Genetic basis of maize whole kernel, embryo, and endosperm oil

by

Karen Elaine Grote

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Breeding

Program of Study Committee:
Michael Lee, Major Professor
William Beavis
Allen Knapp
Daniel Nettleton
Maria Salas-Fernandez

Iowa State University

Ames, Iowa

2011

TABLE OF CONTENTS

ABSTRACT

Maize hybrids with elevated oil content have potential value as livestock feed directly and as a source of oil for human consumption. Selection on embryo and/or endosperm oil content could supplement selection at the whole kernel level and enable the development of hybrids with tissue-specific oil accumulation. The genetic basis of embryo and endosperm oil content was investigated in two separate populations: a set of elite commercial inbreds which varied for whole kernel oil content and in a segregating set of lines derived from a cross between high and low whole kernel oil parents. The traits were repeatable across environments and the impact of genotype by environment was small relative to the effect of genotype. The phenotypic data suggested that there was a relationship between embryo and endosperm oil content, but that relationship was germplasm dependent. Regions of the genome associated with the traits were detected in the inbred population using association mapping and in the segregating population with composite interval mapping. The results of the genetic mapping suggested that embryo and endosperm oil content have some common controlling loci but that the traits are under partially independent genetic control.

CHAPTER 1

INTRODUCTION

Maize hybrids with elevated oil content have potential value as livestock feed directly and as a source of oil for human consumption. Selection upon embryo and/or endosperm oil content could supplement selection at the whole kernel level and could enable the development of hybrids with tissue-specific oil accumulation. The addition of screening for embryo and/or endosperm oil content in the development of lines with elevated whole kernel content could also prevent the unintended but frequently correlated changes in embryo and endosperm size.

Multiple studies have been done in which the genetic basis of whole kernel oil concentration has been analyzed (Alrefai et al., 1995; Clark et al., 2006; Dudley, 2008; Goldman et al., 1994; Laurie et al., 2004; Mangolin et al., 2004; Song et al., 2004;Wassom et al., 2008; Yang et al., 2010; Zhang et al, 2008), but few studies have been conducted for oil content of the embryo (Beló et al., 2008; Zheng et al., 2008) and there are no known studies conducted for endosperm oil content. The repeatability of the embryo and endosperm oil content across multiple environments and the impact of interactions between genotype and environment, both of which affect the ability to identify quantitative trait loci (QTL), have not been investigated to our knowledge. Additionally, phenotypic and genotypic relationships among whole kernel, embryo, and endosperm oil contents are not well known. It is critical that such investigations be conducted since breeding strategies and selection methodologies used will be dictated by the relationships among the traits and the underlying genetic architecture.

The study of whole kernel, embryo, and endosperm oil contents in commercially relevant germplasm could generate results which are more applicable to elite material than studies of these traits in populations developed solely for increased oil content (Holland, 2004), such as the Illinois High Oil population (Dudley and Lambert, 1992)or the Beijing High Oil population (Zhang, 2008). Investigation into the phenotypic relationships and genetic basis of these traits can be approached with at least two different methodologies. Screening a broad set of germplasm would take advantage of historical genetic recombinations and sample multiple genetic backgrounds, whereas analysis within a segregating set of lines derived from a single cross between two inbreds would enable estimation of these parameters in population structure typical of a maize breeding program. Each method has unique strengths and weaknesses. Frequently, the strengths and weaknesses of the two approaches are complementary, and both methodologies are valuable when trying to understand phenotypic relationships and connect genotype and phenotype.

The following research objectives were established 1) determine phenotypic relationships among whole kernel, embryo, and endosperm oil content, 2) assess repeatability of the traits across multiple environments and the relative effects of genotype, environment, and the interaction between genotype and environment, and 3) identify loci associated with the traits by genetic mapping.

DISSERTATION ORGANIZATION

This dissertation is organized into four chapters. Chapter 1 includes the general introduction and a review of the literature. Chapters 2 and 3 are devoted to the experimental results. Chapter 2 focuses on the investigation of the genetic basis of whole kernel, embryo, and endosperm oil content in a set of 208 inbreds. This chapter presents results related to the phenotypic relationship among traits and the effect of genotype, environment, and the genotype by environment interaction on phenotype. Loci associated with the traits were identified using association analysis. Chapter 3 focuses on the investigation of the genetic basis of these traits in a segregating population derived from two inbred lines. Chapters 2 and 3 are written as journal papers and will be submitted to refereed journals for publication. Each of these chapters has its own introduction, materials and methods, results, and discussion sections including separate references at the end of each chapter. Chapter 4 presents general conclusions.

LITERATURE REVIEW

Plant Breeding for Increased Maize Kernel Oil Content

Maize inbreds and hybrids with elevated oil content have value for human consumption, livestock feed, and energy production (Lambert and Hallauer, 1994). The Illinois High Oil (IHO) (Dudley and Lambert, 1992) and Beijing High Oil (BHO) (Song and Chen, 2004) populations were developed specifically for elevated whole kernel oil and levels greater than 200 mg g^{-1} have been achieved (Dudley and Lambert, 1992). Phenotypic selection for oil content has been successful, but they are not commercially relevant because other agronomic characteristics such as yield are sufficient for commercial inbreds and hybrids (Lambert and Hallauer, 1994). Other strategies, including marker assisted selection

(Dudley and Johnson, 2009) and transgenic approaches (Shen et al., 2010; Zheng et al., 2008)could supplement standard breeding practices but require that the genetic factors which influence kernel oil content be understood. The phenotypic or genotypic screening for embryo and endosperm oil content could mitigate some of the frequently observed changes in embryo and endosperm size which can occur during selection for whole kernel oil content. Selection for embryo and/or endosperm oil content could also enable the production of inbreds and hybrids with tissue-specific oil accumulation.

Physiology and Biochemistry of Oil Accumulation

Whole kernel oil content is affected by changes in embryo and endosperm size in addition to changes in the concentration of oil within the embryo and endosperm (Curtis et al., 1968). The maize kernel consists of an embryo, an endosperm, and a pericarp. The endosperm typically comprises at least 80% of kernel mass; the embryo and the pericarp each contribute 10% or less to kernel mass (Val et al. 2008). On a whole kernel basis, about 30 to 40 milligrams per gram (mg g^{-1}) of the kernel dry weight is oil. Oil accumulates primarily in embryos and serves as a high energy food source during germination and heterotrophic growth. In most maize inbreds and hybrids, approximately 330 mg g^{-1} of the embryo's dry matter is oil (Watson, 1987), either as phospholipids or triacylglycerol (TAG) at maturity. Phospholipids are components of cellular membranes and also encapsulate TAG to form small spherical oil bodies, which are the major storage units for oil within the embryo (Tzen et al., 1993). Oil concentration in the starch-rich endosperm is typically less than 10 mg g^{-1} on a dry matter basis. Approximately half of the oil found in the endosperm is associated with starch molecules and the other half is more similar to the TAG found in the embryo and is found closer to the outer endosperm aleurone layer. At maturity, the pericarp typically

contains 8-25 mg g^{-1} oil and includes less than 2% of all of the oil in the kernel (Tan and Morrison, 1979).

Oil production and accumulation are influenced by metabolic processes within the kernel and at the whole plant level. In addition to the reactions immediately involved in fatty acid biosynthesis and modification, the major metabolic processes involved in oil accumulation include photosynthesis, sucrose transport, starch biosynthesis, the glycolytic pathway, and the oxidative pentose phosphate cycle. Several reviews of general and maize-specific lipid biosynthesis (Ohlrogge and Browse, 1995; Val et al., 2008) are available; a brief overview of the process is provided here.

