Grinder®. The buffer solution includes an extraction buffer and a lysis buffer. The extraction buffer consisted of 0.35 M sorbitol, 0.1 M Tris (pH 7.5), and 5mM EDTA. The lysis buffer consisted of 0.2 M Tris (pH 7.5), 0.05 M EDTA, 2 M NaCl, and 2% CTAB. The DNA was purified using a series of phenol/chloroform/isoamyl alcohol and chloroform/isoamyl alcohol steps and precipitated with ethanol. The final step involved the use of a DNA Clean and Concentrator 96-well Kit from Zymo Research (Irvine, CA, USA).

The second DNA isolation followed the ZR-96 Plant/Seed DNA Kit™ protocol (Zymo Research, Irvine, CA, USA). Six kernels of each individual from the College Station 2013 nursery were planted in Metro-Mix 300 potting soil within RLC7 (3.8 cm diameter) UV-stabilized cone-tainers ™ (Stuewe and Sons, Inc., Tangent, OR, USA). Seeds were germinated under greenhouse conditions and bulk tissue samples from the five to six plants were taken from maize seedlings approximately one week post planting so that any residual heterozygosity within each individual was captured. Similar amounts of plant leaf tissue were placed into 96-well racks of 1.2 ml tubes as in the first isolation protocol. A detailed protocol for this method is available in Appendix 1.

To determine the DNA concentration of each sample, an AccuBlue Broad Range assay (Biotium, Hayward, CA, USA) was used in conjunction with a fluorescence microplate reader. A detailed protocol is available in Appendix 1.

4.2.3 Genotyping-by-Sequencing Methodology

Genotyping of the population followed a restriction endonuclease-assisted, reduced-representation marker protocol called digital genotyping (DG) (Morishige et al., 2013). This method was developed for genotyping C4 grasses on the Illumina GAIIx/HiSeq2500 platforms which utilize short read sequencing. The restriction enzyme *Ngo*MIV (GCCGGC) was used and its sensitivity to methylation allowed for a reduction in complexity, an increase in hypomethylated genomic regions and a reduction in hypermethylated regions most often correlated with repetitive DNA around centromeric regions, pseudogenes, transposons, and retrotransposons (Davey et al., 2011). Due to the large amount of methylation and repetitive sequences known in grasses which can create

challenges during bioinformatic analysis (Larrinua and Belmar, 2008), the use of a methylation-sensitive restriction enzyme such as *Ngo*MIV was a crucial step to aid in the downstream analysis of sequence data.

Samples of DNA at a concentration of 250 ng were restriction enzyme digested followed by ligation to a set of Illumina-compatible adapters with individual 12 bp in-line forward read barcodes that upon ligation, allowed for samples to be pooled into groups of 48 individual DNA samples to be run in single lanes on the Illumina flow cell. Pooled samples were randomly sheared to a target size of 250-300 bp using a Covaris® S2 sonicator and size selected on a 2% agarose gel to confirm adequate sonication and eliminate those fragments which were too large or small. Once DNA was precipitated from the agarose gel slices, overhang fill-in, blunting, and 3' adenylation was performed prior to another ligation to an Illumina-specific adapter. Ligated pools were PCR-amplified for 18 cycles using Phusion® High-Fidelity Polymerase. Single strand products were obtained and then underwent a second PCR amplification for 6 cycles which added the Illumina bridge amplification sequence. Using this method, pooled samples of 48 individual DNA samples represented one lane of the Illumina flow cell. Single-end sequencing was performed for 107 cycles on an Illumina HiSeq2500.

4.2.4 Bioinformatics Data Processing

The individual 12 bp barcodes and partial *Ngo*MIV restriction site were used to sort the FASTQ sequences obtained from the Illumina HiSeq2500. An absolute, 100% match to both the barcode and partial restriction site was necessary for each sequence to be retained. Sequences were trimmed of both the 12 bp barcode from the 5' end as well

as 8 bp from the 3' end of each sequence prior to being imported into the CLC Genomics Workbench (Qiagen, Boston, MA). Sets of filtered reads for each individual line were mapped to the B73 maize genome (ZmB73_RefGenV2_masked) (Schnable et al., 2009) and analyzed for SNPs and INDELs using the software from the CLC Bio Genomics Workbench (version 6.5.1). Parameters for read mapping were set to insertion and deletion cost = 3, mismatch cost = 2, 50% minimum read length required to match the reference, and a minimum of 90% similarity between the read and the reference sequence. Any reads that failed to align to the maize reference genome or those reads that aligned to more than one position (repetitive regions) were discarded. SNP detection parameters in the CLC Bio Genomics Workbench were set to a neighborhood radius of 5, a maximum gap and mismatch count of 3, a minimum quality of the SNP base of 20, a minimum average quality of the nucleotides surrounding the SNP of 15, and a minimum read coverage of 10 to call a SNP. These levels of parameters were applied to each SNP to aid in determining whether a SNP was legitimate or a sequence error.

Exported data files from the read mapping and SNP discovery analyses were formatted as SAM files and comma-separated-value (.csv) respectively. These exported files were reformatted for further analysis using custom scripts written in perl and python. The source code and descriptions of these scripts are located at the static url: http://hortsciences.tamu.edu/departmental/klein/. The final reformatted output file combined the data for all sequenced individuals, and connected the SNP or INDEL identity to a specific, physical location in the maize genome. In addition, a separate .csv file was generated containing the number of reads for each variant. Markers with up to

90% missing data were imputed using the program fastPHASE (Scheet and Stephens, 2006). Once imputation was complete, only SNPs with a minor allele frequency greater than 5% were used in downstream analysis. Additionally, only SNPs with less than 10% heterozygosity were retained. SNP names were unique to the chromosome they were on and their physical location based on the maize reference genome ZmB73_RefGenV2_masked (Schnable, 2009). For example, SNP chr1_46978665 is located on chromosome 1 at position 46,978,664 bp.

4.2.5 Phenotyping Yellow Endosperm Red Cob, and Blue Aleurone

Phenotypes used for QTL analysis were taken for kernel (yellow or white endosperm, clear or blue aleurone) and cob color (red or white) from individuals harvested from the College Station 2013 nursery. To determine if there was significant variation in the intensity of yellow endosperm and blue aleurone, a Chroma Meter CR-410 (Konica Minolta, Ramsey, NJ) was also used on kernels from the CS 2013 summer nursery. A colorimeter reports a set of three measurements to represent the color of each sample including "L" (lightness) which measures how white or black the sample is, "a" (red-green), and "b" (blue-yellow) (Jha, 2010). Additional phenotyping and research trial information is detailed in Chapter 2.

4.2.6 Association Analysis

Association mapping of blue aleurone, yellow endosperm, and red cob was performed by the compressed mixed linear model (Zhang et al., 2010) implemented in GAPIT (Lipka et al., 2012). The significant p-value threshold for associations between a marker and trait was adjusted using a conservative Bonferroni correction for multiple

testing (Murray et al., 2009) calculated by dividing 0.05 by the number of markers. Gene locations were downloaded from Phytozome to determine the number of genes within a mapping confidence interval.

To compare the accuracy and resolution of QTL mapping between the various sub-populations 10 random subsets of individuals were sampled from the 4way3sib subpopulation with replacement between each sampling. These random subsets were used for analyzing all phenotypic traits and varied slightly in size based on the level of segregation and/or missing phenotype for each trait, since these segregating individuals were not used in association mapping analysis.

4.3 Results and Discussion

During the early stages of DNA isolation, two methods were tested (CTAB/phenol-chloroform and the FastDNA-96 Plant and Seed DNA KIT from Zymo Research). Early sequence results from the Illumina HiSeq2500 on DNA from field grown tissue isolated with the phenol-chloroform method were of lower quality than what is typically acceptable and a portion of the sequences were of fungal or bacterial origin. Additionally, an issue termed off-site ligation was prevalent at a high level, which is observed when the sequences do not contain the barcode adaptor adjacent to the *Ngo*MIV partial restriction site. It was determined that the extracted DNA was not of adequate purity and had undergone too much shearing during the extraction process. Because of this, the FastDNA-96 Plant and Seed DNA Kit from Zymo Research was tested on DNA extracted from young greenhouse grown seedlings. This resulted in an increase in DNA quality and purity and offsite ligation was reduced. A subset of 159

sequenced samples which were isolated from field grown tissue using the CTAB/phenol chloroform method were retained and combined into the downstream analysis with those individuals isolated with the more suitable FastDNA-96 Plant and Seed DNA Kit due to the cost associated with re-sequencing these individuals. The bulk of individuals were genotyped from kernels harvested from the College Station 2013 nursery. Table 10 details the number of individuals in each subpopulation (mating design), the generation at which they were genotyped, and their tissue source.

Table 10 Multi-parent mapping population specifications for each mating design and the respective population size of genotyped individuals.

Population code	**Number of parents**	**Generations of intermating**	**Generations of inbreeding**	**Number of individuals**
B73*Olc1* x Tx903	2	0	F_4 (4), F_6 (39)	43
B73*Olc1*-1self-2sib	2	2	F_3 (11), F_5 (75)	86
Tx772xTx906	2	0	F_4 (22), F_6 (108)	130
4way	4	0	F_4 (17), F_6 (100)	117
4way1sib	4	1	F_4 (21), F_6 (192)	213
4way2sib	4	2	F_4 (14), F_6 (83)	97
4way3sib	4	3	F_3 (68), F_5 (447)	515
4way*	4	?	F_5 (2), F_6 (4)	6
Total				1207

* Six individuals that were mislabeled as bi-parental crosses, with unknown amount of intermating.

The lower vigor of progeny from the B73*Olc1* x Tx903 cross resulted in a smaller number of individuals after several generations of selfing. Eighty-six progeny were created with B73*Olc1* xTx903 individuals that were inadvertently intermated for two generations and these were added to the B73*Olc1* x Tx903 subpopulation to increase the final population size to 129 individuals. The majority of the genotyped individuals were F_5 or F_6 (87%) greenhouse grown seedlings from seed harvested from the College Station 2013 summer nursery plots. A subset of 159 genotyped individuals stemming from field grown tissue collected from the Weslaco 2011 winter F_3 and F_4 generation. Six individuals originally labeled as progeny from the bi-parental populations were later found to share alleles from all four parents and were included in association mapping of all four parent subpopulations combined but were not included in any of the individual four parent subpopulations because the level of intermating was unknown.

4.3.1 Genotyping the Four Parent Maize Population

The 1207 lines from the four parent maize population (FPM) were genotyped using a genotyping-by-sequencing (GBS) method referred to as digital genotyping (Morishige et al., 2013). Sequence libraries were generated using *Ngo*MIV, a restriction enzyme which has methylation sensitivity. This enzyme cuts at the GC-rich sequence GCCGGC which is in or around many genes. Forty-eight lines, each with a unique 12 bp barcode tag, were pooled to represent a single lane on the Illumina HiSeq2500 flow cell. On average, each individual was represented by 1,821,444 reads each 107 bp in length when including the 12 bp barcode as well as the partial *Ngo*MIV restriction site. After removal of the barcode and trimming from the 3' end of each read, the 87 bp reads were

mapped to the maize reference genome using the CLC Bio Genomics Workbench. Following mapping to the maize reference genome, approximately 80% of the reads mapped to one unique location and were retained for SNP/INDEL discovery. These reads and their respective positions are indicative of gene rich portions of the maize genome since *Ngo*MIV will only restrict non-methylated recognition sequences. An average of 1,464,835 unique reads were obtained per maize line representing ~127 M bp of the maize genome.

The unique reads were used to discover SNPs and INDELs using the Variant Detection Tool in the CLC Bio Workbench. The number of SNPs/INDELs discovered between each individual and the B73 maize genome ranged from 69 to 93,552 with an average of 35,769 (data not shown). Comparing the SNPs/INDELs across the 1,207 individuals identified 170,917 that were found in at least two maize lines. Several stages of quality control had already been taken to maintain a high quality set of SNPs/INDELs for downstream association analysis. When SNPs were called using the CLC Bio Workbench, a minimum read depth of 10 was required as coverage of the SNP in order to be reliably called.

