Characterization of the 22kDa alpha zein gene family and determination of the impact of *opaque2* on two transgenes containing zein promoters

by

Megan Harvey

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Genetics

Program of Study Committee:
M. Paul Scott, Co-major Professor
Michael Lee, Co-major Professor
Carolyn J. Lawrence

Iowa State University

Ames, Iowa

2007

TABLE OF CONTENTS

ABSTRACT

Zeins are the most abundant seed storage protein found in the maize kernel. There are multiple classes with some classes encoded by large gene families. The first part of this research attempted to identify new members of the 22kDa alpha zein gene family. Conserved sequences that were identified in many of the published zein-like sequences were used to create PCR primers that amplified zein-like sequences in three inbred lines, Ill12E, B101, and B73. Sequences found were subjected to CAP3 analysis, and the results of that analysis were used to estimate the number of zein-like sequences expected in the three inbred lines. The second part of this research examined how transgenes with two different zein promoters were impacted by the *opaque2* mutation. Transgenes expressing green fluorescent protein (GFP) using either a 19kDa zein promoter or 27kDa zein promoter were compared in *opaque2* and wild type kernels. The expression of the native zein genes are impacted differently by the *opaque2* mutation so this study investigated whether or not the transgenes behaved similarly to the native genes that the promoters were derived from. The *opaque2* mutation had a variable impact on expression of both transgenes, and variable transgene transmission was observed as well.

CHAPTER 1: GENERAL INTRODUCTION

1. Literature Review

1.1. The Prolamins of Maize

The seed is the organ through which plants propagate. In maize, the kernel, which is a seed is encased in a husk, is the means of propagation. The kernel provides the nutritional support for growth of a new plant. Nutrition is provided is through seed storage proteins which are degraded during germination. A storage protein is defined as a protein whose main function is to store nitrogen, carbon, and sulfur (Shewry and Casey 1999). They are found in the cell in structures called protein bodies, and the major storage proteins in maize, the zeins, are in protein bodies formed in the lumen of the endoplasmic reticulum (Lending et al. 1988). The endosperm accounts for 90% of the kernel mass at maturity, making it the source of most of the nutrition of the seed (Coleman and Larkins 1999).

Kernel development begins with embryogenesis and the fertilization of the ovule. The ovule and structures associated with it form the four seed tissues: the testa, perisperm, endosperm, and embryo (Shewry and Casey 1999). The testa, which does not become a storage tissue, is the seed coat and is sometimes found to be fused with the pericarp. The perisperm comes from the nucellus, a maternal tissue. The endosperm and embryo tissues are the result of the double fertilization found in flowering plants (Figure 1). The endosperm is the result of the fusion between a male generative nucleus and two polar nuclei, which makes it triploid. This is the main seed storage tissue for many species, including maize. The embryo results from the fusion of a male nucleus and an egg cell, creating a diploid tissue. The perisperm, endosperm, and embryo can all become storage tissues, either alone or in combination.

Storage proteins present in kernels are important in the growth and development of plants. In species such as *Zea mays*, they provide much of the nutritional value of the kernel. Seed proteins have been divided into four main classes based on their solubility (Osborne 1924). These classes are the water-soluble albumins, salt-soluble globulins, aqueous alcohol-soluble prolamins, and the glutelins, which are not soluble in water, saline solutions, or aqueous alcohol. These groups have more recently been further divided and maize kernels contain seed storage proteins from the following three groups: the 2S albumin family, the 7S and 11S families of storage globulins, and the alcohol-soluble prolamin family (Shewry and Casey 1999).

The prolamin class of seed proteins is called zein in maize. They are the major seed storage proteins in the endosperm and are deposited during kernel development (reviewed in Heidecker and Messing 1986). Prolamins are so named because of their large amounts of proline and glutamine; however, zeins do not contain a balanced amino acid profile. Corn does not provide equal amounts of amino acids and some are found in excess while others are deficient according to nutritional needs for humans and animals. With the exception of the 18kDa delta zein, all of the zein classes completely lack the essential amino acids lysine and tryptophan and most are low in methionine as well (reviewed in Coleman and Larkins 1999). Essential amino acids are defined as those amino acids that an animal does not naturally produce and so must be acquired through food (Berg et al. 2002). Zeins are still a source of essential amino acids other than lysine, tryptophan, and methionine in the kernel and are therefore important for both human and animal nutrition.

The different zein fractions have been separated using SDS-PAGE revealing seven different polypeptide sizes and four classes of zeins, alpha-, beta-, gamma-, and delta-zeins

(Esen 1987; Wallace et al. 1990). The alpha zeins are the most abundant of the maize prolamins (Coleman and Larkins 1999) making up 70% of the total zein content in the kernel and can be broken into the following four families: the 19kDa A, the 19kDa B, 19kDa D and the 22kDa gene families. The alpha zein polypeptides are from a complex multigene family containing 70-100 members (Hagen and Rubenstein 1981; Wilson and Larkins 1984) with mRNA showing a complexity consistent with zein protein complexity (Park et al. 1980). Related grasses have also been shown to contain the 19kDa and 22kDa prolamin gene families, suggesting an evolutionary relationship in the existence of these multigene families (Wilson and Larkins 1984). The complexity of the polypeptide products causes variation between inbreds and hybrids and the differences can be used in genotyping (Hartings et al. 1989; Smith and Smith 1988; Wilson 1985). The alpha zein genes have been mapped to the long and short arms of chromosome 4, the short arm of chromosome 7, the long arm of chromosome 10, and near the centromere of chromosome 1 (Hartings et al. 1989; Shen et al. 1994; Song et al. 2001).

Although the 22kDa and 19kDa zein gene clusters occupy separate clades in phylogenetic analyses, they have been shown to share a common ancestor (Song and Messing 2002). The 22kDa zein gene family occupies a single clade, and that family is termed z1C. The 19kDa zein gene family has evolved into three subfamilies occupying two clades. The three subfamilies, z1A, z1B, and z1D, are found in five separate genomic regions (Song and Messing 2002). This is the result of 19kDa and 22kDa gene families amplifying in different manners, with the 19kDa gene family showing more far-distance movement within the maize genome. While the 22kDa zein gene cluster, with the exception

of *azs22-16* which is the normal copy of the *floury2* allele, is found in tandem array, the three subfamilies of the 19kDa zein gene family are located in noncontiguous genomic locations.

The z1C gene was originally duplicated 11.5 million years ago (Song et al. 2001). During the evolution of the 22kDa alpha zein gene family, several genes diverged at different times but were later duplicated as a group (Song et al. 2001). Some genes arose as recently as two million years ago and 0.5 million years ago making the expansion of this gene family a relatively recent event. Research has also suggested that a number of zein gene copies may have been inactivated recently and could be easily restored through recombination depending on the inbred line (Song et al. 2001).

Although the alpha zein gene families are large and complex, only a small number of identified genes are transcribed in detectable amounts in the endosperm (Song et al. 2001; Song and Messing 2003; Woo et al. 2001). Several of the genes identified have in-frame stop codons (Song et al. 2001; Song and Messing 2003) and it is postulated that many of the alpha zein genes are actually pseudogenes. In the inbred BSSS53, 23 genes have been identified in the 22kDa alpha zein gene cluster (Song et al. 2001). Twenty-two of the genes are in a tandem array on the short arm of chromosome 4S while the twenty-third gene is the normal allele of the *floury2* locus and is located approximately 20 cM from the others (Coleman et al. 1997; Song and Messing 2002). Of the 23 genes, eight are intact, 13 have a premature stop codon, one has a 3' truncation, and one contains a 714 base pair insertion (Song and Messing 2003). Of the eight intact genes, only seven have detectable levels of mRNA (Song et al. 2001).

A similar situation is seen in the 22kDa alpha zein gene copies from the inbred line B73. Of the 15 genes identified, six have intact coding regions, six have premature stop

codons in the coding region and three have large rearrangements caused by either a transposition insertion or a large truncation (Song and Messing 2003). A single gene resembling the *floury2* allele was also identified in the z1C-2 locus. When comparing sequences in the z1C family, genes in orthologous positions have been shown to be more conserved than genes in paralogous positions (Song and Messing 2003). The sequence differences seen between inbreds are interesting because within the 22kDa zein gene family sequences are conserved by at least 97% identity.

The 19kDa alpha zein gene cluster also contains a number of genes with no detectable levels of mRNA. Of the 25 19kDa zein gene sequences identified in the inbred B73, 12 have intact coding regions while 13 were not intact due to truncation, internal deletion, or stop codons (Song and Messing 2002). All of the genes identified as intact were found to accumulate mRNA although the amount varied between the genes. Additionally, three of the genes with in-frame stop codons were discovered to have detectable levels of mRNA so they may produce truncated versions of the gene product.

Knowledge of these gene families is important in understanding the seed storage proteins of maize as the alpha zeins are the most abundant storage proteins in the kernel. The differences seen in these families between different inbred lines will prove challenging when trying to sequence the maize genome. Many inbreds and hybrids will need to be studied to fully understand the impact and evolution of these gene families in maize.

The cysteine-rich gamma zeins can be mapped to a single locus on the long arm of chromosome 7. There is a single copy of the gene for the 16kDa gamma zein and either one or two gene copies present for the 27kDa gamma zein protein (Das and Messing 1987).

The beta zein contains high levels of methionine (11%) (Pedersen et al. 1986) and cysteine. It contains less glutamine, leucine, and proline than the alpha zeins, but, like them, lacks lysine and tryptophan. The 15kDa beta zein protein is encoded by a single copy gene on the short arm of chromosome 6 (reviewed in Coleman and Larkins 1999).

The delta zeins are not structurally related to the other zeins but they have an amino acid composition similar to the beta zeins (Kirihara et al. 1988a; Kirihara et al. 1988b). The 10kDa delta zein has the highest methionine content of any zein, with 22.5% of the protein consisting of methionine residues (Kodrzycki et al. 1989) . The gene encoding it is present as a single copy gene located on the short arm of chromosome 7 (Swarup et al. 1995). The 18kDa delta zein is unique in that it is the only zein protein known to contain lysine and tryptophan and is located as a single copy gene on the long arm of chromosome 6 (Chui and Falco 1995; Swarup et al. 1995).

1.2. *Dzr1* and the high methionine phenotype of BSSS53

Iowa Stiff Stalk Synthetic (BSSS) was developed in the early 1930s by intermating 16 inbred lines with above average stalk quality (reviewed inTroyer 1999). The progeny were then subject to selection experiments from which came several valuable inbred lines (Figure 2). One such line, B73, was developed from Cycle 5 and came from one of the S2 lines developed for further selection and inbreeding. This inbred was actually initially discarded but further developed when it was discovered to have good yield in hybrids. The inbred line was then released in 1972. Inbred line B101 is another useful line developed from BSSS. It was developed through studies aimed at determining genetic variation and rates of inbreeding depression within BSSS. B101 was originally distributed as BSSS53 and

arose through single-seed descent from BSSS (Hallauer and Wright 1995). B101 is characterized by having excellent grain yield, high tassel branch numbers, and an elevated level of methionine. It was released in 1994.

The nutritional value of maize is strongly influenced by the zeins which are lacking in lysine, tryptophan, and to a lesser extent, methionine. Animal feed is often a corn/soy mixture to make up for the amino acid deficiencies in corn, but this mixture still is still deficient in methionine relative to the needs of some animals (Paulis et al. 1978). In a study searching for lysine plus threonine resistant lines, inbred line BSSS53 was found to be resistant and had a 30% increase in methionine content when compared to other resistant inbreds (Phillips et al. 1981). This increase was determined to be the result of overexpression of the 10kDa delta-zein protein (Phillips and McClure 1985). In other mutations effecting kernel nutritive value, such as *opaque2* and *floury2*, zein content is decreased (Mertz et al. 1964; Nelson et al. 1965), however in BSSS53 the 10kDa delta zein protein is increased while the concentration of the other zeins and total protein content remains similar to other inbreds (Phillips and McClure 1985).

Methionine is the most abundant amino acid in the 10kDa zein gene and makes up 22.5% of the amino acids found in the protein (Kirihara et al. 1988a). Although the increased methionine in BSSS53 has been shown to be the result of overexpression of the 10kDa zein protein (Phillips and McClure 1985), this is not due to an increase in gene copy number (Kirihara et al. 1988b). The structural gene, *delta zein structural10* (*dzs10*), is found as a single copy (Kirihara et al. 1988b) on chromosome 9 but is not the gene responsible for the overexpression of the zein protein, DZS10 (Benner et al. 1989). This overexpression is regulated by the gene, *delta zein regulator1* (*dzr1*), located on chromosome 4. *dzr1* operates

posttranscriptionally with a *trans* acting mechanism (Cruz-Alvarez et al. 1991; Schickler et al. 1993). The result is an increase in the stability of mRNA that results in the increase of the steady state level of 10kDa delta zein mRNA in BSSS53 (Cruz-Alvarez et al. 1991).

While the genetic location of *dzr1* is known, its sequence has not yet been determined. The regulatory gene has been shown to be tightly linked to the 22kDa alpha zein gene cluster, in particular to the gene Zp22/6. Zp22/6 cosegregated tightly with the high 10kDa zein phenotype and no recombination was observed between the two, meaning that *dzr1* is located less than 0.4 cM from Zp22/6 (Chaudhuri and Messing 1995). The regulatory locus does not appear to lower synthesis of any other storage protein in the process (Benner et al. 1989), which differs from other amino acid-altering mutations such as *opaque2* and *floury1*.

The 10kDa delta zein fraction is not the only zein that contains a larger amount of methionine. The 15kDa beta zein and the 18kDa delta zein also contain a larger percentage of methionine than the alpha and gamma zeins. The 18kDa zein, DZS18, is not seen in all maize lines and in fact cannot be detected in BSSS53 (Swarup et al. 1995). The mRNA levels of *dzs18* in BSSS53 were the lowest of all lines tested and the variability of DZS18 suggests that it is not regulated in the same way as DZS10.

dzr1 has also been shown to have a maternal effect (Messing and Fisher 1991) which is essential knowledge for the development of high methionine BSSS53 hybrids. Reciprocal crosses were done and when *dzr1* was passed through the female, the increase in methionine content was found to be over five times more than the increase when *dzr1* was passed through the male. Based on gene copy number, the methionine level was expected to be only twofold higher when passed maternally.

Although *dzr1* has a maternal effect, there is also allele specific imprinting of *dzr1*, specifically in the inbred, Mo17. Unlike BSSS53, there is an unusually low amount of the 10kDa delta zein in Mo17. This was determined to be caused by the fact that the *dzr1+Mo17* allele is parentally imprinted (Chaudhuri and Messing 1994). The allele is dominant when passed through the female and recessive if passed paternally. This knowledge is essential to programs attempting to improve methionine content through selective breeding.