Since kernels are not photosynthetically active, carbon fixed in other parts of the plant must be transported to the developing kernel for use as the substrate and energy source for oil production. Sucrose is the primary carbon-transport molecule and is imported into the endosperm before it goes to the embryo (Hills, 2004). The sucrose is broken down though a series of reactions into acetyl CoA which is used for de novo fatty acid biosynthesis. Several of the steps occur solely in the cytosol, but several – including the reactions specific to fatty acid biosynthesis – are localized to the plastid.

Within the plastid, the first step in fatty acid synthesis is catalyzed by acetyl-CoA carboxylase (ACCase). ACCase catalyzes the carboxylation of acetyl-CoA to form malonyl-CoA which serves as the substrate for fatty acid synthase (FAS). FAS is responsible for condensation of additional acetyl-CoA molecules to malonyl-CoA to form new 16- or 18-carbon fatty acids. The reducing equivalents for that process are generated by glycolysis and the oxidative pentose phosphate pathway in the plastids (Val et al., 2008).

From the plastid, fatty acids are transported to the endoplasmic reticulum (ER) where they are processed to form TAG. The final step of TAG biosynthesis is catalyzed by diacylglycerol acyltransferase (DGAT), an important regulator of oil content (Zheng et al., 2008). Diacylglycerol – the substrate of DGAT – is the branch point between TAG and phospholipid biosynthesis. Also, DGAT may be the rate -limiting step in the formation of TAG (Lung and Weselake, 2006). After synthesis, TAGs are localized to the ER membrane from where oil bodies, consisting of a TAG core, a layer of phospholipids, and an outer layer of oleosin proteins, are formed (Hsieh and Huang, 2004).

Genetic Components of Oil Content in the Maize Kernel

Maize kernel oil is typically considered a quantitative trait affected by a large number of loci (Dudley and Lambert, 1992), environmental effects and interactions but most of the phenotypic variation is due to genetic sources (Clark et al., 2006; Goldman et al., 1994; Laurie et al., 2004). Multiple studies have been conducted to identify regions of the genome associated with increased whole kernel oil concentration in maize and to assess the additive or dominant action of these regions. Among the studies that have been published, there are differences in populations analyzed which will influence the QTL detected and estimated genetic effects. Several studies have focused on the IHO and Illinois Low Oil (ILO) populations (Clark et al., 2006; Laurie et al., 2004; Wassom et al., 2008b). Other experiments have used the Beijing High Oil population (Song et al., 2004; Yang et al., 2010; Zhang et al., 2008), tropical germplasm (Mangolin et al., 2004) or commercial inbreds (Beló et al., 2008). Differences in inbred or hybrid progeny and in the use of self or open pollinated ears of maize can also affect the results of QTL analysis. Whole kernel oil concentration on a dry matter basis was the analyzed phenotype in most studies, but several analyzed whole

kernel concentration of specific fatty acids. Among the studies, all forms of gene action were observed and common and unique QTL regions were identified.

Gene Action and Heritability

All forms of gene action for whole kernel oil concentration have been reported both in inbred and testcross progeny, but the most frequently reported mode is partial dominance (Clark et al., 2006; Mangolin et al., 2004; Moreno-Gonzalez et al., 1975; Zhang et al., 2008). Other estimates of genetic effects in relation to whole kernel oil concentration have indicated that genes function additively (Alrefai et al., 1995; Silvela et al., 1989; Wang et al., 2009). Overdominance was reported in the IHO x ILO population (Alrefai et al., 1995) and in a tropical maize population (Mangolin et al., 2004). All forms of gene action contribute to repeatability and broad-sense heritability; additive genetic effects are the primary contributors to realized heritability.

Broad-sense heritability for whole kernel oil concentration is typically high both in inbreds (Clark et al., 2006; Goldman et al., 1994; Laurie et al., 2004; Mangolin et al., 2004; Wang et al., 2009) and in test crosses (Clark et al., 2006; Laurie et al., 2004) and ranged from 0.80 to 0.99. Realized heritability estimates were considerably lower, ranging from approximately 0.30 (Song and Chen, 2004; Wang et al., 2009) to almost 0.70 (Song and Chen, 2004). Broad-sense heritability is a measure of the genetic variability relative to the phenotypic variability, can be estimated in multi-environment replicated trials (Bernardo, 2002), and is equivalent to repeatability or the intra-class correlation (Kempthorne, 1969). In comparison, realized heritability is measured as a response to selection and will be lower than broad sense heritability because it is a measure of additive genetic variance in relation to the phenotypic variance.

High broad-sense heritability across environments indicates that the effect of the genotype on phenotypic variability is large relative to any environment or genotype by environment effects, and that the phenotype is predominantly determined by genotype. Several studies have found that genotype by environment interaction may be significant for whole kernel oil concentration, but its contribution to the total variance was small relative to genetic variance (Clark et al., 2006; Laurie et al., 2004), or it was due to changes in magnitude rather than changes in rank (Goldman et al., 1994). Overall, whole kernel oil concentration is a highly heritable trait which is a desirable characteristic for QTL mapping. An additional characteristic of whole kernel oil content which are favorable for QTL mapping include the ability to phenotype in a high-throughput manner

QTL Analysis and Putative Genes

Differences in germplasm, population size and structure, progeny type, phenotyping, and QTL mapping methodology can contribute to differences in the number, position, and estimated effect of QTL detected within single analyses. The majority of studies conducted for QTL mapping of oil content in maize kernels have focused on whole kernel total oil content (Alrefai et al., 1995; Clark et al., 2006; Laurie et al., 2004; Song et al., 2004, Mangolin, 2004; Wassom et al., 2008b; Yang et al., 2010; Zhang et al., 2008), but several have analyzed the content of specific types of fatty acids, such as oleic or linoleic (Alrefai et al., 1995; Wassom et al., 2008a; Willmot et al., 2006; Zheng et al., 2008). These differences between the populations, progeny, and phenotypes of the QTL mapping experiments can make direct comparisons among studies difficult. The QTL identified for embryo or endosperm oil content are not be directly comparable to the published QTL for whole kernel oil content because of the differences in the phenotypes.

Large effect loci have been associated with embryo oil content and putative causal genes have been identified. In an association mapping analysis of 553 inbred lines, a single QTL for oleic acid content was identified on chromosome 4 (Beló et al., 2008) A nearby fatty acid desaturase, *fad2*, maps within 2 kilobases to the marker associated with elevated oleic acid and is the probable causative gene. A large-effect QTL identified in a segregating population derived from a bi-parental cross was found to encode an acyl-CoA:diacylglycerol acyltransferase (DGAT), which catalyzes the final step in TAG biosynthesis (Zheng et al., 2008). Ectopic expression of the allele resulted in an almost 50% increase in whole kernel oil concentration.

In spite of the difficulty in identifying the genes underlying QTL for most QTL associated with maize kernel oil, genetic mapping is a valuable tool in plant breeding both for the dissection of complex traits and in the direct application to breeding populations (Bernardo, 2008).

Mapping Quantitative Trait Loci

Linkage and Association Mapping as Complementary Methods of QTL Detection

Linkage and association analysis are the most frequently used methods of QTL mapping in plant populations. Each methodology contains unique sets of strengths and weaknesses which will be described in the following sections. Frequently, the strengths and weaknesses of the two approaches are complementary, and both methodologies are valuable when trying to connect genotype and phenotype.