After the SNPs were filtered using the CLC Bio Workbench, markers with more than 90% missing data across all 1207 lines (34,217) were removed from further analysis. The missing data in the remaining markers was imputed using the program fastPHASE. Markers with a minor allele frequency less than 5% (9,083) and those with heterozygosity above 10% were also discarded (20,309), leaving 107,308 SNPs in the final data set. An average of 10,731 SNPs were present on each chromosome (from

17,225 on chromosome 1 to 7,636 on chromosome 10; Table 11). A higher density of SNPs were found in the euchromatic arms of each chromosome away from the region of the centromere, as was expected due to centromeric regions having higher levels of methylated DNA (Wolfgruber et al., 2009). The majority of SNPs (85%) were within 10,000 bp of the adjacent SNP.

Table 11. Number of SNPs per chromosome across all four parents used in association mapping analysis based on AGPV2 of the B73 maize genome.

Chromosome	SNPs	First SNP position	Last SNP position	Physical distance of marker coverage	Average distance between markers
1	17,225	458,242	300,925,870	300,467,628	17,444
2	13,450	87,908	236,971,845	236,883,937	17,612
3	12,268	1,103,819	232,096,209	230,992,390	18,829
4	9,462	29,275	241,025,655	240,996,380	25,470
5	12,078	402	217,416,788	217,416,386	18,001
6	8,774	272,387	168,974,423	168,702,036	19,227
7	8,582	14,910	176,390,607	176,375,697	20,552
8	9,563	96,774	175,365,331	175,268,557	18,328
9	8,270	169,418	155,975,958	155,806,540	18,840
10	7,636	4,610,840	150,091,634	145,480,794	19,052

Comparing the allele frequencies with the B73 reference genome, it was not surprising that B73*Olc1* shared nearly all of the same alleles with the reference genome (Table 12). Because Tx903 has Tx114 (Betran et al., 2004), a white conversion of B73 within its pedigree, it had higher similarity to the reference genome than Tx772 and Tx906. Tx903 had the same similarity to the B73 reference genome and to the B73*Olc1*

genome. Because Tx772 and Tx906 are substantially less similar with one another than B73*Olc1* and Tx903, larger phenotypic and genotypic variation was observed within this bi-parental population. Very little of the genotypic difference between the parents and the reference genome was attributable to heterozygote alleles as each parent had ~0.40% heterozygous alleles. On an individual basis across the population, the average heterozygosity was 5.7%, with the F_4 (14.1%), F_5 (7.3%), and F_6 (3.3%) individuals having expected amounts of heterozygous SNPs; however, F_3 individuals had a lower amount of heterozygosity (14.5%) than would be expected.

Table 12. Sequence similarity between the maize reference genome and the four founder parents.

	Reference	**B73*Olc1***	**Tx903**	**Tx772**	**Tx906**
Reference	100%	98.5%	64.7%	40.0%	44.9%
B73*Olc1*		100%	64.7%	40.1%	44.9%
Tx903			100%	40.9%	42.9%
Tx772				100%	40.7%
Tx906					100%

Table 13. Segregation data for cob, endosperm, and aleurone color for all sub-populations. The number in parenthesis indicates sub-population size.

	B73*Olc1* x Tx903 (129)	**Tx772 x Tx906 (130)**	**4way0sib (117)**	**4way1sib (213)**	**4way2sib (97)**	**4way3sib (515)**	**All RILs (1201)**
Red cob	56	0	36	86	33	159	370
White cob	70	130	76	124	58	342	798
Segregating cob	3	0	5	3	6	14	33
Yellow endosperm	67	42	46	74	43	259	531
White endosperm	55	78	63	129	46	229	600
Segregating endosperm	7	10	8	10	8	27	70
Blue aleurone	32	31	25	56	26	117	287
Clear aleurone	92	92	83	144	64	357	832
Segregating aleurone phenotype	5	7	9	13	7	41	82

4.3.2 Phenotyping Cob, Endosperm, and Aleurone Color

The color traits were fixed in the individual populations consistent with the level of inbreeding as would be expected in a RIL population (Table 13). Individuals expressing red cob color were a smaller percentage of the populations since only B73*Olc1* had the dominant red cob color. Yellow endosperm was observed anywhere from 30-50% within each subpopulation, and blue aleurone was consistently observed in 20-25% of individuals, indicative of a two gene epistatic control of blue aleurone.

4.3.3 High Resolution and Accurate Confirmation of Major Alleles for Yellow Endosperm and Red Cob Color Across the Entire Population

An association mapping approach was performed on this population with an original intent to construct a genetic map and analyze the population with linkage mapping strategies. Software such as JoinMap (Kyazma, Wageningen, Netherlands) is commonly used to construct genetic maps for bi-parental crosses. In addition to not being able to handle SNP markers that are polymorphic in more than two parents, JoinMap has a ~500 marker analysis limit per linkage group (chromosome) which restricts its use for studies with large marker data sets. A statistical package within the R environment called Mp/Map can handle four- or eight-parent mapping populations and has been used previously to analyze MAGIC populations in wheat (Huang and George, 2011). Like JoinMap, Mp/Map is currently restricted to a low number of markers, and also cannot handle heterozygous alleles, resulting in a loss of statistical power if heterozygous calls must be changed to missing data points or if the entire marker must be removed. As multi-parent populations coupled with GBS data become more popular,

improved software will be necessary and will likely be developed. At this stage of analysis of this novel maize population, association mapping, essentially t-tests of marker phenotype prediction, provided a benchmark for future results.

To adjust for multiple testing over the large number of markers the Bonferroni correction was implemented for the 4-parent populations by dividing 0.05 by 107,308 markers, for a significance threshold of 6.33 LOD. The bi-parental populations required separate Bonferroni corrections due to the fact that each bi-parental population had a smaller number of markers segregating within their respective populations. The B73*Olc1* x Tx903 population contained 44,581 polymorphic markers and the Tx772xTx906 population contained 73,717. This resulted in LOD thresholds of 5.95 and 6.17, respectively.

Yellow endosperm and red cob were used to validate that the data was correct and evaluate the mapping accuracy and resolution possible in the population. The loci most strongly associated with these traits were *yellow endosperm1* (*y1*) and *pericarp color1* (*p1*) respectively. The 3,731 bp *y1* gene has been cloned (Buckner et al., 1990) and is located on chromosome 6 (82,017,148-82,020,879 bp). The 10,550 bp *p1* gene has also been cloned (Lechelt et al., 1989) and is located on chromosome 1 (48,117,497-48,128,047 bp). Association analysis of the entire population detected a highly significant (107.08 LOD) peak QTL positioned at 82,017,348 bp on chromosome 6 (Figure 3; Table 14), within the *y1* gene for the yellow endosperm trait, and a highly significant (65.50 LOD) peak QTL positioned at 47,994,162 bp on chromosome 1 for the red cob trait (Figure 3; Table 14). The peak QTL detected by combining the entire

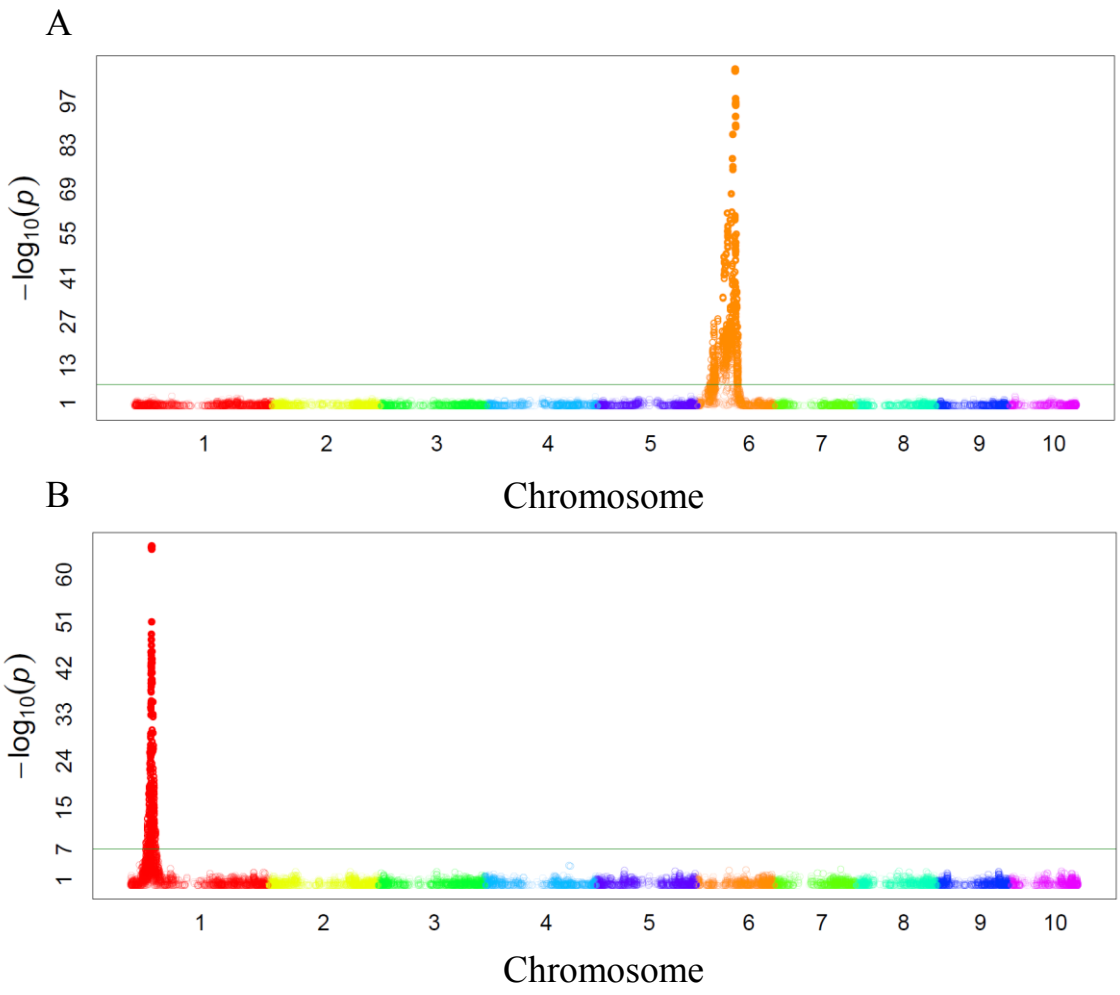

Figure 3. Graphical summary (Manhattan plot) of genome-wide association results for endosperm color (A) and cob color (B). The *x* axis represents the maize genome in physical order; the *y* axis shows $-\log_{10} P$ for all SNPs.

Table 14. Peak QTL position and LOD values for yellow endosperm, red cob, and blue aleurone for the entire population.

Trait	Individuals	Chromosome	LOD	Peak position (bp)	1-LOD interval (bp)
Yellow endosperm: *yellow endosperm1* (*y1*)	1141	6	107.1	82,017,348	82,017,294 – 82,017,402
Red cob: *pericarp color1* (*p1*)	1174	1	65.5	47,994,162	48,082,512 – 47,905,812
Blue aleurone: *colored aleurone1* (*c1*)	1128	9	47.96	9,397,546	9,053,844 - 9,741,248
Blue aleurone: *colored1* (*r1*)	1128	10	32.6	138,468,760	138,327,025 – 138,610,495

population for red cob did not capture the *p1* gene within the 1-LOD confidence interval (Table 14) but did come within 35,000 bp. An explanation for this may be that there were not enough polymorphism to get closer to the gene and not enough recombination occurred in the region to identify the cloned gene, as the two closest flanking markers were 121,417 and 252,706 bp away from *p1*.