A recent backcrossing experiment with the objective of increasing the methionine content in maize inbreds and hybrids involved BSSS53 as the high methionine donor (Olsen et al. 2003). There was an increase in the total methionine levels in both inbreds and hybrids although a significant portion of the increase came as a result of increase in total protein, not just increase in methionine as a percentage of total protein. None of the backcrosses yielded plants with methionine content as high as BSSS53 but the increases that did occur showed that it was possible to increase methionine in inbreds through backcrossing. The high methionine inbreds then produced hybrids with higher amounts of methionine as well. Previous experiments have demonstrated that the higher amounts of methionine in BSSS53 hybrids are nutritionally available and able to improve growth in chicks (Messing and Fisher 1991). The results of the backcross experiments show great promise for nutritionally improving animal feed as well.

1.3. Opaque2

Amino acid composition has long been a problem in maize as the major storage protein fraction, the zeins, with the exception of the 18kDa delta zein, is completely lacking in lysine and tryptophan (Chui and Falco 1995; Swarup et al. 1995). Several mutations that

decrease the amount of zeins present in the maize kernel have been documented, and one of great interest is the *opaque2* mutation. This mutation causes an increase in the lysine concentration of the kernel (Mertz et al. 1964) through the decrease of one of the major storage protein fractions, the 22kDa alpha-zeins (Jones et al. 1977). The *opaque2* kernels have a reduced ratio of zein to glutelin and nearly 70% more lysine than normal endosperm.

The *opaque2* mutation partially inhibits zein synthesis thereby altering the distribution of the protein fractions (Landry and Moureaux 1982). A decrease in zein mRNA was noticed in *brittle2-opaque2* double mutants (Tsai et al. 1978) and further research showed the zein mRNA greatly reduced in *opaque2* mutants and no 22kDa mRNA was detected (Pedersen et al. 1980). While the 22kDa mRNA could not be detected, the complexity of the zein mRNAs remained similar to wild type maize, suggesting that the mRNA levels were likely reduced but the mRNAs were not completely absent. Two other mutations, *opaque7* and *floury2* have also been shown to alter protein content by lowering the level of zeins present (Nelson et al. 1965; Soave et al. 1976). They differ from *opaque2* in that they reduce the content of all zein classes equally, rather than reducing a specific fraction (Soave et al. 1976). While the *opaque2* and *opaque7* mutations have shown an additive effect on zein expression, they appear to be epistatic to *floury2* (Schmidt et al. 1987). The research done on these three mutations suggests that the *opaque2* product is not regulated by either *opaque7* or *floury2*.

The *opaque2* mutation is unique in that its main effect is a decrease in the alpha zeins, specifically the 22kDa alpha zein gene faction, as the result of reduced 22kDa zein mRNA levels (Langridge et al. 1982). Kodrzycki et al. (1989) found that the 19kDa alpha zein protein was delayed in production and the 22kDa zein protein was only found in trace

amounts in *opaque2* endosperm. The transcription of the alpha zeins was reduced by over 90% for the 22kDa gene and 40 to 75% for the 19kDa zein gene, depending on the subfamily. There were similar transcription levels found in both alpha zeins but in the endosperm much less 22kDa protein was present compared to the 19kDa protein. It is likely that there is a posttranscriptional mechanism controlling the level of protein to account for the difference in protein found in the endosperm. While the alpha zeins are greatly affected, the transcription of the other zein fractions, beta, delta, and gamma are altered very little (Kodrzycki et al. 1989).

The *Opaque2* gene has been isolated and characterized and found to encode a bZip transcription factor that can bind to the promoter region of a 22kDa zein, thereby enabling it to activate zein gene transcription (Hartings et al. 1989; Schmidt et al. 1990). This research supports a model of O2 transcriptional regulation that occurs by direct interaction with zein gene promoters. However, the specific, unique sequence that binds to a 22kDa zein promoter is not found in the 19kDa zein gene promoter or any other zein promoter (Schmidt et al. 1992). Studies involving *opaque2* mutants have also identified two conserved amino acids in the basic region of O2 that are crucial to the DNA binding activity of the bZIP protein (Aukerman et al. 1991). One of the mutants discovered had a greater decrease in 19kDa alpha zein than 22kDa alpha zein, presumably because of alterations in those two amino acids. The *opaque2* mutation usually causes a greater decrease of 22kDa zeins than 19kDa zeins so this result was unexpected. The mechanism by which the 19kDa zeins are reduced in *opaque2* kernels is not fully understood as 19kDa zein promoter sequences have not been found to bind to O2. This suggests that the *opaque2* mutation affects the 19kDa zein

promoters through an indirect interaction. These studies indicate that O2 transcriptional regulation may occur through both direct and indirect interaction with zein gene promoters.

While the initial discovery of the high lysine mutant was of great interest, the *opaque2* kernels were found to have several drawbacks including increased damage during harvest, greater susceptibility to pathogens, and reduced yield. These problems were partially due to the opaque nature of the kernels in that they had a soft, floury endosperm (reviewed in Vasal 2001). Much research has gone into developing varieties with *opaque2* modifier genes which return *opaque2* kernels the normal vitreous phenotype. These varieties are referred to as Quality Protein Maize (QPM).

The modifiers in QPM have also been discovered to cause two to four times higher levels of gamma zein than normal or unmodified *opaque2* kernels with the increase seen in all QPM lines tested (Wallace et al. 1990). The increase in gamma zein is not seen in unmodified *opaque2* kernels, where the gamma zein transcript is similar to normal phenotype (Kodrzycki et al. 1989). However, the drastic decrease in alpha zeins in *opaque2* maize caused the gamma zeins to show the highest transcription rate of all the zeins. Geetha et al. (1991), found that gamma zein transcription levels were not altered, suggesting that the increase may be affected instead by mRNA stability. The gamma zein increase was later determined to be the result of a post-transcriptional modification which caused an increase in the gamma zein mRNA level (Or et al. 1993). This led to an increase in the synthesis of the protein. The modifiers are not specific to the *opaque2* background and have also been seen in *floury2* and normal backgrounds (Lopes and Larkins 1991). Further research on the role of gamma zeins and other aspects of QPM maize holds great promise for improving the

nutritional value of food for the many countries where people receive their primary nourishment from maize.

1.4. Green Fluorescent Protein

Green fluorescent protein (GFP) is produced by *Aequorea victoria*, a bioluminescent jellyfish, and it produces a strong green fluorescence when excited by blue light (Chalfie et al. 1994). Several aspects of GFP make it ideal as a reporter gene in plants. It is active without the use of exogenous substrates (Chalfie et al. 1994) and the GFP protein can be quantified using GFP fluorescence (Niwa et al. 1999; Remans et al. 1999). GFP was shown by Chalfie et al. (1994) to express in prokaryotic cells using a bacteriophage promoter and in eukaryotic cells via a native promoter without interfering with cell growth and function. These discoveries led researchers to suggest that GFP could be used as a non-invasive marker, revolutionizing its use. GFP has now been expressed in many organisms, including maize (Cubitt et al. 1995), where it has been successfully produced in both maize callus and plants (van der Geest and Petolino 1998).

Research has been done recently using maize embryo and endosperm specific GFP constructs as tissue markers (Shepherd et al.). The zein promoters are of interest for endosperm specific markers as they are primarily found in the endosperm tissue in maize. The 27kDa zein promoter has been previously shown to express a marker transgene (Russell and Fromm 1997), a bacterial protein in plants (Chikwamba et al. 2003), and a milk protein in maize endosperm (Bicar et al. 2007). In maize, both the 19kDa alpha zein promoter and the 27kDa gamma zein promoter have been successfully used to create constructs that stably express GFP through at least three generations (Shepherd et al.). These two transgenes were

shown to function exclusively in the endosperm. Transient expression experiments showed that both constructs were significantly different from the negative control in B73 and Va26 endosperm tissue but not significantly different than the negative control in OH43 genotype endosperm tissue. This suggests that genetic background may have an effect on the expression of these GFP constructs. The study shows the usefulness of GFP as a tissue marker and provides insights about how to best determine transgene expression in future experiments.

2. Author's Role

The chapters were written by the primary author with the guidance and assistance of Dr. Scott. For the first study (Chapter 2) Dr. Scott identified conserved sequences used to design initial primers and using these primers contributed 218 sequences to the analysis. The primary author designed different primers using the same conserved sequences. The PCR and sequencing using these primers was conducted by the primary author. All reported sequence analysis was conducted by the primary author.

For the second study (Chapter 3) the cross of *opaque2* plants and plants containing the GFP transgene, which yielded F_1 seed was done by Dr. Scott. The F_1 seed were planted, selfed, and harvested by the primary author. Initial GFP screening and shelling was also done by the primary author. Separation of opaque and translucent kernels was conducted by both the primary author and Dr. Moran-Lauter. The primary author designed the fluorometer experiment. Sanding kernels was carried out by the primary author, Dr. Moran-Lauter, and Chase Lucas. Plate setup and fluorometer readings were conducted by the primary author

and Dr. Moran-Lauter using the parameters selected by the primary author. Data analysis was done by the primary author.

3. References

Aukerman MJ, Schmidt RJ, Burr B, Burr FA. 1991. An arginine to lysine substitution in the bZIP domain of an opaque-2 mutant in maize abolishes specific DNA binding. Genes and Development 5:310-320.

Benner MS, Phillips RL, Kirihara JA, Messing JW. 1989. Genetic analysis of methionine-rich storage protein accumulation in maize. Theor Appl Genet 78:761-767.

Berg JM, Tymoczko JL, Stryer L. 2002. Biochemistry: W. H. Freeman & Co. 974 p.

Bicar EH, Woodman-Clikeman W, Sangtong V, Peterson JM, Yang SS, Lee M, Scott MP. 2007. Transgenic maize endosperm containing a milk protein has improved amino acid balance. Transgenic Research Online.

Chalfie M, Tu Y, Euskirchen G, Ward William W, Prasher Douglas C. 1994. Green fluorescent protein as a marker for gene expression. Science 263(5148 802-805):802-805.

Chaudhuri S, Messing J. 1994. Allele-specific parental imprinting of dzr1, a posttranscriptional regulator of zein accumulation. Proc Natl Acad Sci U S A 91(11):4867-4871.

Chaudhuri S, Messing J. 1995. RFLP mapping of the maize *dzr1* locus, which regulates methionine-rich 10 kDa zein accumulation. Mol Gen Genet 246(6):707-715.

Chikwamba RK, Scott MP, Mejia LB, Mason HS, Wang K. 2003. Localization of a bacterial protein in starch granules of transgenic maize kernels. Proceedings of the National Academy of Sciences of the United States of America 100(19):11127-11132.

Chui CF, Falco SC. 1995. A new methionine-rich seed storage protein from maize. Plant physiol 107(1):291.

Coleman CE, Clore AM, Ranch JP, Higgins R, Lopes MA, Larkins BA. 1997. Expression of a mutant alpha-zein creates the *floury2* phenotype in transgenic maize. Proc Natl Acad Sci U S A 94(13):7094-7097.

Coleman CE, Larkins BA. 1999. The prolamins of maize. In: Shewry PR, Casey R, editors. Seed Proteins. Dordrecht, Netherlands: Kluwer Academic Publishers. p 109–139.

Cruz-Alvarez M, Kirihara JA, Messing J. 1991. Post-transcriptional regulation of methionine content in maize kernels. Mol Gen Genet 225(2):331-339.

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY. 1995. Understanding, improving and using green fluorescent proteins. Trends in Biochemical Sciences 20:448-455.

Das OP, Messing JW. 1987. Allelic variation and differential expression at the 27-kilodalton zein locus in maize. Molecular and Cellular Biology 7(12):4490-4497.

Esen A. 1987. A Proposed Nomenclature for the Alcohol-soluble Proteins (Zeins) of Maize (Zea mays L.). Journal of Cereal Science 5:117-128.

Geetha KB, Lending CR, Lopes MA, Wallace JC, Larkins BA. 1991. opaque2 modifiers increase gamma-zein synthesis and alter its spatial distribution in maize endosperm. Plant Cell 3(11):1207-1219.

Hagen, Rubenstein. 1981. Complex organization of zein genes in maize. Gene 13(3):239-49.

Hallauer AR, Wright AD. 1995. Registration of B101 maize germplasm. Crop Sci 35(4):1238-1239.

Hartings H, Maddaloni M, Lazzaroni N, Di Fonzo N, Motto M, Salamini F, Thompson R. 1989. The O2 gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. EMBO J 8(10):2795-2801.

Heidecker G, Messing J. 1986. Structural analysis of plant genes. Annu Rev Plant Physiol 37:439-466.

Jones RA, Larkins BA, Tsai CY. 1977. Storage Protein Synthesis in Maize. Plant Physiology 59:525-529.

Kirihara JA, Hunsperger JP, Hahoney WC, Messing JW. 1988a. Differential expression of a gene for a methionine-rich storage protein in maize. Mol Gen Genet 211(3):477-484.

Kirihara JA, Petri JB, Messing J. 1988b. Isolation and sequence of a gene encoding a methionine-rich 10-kDa zein protein from maize. Gene 71(2):359-370.

Kodrzycki R, Boston RS, Larkins BA. 1989. The opaque-2 mutation of maize differentially reduces zein gene transcription. Plant Cell 1(1):105-114.

Landry J, Moureaux T. 1982. Distribution and amino acid composition of protein fractions in opaque-2 maize grains Zea mays. Phytochemistry 21(8):1865-1869.

Langridge P, Pintor-Toro JA, Feix G. 1982. Transcriptional effects of the opaque-2 mutation of Zea mays L. Planta 156:166-170.

Lending CR, Kriz AL, Larkins BA, Bracker CE. 1988. Structure of maize protein bodies and immunocytochemical localization of zeins. Protoplasma 143(1):51-62.

Lopes MA, Larkins BA. 1991. Gamma-zein content is related to endosperm modification in quality protein maize. Crop Sci 31(6):1655-1662.

Mertz ET, Bates LS, Nelson OE. 1964. Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145:279-280.

Messing J, Fisher H. 1991. Maternal effect on high methionine levels in hybrid corn. J Biotechnol 21(3):229-238.

Nelson OE, Mertz ET, Bates LS. 1965. Second Mutant Gene Affecting the Amino Acid Pattern of Maize Endosperm Proteins. Science 150:1469-1470.

Niwa Y, Hirano T, Yoshimoto K, Shimizu M, Kobayashi H. 1999. Non-invasive quantitatived detection and applications of non-toxic, S56T-type green fluorescent protein in living plants. The Plant Journal 18(4):455-463.

Olsen MS, Krone TL, Phillips RL. 2003. BSSS53 as a donor source for increased whole-kernel methionine in maize: selection and evaluation of high-methionine inbreds and hybrids. Crop Sci 43:1634-1642.

Or E, Boyer SK, Larkins BA. 1993. Opaque2 modifiers act post-transcriptionally and in a polar manner on gamma-zein gene expression in maize endosperm. Plant cell 5(11):1599-1609.

Park WD, Lewis ED, Rubenstein I. 1980. Heterogeneity of zein mRNA and protein in maize. Plant Physiology 65:98-106.

Paulis JW, Wall JS, Sanderson J. 1978. Origin of high methionine content in sugary-1 corn endosperm. Cereal Chemistry 55:705-712.