Bi-parental Mapping Populations and Linkage Mapping

In bi-parental QTL mapping populations, two inbred lines that differ for the trait are crossed to form the F1 generation. Progeny for mapping are produced by self-pollination for

one to several generations or through the doubled haploid program to obtain a population segregating for the phenotype of interest (Lander and Botstein, 1989; Soller et al., 1976). Any loci that were heterozygous in the F1 generation would be expected to segregate. After genotyping and phenotyping, statistical approaches may detect associations between markers and the trait. Single-marker analysis (Soller et al., 1976), interval mapping (Lander and Botstein, 1989), and composite interval mapping (Jansen and Stam, 1994; Zeng, 1994) have been common procedures used for assessing marker-trait associations (Doerge, 2002) in bi-parental populations.

The design and analysis of bi-parental populations for QTL detection is well established (Bernardo, 2008); however, this method of linkage analysis does have several weaknesses and limitations in addition to the strengths that contributed to its popularity. Allelic diversity is restricted to the inbreds used to create the segregating population, and the applicability of a trait-associated marker identified in one population to QTL in others is limited (Holland, 2007) because trait-associated regions identified tend to be population specific (Lubberstedt et al., 1998). The resolution of linkage mapping is limited by the number of recombination events that occur during the development of the segregating population (Zhu et al., 2008). Typically, QTL are localized to a 10 to 20 cM region of the genetic map (Zhu et al., 2008) and the marker may not be in close proximity to the causative allele (Kearsey and Farquhar, 1998). In a simulation study, Hyne et al. (1995) found that the confidence interval around the location of a putative QTL was approximately 35 cM, or almost ¼ the genetic length of the simulated chromosome (Hyne et al., 1995). Several of the strengths of linkage analysis are that a relatively small number of markers are needed to

cover the genome and trait-associated markers can be directly used for selection within the mapping population (Lande and Thompson, 1990).

Broad-based Mapping Populations and Association Mapping

Linkage disequilibrium (LD), or association, mapping exploits linkage disequilibrium to analyze the relationship between molecular markers and phenotypic variation and has several advantages over linkage mapping. Unlike linkage mapping, which requires the development of a segregating bi-parental population, association mapping can utilize a genetically broader set of germplasm and takes advantage of the previous recombination that occurred during germplasm development (Holland, 2007). Mapping resolution may be improved (Yu and Buckler, 2006), and the statistical power to detect QTL is higher (Yu et al., 2006). Theoretically, trait-associated markers identified in an association mapping population can be more easily applied to pedigrees outside of the original set than those identified in bi-parental mapping populations (Zhu et al., 2008). The success application of association analysis to QTL mapping is related to several population characteristics including linkage disequilibrium, population structure, and pedigree relationships within the population.

Population LD level is the non-random association between alleles at different loci; it can be affected by biological factors such as physical linkage on a chromosome and recombination but is also influenced by population history such as population subdivision, admixture, and mating system (Gaut and Long, 2003). LD is expected to be low in outcrossing species such as maize and in diverse maize germplasm, intragenic LD declines within 100-200 base pairs (Tenaillon et al.). In elite lines, LD does not decline as rapidly as in diverse germplasm. Several studies have shown that genome-wide LD can be substantial

in elite germplasm used in breeding programs (Remington et al., 2001) and can extend over 500 kilobases in maize (Ching et al., 2002; Jung et al., 2004).

Elevated LD in elite maize inbred lines could be due to processes such as selection, inbreeding, and population bottlenecks (Flint-Garcia et al., 2003). Although high levels of LD in elite germplasm can reduce resolution, there are still advantages to association mapping in elite germplasm; elite lines are preferable for assessing low heritability agronomic traits such as yield, and favorable alleles are detected in the target population (Breseghello and Sorrells, 2006). Some of the forces that affect LD in elite maize inbreds can also contribute to population structure.

Association mapping in populations with low LD requires a dense molecular marker map in order to have LD between a potential QTL and a genotyped molecular marker. Populations with high LD require a less dense marker map than does mapping in a population with low LD, but resolution will be lower due to extensive LD and it is not possible to detect and isolate multiple QTL that are in LD with one another.

Population structure is the presence of subpopulations in which members are more closely related to one another than to an average pair of individuals taken at random from the population (Breseghello and Sorrells, 2006). Population structure is expected in plant breeding programs because of selection for specific breeding goals and extensive pedigree relationships due to line recycling. Association mapping does not require the development of a bi-parental segregating population; this saves time and resources and allows for the use of lines developed within any plant breeding program but does induce population structure because relationships among the lines are not equal. Such structure can result in spurious associations due to unequal allele frequencies among sub-populations; thus it is very

important that population structure is accounted for in association mapping for QTL detection. Population structure can be accounted for using several methods, including pedigree records, STRUCTURE analysis (Pritchard et al., 2000), and principal components (Price et al., 2006)

Family relationships occur when the phenotypes of two genetically similar individuals within a subpopulation tend to be more highly correlated than the phenotypes of two genetically dissimilar individuals. Accounting only for population structure when family relationships are prevalent can result in an elevated risk of false positives or a loss of statistical power (Yu et al., 2006). Relative kinship is a measure of family relatedness and can be estimated by background molecular markers (Lynch and Ritland, 1999). The use of kinship information has been shown to increase accuracy (George et al., 2000), increase statistical power, and reduce false positives (Yu et al., 2006; Zhao et al., 2007) of quantitative trait analyses when family relationships are found within subpopulations. In plant breeding, population structure is often compounded by complex pedigree relationships within each subpopulation; thus inclusion of kinship can decrease the frequency of false positive associations and increase power.

General Hypotheses

Multiple studies have been done in the public sector in which the genetic basis of whole kernel oil concentration has been analyzed and QTLs have been identified. Whole kernel oil content is well-suited for QTL analysis due to it high repeatability and the relatively low contribution of genotype by environment interaction to overall phenotypic variance. Although these studies have produced a large number of QTLs, their application to commercial maize breeding programs is limited. Additionally, very few studies have been

conducted for separate oil contents of the embryo or endosperm tissues. The ability to screen for embryo and endosperm oil content, and the feasibility of producing hybrids with tissue-specific oil accumulation, warrants further investigation of these traits.

General hypotheses include:

1) Selection for embryo and endosperm oil content could supplement whole kernel oil content in the development of high oil maize inbreds
2) Oil contents of the embryo and the endosperm are genetically controlled and are repeatable across environments
3) Oil contents of the embryo and endosperm tissues are under independent genetic control and different genomic regions will contribute to oil contents of the separate tissues

REFERENCES

Alrefai R., Berke T.G., Rocheford T.R. (1995) Quantitative trait locus analysis of fatty acid concentrations in maize. Genome 38:894-901.

Beló A., Zheng P., Luck S., Shen B., Meyer D.J., Li B., Tingey S., Rafalski A. (2008) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Molecular Genetics and Genomics 279:1-10.

Bernardo R. (2002) Breeding for quantitative traits in plants Woodbury, Minn.(EUA). Stemma Press.

Bernardo R. (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Science 48:1649-1664.

Breseghello F., Sorrells M.E. (2006) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Science 46:1323-1330.

Ching A., Caldwell K., Jung M., Dolan M., Smith O., Tingey S., Morgante M., Rafalski A. (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:19-33.