The LOD score was likely lower for red cob than for yellow endosperm because only one of the four parents, B73*Olc1*, had the dominant red cob color phenotype and this resulted in only 30% of the lines having red cob, while 47% of the lines had yellow endosperm. Both yellow endosperm and red cob served as validation of the genotyping data in this study and these observed locus peaks from analysis of the entire population

served as a "gold standard" for comparing the genetic mapping resolution and accuracy of the sub-populations.

4.3.4 Identifying Two Major Alleles with Kernel Expression for Blue Aleurone Across the Entire Population

Highly significant loci peaks co-localizing with two cloned genes, *colored aleurone1* (*c1*) (Cone et al., 1986) and *colored1* (*r1*) (Dellaporta et al., 1988) were discovered segregating in this population for blue aleurone. The cloned *c1* gene is 1,073 bp long and is located on chromosome 9 between 9,740,803 and 9,741,876 bp. The *r1* gene is 8,820 bp in length and is located on chromosome 10 between 138,462,252 and 138,471,072 bp. When all subpopulations were combined (n= 1128), the QTL peak associated with *c1* on chromosome 9 was shared by two positions, one at 9,397,546 bp and one at 9,741,248 bp each with a LOD of 47.96 (Figure 4; Table 14), with the next peak 18 LOD less significant. The peak associated with *r1* was observed at position 138,468,760 on chromosome 10 with a LOD of 32.60 (Figure 4; Table 14). Analysis of the peak QTL for *c1* and *r1* across all individuals clearly showed that both genes were required for blue aleurone expression and both came from blue parent lines Tx903 and Tx906. The LOD scores of both blue alleles were likely lower than that for cob color and yellow endosperm because of the epistasis between the two blue aleurone genes and only 25.6% of the lines were blue. The LOD scores of *r1* could have had a lower association to blue aleurone in this population than *c1* because although both are regulators in the anthocyanin pathway, *c1*is upstream of *r1* in the anthocyanin pathway (Petroni et al., 2000), with *c1* regulating the production of pigment in the seed tissue and *r1* regulating

which tissues will express the pigment that is produced (Hanson et al., 1996). The visual qualitative approach to labeling the kernels blue or yellow yielded highly significant peaks with no signs of smaller effect, quantitative variation. The hypothesis to be tested was that there likely was quantitative variation for blue color and intensity but that it could not be visually scored and thus another quantitative method was needed to determine if smaller effects could be associated with kernel color.

Therefore, a colorimeter was used to observe whether other smaller effect QTL could be detected for kernel color. Association analysis of the level of black or white (L), and blue or yellow (b) colorimeter values provided the same peak loci positions on chromosomes 9 and 10 as the qualitative blue phenotype data (data not shown).

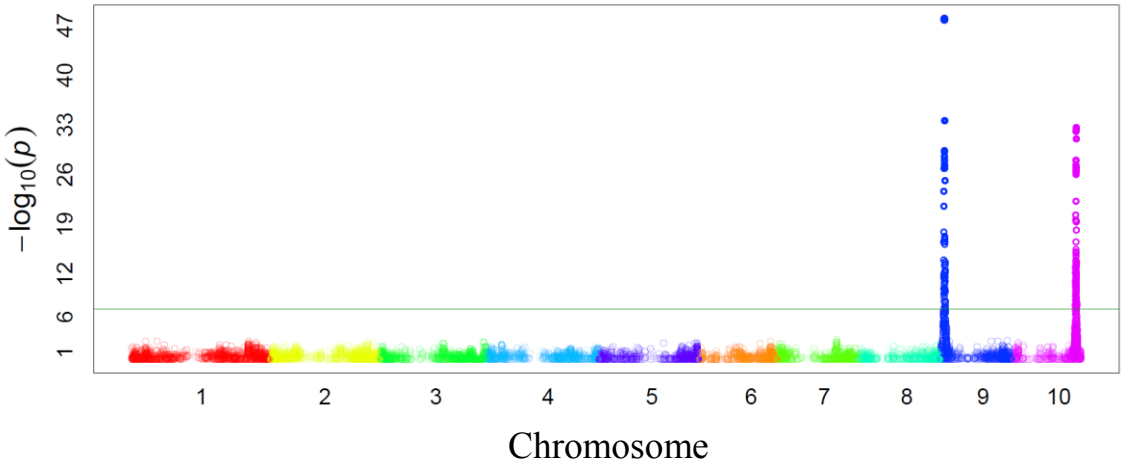

Figure 4. Graphical summary (Manhattan plot) of genome-wide association results for blue aleurone. The *x* axis represents the maize genome in physical order; the *y* axis shows -$\log_{10}$ *P* for all SNPs.

Interestingly, association analysis of the level of red or green (a) colorimeter value resulted in detection of all four of the peaks on chromosomes 1, 6, 9, and 10, co-localizing with color genes *p1*, *y1*, *c1*, *r1* respectively (data not shown). While the detection of the yellow endosperm and blue aleurone alleles may not be surprising, the detection of the cob color gene was unexpected. This was likely the result of red glumes remaining near the tip caps of kernels shelled from red cobs. The inability to detect additional significant loci or even the lack of near-significant peaks further reinforces the finding that *c1* and *r1* were the principle alleles responsible for blue aleurone in this population and would be the only loci needed to convert B73 to blue aleurone. This was surprising because as many as 10 genes have been hypothesized to contribute to blue aleurone expression, including those expressed in both the maize kernel and plant (Betran, 2000). Given the diversity of the parents, it seems unlikely that this many genes are needed for blue color. Additional surveys of the literature show that although these genes have been cloned using methods of transposon tagging, but this is the first time, to our knowledge, that association analysis has been used to identify the genes conditioning blue color, and specifically the first time that this has been done in breeding material relevant to improving blue corn production for growers.

4.3.5 Evaluating and Comparing Locus Detection Abilities of Sub-populations

4.3.5.1 Yellow Endosperm

Every subpopulation had significant (LOD>6.33) peaks within varying distances of *y1* on chromosome 6 (Table 15) for yellow endosperm. Not surprisingly, the three largest groupings which included the entire population (n=1141), all of the four-way

crosses (n=899), and the 4way3sib (n=488), had the smallest confidence interval, although it was surprising that this 54 bp interval landed within the gene region of *y1*. In fact, population size was nearly perfectly correlated (R^2= 0.998) to the respective LOD significance score as would be expected from a Mendelian inherited trait. An interesting observation was the extremely large confidence intervals and poor resolution for the 4way2sib and 4way1sib populations, which would be expected to have more effective recombination than the bi-parental populations. These results demonstrated that the ability to detect this locus was most improved with more (~500 individuals), but that intermating alone did not seem to result in much improvement. Only the 4way0sib subpopulation failed to capture the *y1* allele within a 1-LOD confidence interval of the peak QTL, likely due to stochastic chance.

While it would be expected that the two-parent populations would have the lowest QTL resolution, the population of Tx772 x Tx906 had similar or smaller confidence intervals than four of the sub-populations. While this could be due to stochastic chance, it could also be since the Tx772 x Tx906 population has ~30,000 more polymorphisms than the B73*Olc1* x Tx903 population.

With the overall high significance, high resolution, and high accuracy of the larger 4way populations, only the *y1* gene was implicated. In contrast the confidence interval in the 4way1sib population implicated 251genes and in the bi-parental populations 32 to 88 genes were implicated under the confidence interval (Table 15). If the y1 gene had been unknown before this study, as genes being mapped often are, these substantially greater number of candidate genes would require a much larger amount of

follow-up work to identify the causal candidate. Furthermore in the case of the 4way0sib, the 29 genes identified would all have been incorrect.

Table 15. Sub-population statistics for yellow endosperm (*y1*) peak locus position and 1-LOD confidence interval.

Group	n	Correctly identified gene	Peak position	LOD	Conf. int. pos.	Conf. int. dist.	Genes within conf. int.
all subpops	1141	Yes	82,017,348	107.08	82,017,294 – 82,017,402	54	1
all4ways	899	Yes	82,017,348	87.2	82,017,294 - 82,017,402	54	1
4way3sib	488	Yes	82,017,402	48.95	82,017,348 – 82,017,456	54	1
4way2sib	89	Yes	83,621,389	11.89	80,909,251 - 86,333,527	2,712,138	104
4way1sib	203	Yes	82,017,348	26.85	74,708,450 – 89,321,337	7,303,989	251
4way0sib	109	No	83,621,056	14.74	82,863,651 – 84,378,461	757,405	29*
B73*Olc1* x Tx903	121	Yes	78,735,091	16.90	75,116,552 – 82,353,630	3,618,539	88
Tx772 x Tx906	121	Yes	82,764,656	14.81	82,017,402 – 83,511,910	747,254	32
all2ways	242	Yes	82,017,348	29.41	81,171,679 – 82,863,017	845,669	35

*Gene total does not include the target gene, *y1*.

4.3.5.2 Red Cob

With only one of the four parents expressing red cob instead of white (B73*Olc1*) the analysis of cob color provided an interesting contrast to *y1* (Table 16). Reducing the number of individuals (n=920) to only include the four parent subpopulations increased the significance of the peak QTL position instead of decreasing it as was observed for yellow endosperm. Considering that the B73*Olc1* x Tx903 population and the combined

bi-parental populations had much larger confidence intervals and failed to detect the *p1* allele, it was not surprising that removing them and analyzing only the 4-parent populations resulted in improved accuracy and resolution of the peak associated with the *p1* gene. This increased significance with 920 individuals shared the same confidence interval size as the 4way3sib population which had a little more than half of the number of individuals (n= 501). An interesting observation in regards to the bi-parental populations was that the B73*Olc1* x Tx903 population and the Tx772 x Tx906 populations combined had similar confidence intervals but produced a two-fold increase in significance (LOD= 15.38). Because the addition of the Tx772 x Tx906 population only added white cob data it was surprising that it resulted in an increased LOD. The 4way0sib (n= 112) population detected *p1* and also had a LOD score of 14.35 compared to 7.82 for the B73*Olc1* x Tx903 population (n= 126), and suggested an opportunity to increase the significance of QTL peaks by including additional parents and/or intermating in the population design; at least when a marker closer to the gene of interest is not available. In addition, the four-parent populations all had smaller confidence intervals (better resolution) than the B73*Olc1* x Tx903 population alone or when combined with the Tx772 x Tx906 population. This highlighted the advantage of including more parents in a mapping population than a traditional bi-parental model as reported with other multi-parent designs (Huang et al., 2012; Mackay et al., 2014), especially in the case of the association mapping results for all of the 4-parent populations and the 4way3sib and 4way0sib subpopulations. These results suggest that an uncommon trait that is only represented in one or a limited number of inbred lines

may be more easily mapped with increased resolution if multiple parents are used, even if the additional parents are not themselves expressing the target trait. This finding will require further research however since the 4way1sib and 4way2sib subpopulations failed to detect *p1*.

Table 16. Sub-population statistics for red cob (*p1*) peak locus position and 1-LOD confidence interval.

Group	n	Correctly identified gene	Peak position	LOD	Conf. int. pos.	Conf. int. dist.	Genes within conf. int.
all subpops	1174	No	47,994,162	65.50	47,905,812 – 48,082,512	88,350	7*
all4ways	920	Yes	47,994,162	69.00	47,775,210 – 48,213,114	218,952	15
4way3sib	501	Yes	47,994,162	44.40	47,775,210 – 48,213,114	218,952	15
4way2sib	91	No	46,865,045	11.04	46,359,308 – 47,370,782	505,737	27*
4way1sib	210	No	47,775,210	23.46	47,644,608 – 47,905,812	130,602	4*
4way0sib	112	Yes	47,905,812	14.35	46,628,967 – 49,182,657	1,276,845	58
all2ways	253	No	51,293,959	15.38	48,652,544 – 53,935,374	2,641,415	104*
B73*Olc1* x Tx903	126	No	51,293,990	7.82	49,000,045 – 53,587,935	2,293,945	89*

*Gene total does not include the target gene, *p1*.