Pedersen K, Argos P, Naravana S, Larkins B. 1986. Sequence analysis and characterization of a maize gene encoding a high- sulfur zein protein of Mr 15,000. J. Biol. Chem. 261(14):6279-6284.

Pedersen K, Bloom KS, Anderson JN, Glover DV, Larkins BA. 1980. Analysis of the Complexity and Frequency of Zein Genes in the Maize Genome. Biochemistry 19:1644-1650.

Phillips RL, McClure BA. 1985. Elevated protein-bound methionine in seeds of a maize line resistant to lysine plus threonine. Cereal Chem 62(3):213-218.

Phillips RL, Morris PR, Wold F, Gengenbach BG. 1981. Seedling screening for lysine-plus-threonine feedback resistant maize. Crop Sci 21(4):601-607.

Remans T, Schenk PM, Manners JM, Grof CPL, Elliott AR. 1999. A Protocol for the Fluorometric Quantification of mGFP-ER and sGFP(S65T) in Transgenic Plants. Plant Molecular Biology Reporter 17:385-395.

Russell DA, Fromm ME. 1997. Tissue-specific expression in transgenic maize of four endosperm promoters from maize and rice. Transgenic res 6(2):157-168.

Schickler H, Benner MS, Messing J. 1993. Repression of the high-methionine zein gene in the maize inbred line Mo17. Plant J 3(2):221-229.

Schmidt RJ, Burr FA, Aukerman MJ, Burr B. 1990. Maize regulatory gene *opaque2* encodes a protein with a "leucine-zipper" motif that binds to zein DNA. Proc Natl Acad Sci 87:46-50.

Schmidt RJ, Burr FA, Burr B. 1987. Transposon tagging and molecular analysis of the maize regulatory locus opaque-2. Science 238(4829):960-963.

Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G. 1992. *Opaque2* Is a Transcriptional Activator That Recognizes a Specific Target Site in 22-kD Zein Genes. Plant Cell 4(6):689-700.

Shen B, Carneiro N, Torres-Jerez I, Stevenson B, McCreery T, Helentjaris T, Baysdorfer C, Almira E, Ferl RJ, Habben JE and others. 1994. Partial sequencing and mapping of clones from two maize cDNA libraries. Plant Molecular Biology 26:1085-1101.

Shepherd CT, Vignaux N, Peterson JM, Johnson LA, Scott MP. Green Fluorescent Protein as a tissue marker in transgenic maize seed. Cereal Chem: Not yet published.

Shewry PR, Casey R. 1999. Seed Proteins. Shewry PR, Casey R, editors. Dordrecht, The Netherlands: Kluwer Academic Publishers. 1-10 p.

Smith JSC, Smith OS. 1988. Comparisons of zein profiles from inbred, F1, and F2 generations of maize as revealed by reversed-phase high-performance liquid chromatography. Theoretical and Applied Genetics 76:244-252.

Soave C, Righetti PG, Lorenzoni C, Gentinetta E, Salamini F. 1976. Expressivity of the *opaque2* gene at the level of zein molecular components. Maydica 21:61-75.

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

Song RT, Llaca V, Linton E, Messing J. 2001. Sequence, regulation, and evolution of the maize 22-kD alpha zein in gene family. Genome Research 11(11):1817-1825.

Song RT, Messing J. 2003. Gene expression of a gene family in maize based on noncollinear haplotypes. Proceedings of the National Academy of Sciences of the United States of America 100(15):9055-9060.

Swarup S, Timmermans MCP, Chaudhuri S, Messing J. 1995. Determinants of the high-methionine trait in wild and exotic germplasm may have escaped selection during early cultivation of maize. Plant J 8(3):359-368.

Troyer AF. 1999. Background of U.S. Hybrid Corn. Crop Science 39:601-626.

Tsai CY, Larkins BA, Glover DV. 1978. Interaction of the opaque-2 gene with starch-forming mutant genes on the synthesis of zein in maize endosperm. Biochem Genet 16(9/10):883-896.

van der Geest AHM, Petolino JF. 1998. Expression of a modified green fluorescent protein gene in transgenic maize plants and progeny. Plant Cell Reports 17:760-764.

Vasal SK. 2001. High Quality Protein Corn. In: Hallauer AR, editor. Specialty Corns. Second Edition ed: CRC Press LLC. p 85-129.

Wallace JC, Lopes MA, Paiva E, Larkins BA. 1990. New methods for extraction and quantitation of zeins reveal a high content of γ-zein in modified *opaque2* maize. Plant Physiol 92(1):191-196.

Wilson CM. 1985. A nomenclature for zein polypeptides based on isoelectric focusing and sodium dodecylsulfate polyacrylamide gel electrophoresis. Cereal chemistry 62(5):361-365.

Wilson DR, Larkins BA. 1984. Zein gene organization in maize and related grasses. J Mol Evol 20(330-340).

Woo Y-M, Hu DW-N, Larkins BA, Jung R. 2001. Genomics Analysis of Genes Expressed in Maize Endosperm Identifies Novel Seed Proteins and Clarifies Patterns of Zein Gene Expression. Plant Cell 13(10):2297-2317.

4. Figures

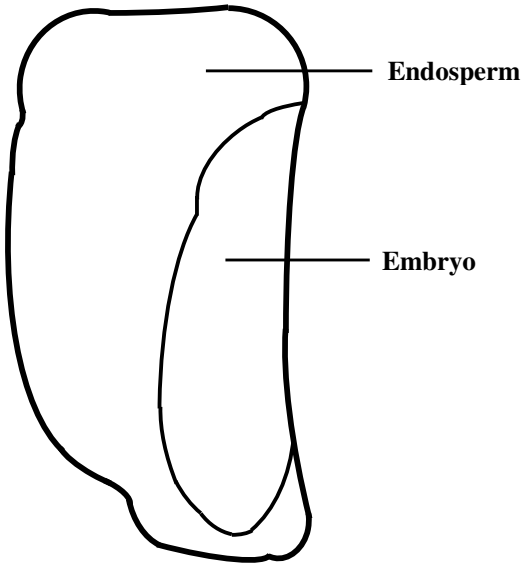

Figure 1. Cross section of a maize kernel. Endosperm and embryo tissues are indicated.

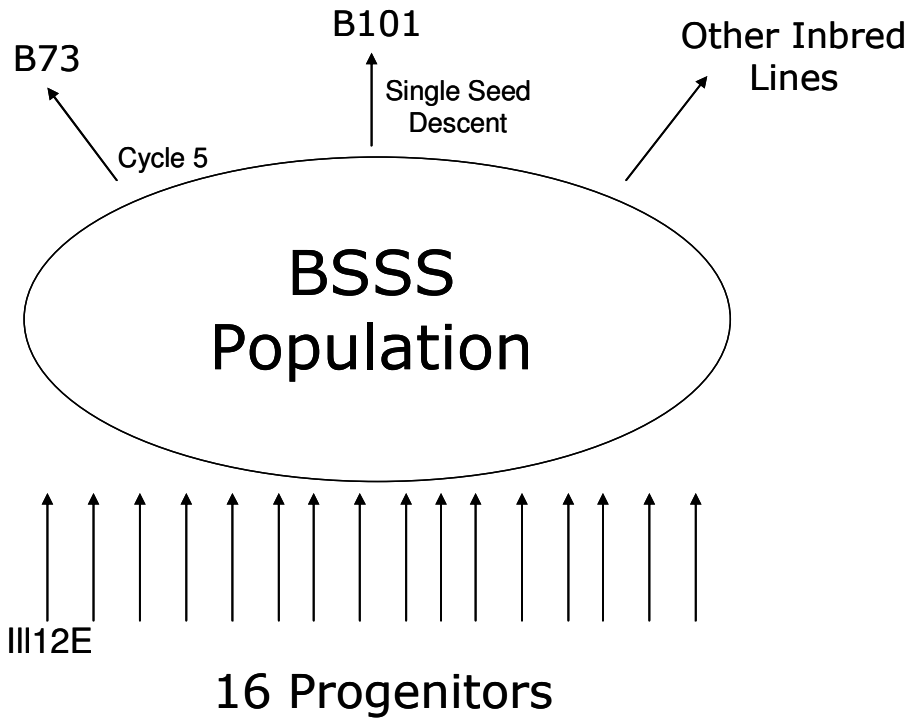

Figure 2. Development of Iowa Stiff Stalk Synthetic (BSSS) population. The relationship of the three inbred lines analyzed in this study, B73, B101, and Ill12E to BSSS are also indicated.

CHAPTER 2: ANALYSIS OF THE 22KDA ALPHA ZEIN GENE FAMILY IN THREE INBRED LINES RELATED TO IOWA STIFF STALK SYNTHETIC

1. Abstract

The 22kDa alpha zein gene family is large and complex. Although a number of genes in this family have been identified, the actual number present is unknown. The main objective of this work is to characterize the 22kDa alpha zein gene family more thoroughly in three inbred lines related to the Iowa Stiff Stalk Synthetic (BSSS) population. Further characterization of this gene family is needed because it is abundant in maize and is therefore economically and nutritionally important. The BSSS population is a major contributor to maize hybrids today with most hybrids having one parent descended from BSSS. Conserved sequences found in a number of previously identified zein-like sequences were used to create PCR primers aimed at amplifying as many 22kDa alpha zein-like genes as possible. Amplified PCR product was then sequenced and analyzed using the DNA sequencing assembly program, CAP3. The results of the CAP3 analysis were then used to estimate the number of zein-like sequences in each inbred and identify zein-like sequences that were present in more than one of the lines examined. The feasibility of using this approach to identify candidate genes for *dzr1* was also evaluated.

2. Introduction

The 22kDa alpha zein gene family in maize is complex with numerous members. The complexity of this family has not been thoroughly characterized, however, 23 22kDa zein-like sequences have been identified in BSSS53 and 15 have been identified in B73 (Song et al. 2001; Song and Messing 2003).

Although maize is a widely used food source for both animals and humans, it is lacking in three essential amino acids, lysine, tryptophan, and methionine. This is largely due to the fact that zeins make up a majority of the protein in the maize kernel and these amino acids are found in small amounts in the zein genes (reviewed in Coleman and Larkins 1999). One exception to this is in the inbred line BSSS53. This line was found to have 30% more methionine compared to normal maize inbred lines (Phillips et al. 1981). The increased methionine is due to overexpression of the 10kDa delta zein gene (Phillips and McClure 1985). The 10kDa zein in normal maize lines contains 22.5% of the methionine found in the kernel (Kirihara et al. 1988). The overexpression is not due to an increase in gene copy number and is actually due to regulation by the *dzr1* gene which maps in the 22kDa alpha zein gene cluster (Schickler et al. 1993). The *dzr1* gene acts on the 10kDa delta zein posttrasncriptionally via a transacting mechanism. The exact physical location of the *dzr1* gene is not known but it has been mapped to within 0.4 cM of at least one known 22kDa zein-like sequence (Chaudhuri and Messing 1995). While its relative genetic location is known, the sequence of the gene has not been determined.

B101 was originally distributed as BSSS53 (Hallauer and Wright 1995) which is derived from Iowa Stiff Stalk Synthetic (BSSS). Phillips et al. (1981) also tested 16 of the BSSS progenitors for lysine plus threonine resistance and only one progenitor, Ill12E, was found to have resistance. This same line was also the only BSSS progenitor shown to have a high methionine phenotype similar to B101. This shared phenotype between B101 and Ill12E suggests that the *dzr1* gene in B101 is also present in Ill12E. As *dzr1* has been mapped to the 22kDa alpha zein gene cluster (Chaudhuri and Messing 1995), it was hypothesized that a 22kDa alpha zein-like sequence that is present in B101 and Ill12E but not

in an inbred with normal methionine levels is a candidate for *dzr1*. Inbred line B73 has normal methionine levels, therefore it was chosen as the third inbred in this study.

The main objective of this work was to characterize the 22kDa alpha zein gene family more thoroughly in three inbred lines related to the Iowa Stiff Stalk Synthetic population. This evaluation included estimation of the number of genes in each inbred and identification of genes that are present in more than one of the lines examined. The feasibility of using this approach to identify candidate genes for *dzr1* was also evaluated.

3. Materials and Methods

DNA was extracted from three inbred lines: B73, B101, and Ill12E using the Epicentre MasterPure Plant Leaf DNA Purification Kit. The extracted DNA was diluted to a concentration of approximately 195 ug/mL for PCR reactions. Primers were designed using conserved sequences found in previously identified zein-like sequences located in the 22kDa zein gene cluster of BSSS53 (Song et al. 2001). The forward conserved sequence is CAAACTTTACATATTCATATCATGTTAA and is found in 15 published zein-like sequences. The reverse conserved sequence, TGTTGYAGGTATGTAGGTGCA, contains a degenerate base and is found in 16 published zein-like sequences. These conserved sequences were used to make several forward and reverse primers with different tails containing restriction sites to facilitate cloning. Figure 1 shows a typical 22kDa zein gene and where the primer sequences are located within the gene.

Three different PCR reactions were used to obtain sequence in an attempt to optimize conditions. The first one was a 25µl reaction which contained 2µl dNTPs (2.5mM), 2µl $MgCl_2$ (25µM), 2ul forward primer (50µM), 2ul reverse primer (50µM), 2.5µl 10x buffer, 1ul

DNA (195 µg/mL), 0.5µl *Taq* (5U/µl) and 13µl water. Cycling conditions consisted of a 94°C, 10 minute denaturation, an annealing temperature of 45°C for 30 seconds, and an extension time of 1.5 minutes at 72°C. The program had 41 cycles. The second PCR reagents varied only in that the $MgCl_2$ concentration was 50mM and *Pfu* polymerase was used instead of *Taq*. The third PCR reaction was deemed optimal and was a 27µl reaction with 1µl dNTPs (2.5mM), 1µl $MgCl_2$ (50mM), 1µl forward primer (50mM), 1µl reverse primer (50µM), 5µl 5x buffer, 1µl DNA (195µg/mL), 1µl Gotaq (Promega) (5U/µl), and 16µl water. The PCR program that was used for the second and third PCR reagents differed from the first PCR program described only in that the annealing temperature was changed to 52°C and the extension time was increased to 3 minutes.

PCR reactions were then run on 0.7% agarose gels and the band of interest, at approximately 800 bp, was cut out and the DNA in the band was purified using the Montage Gel Purification Kit. The purified DNA was then inserted into a plasmid through either directional cloning or TOPO2.1 cloning (Invitrogen). Ligated plasmids were then chemically transformed into OneShot Top10 Chemically Competent Cells (Invitrogen) and grown on LB agarose containing ampicillin and XGal. White colonies were then selected and grown in 100-200µl LB:Amp broth overnight in 96 well plates. Plasmid preparation and sequencing was performed at the Iowa State University DNA Sequencing Facility.