Clark D., Dudley J., Rocheford T.R., LeDeaux J. (2006) Genetic analysis of corn kernel chemical composition in the random mated 10 generation of the cross of generations 70 of IHO x ILO. Crop Science 46:807-819.

Curtis P.E., Leng E.R., Hageman R.H. (1968) Development changes in oil and fatty acid content of maize strains varying in oil content. Crop Science 8:689-693.

Doerge R.W. (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nature Review Genetics 3:43-52.

Dudley J., Lambert R.J. (1992) Ninety generations for selection for oil and protein in maize. Maydica 37:1-7.

Dudley J.W., Johnson G.R. (2009) Epistatic models improve prediction of performance in corn. Crop Science 49:763-770.

Flint-Garcia S.A., Thornsberry J.M., S E., Iv B. (2003) Structure of linkage disequilibrium in plants. Annual review of plant biology 54:357-374.

Gaut B.S., Long A.D. (2003) The lowdown on linkage disequilibrium, Am Soc Plant Biol. pp. 1502-1506.

George A.W., Visscher P.M., Haley C.S. (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156:2081-2092.

Goldman I.L., Rocheford T.R., Dudley J.W. (1994) Molecular markers associated with maize kernel oil concentration in an Illinois High Protein x Illinois Low Protein cross. Crop Science 34:908-915.

Hills M.J. (2004) Control of storage-product synthesis in seeds. Current Opinion in Plant Biology 7:302-308.

Holland J.B. (2007) Genetic architecture of complex traits in plants. Current Opinion in Plant Biology 10:156-161.

Hsieh K., Huang A.H.C. (2004) Endoplasmic reticulum, oleosins, and oils in seeds and tapetum cells. Plant Physiology 136:3427-3434.

Hyne V., Kearsey M.J., Pike D.J., Snape J.W. (1995) QTL analysis: unreliability and bias in estimation procedures. Molecular Breeding 1:273-282.

Jansen R.C., Stam P. (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455.

Jung M., Ching A., Bhattramakki D., Dolan M., Tingey S., Morgante M., Rafalski A. (2004) Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 locus in elite maize germplasm. Theoretical and Applied Genetics 109:681-689.

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

Kempthorne O. (1969) An introduction to statistical genetics, Ames, Iowa: The Iowa State University Press.

Lambert R., Hallauer A. (1994) High-oil corn hybrids. Specialty corns.:123-145.

Lande R., Thompson R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

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

Laurie C.C., Chasalow S.D., LeDeaux J.R., McCarroll R., Bush D., Hauge B., Lai C., Clark D., Rocheford T.R., Dudley J.W. (2004) The Genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141-2155.

Lubberstedt T., Melchinger A.E., Fahr S., Klein D., Dally A., Westhoff P. (1998) QTL mapping in testcrosses of flint lines of maize: III. comparison across populations for forage traits. Crop Science 38:1278-1289.

Lung S.C., Weselake R.J. (2006) Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. Lipids 41:1073-1088.

Lynch M., Ritland K. (1999) Estimation of pairwise relatedness with molecular markers. Genetics 152:1753-1766.

Mangolin C.A., de Souza C.L., Garcia A.A.F., Garcia A.F., Sibov S.T., de Souza A.P. (2004) Mapping QTLs for kernel oil content in a tropical maize population. Euphytica 137:251-259.

Moreno-Gonzalez J., Dudley J.W., Lambert R.J. (1975) A design III study of linkage disequilibrium for percent oil in maize. Crop Science 15:840-843.

Ohlrogge J., Browse J. (1995) Lipid biosynthesis. The Plant Cell 7:957-970.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A., Reich D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904-909.

Pritchard J.K., Stephens M., Donnelly P. (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., Buckler E.S. (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences 98:11479-11484.

Shen B., Allen W.B., Zheng P., Li C., Glassman K., Ranch J., Nubel D., Tarczynski M.C. (2010) Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. Plant Physiology 153:980-987.

Silvela L., Rodgers R., Barrera A., Alexander D.E. (1989) Effect of selection intensity and population size on percent oil in maize, Zea mays L. TAG Theoretical and Applied Genetics 78:298-304.

Soller M., Brody T., Genizi A. (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. TAG Theoretical and Applied Genetics 47:35-39.

Song T.M., Chen S.J. (2004) Long term selection for oil concentration in five maize populations. Maydica 49:9-14.

Song X.F., Song T.M., Dai J.R., Rocheford T., Li J.S. (2004) QTL mapping of kernel oil concentration with high-oil maize by SSR markers. Maydica 49:41-48.

Tan S.L., Morrison W.R. (1979) The distribution of lipids in the germ, endosperm, pericarp and tip cap of amylomaize, LG-11 hybrid maize and waxy maize. Journal of the American Oil Chemists' Society 56:531-535.

Tenaillon M.I., Sawkins M.C., Long A.D., Gaut R.L., Doebley J.F., Gaut B.S. (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proceedings of the National Academy of Sciences 98:9161-9166.

Tzen J.T.C., Cao Y.Z., Laurent P., Ratnayake C., Huang A.H.C. (1993) Lipids, proteins, and structure of seed oil bodies from diverse species. Plant Physiology 101:267-276.

Val D.L., Schwartz S.H., Kerns M.R., Deikman J. (2008) Development of a high oil trait for maize, in: A. L. Kriz and B. A. Larkins (Eds.), Molecular Genetic Approaches to Maize Improvement, Springer-Verlag, Berlin. pp. 303-323.

Wang H.-W., Wu B.-Y., Song T.-M., Chen S.-J. (2009) Effects of long-term selection for kernel oil concentration in KYHO, a high-oil maize population. Crop Science 49:459-466.

Wassom J.J., Mikkelineni V., Bohn M.O., Rocheford T.R. (2008a) QTL for fatty acid composition of maize kernel oil in Illinois High Oil x B73 backcross-derived lines. Crop Science 48:69-78.

Wassom J.J., Wong J.C., Martinez E., King J.J., DeBaene J., Hotchkiss J.R., Mikkilineni V., Bohn M.O., Rocheford T.R. (2008b) QTL associated with maize kernel oil, protein, and starch concentrations; kernel mass; and grain yield in Illinois High Oil x B73 backcross-derived lines. Crop Science 48:243-252.

Watson S. (1987) Structure and composition, in: S. Watson and P. Ramstad (Eds.), Corn Chemistry and Technology, American Association of Cereal Chemists, Inc., St. Paul, MN. pp. 53-82.

Willmot D.B., Dudley J., Rocheford T.R., Bari A. (2006) Effect of random mating on marker-QTL associations for grain quality traits in the cross of Illinois High Oil x Illinois Low Oil. Maydica 51:187-199.

Yang X., Guo Y., Yan J., Zhang J., Song T., Rocheford T., Li J.S. (2010) Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. Theoretical and Applied Genetics120:665-678.

Yu J., Buckler E.S. (2006) Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology 17:155-160.

Yu J., Pressoir G., Briggs W.H., Bi I.V., Yamasaki M., Doebley J.F., McMullen M.D., Gaut B.S., Nielsen D.M., Holland J.B. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203-208.

Zeng Z.B. (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

Zhang J., X.Q. L., X.F. S., Yan J.B., Song T.M., Dai J.R., Rocheford T.R., Li J.S. (2008) Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. Euphytica 162:335-344.

Zhao K., Aranzana M.J., Kim S., Lister C., Shindo C., Tang C., Toomajian C., Zheng H., Dean C., Marjoram P. (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:71-82.