4.3.5.3 Blue Aleurone

The magnitude of the significance of the peak positions for both *c1* on chromosome 9 and *r1* on chromosome 10 were considerably lower than those for yellow endosperm and red cob (Table 17). In fact, the 4way2sib population did not have a

significant QTL peak (LOD<6.33) for either gene. The 4way2sib populations lack of significance can be mainly attributed to a small population size (n= 90). The Tx772 x Tx906 bi-parental subpopulation did not have a significant association for *c1* either. This demonstrates the difficulty of finding significant associations even for a two gene Mendelian trait when using a small population and a large marker set. Neither of the bi-parental populations captured *c1* in their confidence intervals, showing that the ~100 individual mapping populations that have become common throughout the plant breeding community may not be statistically sufficient for even a Mendelian inherited trait if a large number of markers are being used, as is often the case with GBS data. The B73*Olc1* x Tx903 population was able to detect the *r1* gene which suggests that a bi-parental population's ability to detect the association of a trait is also highly dependent on the trait and the distribution of variation between the individuals within the given population.

When the two bi-parental subpopulations were combined (n=247), *c1* was detected, and became significant with an accurate association to the region containing the *c1* gene (Figure 1). Although effective in detecting *c1*, a result of combining the two-parent crosses was a large confidence interval of nearly 1.3 Mb, much larger than any four-parent group (Table 17). Even comparing with the 4way0sib subpopulation that had the largest confidence interval of the 4-parent populations at 733,357 bp, the interval for *c1* captured 24 fewer known genes than the 1-LOD interval from the combined bi-parental populations. The difference in confidence interval size is likely from less effective recombination in the bi-parental populations.

Table 17. Sub-population statistics for blue aleurone peak QTL positions (*c1* and *r1*) and 1-LOD confidence interval.

Group	n	Correctly identified gene	Peak position	LOD	Conf. int. pos.	Conf. int. dist.	Genes within conf. int.
colored aleurone1							
all subpops	1128	Yes	9,397,546	47.96	8,710,142 – 9,741,248	343,702	18
all4ways	881	Yes	9,397,546	44.48	8,710,142 – 9,741,248	343,702	18
4way3sib	474	Yes	9,397,546	27.95	8,710,142 – 9,741,248	343,702	18
4way2sib	90	Yes	9,397,546	4.12	8,710,142 – 9,741,248	343,702	18
4way1sib	200	Yes	9,741,248	13.50	9,397,546 – 10,084,950	343,702	12
4way0sib	108	Yes	9,397,546	7.11	8,664,189 – 10,130,903	733,357	27
all2ways	247	Yes	9,741,248	9.64	8,456,305 – 11,026,191	1,284,943	51
B73*Olc1* x Tx903	124	No	9,147,883	6.26	8,898,220 – 9,397,546	249,663	8*
Tx772 x Tx906	124	No	11,155,191	5.18	10,529,228 – 11,781,154	625,963	43*
colored1							
all subpops	1128	Yes	138,468,760	32.6	138,327,025 – 138,610,495	141,735	6
all4ways	881	Yes	138,044,150	21.48	137,539,749 – 138,548,551	504,401	30
4way3sib	474	Yes	138,468,760	16.68	138,044,150 – 138,893,370	424,610	33
4way2sib	90	No	137,903,207	5.86	137,762,385 – 138,044,029	140,822	6*
4way1sib	200	Yes	138,454,640	11.88	138,028,587 – 138,880,693	426,053	31
4way0sib	108	No	137,193,006	7.39	137,103,610 – 137,282,402	89,396	3*
all2ways	247	Yes	138,454,640	14.83	138,046,877 – 138,862,403	407,763	19
B73*Olc1* x Tx903	124	Yes	138,262,103	10.07	138,051,755 – 138,472,451	210,348	12
Tx772 x Tx906	124	Yes	138,454,640	5.43	138,093,745 – 138,815,535	360,895	21

*Gene total does not include the target gene *c1* or *r1*.

Another interesting observation for the *c1* QTL was that every group excluding the 4way0sib, and both bi-parental populations shared the same confidence interval distance of 343,702 bp. Closer examination of these two shared positions of highest significance (9,397,546 bp and 9,741,001 bp), revealed that of the 30 SNPs physically located between them, none of these were polymorphic in both bi-parental subpopulations. This lack of joint polymorphism between the two sets of parents for other markers is likely the reason that there was an 18 LOD drop off to the next significant SNP for *c1*.

4.3.6 Comparing QTL Resolution Due to Intermating

This population was designed to compare mapping results for known genes across diverse genetic population designs with all factors except size held constant. We chose to further compare the 4way3sib sub-population with the two bi-parental populations. To make this comparison independent of population size, 10 smaller sets were randomly sub-sampled with replacement from the larger 4way3sib subpopulation. In theory, the 4way3sib population should have the most recombination, given the additional generations of intermating prior to inbreeding. The amount of variation between the randomly sampled groups of the 4way3sib subpopulation is evident for all three traits (Figures 5 and 6).

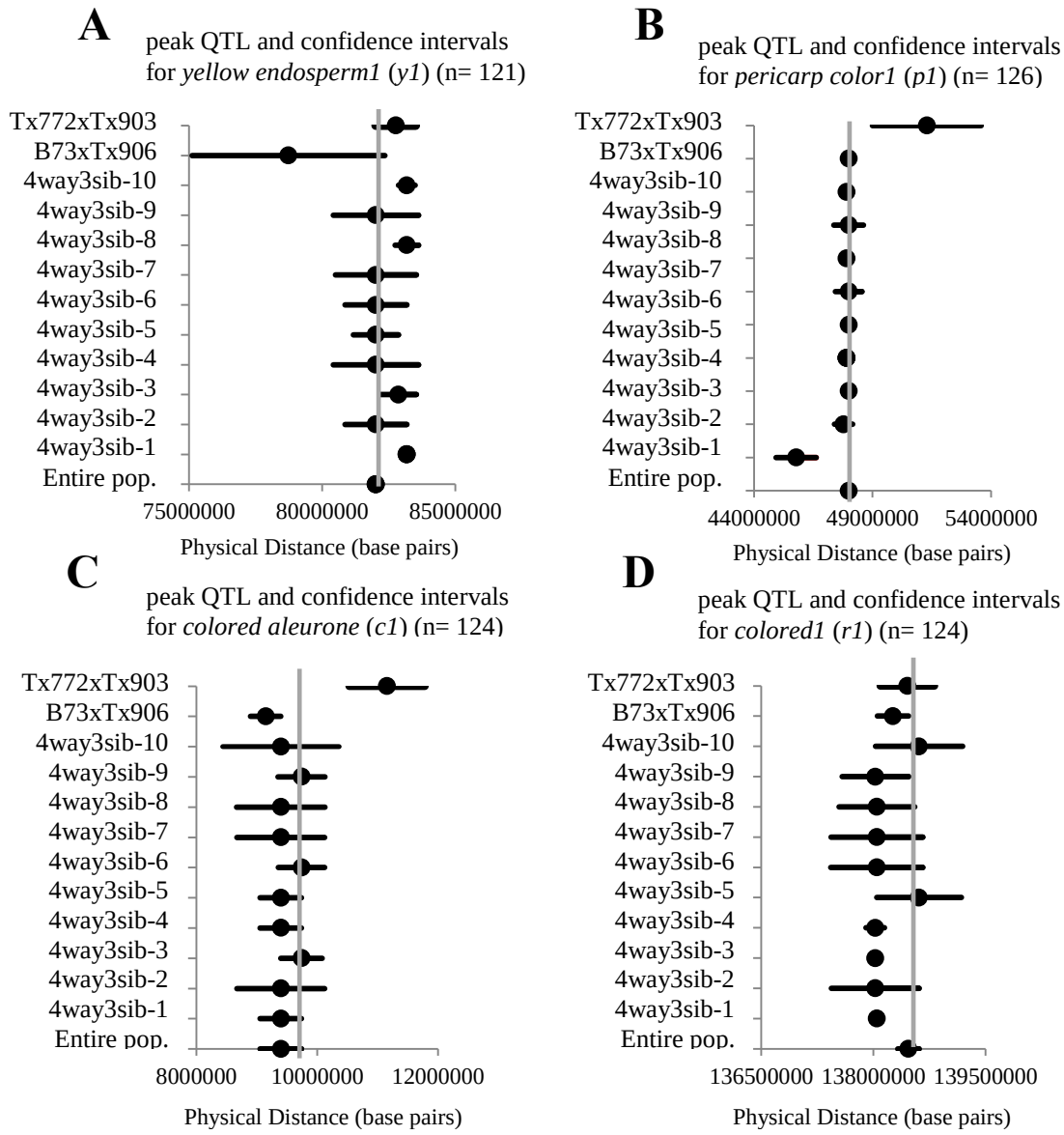

Figure 5. Peak QTL and 1-LOD confidence interval comparison between randomly sampled 4way3sib subpopulation groups and the bi-parental subpopulations for yellow endosperm on chromosome 6 (A), red cob on chromosome 1 (B), and blue aleurone on chromosomes 9 (C) and 10 (D). The *x* axis indicates physical distance (bp) and the *y* axis indicates subpopulation.

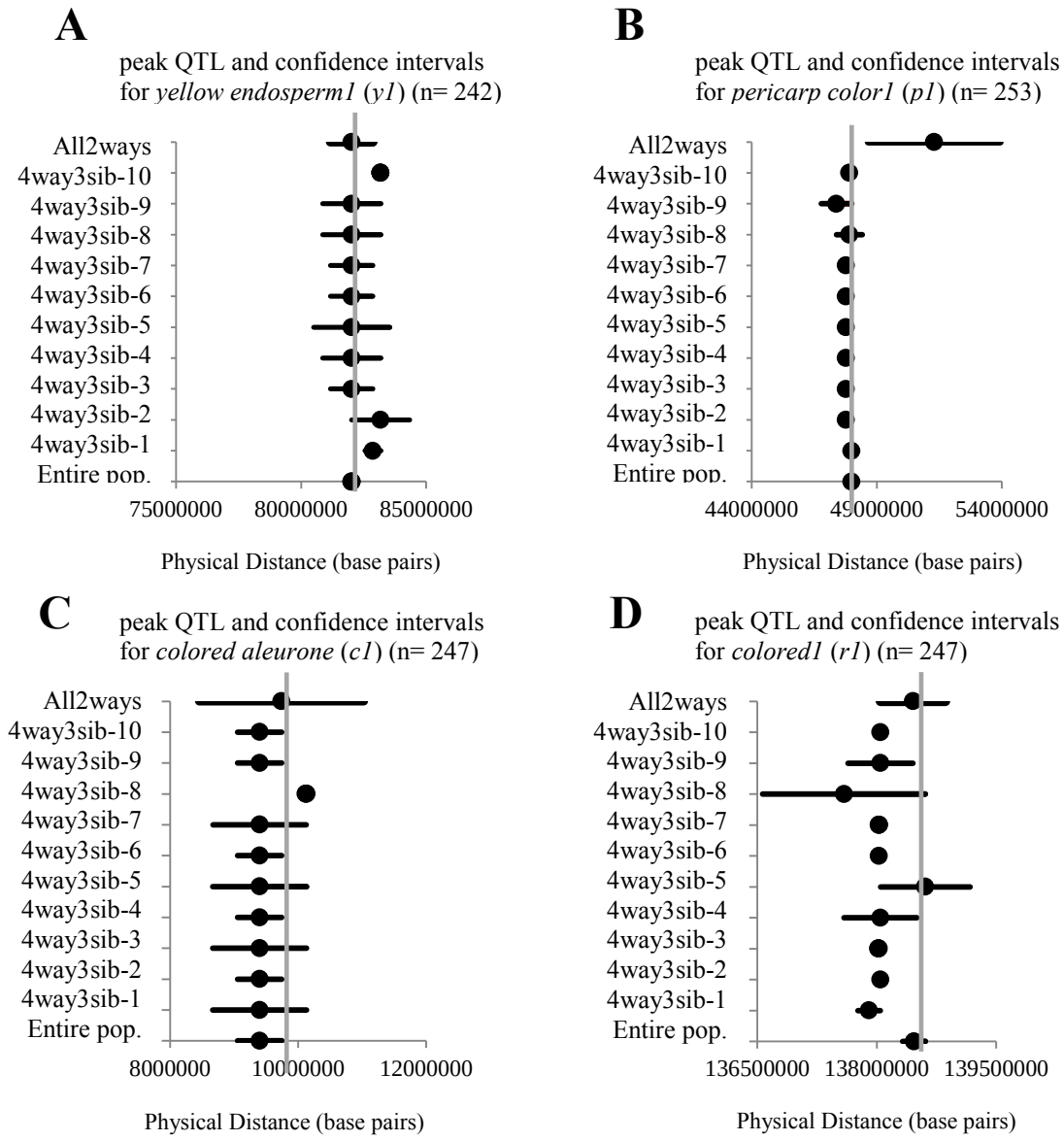

Figure 6. Peak QTL and 1-LOD confidence interval comparison between randomly sampled 4way3sib subpopulation groups and the combined bi-parental subpopulations for yellow endosperm on chromosome 6 (A), red cob on chromosome 1 (B), and blue aleurone on chromosomes 9 (C) and 10 (D). The *x* axis indicates physical distance (bp) and the *y* axis indicates subpopulation.