Sequences obtained were initially aligned in VectorNTI (Invitrogen) using the ClustalW algorithm (Thompson et al. 1994). These alignments were used to identify zein-like sequences. Sequences were trimmed based on the overlapping region for plasmids that were sequenced in both directions. Trimmed sequences range in length from 502bp to 540bp. The relative location of the trimmed sequence in a typical 22kDa zein gene is shown

in Figure 1. Trimmed sequences were then analyzed using the CAP3 program (Huang and Madan 1999). Overlap was set at 500 and percent identity was set at 99 with all other parameters being default values. Overlap was set so high because analyses using lower overlap values grouped published 22kDa zein-like sequences in the same contig.

Results from the CAP3 program were visualized in a 3-circle Venn diagram (Chow and Rodgers 2005). Sequences included were ones found in contigs containing sequences from multiple PCR reactions.

Total 22kDa zein-like sequences were predicted for each inbred line using the number of contigs containing sequence from the inbred and the number of published sequences found for the inbred line. First the number of published sequences found for the inbred line was determined. The percent of published sequence found was determined by dividing the number of published sequence found by the total number of published sequences for the inbred. The number of contigs containing sequence from the inbred was divided by the percent of published sequences found to estimate total zein-like sequence expected. As Ill12E does not have any published 22kDa zein-like sequences the percent published sequence for this inbred was estimated by averaging the percent published sequence for B101 and B73.

Similarity plots were constructed in VectorNTI using the AlignX internal analysis. The window size was set at 1.

4. Results

The 22kDa zein gene family is a complex family of zein-like sequences containing both genes and pseudogenes. Many zein-like sequences have been reported, and it is clear

that zeins are encoded by a large multigene family. In order to characterize this family more thoroughly, known zein-like sequences were used to design degenerate oligonucleotides corresponding to conserved regions of the known zein-like sequences. These primers were designed to amplify genomic DNA, and PCR products produced from these primers were cloned and sequenced with the aim of identifying as many 22 kDa alpha zein-like sequences as possible. The primers only have an exact match to eighteen of the twenty-three previously identified 22kDa zein-like sequences therefore all of the 22kDa zein-like sequences were not expected to be found in the inbred lines studied.

The PCR and cloning yielded 774 sequences from three inbreds: B73, B101, and Ill12E as putative 22kDa alpha zein-like sequences. These sequences potentially contained errors from PCR and sequencing. It was reasoned that the likelihood of the same error occurring in two different PCR and/or sequencing reactions would be small, so in the analysis those sequences that were found in more that one PCR reaction were emphasized. To identify sequences that existed in multiple PCR reactions, the 774 sequences along with 22 published 22kDa alpha zein-like sequences (Song et al. 2001) were all aligned by the DNA sequence assembly program CAP3 (Figure 2). The CAP3 analysis yielded 51 contigs, 32 of which contained sequences from multiple PCR reactions (Table 1). The 51 contigs contained a total of 419 sequences, with 380 in the 32 contigs containing sequences from multiple PCR reactions.

Fourteen of the 22 published 22kDa alpha zein-like sequences appeared in contigs with sequence from one or more of the inbred lines examined. Of these, seven included sequences from all three inbreds tested, two contained sequences from only B101 and B73, and the remaining five contained sequences amplified exclusively from B101. Sequences

from B73 were placed in 32 contigs, 20 of which contained sequences from multiple PCR reactions. Of those 20, nine included known 22kDa alpha zein-like sequence. B101 sequence was present in 23 contigs, 20 of which had sequences from multiple PCR reactions. All 14 contigs containing known 22kDa alpha zein-like sequences contained B101 sequences. Ill12E was represented in 20 contigs, 16 of which included sequences from multiple PCR reactions. Known 22kDa alpha zein-like sequence was found in seven of those 16 contigs.

Sixteen contigs represented previously unidentified 22kDa alpha zein-like sequences. Two of these contigs contained sequence from all three lines and no known 22kDa alpha zein-like sequences. One contig contained only B73 and B101 sequence and two contigs contained sequence exclusively from B101 and Ill12E. Also, one contig containing sequence only from B73 and Ill12E was identified. Twelve contigs containing the sequences of only one inbred and no known 22kDa alpha zein-like sequences were also discovered. Of these, seven contigs contained only B73 sequence, one contig contained only B101 sequence, and four contigs had sequences unique to Ill12E.

Figure 3 depicts how the contigs containing sequences from multiple PCR reactions are shared between the three maize lines. Seventeen, or nearly half, of those 32 contigs contain sequence that is exclusive to one of the inbred lines. Of the remaining 15 contigs, six are shared between two inbreds and nine are shared between all three inbreds.

355 sequences were not placed in contigs and were labeled singlets by CAP3. Additionally, 39 sequences were placed in contigs that only represented one PCR reaction and could potentially represent PCR errors. For inbred B73, 12 contigs contained 25 sequences in which this occurred. B73 had 57% of its sequences identified as singlets or

placed in one-reaction contigs which was the highest percentage of the three inbreds examined. Incidentally, it also had the highest number of sequences included in the analysis. For inbred Ill12E, there were four contigs containing a total of eight sequences that were all from one PCR reaction. For this inbred 48% of the sequences were singlets or in contigs with sequence from only one reaction. Although B101 had the lowest number of sequences included in the analysis there were only two contigs, with four sequences total, where only one PCR reaction was represented. B101 also had the lowest percentage (43%) of singlets or sequences in contigs only containing one reaction. Overall, 51% of the 774 sequences analyzed were not included in contigs identified as known or potential 22kDa alpha zein-like sequences.

Using the results from Table 1 the number of 22kDa zein-like sequences present in all three inbred lines was estimated (Table 2). The number of 22kDa zein-like sequences observed in multiple PCR reactions in B101 and B73 was divided by the percentage of known 22kDa zein-like sequences found per inbred line. As 22kDa alpha zein-like sequences have not been sequenced in Ill12E, the 22kDa zein-like sequences present in this line are unknown. To estimate the percent complete value of the results for this line the percent complete values for B101 and B73 were averaged. Using this information it is estimated that there are approximately 33 22kDa zein-like sequences for B101, 38 22kDa zein-like sequences for B73, and 28 22kDa zein-like sequences for Ill12E.

The similarity between the 32 contigs containing sequence from multiple PCR reactions was also examined (Fig. 4). Three similarity graphs were constructed, one from the alignment of the consensus sequences from the contigs containing published sequence (Fig. 4A), one from the alignment of consensus sequences from the contigs not containing

published sequence (Fig. 4B), and one from the alignment of the consensus sequences from all 32 contigs (Fig. 4C). The similarity graphs were also compared to the sequence location within a typical zein gene. The most conserved regions were near the translation start site with greater differences seen at the ends of the sequences. The ATG was conserved in all consensus sequences and the TATA box was conserved in all but one consensus sequence. This contig from which the consensus sequence was derived contained the published sequence, *azs22-1*, which has been identified as a pseudogene and encodes a truncated protein. Greater differences between sequences were seen within the promoter region and also within the region containing repeated domains.

5. Discussion

While PCR is a valuable tool, it is an imperfect one. The taq polymerase error rate has been reported to be 8.0×10^{-6} errors per base (Cline et al. 1996), however this is an absolute error rate. As errors in earlier cycles will continue to be amplified it is possible for errors to be propagated into the product. The efficiency for the PCR reactions is unknown and without this information the exact frequency of error cannot be determined. It is likely however that some of the sequences may contain errors amplified by PCR.

Sequencing error should also be considered along with PCR error. For these sequences the accuracy is 99% (Gary Polking, pers. comm.) which means up to five errors could be expected in each sequence submitted to CAP3 analysis. The CAP3 analysis yielded a total of 51 contigs, with 32 of those containing sequence from multiple PCR reactions. Nineteen contigs contained sequence from only one PCR reaction so the 39 sequences contained in these contigs could potentially result from PCR errors. Additionally, the

analysis identified 355 singlets which could also represent sequences with PCR or sequencing error. Any future sequencing should be done in both the forward and reverse directions to help alleviate this problem. Further sequencing is also expected to identify sequences that match some of the singlets. Having sequences present in multiple PCR reactions helps ensure that the sequence matches are not merely the product of a PCR error amplified or a sequencing error. Further discussion of contigs will refer only to the 32 representing sequences from multiple PCR reactions.

Comparison of the sequences in this analysis to known 22kDa zein-like sequences allows the number of 22kDa alpha zein-like sequences present in inbred lines to be predicted. Fourteen contigs contained sequence matching to previously identified 22kDa alpha zein-like sequence (Song et al. 2001). Twenty-three 22kDa alpha zein-like sequences were identified in the 22kDa alpha zein gene cluster of BSSS53 but the twenty-third sequence, *azs22*-3, was not included in this analysis because it is truncated. Sequences were trimmed and only those at least 502 basepairs long were retained for CAP3 analysis. It is interesting to note that only 14 contigs were identified with known 22kDa alpha zein-like sequences. Five of the known sequences, *azs22-11*, *azs22-12*, *azs22-13*, *azs22-17*, and *azs22-18* lacked the conserved sequences used in the primers and they were not expected to be identified. As expected, none of the sequences included in the analysis matched to these five zein-like sequences. All 14 contigs containing known 22kDa alpha zein-like sequence also contained sequence from B101. This was anticipated as these sequences were discovered in the inbred BSSS53 from which B101 was derived (Hallauer and Wright 1995). Because it is unlikely that sequences that do not contain the primer sites would be amplified in the experiments, about 26% (6 of 23) of the 22kDa zein fraction in B101 is not expected to amplify in this analysis.

Furthermore, this analysis did not identify three 22kDa zein-like sequences that were expected, therefore only 82% (14 of 17) of the sequences this method should yield were found. Twenty contigs contained sequence from B101, so knowing this is only 82% of expected sequence, therefore at least four more 22kDa zein-like sequences are expected to be found, for a total of 24 22kDa zein-like sequences in this inbred. By dividing the number of contigs found containing B101 by the percent of total published zein-like sequences that were found, a predicted value of 33 can be given for the total number of 22kDa alpha zein-like sequences present in this inbred (Table 2). This suggests that perhaps there are more zein-like sequences, like the *floury2* allele (*azs22-16*) that are not located in the sequenced portion of the 22kDa alpha zein gene cluster.

A similar calculation can be performed for the inbred B73. There have been 15 previously identified zein-like sequences (Song and Messing 2003). Eleven of these contained the conserved primer regions. However, two genes identified in BSSS53 but not reported in B73, *azs22-14* and *azs22-15*, were found in this analysis. The addition of *azs22-14* and *azs22-15* changes the number of known 22kDa zein-like sequences potentially detectable by this method to 13. The percentage of known 22kDa alpha zein-like sequences potentially detectable in B73 is therefore 76%, leaving 24% of the 22kDa zein fraction invisible by this analysis. Sequences from B73 occurred in 20 contigs, including the nine containing known 22kDa zein-like sequences. As only 69% of the expected sequences were found, it is estimated that approximately nine more 22kDa zein-like sequences, for a total of 29, could be found in B73 using this method. Using the same calculations used for B101, it is estimated that there are at least 38 22kDa alpha zein-like sequences in B73 (Table 2).

The 22kDa zein-like sequences *azs22-14* and *azs22-15* have been shown to arise from duplication events of *azs22-4* and *azs22-5* respectively, in BSSS53 (Song et al 2001). As they are not present in the 22kDa zein gene cluster in B73, it is likely that these two sequences, and perhaps others that have yet to be found, were translocated at some point to a different place in the genome. Already one 22kDa zein-like sequence, *azs22-16*, has been found separate from the 22kDa alpha zein-like sequences tandemly arranged on chromosome 4S (Coleman et al. 1997) so this seems a likely explanation for the presence of *azs22-14* and *azs22-15* in B73 in the analysis.

A predicted number of 22kDa alpha zein-like sequences for Ill12E can be estimated by averaging the values in the '% Published sequences found' column of Table 2. The averaged value estimates that only 57% of the 22kDa zein-like sequences have been identified by this analysis for Ill12E. By multiplying this percentage by 16, the number of 22kDa zein-like sequences observed in multiple PCR reactions in Ill12E, it is estimated that there is a minimum of 28 22kDa zein-like sequencess present in this inbred.

These results may be an underestimate of the total number of zein-like sequences found as the CAP3 program did not differentiate between sequences with higher than 99% identity and sequences with exactly 99% identity, grouping them all in the same contig. Closer examination of some sequences revealed that at least one contig, Contig 2, contains sequences that were derived from amplifications of different, but highly similar genomic regions in the same inbred. It is interesting to note that Contig 2 also contains the known 22kDa zein-like sequence, *azs22-2*. Emrich et al. (2007) identified the presence of a number of nearly identical paralogs (NIPs) in the maize genome. NIPs are characterized by having greater than 98% identity, with most having greater than 99% identity. One explanation for

the closely related sequences seen in this analysis is that they are NIPs. Although only one of the contigs has been verified to contain sequences from multiple genomic regions, the large number of components of several other contigs suggests they may contain several NIPs. One way to resolve this discrepancy is to alter the CAP3 program to have greater stringency or use a different program that would allow for higher than 99% identity.

The variation in sequence length may have also caused similar or identical sequences to be grouped in different contigs. Previous analyses indicated that the sequences in Contig 24 were identical to sequences found in Contig 2. The sequences were trimmed to slightly different lengths causing the sequences in Contig 24 to start further upstream than the sequences in Contig 2. It is not known if there were similar cases in other contigs so further scrutiny of the contig sequences and perhaps more uniform sequence lengths are needed. Additionally, there were 355 singlets from the CAP3 analysis that through further sequencing could be found to have matches in multiple PCR reactions, leading to identification of new 22kDa alpha zein-like sequences.

Eighteen contigs contained sequence that can be considered putative new 22kDa alpha zein-like sequences. Among the inbreds, B73 had the most potential new 22kDa alpha zein-like sequences as seven contigs were identified with sequence exclusive to B73. One possible reason for the high number in this inbred compared to the other two is that B73 contributed the most sequences to the CAP3 analysis and had the highest number of sequences placed in contigs. However, this inbred had the highest percentage of singlets compared to sequences included in the analysis and the highest number of sequences present in a contig containing only one PCR reaction.

There was only one putative new 22kDa alpha zein-like sequence identified that contained sequences exclusive to B101. This is somewhat expected as extensive research has been done on the 22kDa alpha zein-like sequences in this inbred (Song et al. 2001; Song and Messing 2003). B101 did have the highest percentage of sequences in contigs compared to the other inbreds as well as the least number of sequences in contigs representing only one PCR reaction.

Four putative new 22kDa alpha zein-like sequences were found in Ill12E. 22kDa zein-like sequences that are unique to Ill12E are of interest in determining how the 22kDa zein-like sequences were inherited through the development of different Iowa Stiff Stalk Synthetic lines. While these putative genes may be unique to Ill12E in this analysis, evaluation of other BSSS lines may show that these sequences were indeed passed on. Further studies may identify more 22kDa alpha zein-like sequences in Ill12E, adding to our knowledge of the inheritance of this gene family.