Zheng P., Allen W.B., Roesler K., Williams M.E., Zhang S., Li J., Glassman K., Ranch J., Nubel D., Solawetz W., Bhattramakki D., Llaca V., Deschamps S., Zhong G.-Y., Tarczynski M.C., Shen B. (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nature Genetics 40:367-372.

Zhu C., Gore M., Buckler E.S., Yu J. (2008) Status and prospects of association mapping in plants. The Plant Genome 1:5-20.

CHAPTER 2: ASSOCIATION MAPPING AND ANALYSIS OF QUANTITATIVE TRAIT LOCI AFFECTING MAIZE WHOLE KERNEL, EMBRYO, AND ENDOSPERM OIL CONTENT

A paper to be submitted to Theoretical and Applied Genetics

Karen Grote, Maria Salas-Fernandez, and Michael Lee

ABSTRACT

High oil maize have value for human consumption, livestock feed, and energy production. Maize breeding for increased oil content has focused on increasing concentration of oil in the whole kernel, which is affected by embryo and endosperm oil content and size. Two hundred eight commercially relevant inbred lines were analyzed per se across multiple environments for seven traits: whole kernel oil in milligrams of oil per gram of kernel (WK), milligrams of oil per gram of embryo (EMB), milligrams of oil per embryo (EMB_MG), milligrams of embryo tissue (EMB_WT), milligrams of oil per gram of endosperm (ENDO), milligrams of oil per endosperm (ENDO_MG), and milligrams of endosperm tissue (ENDO_WT). All traits had broad-sense heritability estimates over 0.75 and the effect of genotype was at least 10 times greater than the effect of environment or the genotype by environment interaction. Correlations between pairs of traits ranged from -0.76 to 0.85. Association mapping identified 173 associations between 129 quantitative trait loci (QTL) and at least one of six traits using the critical threshold of False Discovery Rate (FDR)<0.05. No QTL were shared between embryo and endosperm oil contents. Several of the QTL for which FDR<0.001were located in putative genes which had potential or known function in oil metabolism.

INTRODUCTION

Maize hybrids with elevated oil content have value for human consumption, livestock feed, and energy production (Lambert, 2001). Whole kernel oil in maize can exceed 200 mg g^{-1} with selection (Dudley and Lambert, 1992) and many quantitative trait loci (QTL) affecting the trait have been identified (Laurie et al., 2004; Mangolin et al., 2004; Song et al., 2004; Wassom et al., 2008b; Willmot et al., 2006; Yang et al., 2010; Zhang et al., 2008). These studies have focused on whole kernel oil concentration which is determined by the amount of oil in the kernel divided by total weight. Whole kernel oil concentration is affected by oil concentrations within the embryo and endosperm as well as the size of the embryo and endosperm. Maize embryo and endosperm oil content could be used to supplement selection on a whole kernel basis, but oil contents of the specific tissues have received relatively little attention.

To our knowledge, few investigations and genetic mapping of factors associated with oil contents of the endosperm and embryo, assessed as separate tissues, have been published, and heritability and the genetic basis of the traits is currently unknown. There is no evidence that the oil contents of embryo and endosperm are genetically controlled, or that the oil contents of the separate tissues are under common or independent genetic control. The genetic basis of these traits will influence breeding strategies and development of inbreds and hybrids with tissue-specific oil accumulation may require partially independent genetic control. Further investigation into embryo and endosperm oil contents is warranted to address the basic biological questions related to the genetic basis of the traits as well as to guide plant breeding strategies for the traits.

Previous research has identified QTL with large effects on oil content of the embryo (Belo et al., 2008; Zheng et al. 2008). Analysis of bi-parental backcross-derived lines identified an allele of *acyl-CoA:diacylglycerol acyltransferase* (*DGAT*) which increased embryo oil concentration by approximately 18% (Zheng et al. 2008). Studies which have analyzed whole kernel oil content have typically conducted genetic mapping with bi-parental populations (Alrefai et al., 1995; Clark et al., 2006; Curtis et al., 1968; Dudley and Lambert, 1992; Wassom et al., 2008b; Willmot et al., 2006; Yang et al., 2010; Zhang et al., 2008). The effects of QTL detected in such populations tend to be overestimated and inconsistent across germplasm groups (Melchinger et al., 1998), thus, limiting the utility of loci detected using these methods.

Association mapping, which utilizes linkage disequilibrium (LD), has several advantages over linkage analysis in segregating populations derived from crosses between inbred lines and has rapidly been adopted in many plant species for QTL mapping (Rafalski, 2010). Since association mapping is not restricted to the segregating offspring of a single cross, it is possible to include a larger and more diverse sample of germplasm. Therefore, QTL should be applicable to a wide range of germplasm (Buckler et al., 2009; Zhu et al., 2008). The use of elite germplasm which represents the diversity within a breeding population for association mapping enables detection of QTL which can directly be utilized in marker assisted selection during subsequent breeding cycles (Breseghello and Sorrells, 2006). Several other advantages of association analysis include potential increases in mapping resolution and reduced research time due to the ability to sample from existing germplasm resources (Yu et al., 2006). Association mapping does have several weaknesses including sensitivity to population structure (Pritchard et al., 2000b; Yu et al., 2006), reduced

power for detection of rare alleles (Visscher et al., 2008), and variable resolution and power for QTL detection dependent upon linkage disequilibrium and marker density (Breseghello and Sorrells, 2006).

Whole genome association mapping was used to identify a fatty acid desaturase, *fad2*, on chromosome 4 which affected oleic acid concentration in embryo tissue (Beló et al., 2008). In maize, candidate gene association mapping was used to identify or confirm loci associated with traits such as flowering time (Thornsberry et al., 2001), kernel composition (Wilson et al., 2004), carotenoid content (Yan et al., 2010), and aluminum tolerance(Krill et al., 2010). Candidate gene association mapping has been utilized more extensively than whole genome association mapping, possibly due to rapid LD breakdown in diverse germplasm (Yan et al., 2011).

Breeding strategies for development of inbreds and hybrids with tissue-specific oil accumulation will be influenced by the underlying genetic architecture of the traits. Genetic mapping is a tool that can be used to understand the genetic basis of traits and enhance the understanding of phenotypic stability and trait correlations. The first objective of this study was to assess the broad- sense heritability and determine phenotypic relationships among whole kernel, embryo, and endosperm fractions in 208 commercially relevant inbred lines. The second objective was to identify QTL associated with oil content of the whole kernel, embryo, and endosperm via association mapping. The results of the QTL analysis will be analyzed to reveal the genetic relationship among the traits and to investigate the physiological basis of oil accumulation.

MATERIALS AND METHODS

Genetic Materials

Two hundred eight maize inbred lines representing the genotypic and phenotypic diversity available within the Monsanto High Oil Maize Breeding program were evaluated to determine broad-sense heritabilities, phenotypic correlations, and genetic factors association with oil content in the whole kernel, embryo, and endosperm. All 208 inbreds were adapted to the central US cornbelt. One hundred sixty four inbreds were produced by a breeding program that had a goal of combining elevated whole- kernel oil concentration and superior agronomic traits such as yield and disease tolerance (high oil inbreds, abbreviate O). The other 44 were selected for agronomic traits only (conventional inbreds, abbreviated C). The inbreds could further be divided into groups of males (abbreviated M)or females (abbreviated F)in accordance with their use in the production of hybrid seed and in the breeding program. The selection history and utility in the breeding programs were used to develop four sub-groups: female inbreds selected for high oil content and agronomic traits (oil females or OF), male inbreds selected for high oil and agronomic traits (oil males or OM), female inbreds selected for agronomic traits (conventional females or CF), and male inbreds selected for agronomic traits (conventional males or CM).