Panels A and B in Figure 5 show results that were mostly expected, the doubling of population size causing more groups to capture the *y1* allele and improve the proximity of the peak QTL position to the causative allele. What wasn't expected were the number of sub-sampled sets completely missing the actual position of the *y1* allele as well as the very large 1-LOD confidence interval and thus poor resolution of the B73*Olc1* x Tx903 population. This lower resolution could be related to an identical by decent haplotype between these two related parents.

Small confidence intervals were prevalent in the results of red cob analysis for both population sizes except in the B73*Olc1* x Tx903 population. In panels C and D of Figure 5, only two randomly sampled groups detected *p1*, although nearly all of the peak QTL positions were within ~200,000 bp, including group 12 (all individuals) which was only 34,985 bp away from the *p1* gene. Although the proximity of these peak QTL are promising, the tight confidence intervals did not allow for effective detection of the allele for red cob color. These results show the difficulty of mapping a trait that is not highly represented in the population (in this case only one parent had red cob color), and the results of attempting to accurately detect a gene region when a marker is not available within the region of interest

The observation that small population sizes (n= 100-150) or populations with less opportunity for recombination (bi-parental) may not be sufficient to map Mendelian genes when large marker sets are used was reinforced in this analysis with neither of the bi-parental populations accurately detecting *c1* on chromosome 9. The random samples of the 4way3sib subpopulation had variation for confidence interval size for *c1*, but only

one group, once the sample size was doubled, failed to detect *c1*. For *r1*, the bi-parental populations performed favorably when compared to the 4-parent subsampled populations. In fact, when looking at 124 individuals, three of the ten 4way3sib random sample subpopulations failed to detect *r1*. As noted earlier, the doubling of individuals increased the significance of the QTL peaks and usually improved resolution. Unlike the other color genes, the confidence intervals became too small for *r1*, causing seven of the 4-parent populations, to completely miss the gene. This could be partially due to the difference between the *c1* and *r1* alleles. While *c1* is a single gene, the region harboring *r1* is a complex of multiple *r1* alleles which regulate specific pigmentation patterns (Petroni, 2000). It is possible that *r1* derivatives are segregating within this population in a manner that the blue aleurone phenotyping was not able to observe.

4.4 Conclusion

The novel FPM population has increased our understanding of the simplicity of blue aleurone expression, and the advantages and problems that arise with complex mating designs which incorporate multiple parents and increased recombination. Association analysis determined that *colored aleurone1* (*c1*) and *colored1* (*r1*) were the primary genes segregating for the control of blue aleurone in this population, and by extension are the only two needed for the conversion of an elite yellow dent line, such as B73, to blue. By analyzing the entire population and individual sub-populations for yellow endosperm and red cob color in addition to blue aleurone, the distinct relationship between increased population size and significance (LOD score) was observed, as was the inability of the typical mapping population size (~130) to

accurately detect Mendelian inherited traits when correcting for multiple marker testing. The ability of a smaller four-parent population, in this case the 4way0sib population, to capture the *p1* gene when the bi-parental populations could not, is an approach to improve genetic mapping of a rare trait by including additional parents that may not express the target trait. While many interesting results have, and will continue to come from this unprecedented population, a proper mapping population design recommendation is dependent on the specific crop, crop diversity, the trait of interest, and the difficulty of performing crosses. While it does seem advantageous from a QTL mapping standpoint to incorporate more parents into mapping populations, the effect of intermating was distorted by variation in population size and does not seem as likely to be worth the additional effort. To fully understand the usefulness of these mating designs, proper mapping software that can handle more than two parents as well as large marker sets is needed so that genetic maps and linkage mapping can be performed to compare results with the exploratory association mapping done here.

5. CONCLUSIONS

The FPM population has served as a first insight into the potential of multi-parent populations in maize, especially when coupled with GBS. While the increase to four parents did increase the phenotypic variation of the population, especially for plant height, intermating did not seem to have an overwhelming effect. Similarly in association mapping, the four-parent subpopulations often had smaller QTL confidence intervals than the bi-parental subpopulations, especially the subpopulations with a large number of individuals. The largest single subpopulation also had the greatest amount of intermating (three generations), making it difficult to separate population size effect from an intermating effect. For most traits however, one or more of the 4-parent subpopulations would either fail to detect the gene, or have a large confidence interval around the gene. So while multi-parent populations do seem to have an advantage in both phenotypic and genotypic diversity, a larger population size (~500) of the 4-parent subpopulation without intermating would be required to perform a direct comparison with the 4-parent subpopulation that was intermated three times.

The potential impact of this population goes far beyond what has been covered in this dissertation. The interesting aspects of this population that have yet to be researched include constructing a genetic map and traditionally mapping the phenotypic data, and quantifying and comparing recombination rate differences between each subpopulation. These objectives were originally planned to be part of this dissertation, but it was soon determined that proper analysis tools are not currently available to handle a maize

population of this size and scope. Other projects from this population that are currently underway include continued phenotypic analysis, including a subset of individuals that represent the minimum and maximum values of a variety of traits including flowering time and plant height. Additionally, in another project it was discovered that this population sufficiently segregates for oleic acid content such that this trait could be quantified for the entire population and possibly identify the causative gene region for the increase in oleic acid found in B73*Olc1*. The preliminary results from ground maize samples scanned with near infrared spectroscopy show promise to develop an oleic acid calibration curve, which could be used to screen maize samples across the breeding program.

With any project, there are always things that could have been done differently or problems that arise that become opportunities for learning. The B73*Olc1* x Tx903 bi-parental population exhibited reduced genetic diversity due to the relatedness of the two parents. This became very obvious with the low vigor of the progeny from this cross and the reduced number of polymorphisms, compared to the Tx772 x Tx906 bi-parental population. Performing the initial bi-parental crosses differently at the offset would have made maintaining bi-parental population size easier and very well could have resulted in even greater variation from the 4-parent crosses than what was observed. However, these regions of identity by decent may result in novel discoveries of recombination. With this population, a large importance was placed on randomly selecting individuals for each subsequent generation. While this is an approved practice for the sake of maintaining variation in the population, it can also lead to an overwhelming number of individuals

that either become lost in future generations because they fail to produce seed, or remain difficult to increase seed on. A small degree of selection, as long as sufficient population sizes are maintained, should not have a deleterious effect on genetic diversity in the population as a whole.

Within the plant breeding industry, in terms of maize, an increased emphasis has been put on fast strategies of progeny generation that increase the genetic gain per selection cycle. This has resulted in a reliance on bi-parental populations and double haploid strategies that result in inbred lines within ~2 years. While it is hard to argue that this strategy hasn't been successful, from this population I have learned that multi-parent strategies deserve consideration, depending on the end goal. Because double haploid technology takes the labor and time of manually self-pollinating these crosses to inbred lines, it will likely be worth a breeder's time to have a small portion of their breeding scheme dedicated to exotic mating designs beyond the traditional bi-parental cross.

REFERENCES

Allard, R. 1960. Principles of plant breeding. Wiley, New York.

Alpert, K.B. and S.D. Tanksley. 1996. High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: a major fruit weight quantitative trait locus in tomato. Proc. Natl. Acad. Sci. USA 93:15503-15507.

Anderson, J.A., G.A. Churchill, J.E. Autrique, S.D. Tanksley, and M.E. Sorrels. 1993. Optimizing parental selection for genetic linkage maps. Genome 31:181-186.

Arbelbide, M. and R. Bernardo. 2004. Random mating before selfing in maize BC_1 populations. Crop Sci. 44:401-404.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLos ONE 3:e3376.

Balasubramanian, S., C. Schwartz, A Singh, N. Warthmann, M.C. Kim, J. Maloof, et al. 2009. QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. PLoS ONE 4:e4318.

Bandillo, N., C. Raghavan, P.A. Muyco, M.A.L. Sevilla, I.T. Lobina, C.J. Dilla-Ermita, et al. 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice J. 6:11.

Bentley, D.R., S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53-59.

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits

in maize. Crop Sci. 47:1082-1090.

Betran, F.J., A.J. Bockholt, F. Fojt, D. Pietsch, L. Rooney, and K. Mayfield. 2004. Registration of Tx114 Maize. Crop Sci. 44:1036.

Betran, F.J., A.J. Bockholt, and L.W. Rooney. 2000. Blue corn. p. 293-301. *In* A.R. Hallauer (ed.) Specialty corns. CRC Press, Boca Raton, FL.

Blanc, G., and K. H. Wolfe. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667-1678.

Broman, K.W. 2004. The genomes of recombinant inbred lines. Genetics 169:1133-1146.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889-890.

Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, et al. 2009. The genetic architecture of maize flowering time. Science 325:714-718.

Buckner, B., T.L. Kelson, and D.S. Robertson. 1990. Cloning of the *y1* locus of maize, a gene involved in the biosynthesis of carotenoids. Plant Cell 2:867-876.

Cavanagh, C., M. Morell, I. Mackay, and W. Powell. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr. Opin. Plant Biol. 11:215-221.

Cheema, J.J.S. and J. Dicks. 2009. Computational approaches and software tools for genetic linkage map estimation in plants. Brief. Bioinform. 10:595-608.

Cheema, J.J.S., T.H.N. Ellis and J. Dicks. 2008. Exploring genome wide genetic linkage

manifolds. In: Proceedings of the Genome Informatics Conference. Hinxton, Cambridge, UK.

Chia, J-M., C. Song, P.J. Bradbury, D. Costich, N. de Leon, J. Doebley, et al. 2012. Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. 44:803-809.

Ching, A., K.S. Caldwell, M. Jung, M. Dolan, O.S. Smith, S. Tingey, et al. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet. 3:19.

Collaborative Cross Consortium. 2012. The genome architecture of the collaborative cross mouse genetic reference population. Genetics 2:389-401.

Cone, K.C., F.A. Burr, and B. Burr. 1986. Molecular analysis of the maize anthocyanin regulatory locus *C1*. Proc. Natl. Acad. Sci. USA 83:9631-9635.

Darvasi, A., and M. Soller. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199-1207.

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M Catchen, and M.L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499-510.

Dekalb-Pfizer Genetics. 1988. Corn 'PB80'. Plant Variety Protection Certificate. Issued July 29.

Del Pozo-Insfran, D., S.O. Serna Saldivar, C.H. Brenes, and S.T. Talcott. 2007. Polyphenolics and antioxidant capacity of white and blue corns processed into tortillas and chips. Cereal Chem. 84:162-168.