There are six putative new 22kDa alpha zein-like sequences from contigs that contain sequences from multiple inbreds. While several of the putative new 22kDa zein-like sequence contigs contain sequence only from two inbreds, because the percent published sequences found is 57% on average, which is also an estimate of the percent of total zein-like sequences the analysis has identified, it is possible that the third inbred contains that zein but this analysis is not yet complete enough to identify it. Further studies of these shared sequences will be useful in characterizing the inheritance of the 22kDa alpha zein gene family.

It is also important to note that there can be a number of sources for a single inbred line. Gethi et al. (2002) estimated the level of genetic diversity both within and among

several inbred lines. This study found that there was a significant amount of variation found within the sources for a single inbred and even within one source there may be variation. Complete homozygosity cannot be assumed for an inbred line source; therefore, this variation may cause differences to be seen when identifying 22kDa zein-like sequences in inbreds. Alternate sources of the inbred lines used in this study may contain different sequences than the ones identified here.

The inbred line B101 shows a rare high methionine phenotype. This line was originally distributed as BSSS53 (Hallauer and Wright 1995) which is a high methionine line derived from Iowa Stiff Stalk Synthetic. When the 16 available progenitors of Iowa Stiff Stalk Synthetic were also tested for methionine content, only Ill12E was shown to have a high methionine phenotype similar to B101. When Phillips et al. (1981) discovered the high methionine phenotype of B101 while looking for lysine plus threonine resistant lines, they also tested 16 of the Stiff Stalk Progenitors. Of the progenitors, only Ill12E was found to be resistant. It was believed that a higher methionine to lysine ratio was responsible for the resistance which is supported by the increased level of methionine found in Ill12E. This shared phenotype between B101 and Ill12E suggests that the *dzr1* gene that regulates the methionine increase in B101 is also present in Ill12E. This also suggests that the *dzr1* allele from Ill12E was inherited by B101. As *dzr1* has been mapped to the 22kDa alpha zein gene cluster (Chaudhuri and Messing 1995), this study hypothesizes that a 22kDa alpha zein-like sequence that is present in B101 and Ill12E but not in an inbred with normal methionine levels is a candidate for *dzr1*. Earlier it was mentioned that an objective of this research was to determine the feasibility of this method for identifying *dzr1*. Although two sequences were identified as being present in contigs only containing sequence from Ill12E and B101

(Contigs 24 and 30), the completion rate of this method needs to be much higher to make any definitive conclusions about the identification of putative sequences for *dzr1*. It should also be noted that a 22kDa zein-like sequence could be modified and different between these two lines. If this has occurred, it is possible the modifications would not be detected by this analysis.

The similarity of consensus sequences derived from the contigs was examined for the 32 contigs containing PCR from multiple reactions. The area around the start of translation was highly conserved which is expected. This is important to ensure translation initiation. The regions at either end of the sequences showed fairly high levels of variation. Some of this variation was seen in the promoter region and a large portion of the variation was seen in the region where repeated domains are prevalent. Variation is expected in the repeated domains because the number of repeats can vary between sequences.

6. References

Chaudhuri S, Messing J. 1995. RFLP mapping of the maize *dzr1* locus, which regulates methionine-rich 10 kDa zein accumulation. Mol Gen Genet 246(6):707-715.

Chow S, Rodgers P. Extended Abstract: Constructing Area-Proportional Venn and Euler Diagrams with Three Circles; 2005; Paris.

Cline J, Braman JC, Hogrefe HH. 1996. PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Research 24(18):3546-3551.

Coleman CE, Clore AM, Ranch JP, Higgins R, Lopes MA, Larkins BA. 1997. Expression of a mutant alpha-zein creates the *floury2* phenotype in transgenic maize. Proc Natl Acad Sci U S A 94(13):7094-7097.

Coleman CE, Larkins BA. 1999. The prolamins of maize. In: Shewry PR, Casey R, editors. Seed Proteins. Dordrecht, Netherlands: Kluwer Academic Publishers. p 109–139.

Gethi JG, Labate JA, Lamkey KR, Smith ME, Kresovich S. 2002. SSR Variation in Important U.S. Maize Inbred Lines. Crop Science 42:951-957.

Hallauer AR, Wright AD. 1995. Registration of B101 maize germplasm. Crop Sci 35(4):1238-1239.

Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Research 9:868-877.

Kirihara JA, Hunsperger JP, Hahoney WC, Messing JW. 1988. Differential expression of a gene for a methionine-rich storage protein in maize. Mol Gen Genet 211(3):477-484.

Phillips RL, McClure BA. 1985. Elevated protein-bound methionine in seeds of a maize line resistant to lysine plus threonine. Cereal Chem 62(3):213-218.

Phillips RL, Morris PR, Wold F, Gengenbach BG. 1981. Seedling screening for lysine-plus-threonine feedback resistant maize. Crop Sci 21(4):601-607.

Schickler H, Benner MS, Messing J. 1993. Repression of the high-methionine zein gene in the maize inbred line Mo17. Plant J 3(2):221-229.

Song RT, Llaca V, Linton E, Messing J. 2001. Sequence, regulation, and evolution of the maize 22-kD alpha zein in gene family. Genome Research 11(11):1817-1825.

Song RT, Messing J. 2003. Gene expression of a gene family in maize based on noncollinear haplotypes. Proceedings of the National Academy of Sciences of the United States of America 100(15):9055-9060.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22(22):4673-4680.

7. Figures and Tables

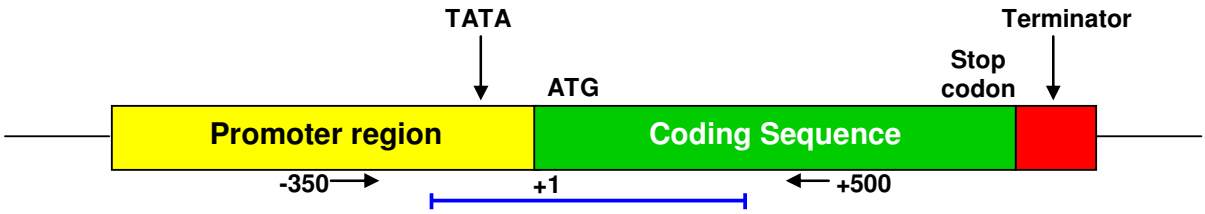

Figure 1. Typical 22kDa zein gene. Locations of the forward and reverse primer conserved sequences indicated by black arrows. The blue bar indicates the approximate location of the trimmed sequences used for CAP3 analysis. +1 indicates the start of translation, with -350 and +500 relative to that position. Figure is relatively drawn to scale.

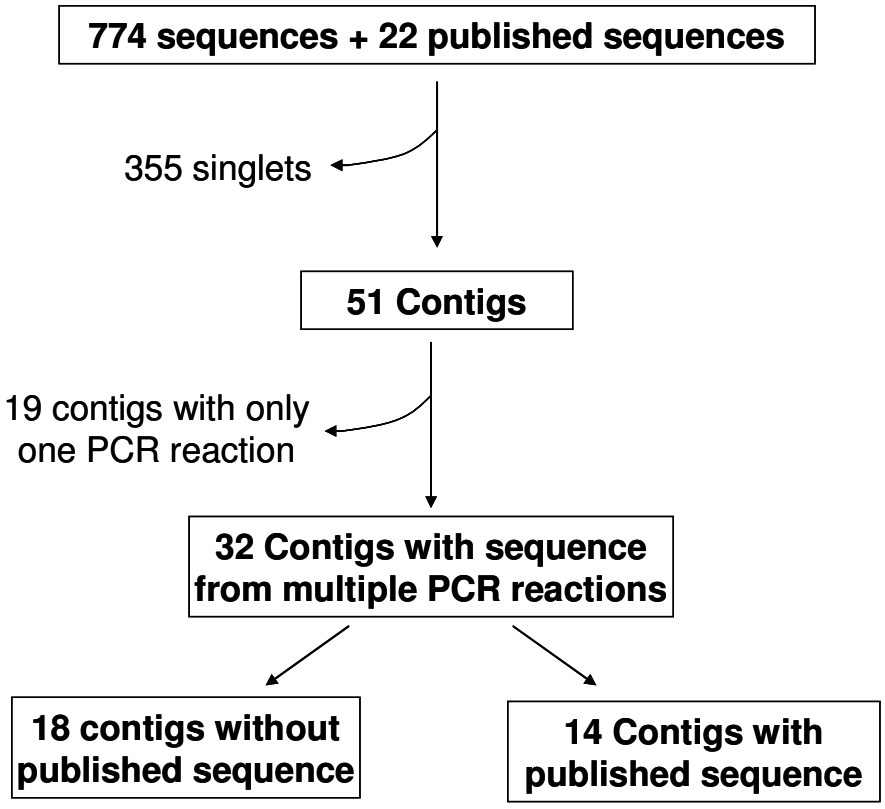

Figure 2. Breakdown of CAP3 analysis. General input and output of analysis is shown. Singlets and contigs containing sequences from only one PCR reaction were not subject to further analysis.

Contig	B73	B101	Ill12E	Known zein sequence	GenBank Accession #	Accession Region
1	**5**	**12**	**7**	azs22-1	AF090447	147795-148594
2	**16**	**5**	**20**	azs22-2	AF090447	151904-152704
3	**12**	**12**	**14**	azs22-6	AF090447	183200-184006
4	**3**	**6**	**8**	azs22-15	AF090447	112406-113206
5	**33**	**10**	**15**	azs22-5	AF090447	169654-170454
6	**2**	**2**		azs22-14	AF090447	108295-109095
7	**3**					
8	**23**	**12**		azs22-16	AF090446	15313-16104
9		**7**		azs22-10	AF090447	220387-221187
10		**3**		azs22-20	AF090447	228600-229397
11	2	1	**7**	azs22-7	AF090447	187312-188814
12	1	1	1	azs22-8	AF090447	199338-200138
13		**4**		azs22-9	AF090447	205153-205948
14	**13**					
15		**2**				
16	**8**		**3**			
17	1	1				
18			**15**			
19		**13**		zp22/D87	AF090447	261759-262472
20	**6**	**9**	**7**			
21	2					
22	**6**					
23		2				
24		1	**2**			
25	1	1	1			
26	**6**					
27			2			
28	2					
29		2				
30		1	**3**			
31			2			
32			**15**			
33			**2**			
34			2			
35			2			
36			**7**			
37	**2**					
38	2					
39	2					
40	2					
41	**2**					
42	2					
43	2					
44	2					
45	2					
46	2					
47	2					
48	3					
49		2				
50	**2**					
51		**3**		zp22/6	AF090447	253535-254335
Total contig sequences	172	112	135			
Singlets	175	73	107			
Total sequence analyzed	347	185	242			

Table 1. Contig and singlet results from CAP3 analysis. The number of sequences contained in each contig are listed. Yellow highlighted boxes indicate contigs with the strongest support because they contain sequences from multiple PCR reactions. Numbers in bold red type indicate contigs that contain sequences from multiple PCR reactions within the same inbred line. The GenBank Accession number and region for known sequences are included.

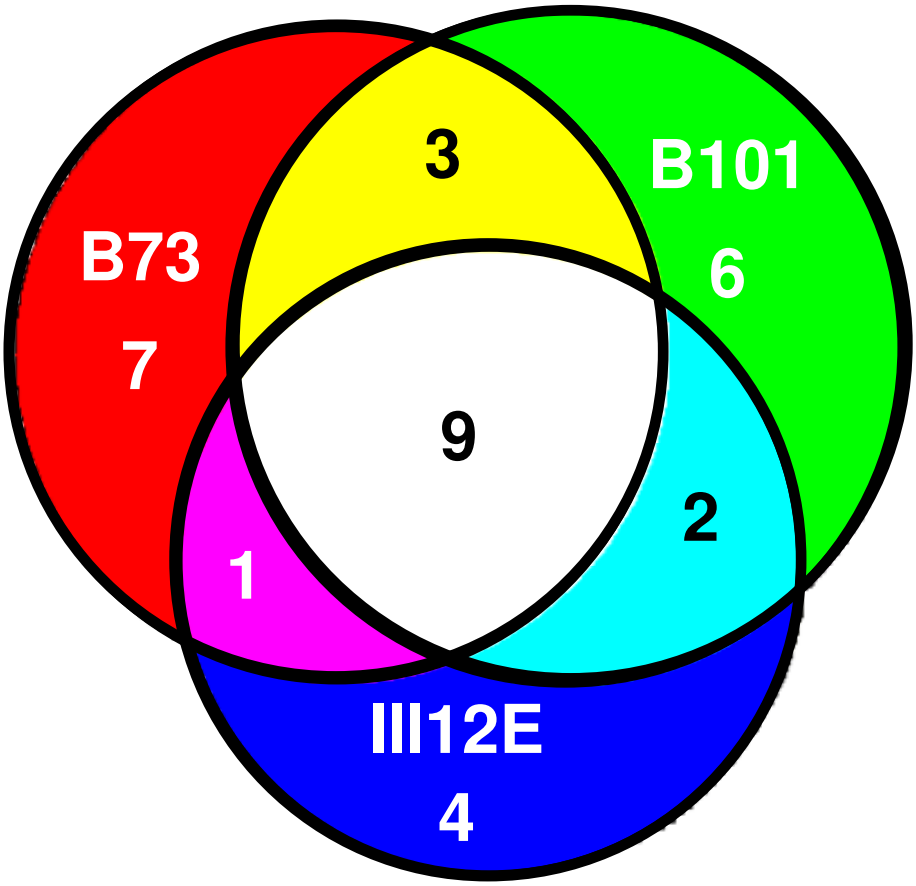

Figure 3. Venn diagram representing 32 contigs containing sequences from multiple PCR reactions.

Inbred	Contigs identified[a]	Published sequences[b]	# Published sequences found	% Published sequences found	Total zein-like sequences expected[f]
B101	20	23	14	61%[d]	33
B73	20	17	9	53%[d]	38
Ill12E	16	N/A[c]	N/A[c]	57%[e]	28

Table 2. Predicted number of 22kDa alpha zein genes in B101, B73, and Ill12E.

[a] Only contigs with sequences from multiple PCR reactions included

[b] (Song et al. 2001; Song and Messing 2003)

[c] No 22kDa alpha zein sequence has been published for Ill12E so the number of genes present is unknown

[d] % Published sequences found = # Published sequences found/Published sequences

[e] The % Published sequences found for Ill12E was determined by averaging that number for B101 and B73.