Genotypic Data and Analyses

Fifteen kernels of each inbred were ground to powder and DNA was extracted using methodology similar to the standard Qiagen protocol (Csaikl et al., 1998). The inbreds were genotyped at 1311 single nucleotide polymorphism (SNP) markers using Illumina GoldenGate technology. Genotyping was conducted by Monsanto in St. Louis, MO. DNA was bound to the BeadArray^{TM} chip, chips were analyzed using the BeadArray Reader, and

genotype scoring was conducted using Illumina software (Fan et al., 2003). The loci included in this set had less than 10% missing data. At each locus, no allele had frequency lower than 5% within this set of 208 inbreds. SNP primers and probes, genetic and physical positions were developed within Monsanto. Physical and genetic position were established within Monsanto. The primer and probe sequences for the genotyping assays are available at the request of K. E. Grote at Monsanto. Access requires agreement that the sequences will be used only for noncommercial research and will not be transferred to a third party.

The polymorphism information content (PIC) and frequency of missing data were calculated for each locus for all inbred lines as well as the sub groups of OF, OM, CF, and CM. PIC was calculated as:

$$PIC = 1 - \sum(p_i^2),$$

where p_i is the frequency of the i^{th} allele. The PIC value is equivalent to the expected heterozygosity and is a measure of the frequency of alleles within the population and is related to the probability that two individuals selected at randoml from the population will have different alleles at a locus (Halliburton, 2004).

A similarity matrix and cladogram were created from all SNP loci using the neighbor-joining algorithm (Saitou and Nei, 1987)and the simple parsimony substitution model (Bradbury et al., 2007). The cladogram was reformatted in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Principal components (PC) were calculated from a randomly selected subset of 500 SNPs using the covariance method of Principal Component Analysis (PCA). For the 500 loci used for PC calculation, missing data were imputed using the k-nearest neighbor algorithm using k=3 (Cover and Hart, 1967). Relative kinship was calculated from the same set of 500 randomly selected loci without imputation of

missing data. For all pairs of loci located on a single chromosome, LD was calculated as r^2. The similarity matrix, PCs, relative kinship matrix, and r^2 were estimated in TASSEL (Bradbury et al., 2007).

Phenotypic Data and Analysis

Two replications of the 208 inbreds were grown single- row plots in 2008 in Huxley, IA and 2009 in Huxley, IA and Williamsburg, IA. In Puerto Vallarta (Mexico), Kunia (Hawaii), and Rangacua (Chile) in the 2009-2010 season, inbreds were evaluated in unreplicated single row plots due to the higher cost associated with growing plots at these locations. Inbreds were blocked by replication and completely randomized within each block. Plants were approximately 15 centimeters apart within a row and the rows were separated by 0.76 meters. At least 3 plants were self- pollinated in each row. The self-pollinated ears were harvested at maturity and dried to 10% moisture.

From each field plot, a balanced bulk of 24 kernels was created from self-pollinated ears to form the samples for kernel dissection and embryo and endosperm analysis. Kernels were imbibed in water to facilitate removal of the pericarp and separation of embryo and endosperm. Following separation, the embryo and endosperm were air-dried overnight at room temperature, weighed and analyzed for oil percentage via nuclear magnetic resonance (Alexander et al., 1967) using a Maran Ultra-20 (Oxford Instruments) at Monsanto in St. Louis, MO. A separate balanced bulk of 200-300 kernels from the same field plot was used for analysis of whole kernel oil. Whole kernel oil percentage was measured by near infrared transmittance (NIT) (Orman and Schumann, 1992) with a Foss Infratec 1221 NIT Grain Analyzer instrument at Monsanto in Ankeny, IA. Both NIT and NMR were calibrated using solvent-extracted samples.

Seven traits were measured on each sample: whole kernel oil in milligrams of oil per gram of kernel (WK), milligrams of oil per gram of embryo (EMB), milligrams of oil per embryo (EMB_MG), milligrams of embryo tissue (EMB_WT), milligrams of oil per gram of endosperm (ENDO), milligrams of oil per endosperm (ENDO_MG), and milligrams of endosperm tissue (ENDO_WT). Descriptions of trait name, tissue type, units of measurement, and calculation used to generate the data for analysis are listed in Table 1.

Data were analyzed by PROC GLM in SAS to test for the effect of environment, genotype, and genotype by environment. Inbred lines were considered fixed and environment and genotype by environment were considered random factors. Broad-sense heritability (H^2) and confidence intervals were calculated for each trait according to Knapp (1985). As calculated, the H^2 estimates are equivalent to repeatability of the phenotypic measurements of inbred lines across environments (Kempthorne, 1969), but will be referred to as H^2 through the remainder of the text. For each trait, within environment and across environment least square means were calculated and phenotypic correlations between traits were calculated. When correlations calculated between traits within single environments were compared, they were not statistically different from the across environment values. Therefore, only the correlations calculated using the across environment least square means were reported. Correlations and tests for significance of correlation were calculated as described (Kutner et al. 2005).

Association Mapping Analysis

Using genotypic data from the 1311 loci and phenotypic values, locus-trait associations were assessed in TASSEL by mixed model analysis (Bradbury et al., 2007) using the efficient mixed-model association algorithm (Kang et al., 2008). PCs and the

kinship matrix were used to correct for population structure and pedigree relationships, respectively. PC1 through PC14 were included in the analysis as fixed effects; random effects were approximates of the identity by descent between two individuals as estimated by the kinship matrix (Allison et al., 2008; Yu et al., 2006; Zhao et al., 2007). The mixed linear model used in TASSEL can be described as:

$$Y = X\beta + Zu + e$$

where y is the vector of phenotypic observations, β is an unknown vector of fixed effects including the locus and the PCs, u is an unknown vector of random additive genetic effects from multiple background QTL, X and Z are known design matrices, and e is the unobserved residual. Best linear unbiased estimates for fixed effects and best linear unbiased predictions for random effects were obtained by solving the mixed-model equation. The association between any SNP/haplotype and the trait is tested by the F-test with denominator degrees of freedom calculated via the Satterthwaite method (Yu et al., 2006).

Association mapping was conducted for all traits using the phenotypic values based upon the least square mean calculated across all environments (across-environment) and the least square means from Hawaii or Williamsburg (within-environment). The two environments used for within-environment analysis were selected because the pair had the lowest correlation for the phenotype under consideration. The loci which were declared significant in across-environment analysis were always declared significant in the within-environment analysis. Therefore, only results obtained from association mapping using the across-environment least square mean are reported.

The p-values from TASSEL were used to calculate the false discovery rate (FDR) (Benjamini and Hochberg, 1995) using SAS PROC MULTTEST. The number of loci which

met thresholds of FDR<0.05 and FDR<0.001 were determined. Benjamini and Yekutieli (2005) suggested that FDR of 0.05 be used for identification of significant QTL. FDR value less than 0.001 is more stringent than suggested by Benjamini and Yekutieli (2005), but was used to narrow the focus of further investigation of putative genes.