De La Fuente, G.N., I. Barrero, S.C. Murray, T. Isakeit, and M.V. Kolomiets. 2014. Improving Maize Production under Drought Stress. Handbook of Plant and Crop Physiology. Third Edition. pp. 891-916

Dellaporta, S.L., I.M. Greenblatt, J.L. Kermicle, J.B. Hicks, and S.R. Wessler. 1988. Molecular cloning of the *R-nj* gene by transposon tagging with *Ac*. In: Chromosome Structure and Function: Impact of New Concepts, J.P. Gustafson and R. Appels, eds. New York: Plenum Press. pp.263-282.

Doebley, J., A. Stec, and L. Hubbard. 1997. The evolution of apical dominance in maize. Nature 386:485-488.

El-Din El-Assal, S., C. Alonso-Blanco, A.J.M. Peeters, V. Raz, and M. Koornneef. 2001. A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. Nat. Genet. 29:435-440.

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

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357-374.

Flint, J. and R. Mott. 2001. Finding the molecular basis of quantitative traits: successes and pitfalls. Nat. Rev. Genet. 2:437-445.

Fridman, E., T. Pleban, and D. Zamir. 2000. A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. USA 97:4718-4723.

Ford, R.H. 2000. Inheritance of kernel color in corn: explanations & investigations. Amer. Biol. Teach. 62:181-188.

Fu, H., and H. K. Dooner. 2002. Intraspecific violation of genetic colinearity and its implications in maize. Proc. Natl. Acad. Sci. USA 99.14: 9573-9578.

Gaut, B.S. 2001. Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. Genome Res. 11:55-66.

Gaut, B.S., and A.D. Long. 2003. The lowdown on linkage disequilibrium. Plant Cell 15:1502-1506.

Gore, M.A., J.M. Chia, R.J. Elshire, Q. Sun, E.S. Ersoz, B.L. Hurwitz, et al. 2009. A first-generation haplotype map of maize. Science 326:1115-1117.

Haldane, J.B.S. and C.A.B. Smith. 1947. A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. Ann. Eugen. 14:10-31.

Haley, C.S. 1999. Advances in QTL mapping. In: J.C.M. Dekkers, S.J. Lamont, and M.F. Rothschild, editor, Proceedings of from lush to genomics: visions for animal breeding and genetics. Iowa State Univeristy, Ames, IA.

Han, L. and S. Xu. A fisher scoring algorithm for the weighted regression method of QTL mapping. 2008. Heredity 101:453-464.

Hanson, M.A., B.S. Gaut, A.O. Stec, S.I. Fuerstenberg, M.M. Goodman, E.H. Coe, et al. 1996. Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. Genetics 143:1395-1407.

Hattori, T., V. Vasil, L. Rosenkrans, L.C. Hannah, D.R. McCarty, and I.K. Vasil. 1992.

The *Viviparous-1* gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during see maturation in maize. Gene Devel. 6:609-618.

Holden's Foundation Seeds, Inc. 1991. Corn 'LH195'. Plant Variety Protection Certificate. Issued May 31.

Holden's Foundation Seeds, Inc. 1985. Corn 'LH82'. Plant Variety Protection Certificate. Issued July 26.

Hu, Z., J.D. Ehlers, P.A. Roberts, T.J. Close, M.R. Lucas, S. Wanamaker, et al. 2012. ParentChecker: a computer program for automated inference of missing parental genotype calls and linkage phase correction. BMC Genet. 13:9.

Huang B.E., and A.W. George. 2011. R/mpMap: A computation platform for the genetic analysis of multi-parent recombinant inbred lines. Bioinformatics 27:727-729.

Huang, B.E., A.W. George, K.L. Forrest, A. Kilian, M.J. Hayden, M.K. Morell, et al. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol. J. 10:826-839.

Hufford, M.B., X. Xu, J. Van Heerwaarden, T. Pyhajarvi, J-M. Chia, R.A. Cartwright, et al. 2012. Comparative population genomics of maize domestication and improvement. Nat.Genet. 44:808-811.

Hung, H-Y., C. Browne, K. Guill, N. Coles, M. Eller, A. Garcia, et al. 2012. The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. Heredity 108:490-499.

Hyun, J.W., and H.S. Chung. 2004. Cyanidin and malvidin from *Oryza sativa* cv. *Heugjinubyeo* mediate cytotoxicity against human monocytic leukemia cells by arrest of G2/M phase and induction of apoptosis. J. Agric. Food Chem. 52:2213-2217.

Jensen, N.F. 1970. A diallel selective mating system for cereal breeding. Crop Sci. 10:629-635.

Jha, S.N. 2010. Colour measurements and modeling. In Nondestructive evaluation of food quality (pp.17-40). Springer Berlin Heidelberg.

Kearsey, M.J. and A.G.L. Farquhar. 1998. QTL analysis in plants; where are we now? Heredity 80:137-142.

Kroymann, J., S. Donnerhacke, D. Schnabelrauch, and T. Mitchell-Olds. 2003. Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc. Natl. Acad. Sci. USA 100:14587-14592.

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

Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Meth. 9:29.

Kozik, A. and R. Michelmore. 2006. MapMapper and CheckMatrix—python scripts to infer orders of genetic markers and for visualization and validation of genetic maps and haplotypes. In: Proceedings of the Plant and Animal Genome XIV Conference, San Diego, CA.

Lander, E.S. and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, et al. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Larrinua, I.M. and S.B. Belmar. 2008. Bioinformatics and its relevance to weed science. Weed Sci. 56:297-305.

Lechelt, C., T. Peterson, A. Laird, J. Chen, S.L. Dellaporta, E. Dennis, et al. 1989. Isolation and molecular analysis of the maize *P* locus. Mol. Gen. Genet. 219:225-234.

Lee, M., N. Sharapova, W.D. Beavis, D. Grant, M. Katt, D. Blair, et al. 2002. Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol. Biol. 48:453-461.

Lewis, Z.A., A.L. Shiver, N. Stiffler, M.R. Miller, E.A. Johnson, and E.U. Selker. 2007. High-density detection of restriction-site-associated DNA markers for rapid mapping of mutated loci in neurospora. Genetics 177:1163-1171.

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 28:2397–239.

Liu G., Liu J., Cui X., Cai L. 2012. Sequence-dependent prediction of recombination hotspots in *Saccharomyces cerevisiae*. J. Theor. Biol. 293:49-54.

Liu, Y. and Z-B. Zeng. 2000. A general mixture model approach for mapping

quantitative trait loci from diverse cross designs involving multiple inbred lines. Genet. Res. 75:345-355.

Llorente, C.F., F.J. Betran, A. Bockholt, and F. Fojt. 2004. Registration of Tx772 Maize. Crop Sci. 44:1036-a-1037.

Lopez-Martinez, L.X., R.M. Oliart-Ros, G. Valerio-Alfaro, C. Lee, K.L. Parkin, and H.S. Garcia. 2009. Antioxidant activity, phenolic compounds and anthocyanins content of eighteen strains of Mexican maize. Food Sci. Technol. 42:1187-1192.

Lu, Y., S. Zhang, T. Shah, C. Xie, Z. Hao, X. Li, et al. 2010. Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc. Natl. Acad. Sci. USA 107:19585-19590.

Mackay, I.J., P. Bansept-Basler, T. Barber, A.R. Bentley, J. Cockram, N. Gosman, et al. 2014. An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: Creation, properties, and validation. G3: Genes, Genomes, Genetics 4:1603-1610.

Mahan, A.L., S.C. Murray, L.W. Rooney, and K.M. Crosby. 2013. Combining ability for total phenols and secondary traits in a diverse set of colored (Red, Blue, and Purple) Maize. Crop Sci. 53:1-8.

Manly, K.F., R.H. Cudmore and J.M. Meer Jr. 2001. Map Manager QTX, cross-platform software for genetic mapping. Mamm. Genome. 12:930-932.

Mathias, R.A., P. Gao, J.L. Goldstein, A.F. Wilson, E.W. Pugh, P. Furbert-Harris, et al.

2006. A graphical assessment of *p*-values from sliding window haplotype tests of association to identify asthma susceptibility loci on chromosome 11q. BMC Genet. 7:38.

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H. Li, Q, Sun, et al. 2009. Genetic properties of the maize nested association mapping population. Science 325:737-740.

Miller M.R., J.P. Dunham, A. Amores, W.A. Cresko, and E.A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240-248.

Moehring, J., E.R. Williams, and H.P. Piepho. 2014. Efficiency of augmented p-rep designs in multi-environmental trials. Theor. Appl. Genet. 127:1049-1060.

Morgante, M., and F. Salamini. 2003. From plant genomics to breeding practice. Curr. Opin. Biotechnol. 14:214-219.

Morishige, D.T., P.E. Klein, J.L. Hilley, S.M.E. Sahraeian, A. Sharma, and J.E. Mullet. 2013. Digital genotyping of sorghum – a diverse plant species with a large repeat-rich genome. BMC Genomics 14:448.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, et al., 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc. Natl. Acad. Sci. USA 110:453-458.

Morton, N.E. 1955. Sequential tests for the detection of linkage. Am. J. Human Genet. 7:277-318.

Mott, R., C.J. Talbot, M.G. Turri, A.C. Collins, and J. Flint. 2000. A method for fine

mapping quantitative trait loci in outbred animal stocks. Proc. Natl. Acad. Sci. USA 97:12649-12654.

Mouchel, C.F., G.C. Briggs, and C.S. Hardtke. 2004. Natural genetic variation in *Arabidopsis* identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. Genes Dev. 18:700-714.

Murray, S.C., W.L. Rooney, M.T. Hamblin, S.E. Mitchell, and S. Kresovich. 2009. Sweet sorghum genetic diversity and association mapping for brix and height. Plant Gen. 2: 48–62.

Paterson, A. H., J. E. Bowers, and B. A. Chapman. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc. Nat. Acad. Sci. USA 101:9903-9908.

Paterson, A.H., J.E. Bowers, R. Bruggman, I. Dubchak, J. Grimwood, H. Gundlach, et al. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457:551-556.

Petroni, K., E. Cominelli, G. Consonni, G. Gusmaroli, G. Gavazzi, and C. Tonelli. 2000. The developmental expression of the maize regulatory gene *Hopi* determines germination-dependent anthocyanin accumulation. Genetics 155: 323-336.

Piazza P., A. Procissi, G.I. Jenkins, and C. Tonelli. 2002. Members of the *c1*/*pl1* regulatory gene family mediate the response of maize aleurone and mesocotyl to different light qualities and cytokinins. Plant Physiol. 128: 1077-1086.

Pioneer Hi-Bred International, Inc. 1988. Corn ‘PHV63’. Plant Variety Protection Certificate. Issued July 29.

Poland, J.A., P.J. Brown, M.E. Sorrels, and J-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253.

Price, A.H, E.M. Young, and A.D. Tomos. 1997. Quantitative trait loci associated with stomatal conductance, leaf rolling and heading date mapped in upland rice (*Oryza sativa*). New Phytol. 137:83-91.

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14:R55.

Rothberg, J.M., W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, et al. 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348-352.

Salinas-Moreno, Y., J.J. Lopez-Reynoso, G.B. Gonzalez-Flores, and G.Vazquez-Carrillo. 2007. Compuestos fenolicos del grano de maiz y surelacion con el oscurecimiento de masa y tortilla. (Phenolic compoundsof maize grain and their relationship with darkening in dough and tortilla.) (In Spanish, with English translation.) Agrociencia 41:295–305.

Scheet P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Human Genet. 78:629-644.

Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, et al. 2009. The

B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115.

Schwartz, D., and E. Dennis. 1986. Transposase activity of the Ac controlling element in maize is regulated by its degree of methylation. Mol. Gen. Genet. 205:476-482.

Selinger, D.A., and V.L. Chandler. 1999. A mutation in the *pale aleurone color1* gene identifies a novel regulator of the maize anthocyanin pathway. Plant Cell 11:5-14.

Shendure, J., and H. Ji. 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26:1135-1145.