[f] Total zein-like sequences expected = Contigs identified/(% Published sequences found)

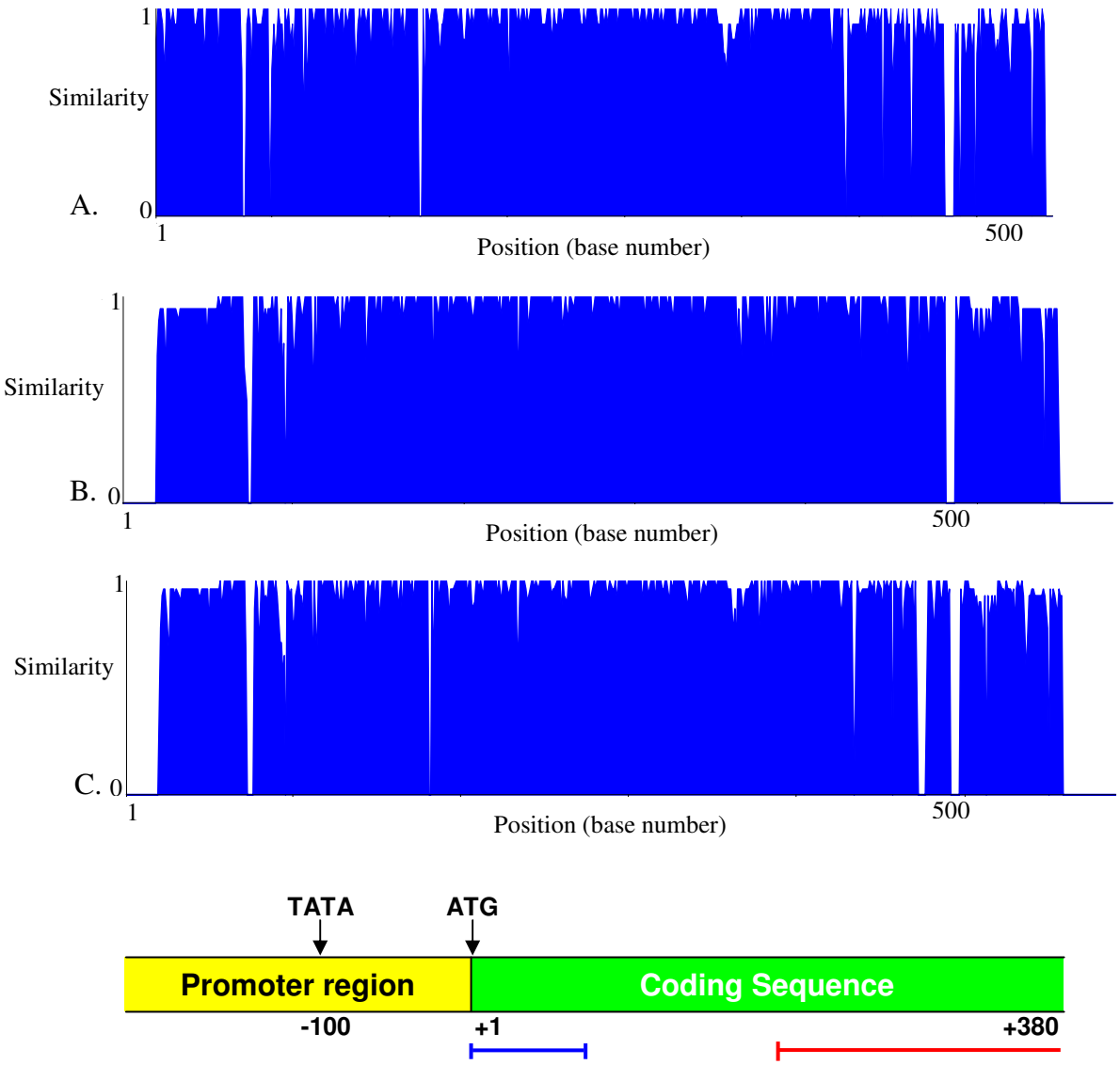

Figure 4. Similarity graphs for the consensus sequences from the 32 contigs containing sequences from multiple PCR reactions and the relation to a typical zein gene. Position refers to the base number within the PCR sequence. In the gene figure the blue bar indicates the approximate location of the signal sequence. The red bar indicates the area where repeated domains are prevalent. +1 in the gene figure indicates the start of translation, with -100 and +380 relative to that position. Figure is approximately drawn to scale. A) 14 contig consensus sequences from contigs containing published zein-like sequence, B) 18 contig consensus sequences from contigs not containing published zein-like sequence, C) All 32 contig consensus sequences.

CHAPTER 3: THE IMPACT OF THE *OPAQUE2* MUTATION ON THE EXPRESSION OF TWO TRANSGENES CONTAINING ZEIN PROMOTERS

1. Abstract

Transgenes are potentially valuable for crop improvement but it is important to develop ways of regulating their expression. Using known mutations can be one way to regulate transgene expression. The objective for this study was to examine the impact of the *opaque2* mutation on expression of two Green Fluorescent Protein (GFP) transgenes with different zein promoters. The transgenes contained either a 19kDa zein or 27kDa zein gene promoter. As the *opaque2* (*o2*) mutation impacts expression of the two zein classes differently, it is of interest to determine whether transgenes with these promoters have expression patterns comparable to the native genes. GFP was used because it is easily quantified using fluorescence values. Kernels from ears containing either the 19zn-GFP or 27zn-GFP transgene were analyzed using a fluorometer. Translucent and opaque kernels were separated prior to fluorescence analysis and the two classes were compared within each ear. For the 19zn-GFP transgene, ears with different transformation events and genome compositions were compared. Although some ears had unexpected transgene phenotypic segregation ratios, epistasis was seen to some extent in all ears. Ear to ear variation made it hard to make definite conclusions about the impact of the *opaque2* mutation on this transgene. The ears containing the 27zn-GFP transgene had the same transformation event but different genome compositions and some ears exhibited epistasis and other did not. Additionally, some ears had unexpected transgene phenotypic segregation ratios. This variation made conclusions about the impact of *opaque2* on this construct difficult as well.

2. Introduction

Transgenes are economically important for a variety of reasons. They are generally thought of as means for crop improvement but they can also be used to produce pharmaceuticals in plants. Maize seeds in particular are of interest for a number of reasons including their ability to be easily stored and transported, ability to stably store protein for long periods of time, and the ability to be easily processed (reviewed in Howard and Hood 2005). Maize is also a widely consumed food source for both humans and animals so it is a suitable vessel for biopharmaceuticals that contain orally delivered products. The increasing prevalence of transgenes for both crop improvement and biopharmaceuticals demands that ways to regulate transgene expression be explored.

One way to regulate transgene expression is with known mutations. By studying the impacts of known mutations on transgene expression it can be determined whether or not these mutations would be useable in the regulation of transgenes. The mutation that this study looked at is the *opaque2* mutation in maize. The *opaque2* mutants are characterized by having a soft, floury endosperm that causes the kernel to appear opaque rather than translucent when placed on a light box (Figure 1). The major impact of this mutation is to decrease the amount of alpha zeins in the kernel (Kodrzycki et al. 1989; Langridge et al. 1982). The *opaque2* gene encodes a bZip transcription factor that binds to a 22kDa alpha zein gene promoter, thereby having a direct impact on gene expression (Hartings et al. 1989; Schmidt et al. 1990). The 22kDa alpha zeins are actually seen to decrease by more than 90% in *opaque2* mutants. The *opaque2* mutation also causes a 45-70% decrease in the 19kDa alpha zeins although the mechanism by which this occurs is not known. The bZip transcription factor has not been found to bind to 19kDa zein promoters (Schmidt et al. 1992)

suggesting that the *opaque2* mutation has an indirect impact on this gene family. Interestingly, the *opaque2* mutation does not impact the levels of the other zein proteins (Kodrzycki et al. 1989). Maize with the *opaque2* mutation is also promising for biopharaceuticals because although overall yield is lowered, recombinant protein can be produced 3-5 times higher than normal maize (Howard and Hood 2005), but the authors do not indicate what promoters were used. Studying how transgene expression is impacted by this mutation may help determine if this mutation is beneficial for transgenic plants.

The transgenes used in this study express Green Fluorescent Protein (GFP) using one of two endosperm specific promoters. The constructs were originally made to evaluate milling processes (Shepherd et al.) but are useful for other purposes as tissue specific transgenes. The constructs contained one of two zein gene promoters, either a 27kDa or a 19kDa zein promoter. These promoters are of interest because the native zein genes from which they are derived are impacted differently by the *opaque2* mutation. The 19kDa zeins are decreased while the 27kDa zeins are unaltered so it is of interest to find out if transgenes carrying their promoters exhibit expression patterns similar to the native genes from which they were derived. The objective for this project was to examine the impact of the *opaque2* mutation on expression of two GFP transgenes with different zein promoters.

3. Materials and Methods

The two GFP transgenes used, 19zn-GFP and 27zn-GFP, were from research done by Shepherd et al. Maize plants homozygous for the *opaque2* mutation and plants homozygous for the GFP transgene were crossed to yield F1 seed. The F1 seed was heterozygous for the *opaque2* mutation and the GFP transgene and plants from F1 seed were grown and self-

pollinated to yield F2 kernels with the two loci segregating. Twenty-one plants containing the 19zn-GFP transgene were selfed and 11 plants containing the 27zn-GFP transgene were selfed. The ears were prescreened for GFP using a blue light and orange filter. Nine ears were chosen that exhibited GFP fluorescence, four contained the 19zn-GFP transgene and five contained the 27zn-GFP transgene. The genome compositions and further description of the ears can be found in Table 1. The ears were then shelled and kernels were separated based on opaque or translucent phenotype using a light box. All ears exhibited a segregation ratio that was not statistically different from the expected ratio of 3 translucent: 1 opaque. Individual kernels were not prescreened for fluorescence.

To quantify fluorescence, the top of each kernel was sanded to expose the endosperm. Forty-eight opaque and 48 translucent kernels from each ear were randomly selected for analysis. Two Costar 48-well plates were spray painted black and used for each ear. In each plate 24 opaque and 24 translucent kernels were placed into wells in a randomized design. One kernel was placed in each well and secured with black clay with the exposed endosperm centered and facing up in the well. Fluorescence was quantified using a Tecan SPECTRAFlourPlus. It was set to measure fluorescence with an excitation wavelength of 485 and an emission wavelength of 535. The gain was manually set at 60. Each well was read once with 30 flashes and an integration time of 40 μs. Each plate was read one time.

4. Results

The *opaque2* mutation has a pronounced effect on the expression of the alpha zeins in maize kernel endosperm. In an *o2/o2* genetic background, the 22kDa alpha zein genes are generally only present in trace amounts and the 19kDa alpha zeins are diminished by up to

70% (Kodrzycki et al. 1989). However, the other zein genes, including the 27kDa gamma zeins, have not shown a pronounced decrease in *opaque2* kernels. Nine maize ears segregating for the *opaque2* mutation and one transgene were analyzed for transgene expression. The ears contained either the 19zn-GFP transgene which has a 19kDa zein promoter or the 27zn-GFP transgene which has a 27kDa zein promoter. Based on the behavior of the native genes as described in the literature, GFP levels were expected to be decreased in *opaque2* kernels compared to translucent kernels in ears containing the 19zn-GFP construct but not in ears containing the 27zn-GFP construct.

F1 ears with a genotype of *o2O2*/GFP(-)GFP(+) were selfed to generate F2 seed with expected segregation ratios listed in Table 2. The four ears that contained transgene 19zn-GFP also had different transformation events and had different genome composition (Table1). The five ears that contained transgene 27zn-GFP had the same transformation event but had different genome compositions. As mutation penetrance can vary based on transformation event or genome composition it is important to study these effects on transgene expression as well.

GFP fluorescence was analyzed in 96 kernels from each ear with 48 having an opaque phenotype and 48 having a translucent phenotype. The fluorescence values were then ranked lowest to highest and plotted as a line graph. Thus, each graph contained two data series, one for opaque kernels and the other for translucent kernels. If the transgene segregated as a single dominant allele according to normal Mendelian ratios, we would expect to see 12 kernels from each set of 48 with no GFP expression because they did not inherit the transgene. The remaining 36 kernels should contain the transgene in either the homozygous or heterozygous state. Figures 3 and 4 show hypothetical data from segregating ears plotted

in this way to illustrate different types of gene action in this presentation format. In all figures, it is indicated with the red boxes where the homozygous GFP null allele genotype kernels are expected to be and with green boxes where the homozygous GFP positive kernels are expected to be. If the transgene has dominant gene action (Figure 3A) the kernels with a GFP(-)GFP(-) genotype should have the lowest fluorescence values and these kernels should comprise one-fourth (12) of the analyzed kernels. Kernels with a GFP(-)GFP(+) or GFP(+)GFP(+) genotype should have higher fluorescence values and these kernels should comprise three-fourths (36) of the analyzed kernels. For co-dominant gene action (Figure 3B) the kernels with a GFP(-)GFP(-) genotype would also have the lowest fluorescence values and again these kernels should comprise one-fourth (12) of the analyzed kernels. Kernels with a GFP(+)GFP(+) genotype should have the highest fluorescence and these kernels should comprise one-forth (12) of the analyzed kernels. Kernels with a GFP(-)GFP(+) should have fluorescence values intermediate between the other two genotypes and these kernels should comprise one-half (24) of the analyzed kernels. If, however, the gene has recessive gene action (Figure 3C), kernels with a GFP(-)GFP(-) or GFP(-)GFP(+) genotype should have the lowest fluorescence values and these kernels should comprise three-fourths (36) of the analyzed kernels. Kernels with a GFP(+)GFP(+) genotype should have the highest fluorescence values and these kernels should comprise one-fourth (12) of the kernels analyzed.

Transgenes occasionally have low pollen transmission or gene silencing, causing higher than expected numbers of kernels negative for the transgene. The expected graph for a transgene with low pollen transmission or gene silencing, is shown in Figure 3D. For this graph the breaks in fluorescence values do not fall in the expected positions. More kernels

are seen with low fluorescence values, extending beyond the homozygous negative region into the heterozygous positive region. Kernels with intermediate fluorescence values likewise are seen in the remainder of the heterozygous positive region and also in the homozygous positive region. This analysis cannot distinguish between low pollen transmission and gene silencing so both must be considered a possibility if a graph similar to Figure 3D is seen.

The *opaque2* mutation is also expected to have an epistatic effect on kernels containing the 19zn-GFP construct as this mutation has been previously reported to significantly lower 19kDa zein expression (Kodrzycki et al. 1989). The graph expected if the *opaque2* mutation has an epistatic effect on the transgene is shown in Figure 4. The transgene is expected to exhibit co-dominance and in this case, the *opaque2* kernels would have diminished fluorescence values because of epistasis. Fluorescence expression is inhibited in opaque kernels but the kernel values fall in the same regions as normally expressing kernels. Opaque kernels with lower fluorescence values are seen in the homozygous negative region, ones with intermediate vales are seen in the heterozygous region and the kernels with the highest fluorescence values are seen in the homozygous positive region, similar to the distribution of translucent kernels. However, the epistatic impact causes the opaque fluorescence values to be lower than the fluorescence values seen in the translucent kernels.

The transgenes in this analysis were expected to show co-dominant gene action. Richards et al. (2003) showed that transgenic plants homozygous for the GFP transgene showed twice the fluorescence as hemizygous plants. In maize containing a GFP transgene, a range of fluorescence levels was seen for GFP positive kernels during visual screening. If

the transgene exhibits co-dominance as expected, the highest 12 kernels are expected to have significantly higher fluorescence values than the other GFP positive kernels.