For the loci declared significantly associated with a trait at FDR<0.001, the base sequence of the loci were located was analyzed by BLAST (McGinnis and Madden, 2004) to determine sequence homology with that of known genes. This is similar to the methodology utilized by Belo et al. (2008) and Jung et al. (2004) to identify putative genes. The location of the loci declared trait associated was also compared to the location of genes which were known to have an effect on kernel oil content based upon IBM2 Neighbors genetic map positions in MaizeGDB (Lawrence et al., 2004; Lee et al., 2002).

RESULTS

Genotypic Data

A summary of locus distribution on the 10 maize chromosomes can be found in Table 2. Average spacing between loci was similar across chromosomes. The average PIC value for all loci was 0.36 and average missing data for a locus was 4%, with ranges of 0.10 to 0.50 and 0 to 10%, respectively. There were differences in missing data and PIC among sub-groups (Table 3), which could be related to population structure (Pritchard et al., 2000a)or the result of variation due to sampling. Missing data occurred randomly and were excluded from association mapping for QTL detection (Balding, 2006) since imputation of genotypes has been shown to bias association analysis results (Lin et al., 2008).

In the cladogram (Figure 1), female inbreds were localized to one branch of the cladogram with the exception of three which were genetically more similar to male inbreds than the other female lines. These three inbreds had been developed early in the history of the Monsanto High Oil Maize breeding program, and it is possible that they had been misclassified as female rather than male. Within the male and female branches of the cladogram, there was some clustering of conventional and high oil inbreds but the separation was not as distinct as that between male and female lines

The division between female and male inbreds also was revealed in PCA. PC1 separated the inbreds into male and female groups and explained 14% of the genetic variation (Figure 2). Eigenvalues for male inbreds clustered between -2 and 2. All female inbreds had PC1 eigenvalues greater than 4 except for 3 OF which had PC1 near 2; the same females which were placed in the male branch of the cladogram. PC2 explained 9% of the genetic variation and had the highest correlation with WK, EMB, EMB_MG, ENDO, and ENDO_MG, of all PCs (data not shown), but the relationship appeared to be driven by the cluster of CM which were separated from the OM. Twenty-four inbreds had PC2 less than -5 and 92% of those were CM. In comparison, male inbreds for which PC2 was greater than -5 contained only 8% CM. Within the female inbreds, there was no clear differentiation between OF and CF. Beyond PC2, the amount of genetic variation explained by each PC decreased rapidly: PC3 explained approximately 5% of the genetic variation, PC4 explained 4%, and anything beyond PC8 explained less than 2% of the genetic variation. Combined, PC1 through PC14 explained 50% of the genetic variation at the 500 randomly selected loci.

These results of the genetic analysis are in agreement with the breeding methods used within the Monsanto High Oil Maize Breeding program in which male and females were not

crossed during inbred development, but conventional and oil inbreds were routinely intermated within the male and female sub-groups. Although the sub-groups of CF, CM, OF, or OM could be generated based upon breeding history, the analysis of the genotypic data did not clearly differentiate OF from CF and there were several CM which were placed within the OM cluster. These results suggested that principal components could be included in the association analysis for correction of population structure and were possibly a better estimate of population structure that sub-population assignment. The use of PCs in place of sub-population assignment via STRUCTURE was used for association analysis in several previous studies and was shown to produce equivalent or improved results (Allison et al., 2008; Yu et al., 2006; Zhao et al., 2007).

Within these 208 inbreds, average r^2 was 0.05 and 94% of loci pairs had r^2 estimates less than 0.20. Overall, r^2 declined as the physical distance between loci increased (Figure 3). However, there were pairs of loci separated by more than 200 megabases which had r^2 greater than 0.20. These results indicate that LD could extend for long distances, which is similar to previous observations in elite maize breeding lines (Ching et al., 2002). Given the variability in r^2 estimates between pairs of loci, the mapping resolution within this set of inbreds should be assessed on a per locus or genomic location basis.

Phenotypic Data

All traits had highly significant differences among environments ($p<0.001$) and genotypes ($p<0.001$) and highly significant genotype by environment interactions ($p<0.001$). Although environment and the interaction between genotype and environment were statistically significant, the variance attributed to genotype was at least 10 times greater than that of either environment or the interaction (Table 4). The relatively high H^2 for all traits

(>0.75) and the amount of variance attributed to genotype, strongly suggests that environment and its interaction with genotype had relatively small effect on the phenotypes (Bernardo, 2002). Across environment least square means, phenotypic ranges, estimates of variances, and H^2 estimates are shown in Table 4.

Phenotypic correlations were positive and statistically different than zero, with the exception of the correlation between EMB_WT and ENDO_WT (Table 5). Correlations between pairs of traits ranged from -0.76 (ENDO_WT-WK) to 0.85 (EMB_MG-EMB_WT). Since previous reports have indicated that the majority of oil is located in the embryo (Watson, 1987), the expectation was that WK would be more strongly influenced by EMB, EMB_MG, and EMB_WT (Zheng et al., 2008). The correlation between WK and ENDO (0.83) was equivalent to that of WK and EMB (0.79). This was unexpected because of the expectation regarding the role of the embryo in determination of WK. That observation led to the investigation of correlations within OF, OM, CF, and CM (Table 5). The EMB-ENDO correlation was not statistically greater than zero and the correlation between EMB_MG and ENDO_MG was relatively low (0.24) in OM which suggests that embryo and endosperm oil contents were affected by independent genetic factors and that it is possible to develop inbreds with tissue-specific oil accumulation.

Phenotypic ranges displayed in Table 4 were divided by sub-group to further explore the relationship among these traits (Figure 4). The phenotypic distributions for WK, EMB, EMB_MG, EMB_WT, ENDO, and ENDO_MG were similar in that the lower boundary was determined by CF and CM and the upper boundary was set by OM inbreds. For WK and EMB, there was a separation in phenotypic values between oil and conventional inbreds. The separation observed in WK was a result of the WK level used in the breeding methods to

declare an inbred high oil. In comparison, EMB had not been a criterion for selection during the breeding process but there was an division between CF-CM and OF-OM at approximately 300 mg g^{-1}. For EMB, more than 95% of OF and OM lines contained greater than 300 mg g^{-1}; in comparison, 90% of CF and CM inbreds contained less than 300 mg g^{-1}. The lower boundary for ENDO and ENDO_MG were similar for all sub-groups so that phenotypic distributions overlapped.

ENDO_WT was higher for CF and CM than in OF and OM, which differed from the pattern for other traits. The tendency for oil inbreds to have smaller endosperms and larger embryos than conventional inbreds could be the result of selection for increased WK, which can be modified by altering the sizes of embryo and endosperm (Curtis et al., 1968). For all traits, phenotypic variation was four to six times greater in OF and OM than in CF and CM.

Association Analysis

Association mapping identified 173 associations between 129 loci and at least one of six traits – WK, EMB, EMB_MG, EMB_WT, ENDO, and ENDO_WT – at FDR<0.05. The number of associations per trait ranged from six for ENDO to 60 for WK. A summary of the number of loci associated with each trait, and the number of loci shared by any pair of traits is in Table 6. A full list of the loci and the traits to which they were associated can be found in the Supplemental Material. There were no loci shared between embryo-based traits (EMB, EMB_MG, and EMB_WT) and endosperm-based traits (ENDO, ENDO_MG, and ENDO_WT).