Springer, N.M., and R.M. Stupar. 2007. Allelic variation and Heterosis in maize: how do two halves make more than a whole?. Genome Res. 17:264-275.

Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J. 3:739-744.

Styles, E.D., and O. Ceska. 1972. Flavonoid pigments in genetic strains of maize. Phytochemistry 11:3019-3021.

Swigoňová, Z., J. Lai, J. Ma, W. Ramakrishna, V. Llaca, J.L. Bennetzen, et al. 2004. Close split of sorghum and maize genome progenitors. Genome Res. 14:1916-1923.

Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley and B.S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1of maize (*Zea mays* ssp. *Mays* L.). Proc. Natl. Acad. Sci. USA 98:9161-9166.

Thudi, M., Y. Li, S.A. Jackson, G.D. May, and R.K. Varshney. 2012. Current state-of-

art of sequencing technologies for plant genomics research. Brief. Funct. Genom. 11:3-11.

Tsuda, T., F. Horio, K. Uchida, H. Aoki, and T. Osawa. 2003. Dietary Cyanidin 3-O-β-D-glucoside-rich purple corn color prevents obesity and ameliorates hyperglycemia in mice. J. Nutr. 133:2125-2130.

Verbyla, A., C.R. Cavanagh, and K.L. Verbyla. 2014. Whole-genome analysis of multienvironment or multitrait QTL in MAGIC. G3: Genes, Genomes, Genetics 4:1569-1584.

Vroh Bi, I., M.D. McMullen, H. Sanchez-Villeda, S. Schroeder, J. Gardiner, M. Polacco, et al. 2005. Single nucleotide polymorphisms and insertion-deletions for genetic markers and anchoring the maze fingerprint contig physical map. Crop Sci. 46:12-21.

Werner, J.D., J.O. Borevitz, N. Warthmann, G.T. Trainer, J.R. Ecker, J. Chory, et al. 2005. Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proc. Natl. Acad. Sci. USA 102:2460-2465.

Wolfgruber, T.K., A. Sharma, K.L. Schneider, P.S. Albert, D-H. Koo, J. Shi, et al. 2009. Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. PLoS Genet. 5:e1000743

Wright, A. 1995. A gene conditioning high oleic maize oil, OLC 1. Maydica 40:85-88.

Wu, Y., P.R. Bhat, T.J. Close and S. Lonardi. 2008. Efficient and accurate construction

of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. 4:e1000212.

Wurschum, T. 2012. Mapping QTL for agronomic traits in breeding populations. Theor. Appl. Genet. 125:201-210.

Xu, S. 1996. Mapping quantitative trait loci using four-way crosses. Genet. Res. 68:175-181.

Xu, C., X. Wang, Z. Li, and S. Xu. 2009. Mapping QTL for multiple traits using Bayesian statistics. Genet. Res. 91:23-37.

Yan, A., H. Zhi-qiu, T. Zai-xiang, W. Xue-feng, Xu Chen-wu. 2008. A general method for QTL mapping in multiple related populations derived from multiple parents. Rice Sci. 16:45-50.

Yan, J., M. Warburton, and J. Crouch. 2011. Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. Crop Sci. 51:433-449.

Yandell, B.S., T. Mehta, S. Banerjee, D. Shriner, R. Venkataraman, J.Y. Moon, et al. 2007. R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. Bioinformatics 23:641-643.

Yi, N. and S. Xu. 2002. Linkage analysis of quantitative trait loci in multiple line crosses. Genetica 114:217-230.

Yisraeli, J., and M. Szyf. 1984. Gene methylation patterns and expression. In: Razin, A., Cedar, H., Rigg, A.D. (eds). DNA methylation: biochemistry and biological significance. Springer, NY, pp 353-376.

Yu, J., and E.S. Buckler. 2006. Genetic association mapping and genome organization of

maize. Curr. Opin. Biotechnol. 17:155-160.

Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178:539-551.

Zeng, Z.B. 1994. A composite interval mapping method of locating multiple QTLs. Proceedings of the 4th World Congress on Genetics Applied to Livestock Production.

Zhang, Y-M. and S. Xu. 2005. A penalized maximum likelihood method for estimated epistatic effects of QTL. Heredity 95:96-104.

Zhang, Z., E. Ersoz, C-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360.

Zhao, C., M.M. Giusti, M. Malik, M.P. Moyer, and B.A. Magnuson. 2004. Effects of commercial anthocyanin-rich extracts on colonic cancer and nontumorigenic colonic cell growth. J. Agric. Food Chem. 52:6122-6128.

APPENDIX

A1. Protocols

Appendix 1. Phenol-Chloroform DNA Extraction Method

1. Put 8 -10 1/4 inch seedlings or 15 leaf punches in 1.2ml racked library tubes containing a ¼ inch steel rod.
2. For dried tissue cut into very fine pieces for about 100 µg.
3. If extracting at a later date place in -80 freezer, let tissue warm to RT before adding buffer.
4. Add 430 µl fresh buffer working solution at 65° to each tube, attach caps.
5. Use 450 µl for dried tissue to rehydrate then heat for 15-30 minutes at 65°C.
6. Using the Geno/Grinder® (Zymo Research, Irvine, CA), grind the samples for 1 minute at 1400 strokes (setting 400), repeating as needed.
7. Spin in centrifuge briefly to get ground tissue off of the strip caps.
8. Incubate in oven or water bath at 65°C for 30 minutes.
9. The phenol and chloroform steps should be done in the fume hood.
10. Add 430 µl phenol:chloroform + isoamyl alcohol (25:24:1 Amresco, Solon, OH) to each tube, attach new strip caps, place a trifold paper towel sheet on top of tubes, add lid, secure tightly in plexi-glass clamp apparatus, shake vigorously for 5 minutes.
11. Spin in centrifuge for 30 min at 3250 rpm, transfer supernatant to new tubes.
12. Add 430 µl chloroform + isoamyl alcohol (24:1 Amresco, Solon, OH) to each tube, attach new caps, paper towel, secure in clamp apparatus, invert ~ 40 times.

13. Spin in centrifuge for 30 min at 3250 rpm, transfer the supernatant to a 96 deep well plate.
14. Add 450 µl cold 100% isopropanol, pipette up and down 5X to mix. Cover with aluminum tape (if needed plates can be placed in freezer overnight).
15. Spin for 60 min at 3250 rpm in cold centrifuge then carefully pour off the supernatant.
16. Add 300µl 70% ETOH.
17. Spin in centrifuge for 15 min at 3250 rpm then carefully pour off the ethanol and blot on paper towel.
18. Air dry the pellet for 1-2 hours, resuspend the pellet in 100 µl sterile milliQ water, heat at 65°C for 10 minutes, then lightly vortex.
19. Add 1µl RNase cocktail (Ambion) to each well. Incubate at 37°C for 45 min.
20. Proceed to clean up step using Zymo Clean and Concentrator 96 kit.

To make the extraction buffer combine 31.9 g Sorbitol, 50 ml of 1 M Tris, and 5 ml of 0.5 M EDTA (pH 7.5) and fill completely to 500 ml with sterilized and distilled H_2O. To make the Lysis buffer combine 100 ml of 1 M Tris, 50 ml of 0.5 M EDTA, 200 ml of 5 M NaCl, and 10 g of CTAB and fill completely to 500 ml with sterilized and distilled H_2O. To make 5% Sarcosyl solution, add 0.5 g of Sarcosyl to 10 ml of sterilized and distilled H_2O.

Appendix 2. DNA Clean and Concentrator 96-well Kit (Zymo Research, Irvine, CA)

1. Before starting, add 96 ml 100% ethanol to the 24 ml DNA Wash Buffer concentrate (192 ml 100% ethanol to the 48 ml DNA Wash Buffer concentrate) to obtain the final DNA Wash Buffer solution.
2. Add two volumes of DNA Binding Buffer to each volume of DNA sample and transfer sample mixtures to the wells of a Zymo-Spin™ I-96 Plate mounted on a Collection Plate.
3. Centrifuge at ≥ 3,000 x *g* (5,000 x *g* max.) for 5 minutes until samples have been completely filtered. Discard the flow-through.
4. Add 500 µl Wash Buffer to each well of the Zymo-Spin™ I-96 Plate. Centrifuge at ≥ 3,000 x *g* for 5 minutes. Spin an additional 5 minutes to dry. Air dry for 5 minutes.
5. Add 30 µl water directly to the column matrix in each well. Transfer the Zymo-Spin™ I-96 Plate onto an Elution Plate and centrifuge at ≥ 3,000 x *g* for 5 minutes to elute the DNA. Ultra-pure DNA in water is now ready for use.

Appendix 3. FastDNA-96 Plant and Seed DNA Kit (Zymo Research, Irvine, CA)

Before beginning protocol, prepare the Inhibitor Removal Plate as follows:

1. Place the Inhibitor Removal Plate on an Elution Plate.
2. Add 150µL of Prep Solution to each well of a Silicon-A™-HRC Plate by puncturing the foil covering with the pipette tip. (It takes a bit of force to puncture the foil, but try not to poke the bottom of the well too hard. Getting the force just right may take some practice.)
3. Incubate at room temperature on the bench for 5 minutes.
4. Centrifuge at 3500rpm for 10 minutes at 20C.
5. The plate is now ready to be used. It can be stored on the bench top until the step where it is needed.

For optimal performance, add beta-mercaptoethanol to the Plant/Seed DNA binding buffer to a final dilution of 0.5% (v/v) i.e., 750µl per 150ml.

1. Dump the spheres out of the wells of the ZR BashingBead™ Lysis Rack.
2. Load tissue into the wells of the ZR BashingBead™ Lysis Rack. Cut tissue into smaller pieces to load into well. Fill the well, but do not jam the tissue in such that it globs up into a ball.
3. Add one autoclaved metal rod.
4. Add 400µL of Lysis Solution to each well.
5. Stretch parafilm over the wells and cap with cap strip. Place a folded paper towel in between the tube caps and plate lid to fill the space.

6. Grind tissue once at 1400rpm for 30s. Check that the tissue ground well. If not, grind for another 30s at 1600rpm.
7. Centrifuge at 3500rpm for 10 minutes at 20C. Check that the tissue has pelleted well enough to remove pure supernatant. If not, centrifuge for another 10 minutes.
8. Transfer up to 250μL of the supernatant to the wells of a clean Deep-Well Plate 96-Well Block. It is recommended that this step be done with a single channel pipette, as that allows for more pure supernatant to be transferred.
9. Add 750μL of Plant/Seed DNA Binding Buffer to each well of the Deep-Well Plate. Cover the plate with a Cover foil. (Make sure each well is sealed.)
10. Briefly vortex at low speed. Incubate on the bench top for 2 minutes.
11. Centrifuge at 3500rpm for 10 minutes at 20C.
12. Put a Silicon-A™ Binding Plate on top of a Collection Plate.
13. Remove the foil from the 96-Well Block and transfer up to 450uL of supernatant to the Binding Plate. Try not to disturb the pellet at the bottom of the well.
14. Centrifuge at 3500rpm for 10 minutes at 20C. Discard the flow-through from the Collection Plate.
15. Repeat steps 13 & 14 until the supernatant has all been transferred.
16. Add 200μL of DNA Pre-Wash Buffer to the Silicon-A™ Binding Plate (which is still on the Collection Plate.)
17. Centrifuge at 3500rpm for 10 minutes at 20C. Discard the flow-through from the Collection Plate.
18. Add 500μL of Plant/Seed DNA wash buffer to each well of the Binding Plate.

19. Centrifuge at 3500rpm for 10 minutes at 20C.
20. Place the Silicon-A™ Binding Plate on a clean Elution Plate.
21. Add 60-70μL of DNA Elution Buffer directly to the matrix of the Binding Plate and let set for 1 minute.
22. Centrifuge at 3500rpm for 10 minutes at 20C.
23. Place the Silicon-A™-HRC Plate prepared at the beginning on a clean Elution Plate.
24. Transfer all of the eluent to the Inhibitor Plate.
25. Centrifuge at 3500rpm for 10 minutes at 20C.
26. The flow-through contains the DNA. Either cover the Elution Plate with a Foil Cover and put in the refrigerator or transfer it to a PCR plate cover with aluminum tape and place in the refrigerator. DNA quantification should be done as soon as possible. After quantification, store the DNA at -20C. For long-term storage, transfer the DNA to 0.5mL tubes. Improper storage will lead to evaporation and lose of samples.