Four ears analyzed contained the 19zn-GFP transgene (Fig. 5). Three of the ears had the same genome composition and transformation event (Fig. 5B-D), while the fourth ear had a different genome composition and transformation event from the other three (Figure 5A). The *opaque2* mutation appears to have an epistatic impact on all four ears but the effect is seen in different classes of kernels. Ears 9141-04 (Fig. 5A) and 9145-02 (Fig. 5C) only appeared to have epistatic impact in the homozygous positive region. In both of those ears there were a high number of kernels with low fluorescence values which could indicate low pollen transmission or gene silencing. Previous research has shown that low pollen transmission results in excess numbers of progeny homozygous null for the transgene (Sangtong et al. 2002). However, for ears 9145-01(Fig. 5B) and 9145-04 (Fig. 5D) the epistatic impact of *opaque2* on the transgene was seen in both the heterozygous positive and homozygous positive regions. These ears also appeared to segregate and to be co-dominant for transgene expression in translucent kernels, as expected. The translucent kernels in the heterozygous positive region had intermediate fluorescence values while the translucent kernels in the homozygous positive region had markedly higher fluorescence values. Ear to ear variation made it difficult to draw conclusions about gene action and transgene inheritance.

Five ears analyzed contained the 27zn-GFP transgene from the same transformation event (Fig. 6). While all ears contained the transgene from the same transformation event they did not all have the same genome composition. Ears 9143-01 and 9143-03 had one genome composition and ears 9148-01, 9148-02, and 9148-03 had another genome

composition (Table 1). The *opaque2* mutation had an impact on transgene expression in some ears and not in others, but this did not appear to be dependent on genome composition. The *opaque2* mutation did not appear to have an epistatic impact on transgene expression for ears 9143-03 (Fig. 6B) and 9148-02 (Fig. 6D) although these ears had different genome compositions. The fluorescence values for both opaque and translucent kernels were nearly identical. Both ears had a slightly larger number of kernels with low fluorescence values than expected, which could be explained by low pollen transmission or gene silencing.

The *opaque2* mutation appeared to have an epistatic impact, in varying degrees, on the other three ears, 9143-01 (Fig. 6A), 9148-01 (Fig. 6C), and 9148-03 (Fig. 6E). Again the ears did not all have the same genome composition. Ears 9148-01 and 9148-03 had the same genome composition but did not show the same degree of epistasis. They also did not show the same segregation ratios as ear 9148-01 had a higher number of translucent kernels with low fluorescence values, again suggesting low pollen transmission or gene silencing. The epistasis in this ear only appeared to occur for kernels with intermediate fluorescence values as both opaque and translucent kernels had similar fluorescence values for the kernels with the highest fluorescence values. However, the epistasis seen in ears 9143-01 and 9148-03 was very similar with inhibition of transgene expression occurring in both the heterozygous positive and homozygous positive regions. Transgene expression is therefore impacted by the *opaque2* mutation for some ears but there is variation seen within genome compositions. The ear to ear variation seen regarding transgene expression made it difficult to draw conclusions about gene action and transgene inheritance.

5. Discussion

The *opaque2* mutation has differing degrees of severity depending on genome composition, which impacts the penetrance of the *opaque2* phenotype. It is actually this phenomenon that allows for the development of QPM maize as genome compositions with less severe *opaque2* phenotypes are chosen. Jia et al. (2007) suggested a novel hypothesis involving transcripts upregulated and downregulated in *opaque2* mutants to explain the variation seen among mutants with different genome compositions. This hypothesis states that transcripts that are downregulated are usually directly affected by *opaque2* and these are seen in multiple genome compositions. However, upregulated transcripts tend to be indirectly affected by *opaque2* and are present only in a particular genome composition. It is the upregulated transcripts that are responsible for the degree to which the kernels express the mutation. As both transgenes were tested in multiple genome compositions, varying impact on transgene expression might be expected.

The *opaque2* mutation impacted the 19zn-GFP transgene in all ears but with no apparent regard to genome composition or transformation event. It was expected that *opaque2* would impact transgene expression for ears containing the 19zn-GFP construct. This construct contained a 19kDa zein promoter and as 19kDa zein transcription is decreased in *opaque2* mutants, it was expected that transgene expression would also be decreased. The three ears from row 9145 (Fig. 5B-D) had the same genome composition and transformation event but showed variation in transgene expression for both translucent and opaque kernels. Both ear 9141-04 (Fig. 5A) and ear 9145-02 (Fig. 5C) had a high number of both opaque and translucent kernels with low fluorescence values indicating that there was low pollen transmission of the transgene or gene silencing. However, they had different genome

compositions and transgenic events. The other two ears with the same genome composition and transformation event as 9145-02, 9145-01 (Fig. 5B) and 9145-04 (Fig. 5D), appeared to have transgene segregation indicative of co-dominant gene action. Additionally, the translucent kernels with the highest fluorescence values for ear 9145-02 had significantly lower fluorescence values than the highest values for the other three ears containing this transgene. While the *opaque2* mutation impacted transgene expression in all ears tested, the variation seen for this impact within genome compositions and transformation events suggested that the specific impact of this mutation cannot be determined from these results. It also suggested that results from a single ear are not sufficient to determine how the mutation impacts transgene expression.

Based on research showing no decrease in 27kDa zeins in *opaque2* mutants (Kodrzycki et al. 1989), the ears with the 27zn-GFP construct were not expected to show a significant difference in fluorescence values between opaque and translucent kernels. However, this research showed that the *opaque2* mutation does impact 27zn-GFP transgene expression in some ears but it did so with no apparent regard for genome composition. There were two genome compositions containing the 27zn-GFP transgene with multiple ears with each composition. However, there was variation in transgene expression seen within both genome compositions. While ears 9143-01 and 9143-03 had the same genome composition, *opaque2* epistatically impacted transgene expression only in ear 9143-01. The results seen for 9143-03 most closely matched the results for ear 9148-02 which had a different genome composition. Likewise, 9143-01 had kernels with fluorescence values similar to ears 9148-01 and 9148-03 although they too had a different genome composition. Additionally, although 9148-01 and 9148-03 both showed epistasis, the epistatic impact was not the same

for the two ears. The *opaque2* mutation appeared to only impact transgene expression for kernels with intermediate fluorescence values for ear 9148-01 while it impacted kernels with both intermediate and high fluorescence values for ear 9148-03.

In addition to the different degrees of epistasis seen among ears with the same genome composition, some ears appeared to have low pollen transmission or possibly gene silencing while others did not. Ear 9143-03 had higher numbers of kernels with low fluorescence values than expected indicating low pollen transmission or gene silencing, while ear 9143-01, which had the same genome composition, did not have excess numbers of kernels with low fluorescence values. The low fluorescence values seen in opaque kernels are explained by epistasis, not low pollen transmission. A similar scenario is seen in ears 9148-01, 9148-02, and 9148-03 which all had the same genome composition. Ears 9148-01 and 9148-02 both appeared to have larger numbers of kernels with low fluorescence values than expected while ear 9148-03 did not. The *opaque2* mutation did not impact transgene expression for all ears tested and in ears where there appeared to be an epistatic impact, the degree of epistasis varied both between and within genome compositions. The occurrence of epistasis also varied within genome compositions. This wide level of variation in *opaque2* impact on transgene expression indicated that the specific impact of the mutation cannot be determined. Again this suggested that results from one ear are not sufficient to determine the mutation's impact on transgene expression.

In this study it was not possible to distinguish between recessive gene action, low pollen transmission, and gene silencing for ears that have high numbers of kernels with low fluorescence values. Further studies need to be done to determine which of these was actually occurring. Genotyping can be done to test for presence of the transgene in kernels

with low fluorescence values. If the kernels do not contain the transgene these kernels would likely be the result of low pollen transmission of the transgene. If they did contain the transgene the kernels could be germinated and then outcrossed and selfed to test for recessive gene action. If the transgene has recessive gene action, no expression should be seen in outcrossed progeny while selfed progeny should have one-forth of the kernels expressing the transgene. If this result occurred from multiple outcrossings, gene silencing could be ruled out. If the kernels with low fluorescence were the result of gene silencing the outcrossing and self experiment would not have consistent results.

6. References

Hartings H, Maddaloni M, Lazzaroni N, Di Fonzo N, Motto M, Salamini F, Thompson R. 1989. The O2 gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. EMBO J 8(10):2795-2801.

Howard JA, Hood E. 2005. Bioindustrial and biopharmaceutical products produced in plants. Advances in Agronomy 85:91-124.

Jia H, Nettleton D, Peterson JM, Vazquez-Carrillo G, Jannik J-L, Scott MP. 2007. Comparison of transcript profiles in wild-type and o2 maize endosperm in different genetic backgrounds. Crop Science 47(Supplement 1):S-45-59.

Kodrzycki R, Boston RS, Larkins BA. 1989. The opaque-2 mutation of maize differentially reduces zein gene transcription. Plant Cell 1(1):105-114.

Langridge P, Pintor-Toro JA, Feix G. 1982. Transcriptional effects of the *opaque2* mutation of Zea mays L. Planta 156:166-170.

Richards HA, Halfhill MD, Millwood RJ, Stewart CN. 2003. Quantitative GFP fluorescence as an indicator of recombinant protein synthesis in transgenic plants. Plant Cell Reports 22:117-121.

Sangtong V, Moran DL, Chikwamba R, Wang K, Woodman-Clikeman W, M.J. L, Lee M, Scott MP. 2002. Expression and Inheritance of the wheat Glu-1Dx5 gene in transgenic maize. Theor Appl Genet 105:937-945.

Schmidt RJ, Burr FA, Aukerman MJ, Burr B. 1990. Maize regulatory gene *opaque2* encodes a protein with a "leucine-zipper" motif that binds to zein DNA. Proc Natl Acad Sci 87:46-50.

Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G. 1992. *Opaque2* Is a Transcriptional Activator That Recognizes a Specific Target Site in 22-kD Zein Genes. Plant Cell 4(6):689-700.

Shepherd CT, Vignaux N, Peterson JM, Johnson LA, Scott MP. Green Fluorescent Protein as a tissue marker in transgenic maize seed. Cereal Chem:Not yet published.

7. Figures and Tables

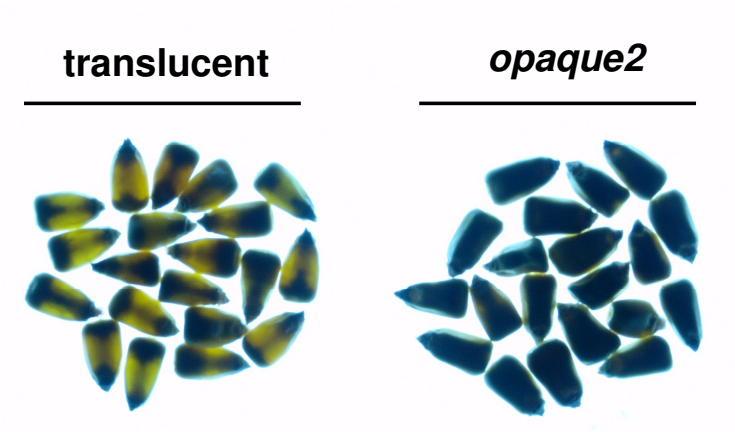

Figure 1. Maize kernels having translucent or opaque phenotype. Opaque kernels shown have *opaque2* mutation.

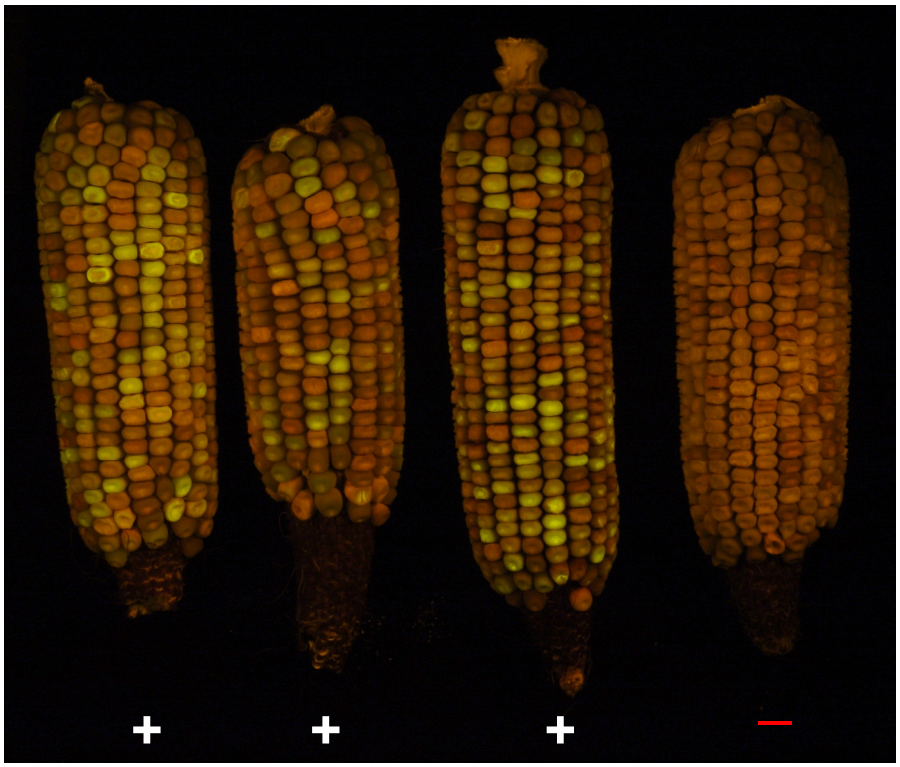

Figure 2. Maize ears with fluorescing kernels and non-fluorescing kernels. Three ears on left contain kernels that are expressing green fluorescent protein (GFP) while the ear on the far right does not have any kernels expressing GFP.