Appendix 4. AccuBlue™ Broad Range dsDNA Quantitation Assay Using a Fluorescence Microplate Reader

Based on the number of samples to quantify, calculate out how much AccuBlue Quantitation Solution is needed and place in a 15 mL or 50 mL conical tube (10 mL per 96 well plate). Shake bottle and warm to room temperature in the dark fairly quickly. Thaw a tube of the AccuBlue Dye, so that the DMSO will thaw. Place foil on top of the dye as it is photosensitive. Vortex the enhancer vial before preparing the mix. The enhancer and dye are light sensitive.

For each 96 well plate, add 100 µL of 100X AccuBlue Enhancer and 100 µL of AccuBlue Dye to 10 mL of AccuBlue Quantitation Solution and mix well by vortexing or shaking. Prepare the working solution immediately before use.

Add 5 µL of each of the dsDNA standards into each of the separate wells (9 total wells) and mix well by pipetting up and down. It is recommended to include a standard curve on each 96-well plate that is used to minimize variability between plates. Accurate multi-channel pipettes and reservoirs can be used to facilitate this process. Black plates are recommended to minimize fluorescence bleed-through from other wells. Add 5 µL of the unknown DNA into each of the separate wells. Add 100 µL of the AccuBlue working solution into each of the separate wells and mix well by pipetting up and down. Incubate the microplate at room temperature for 5 minutes in the dark. Once the plate has incubated measure the fluorescence using the PE 2030 Victor X3 micro plate reader.

Generate a standard curve to determine the unknown DNA concentration. For the DNA standards, plot the amount of DNA vs. Fluorescence, and fit a trend line through

these points. This can be done in Excel by inserting a scatter plot and right-clicking on any of the data points to add a trend line. Be sure to select to display equation, so you can use that equation to calculate all of your DNA concentrations.

A2. Tables

Appendix 5. Sub-population statistics for *yellow endosperm1* (*p1*), *pericarp color1* (*p1*), *colored aleurone1* (*c1*), and *colored1* (*r1*) including peak locus position and 1-LOD confidence interval.

Group	n	Identified Gene	Peak position	Conf. Int. Pos.	Conf. Int. Dist.
Yellow Endosperm1 (y1)					
4way3sib-1	121	No	83,177,661	83,177,255 - 83,178,067	406
4way3sib-2	121	Yes	82,017,348	80,857,035 - 83,177,661	1,160,313
4way3sib-3	121	No	82,863,017	82,188,139 - 83,537,895	674,878
4way3sib-4	121	Yes	82,017,348	80,413,307 - 83,621,389	1,604,041
4way3sib-5	121	Yes	82,017,348	81,171,679 - 82,863,017	845,669
4way3sib-6	121	Yes	82,017,348	80,857,467 - 83,177,229	1,159,881
4way3sib-7	121	Yes	82,017,348	80,496,801 - 83,537,895	1,520,547
4way3sib-8	121	No	83,177,661	82,734,266 - 83,621,056	443,395
4way3sib-9	121	Yes	82,017,348	80,413,640 - 83,621,056	1,603,708
4way3sib-10	121	No	83,177,661	82,863,017 - 83,492,305	314,644
B73*Olc1* x Tx903	121	Yes	78,735,091	75,116,552 - 82,353,630	3,618,539
Tx772 x Tx906	121	Yes	82,764,656	82,017,402 - 83,511,910	747,254
4way3sib-1	242	No	82,863,017	82,548,373 - 83,177,661	314,644
4way3sib-2	242	Yes	83,177,661	82,017,348 - 84,337,974	1,160,313
4way3sib-3	242	Yes	82,017,348	81,171,679 - 82,863,017	845,669
4way3sib-4	242	Yes	82,017,402	80,857,143 - 83,177,661	1,160,259
4way3sib-5	242	Yes	82,017,348	80,496,801 - 83,537,895	1,520,547
4way3sib-6	242	Yes	82,017,348	81,171,679 - 82,863,017	845,669
4way3sib-7	242	Yes	82,017,348	81,171,045 - 82,863,651	846,303
4way3sib-8	242	Yes	82,017,348	80,857,035 - 83,177,661	1,160,313
4way3sib-9	242	Yes	82,017,348	80,857,035 - 83,177,661	1,160,313
4way3sib-10	242	No	83,177,661	83,177,229 - 83,178,093	432
all2ways	242	Yes	82,017,348	81,171,679 - 82,863,017	845,669
all subpops	1141	Yes	82,017,348	82,017,294 - 82,017,402	54

Appendix 5. Continued.

Group	n	Identified Gene	Peak position	Conf. Int. Pos.	Conf. Int. Dist.
***pericarp color1* (*p1*)**					
4way3sib-1	126	No	45,777,986	44,927,230- 46,628,742	850,756
4way3sib-2	126	Yes	47,775,210	47,389,045 - 48,161,375	386,165
4way3sib-3	126	No	47,905,812	47,817,462- 47,994,162	88,350
4way3sib-4	126	No	47,905,812	47,904,957 - 47,906,667	855
4way3sib-5	126	No	47,994,162	47,906,667 - 48,081,657	87,495
4way3sib-6	126	Yes	47,994,162	47,434,182 - 48,554,142	559,980
4way3sib-7	126	No	47,905,812	47,817,462 - 47,994,162	88,350
4way3sib-8	126	Yes	47,994,887	47,370,810 - 48,618,964	624,077
4way3sib-9	126	No	47,905,812	47,816,737 - 47,994,887	89,075
4way3sib-10	126	No	47,994,887	47,906,667 - 48,083,107	88,220
B73*Olc1* x Tx903	126	No	51,293,990	49,000,045 - 53,587,935	2,293,945
4way3sib-1	253	Yes	47,994,162	47,775,210 - 48,213,114	218,952
4way3sib-2	253	No	47,775,210	47,556,258 - 47,994,162	218,952
4way3sib-3	253	No	47,775,210	47,644,608 - 47,905,812	130,602
4way3sib-4	253	No	47,775,210	47,555,533 - 47,994,887	219,677
4way3sib-5	253	No	47,775,210	47,643,753 - 47,906,667	131,457
4way3sib-6	253	No	47,775,210	47,644,608 - 47,905,812	130,602
4way3sib-7	253	No	47,775,210	47,556,258 - 47,994,162	218,952
4way3sib-8	253	Yes	47,905,812	47,389,045 - 48,422,579	516,767
4way3sib-9	253	No	47,389,045	46,783,203 - 47,994,887	605,842
4way3sib-10	253	No	47,905,812	47,775,210 - 48,036,414	130,602
all2ways	253	No	51,293,959	48,652,544 - 53,935,374	2,641,415
all subpops	1174	No	47,994,162	47,905,812 - 48,082,512	88,350

Appendix 5. Continued.

Group	n	Identified Gene	Peak position	Conf. Int. Pos.	Conf. Int. Dist.
***colored aleurone1* (*c1*)**					
4way3sib-1	124	Yes	9,397,546	9,054,091 - 9,741,001	343,455
4way3sib-2	124	Yes	9,397,546	8,669,930 - 10,125,162	727,616
4way3sib-3	124	Yes	9,741,248	9,397,546 - 10,084,950	343,702
4way3sib-4	124	Yes	9,397,546	9,053,844 - 9,741,248	343,702
4way3sib-5	124	Yes	9,397,546	9,053,844 - 9,741,248	343,702
4way3sib-6	124	Yes	9,741,248	9,357,334 - 10,125,162	383,914
4way3sib-7	124	Yes	9,397,546	8,669,930 - 10,125,162	727,616
4way3sib-8	124	Yes	9,397,546	8,664,189 - 10,130,903	733,357
4way3sib-9	124	Yes	9,741,248	9,351,593 - 10,130,903	389,655
4way3sib-10	124	Yes	9,397,546	8,433,164 - 10,361,928	964,382
B73*Olc1* x Tx903	124	No	9,147,883	8,898,220 - 9,397,546	249,663
Tx772 x Tx906	124	No	11,155,191	10,529,228 - 11,781,154	625,963
4way3sib-1	247	Yes	9,397,546	8,664,189 - 10,130,903	733,357
4way3sib-2	247	Yes	9,397,546	9,053,844 - 9,741,248	343,702
4way3sib-3	247	Yes	9,397,546	8,664,189 - 10,130,903	733,357
4way3sib-4	247	Yes	9,397,546	9,053,844 - 9,741,248	343,702
4way3sib-5	247	Yes	9,397,546	8,664,189 - 10,130,903	733,357
4way3sib-6	247	Yes	9,397,546	9,053,844 - 9,741,248	343,702
4way3sib-7	247	Yes	9,397,546	9,053,844 - 9,741,248	343,702
4way3sib-8	247	No	10,125,162	10,119,421 - 10,130,903	5,741
4way3sib-9	247	Yes	9,397,546	9,053,844 - 9,741,248	343,702
4way3sib-10	247	Yes	9,397,546	9,053,844 - 9,741,248	343,702
all2ways	247	Yes	9,741,248	8,456,305 - 11,026,191	1,284,943
all subpops	1128	Yes	9,397,546	9,053,844 - 9,741,248	343,702

Appendix 5. Continued.

Group	n	Identified Gene	Peak position	Conf. Int. Pos.	Conf. Int. Dist.
***colored1* (*r1*)**					
4way3sib-1	124	No	138,04,6877	138,027,340 - 138,066,414	19,537
4way3sib-2	124	Yes	138,024,459	137,438,408 - 138,610,500	586,051
4way3sib-3	124	No	138,024,459	138,004,768 - 138,044,150	19,691
4way3sib-4	124	No	138,024,459	137,901,401 - 138,147,517	123,058
4way3sib-5	124	Yes	138,610,495	138,044,029 - 139,176,961	566,466
4way3sib-6	124	Yes	138,046,234	137,429,942 - 138,662,526	616,292
4way3sib-7	124	Yes	138,046,877	137,431,228 - 138,662,526	615,649
4way3sib-8	124	Yes	138,046,877	137,539,749 - 138,554,005	507,128
4way3sib-9	124	Yes	138,024,459	137,580,158 - 138,468,760	444,301
4way3sib-10	124	Yes	138,610,495	138,028,587 - 139,192,403	581,908
B73*Olc1* x Tx903	124	Yes	138,262,103	138,051,755 - 138,472,451	210,348
Tx772xTx906	124	Yes	138,454,640	138,093,745 - 138,815,535	360,895
4way3sib-1	247	No	137,903,207	137,762,264 - 138,044,150	140,943
4way3sib-2	247	No	138,046,877	138,027,274 - 138,066,480	19,603
4way3sib-3	247	No	138,024,459	138,020,475 - 138,028,443	3,984
4way3sib-4	247	Yes	138,046,877	137,588,327 - 138,505,427	458,550
4way3sib-5	247	Yes	138,610,500	138,046,877 - 139,174,123	563,623
4way3sib-6	247	No	138,027,274	138,010,519 - 138,044,029	16,755
4way3sib-7	247	No	138,028,443	138,027,340 - 138,029,546	1,103
4way3sib-8	247	Yes	137,588,327	136,566,154 - 138,610,500	1,022,173
4way3sib-9	247	No	138,046,234	137,637,828 - 138,454,640	408,406
4way3sib-10	247	Yes	138,046,234	138,028,443 - 138,064,025	17,791
all2ways	247	Yes	138,454,640	138,046,877 - 138,862,403	407,763
all subpops	1128	Yes	138,468,760	138,327,025 - 138,610,000	141,735