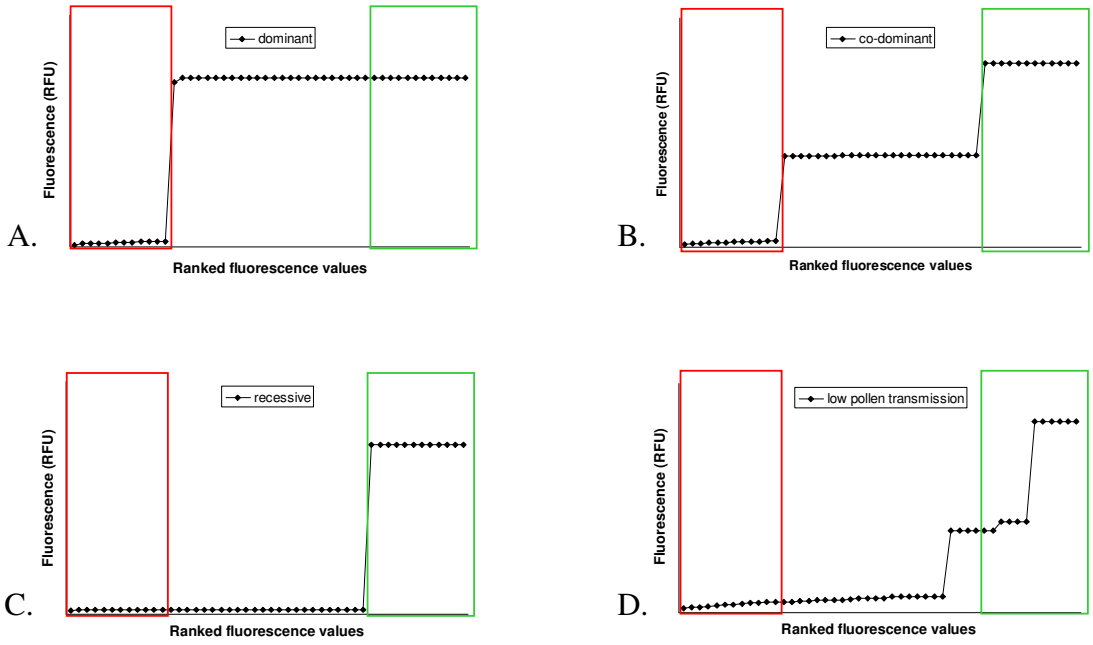

Figure 3. Hypothetical fluorescence segregation based on type of gene action. The red boxes indicate the hypothetical region for homozygous GFP negative kernels if the trait segregates normally. The green boxes indicate the hypothetical region for homozygous GFP positive kernels if the trait segregates normally. A) Dominant gene action, B) Co-dominant gene action, C) Recessive gene action, D) Result from low pollen transmission or gene silencing and co-dominant gene action.

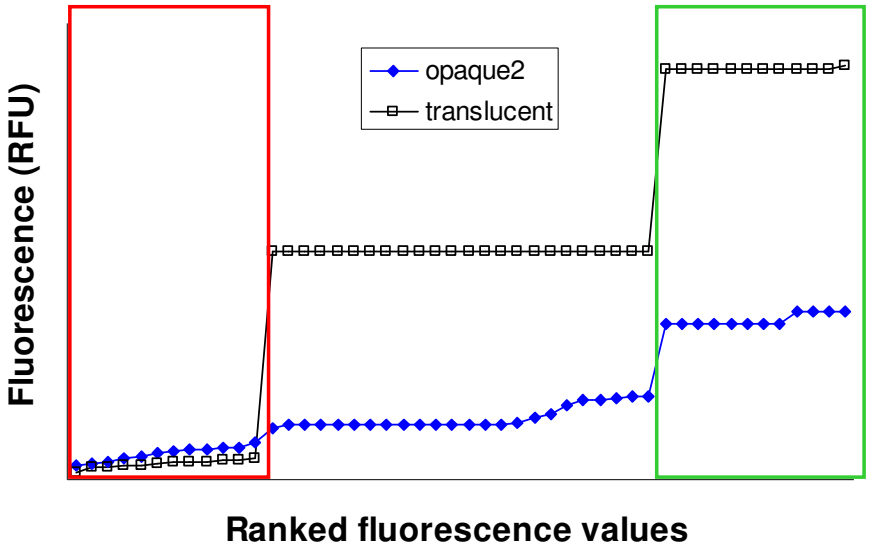

Figure 4. Hypothetical fluorescence segregation based epistasis with co-dominant gene action. The red boxes indicate the hypothetical region for homozygous GFP negative kernels if the trait segregates normally. The green boxes indicate the hypothetical region for homozygous GFP positive kernels if the trait segregates normally.

Source[a]	Construct	Event	Genome composition[b]
9141-04	19zn-GFP	29-1	50% B114A, 43.75% B73, 6.25% A188
9145-01	19zn-GFP	3-2	50% B45, 43.75% B73, 6.25% A188
9145-02	19zn-GFP	3-2	50% B45, 43.75% B73, 6.25% A188
9145-04	19zn-GFP	3-2	50% B45, 43.75% B73, 6.25% A188
9143-01	27zn-GFP	71-1	50% B45, 43.75% B73, 6.25% A188, 25% B110
9143-03	27zn-GFP	71-1	50% B45, 43.75% B73, 6.25% A188, 25% B110
9148-01	27zn-GFP	71-1	50% B14A, 43.75% B73, 6.25% A188
9148-02	27zn-GFP	71-1	50% B14A, 43.75% B73, 6.25% A188
9148-03	27zn-GFP	71-1	50% B14A, 43.75% B73, 6.25% A188

Table 1. Construct, transformation events, and genome composition for the nine ears analyzed.

[a] Source is given as row-plant number. Plants in the same row are the same genotype. All plants are self pollinated.
[b] Genome composition estimated from the pedigree of the source

Segregation ratio	Genotype
1/16	*o2o2*/GFP(-)GFP(-)
1/16	*o2o2*/GFP(+)GFP(+)
1/16	*O2O2*/GFP(-)GFP(-)
1/16	*O2O2*/GFP(+)GFP(+)
2/16	*o2o2*/GFP(-)GFP(+)
2/16	*o2O2*/GFP(-)GFP(+)
2/16	*o2O2*/GFP(-)GFP(-)
2/16	*O2O2*/GFP(-)GFP(+)
4/16	*o2O2*/GFP(-)GFP(+)

Table 2. Expected genotypic ratios for F2 seed.

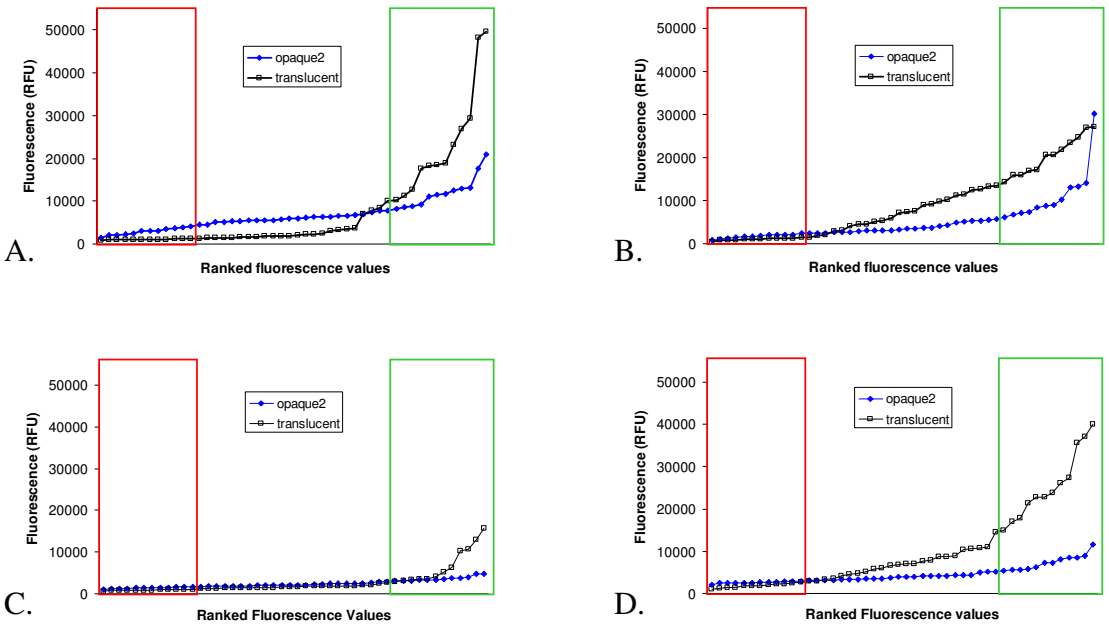

Figure 5. Fluorescence results for four ears segregating for *opaque2* and 19zn-GFP expression. The red boxes indicate the hypothetical region for homozygous GFP negative kernels if the trait segregates normally. The green boxes indicate the hypothetical region for homozygous GFP positive kernels if the trait segregates normally. A) Kernels from ear 9141-04: event 29-1, B) Kernels from ear 9145-01: event 3-2, C) Kernels from ear 9145-02: event 3-2, D) Kernels from ear 9145-04: event 3-2.

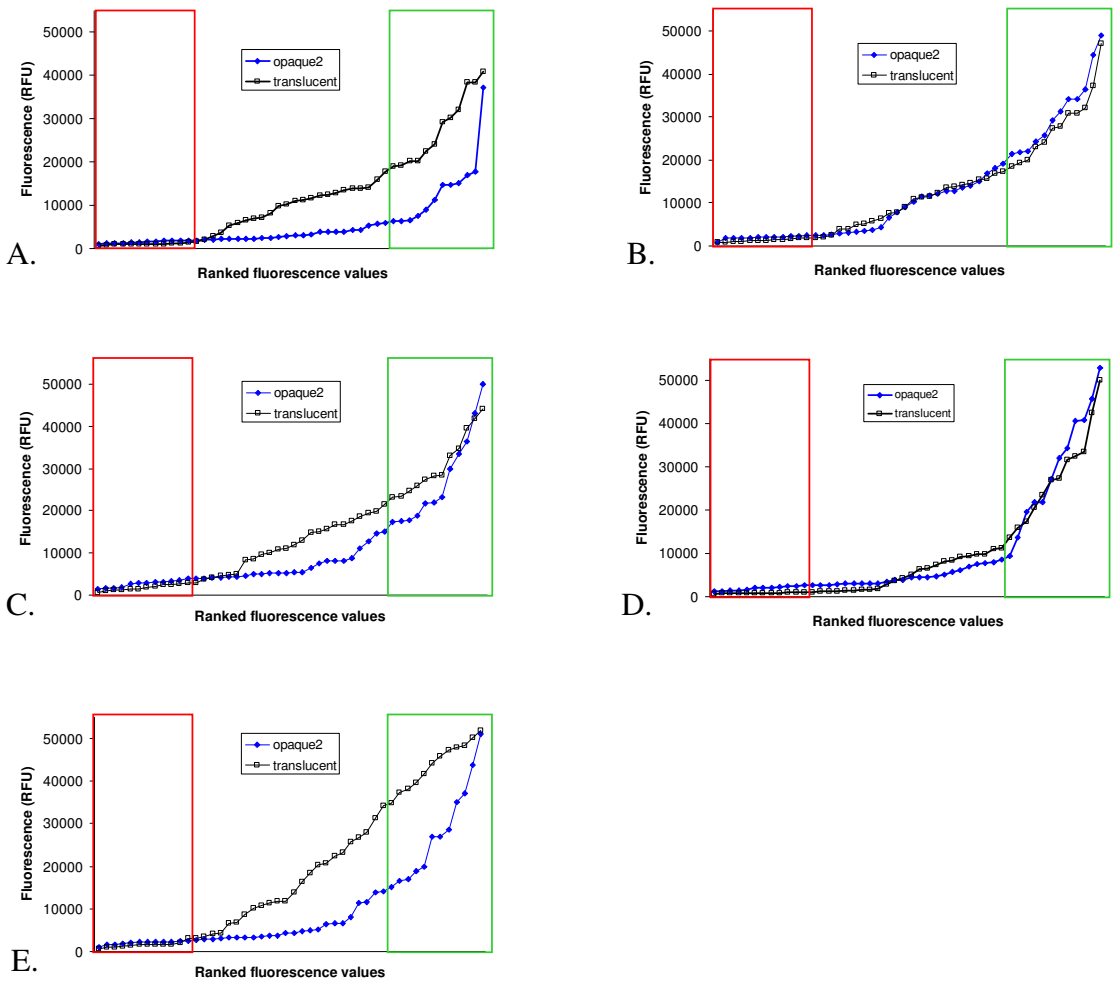

Figure 6. Fluorescence results for five ears segregating for *opaque2* and 27zn-GFP expression. The red boxes indicate the hypothetical region for homozygous GFP negative kernels if the trait segregates normally. The green boxes indicate the hypothetical region for homozygous GFP positive kernels if the trait segregates normally. A) Kernels from ear 9143-01, B) Kernels from ear 9143-03, C) Kernels from ear 9148-01, D) Kernels from ear 9148-02, E) Kernels from ear 9148-03.

CHAPTER 4: GENERAL CONCLUSIONS

1. General Conclusions

The zeins are the most prevalent protein found in maize kernels. The different classes are found in varying amounts with some classes having multiple genes contained within large gene families. The research contained in this manuscript sought to both identify new 22kDa zeins and also examine the use of promoters from different zein classes on transgene expression.

The first paper (Chapter 2) looked at a method of identifying new 22kDa zein-like sequences in B73, B101, and Ill12E. The results indicated that there are more 22kDa zein-like sequences present in the maize genome than currently published. This analysis identified several putative 22kDa zein-like sequences for all three inbreds and indicated some 22kDa zein-like sequences are found in multiple inbreds. The method used also allowed for the estimation of total number 22kDa zein-like sequences in all three inbreds. The results indicated that B73 contains at least 38 22kDa zein-like sequences, B101 contains at least 33 22kDa zein-like sequences, and Ill12E contains at least 28 zein-like sequences. A number of highly similar sequences were found indicating that greater stringency may be needed to definitively identify 22kDa zein-like sequences. The study also looked at the feasibility of using this method to identify gene candidates for *dzr1*. The low percent of published sequences found, which was also an estimate of the percentage of total zein-like sequences found, indicated that more sequences are needed before attempting to identify gene candidates for this gene.

The second study (Chapter 3), looked at the impact of the *opaque2* mutation on expression of two GFP transgenes. The transgenes contained either a 19kDa or 27kDa zein

promoter, and these two zein classes have different expression in *opaque2* kernels. The study examined whether the transgenes expressed similarly to the native protein. For the 19zn-GFP transgene, ears with different genetic backgrounds and transformation events were analyzed. All ears showed some degree of epistatic impact of *opaque2* on the transgene, although there were unexpected transgene phenotypic segregation ratios for some ears. For 27zn-GFP transgene, ears with different genetic backgrounds were analyzed. Again unexpected transgene phenotypic segregation ratios were seen for some ears. Additionally, *opaque2* was seen to have epistatic impact on several ears although this result was unexpected. These results suggest that transgene expression cannot be determined on a per ear basis and the same genetic background or transformation event does not dictate that transgene expression will be the similar.

Knowledge of the number of zeins and how they are regulated is important in maize research. This research suggested that within the 22kDa alpha zein gene family there are more members than currently published and possibly in locations not yet identified. The transgene expression study indicated that the impacted of *opaque2* on zein expression may be more complicated than previous research indicated. Further studies into both these research areas are important in understanding these important and beneficial proteins.

2. Acknowledgements

I wish to express my gratitude and appreciation to my major professor, Dr. Scott. Thank you for all your guidance and insights, in both my research and in writing this manuscript. Thank you to Dr. Moran-Lauter and Chase Lucas for sanding many of my seeds.

Also thank you to Dr. Moran-Lauter for doing many of the fluorometer readings used in Chapter 3, allowing me more time to work on this manuscript.

Thank you to the members of the Scott lab, past and present, for their insights, explanations, and friendship.

To my friends and family, thank you for your never-ending love and support with special thanks to my parents who have always encouraged and supported me in everything I have done.