At FDR<0.001, 11 loci were associated with at least one of four traits; WK, EMB, EMB_MG, and ENDO (Table 7). The number of SNPs associated with each trait ranged from one for ENDO to five for WK. No loci were declared statistically associated with

ENDO_MG at either FDR level. M09-17 had the largest effect on WK and was one of the loci associated with ENDO. This locus was also associated with ENDO_WT at FDR<0.05, but the allele which was associated with higher oil content in both WK and ENDO was associated with smaller endosperm size. Seven of the 11 loci were placed in putative genes based upon BLAST alignment (Table 7). Two loci which were not placed into putative genes, M08-020 and M08-023, flanked a malate dehydrogenase gene (*mdh1*). Malate dehydrogenase enzyme activity had been shown to be correlated with embryo oil (Doehlert and Lambert, 1991).

LD between loci on chromosome 8 between M08-072 and M08-085 (Figure 5) and in the region on chromosome 9 between M09-006 and M09-035 (Figure 6) were plotted against the physical distance between loci. These regions were selected for additional LD analysis because the function of the putative genes identified via association mapping suggested a direct role in oil metabolism and accumulation and the identification of potential causative loci. M09-021 and M09-022 were in high LD (0.90) with one another and with M09-027 (0.76 and 0.68, respectively). M09-012 and M09-017 had relatively low r^2 estimates (<0.40) with the other loci associated with WK at FDR<0.001. Overall, the QTL on chromosome 9 which met the critical threshold of FDR<0.001 had higher LD estimates with one another with an average of 0.38 than did a randomly selected pair of loci with an average of 0.06. Selection for WK could have created LD among these loci if all are truly associated with WK (Flint-Garcia et al., 2003). LD between M08-080 and M08-079 and M08-078 was 0.42. Comparisons between M08-080 and any other loci on chromosome 8 was less than 0.20.

DISCUSSION

To our knowledge, this is the first known investigation into the genetic basis of oil content in the embryo and endosperm of the maize kernel and identification of QTL associated with tissue specific oil content in elite germplasm. This investigation was initiated to determine if oil accumulation within the two tissues was regulated by the same or different genetic factors. Such information could affect the feasibility and strategy for breeding germplasm with elevated oil content within specific tissues. In this set of inbreds, the oil contents of the separate tissues were correlated with one another, but the magnitudes of the correlations were germplasm-dependent. Different QTL were associated with elevated oil content in the separate tissues. The combined results of the phenotypic and QTL analysis support the hypothesis that the oil contents of embryo and endosperm tissues are under partially independent genetic control.

Reliability of Phenotypic Data

Phenotypic data quality, and the importance of the genotype in determination of the phenotype, is one factor which impacts identification of trait associated loci. For this set of inbred lines in this set of environments, the effect of the genotype was much greater than either the effect of the environment or the interaction between genotype and environment for all traits, which should increase power of QTL detection (Holland, 2004; Lande and Thompson, 1990). The estimate of WK H^2 in this experiment is higher than most previously reported values (Clark et al., 2006; Goldman et al., 1994; Laurie et al., 2004; Mangolin et al., 2004; Song et al., 2004). There are no previous reports of H^2 for EMB, EMB_MG, EMB_WT, ENDO, ENDO_WT, or ENDO_MG to compare the estimates reported here but all were relatively high at greater than 0.75. The use of homozygous, homogeneous inbreds

could have increased H^2 estimates by reducing the within line genetic variability and improving precision of inbred means (Bernardo, 2002). Additional replication across multiple environments also could have reduced the effect of environment and improved precision of the inbred means (Buckler et al., 2009).

Phenotypic Correlations and Genotypic Factors Associated with Traits

The phenotypic correlation between embryo and endosperm oil contents suggested that there was a relationship between oil contents of embryo and endosperm. The variation in phenotypic correlation estimates among sub-groups supported the hypothesis that they were under partially independent genetic control. This was most evident in OM, which had undergone extensive selection for WK in addition to selection for agronomic traits. The lack of correlation between EMB-ENDO and the low EMB_MG-ENDO_MG correlation in OM suggests high oil content in one tissue was not correlated with high oil content in the other. The OM sub-group was also unique in that the correlation between ENDO-WK was relatively high (0.70) and EMB-WK was relatively low (0.26). This differs from the results of Zheng et al. (2008) in which the embryo was presumed to have a much larger effect on WK than does endosperm.

The phenotypic distributions among sub-groups also provided evidence that EMB/EMB_MG was not directly related to ENDO/ENDO_MG and provided insight into the feasibility of developing inbreds that combine high embryo oil with low endosperm oil content. The lower boundary for ENDO and ENDO_MG was equal across all sub-groups, whereas there was a clear division between OF or OM and CF or CM for WK and EMB. This indicated that there were inbreds within the OF and OM sub-groups which contained high embryo and low endosperm oil content, and suggested that there could be different

controlling genetic factors influencing oil concentration in those tissues. Additionally, several OF or OM combined ENDO and ENDO_MG equivalent to CF or CM with EMB greater than CF and CM. These observations, in conjunction with the high correlation between WK and ENDO suggest elevated EMB were required for elevated WK, but that differences in ENDO could have a substantial impact on WK.

The QTL associated with WK, EMB, EMB_MG, EMB_WT, ENDO, and ENDO_WT, specifically the lack of common QTL between embryo-based and endosperm-based traits, provided direct evidence that unique loci separately influence the oil contents of embryo and endosperm and corroborates the phenotypic relationships among traits. EMB and ENDO and EMB_MG and ENDO_MG, which had a lower correlation than either EMB-WK or ENDO-WK in the whole population and had no or low correlation in the OM sub-group, had no common QTL at FDR<0.05. In comparison, WK and ENDO had correlation of 0.83 and were the only traits which had a common QTL at FDR<0.001. Several other studies have shown that traits with high phenotypic correlations tend to also share more QTL or have correlated effects than do traits which have lower correlation (Buckler et al., 2009; Veldboom et al., 1994; Yan et al., 2004). WK and ENDO_WT, which had a strong negative correlation (-0.76) shared two common QTL at FDR<0.05. At both loci, the allele which was associated with increased WK was associated with reduced endosperm size.

QTL and Putative Genes

The majority of QTL identified in this study are novel since oil content of the whole maize kernel has been the focus of most previous research (Berke and Rocheford, 1995; Clark et al., 2006; Laurie et al., 2004; Song et al., 2004; Yang et al., 2010; Zhang et al., 2008), there are only two studies to which these results can be compared. The EMB

associated loci at M04-054 which was in the same bin as a QTL associated with embryo oleic acid concentration (Beló et al., 2008). Although oleic acid accounts for 25-40% of fatty acids in maize (Wassom et al., 2008a; Weber, 1987), the traits are not equivalent since EMB would be determined by other fatty acids in addition to oleic. M06-042, one of the loci associated with EMB_MG at FDR<0.05, was located in the same bin as the acyl-CoA:diacylglycerol acyltransferase (DGAT) identified by Zheng et al. (2008).

The loci associated with WK at FDR<0.001 were located in chromosomal bins in which WK QTL had previously been identified (Mangolin et al., 2004; Song et al., 2004; Yang et al., 2010). M09-012 co-located with *bnlg1401*, a microsatellite locus associated with oil content in the Beijing High Oil (BHO) population (Yang et al., 2010) Although the estimated effect of this region on WK was relatively small relative to other QTL detected in the BHO population, M09-012 had the second largest effect on WK in this study. Locus *bnlg127* was within 0.1 cM of M09-027 on the Monsanto genetic map and was associated with oil content of testcrosses generated from the IHO population (Willmot et al., 2006).

Several of the putative genes in which the trait-associated loci were located had known effect on oil metabolism. On chromosome 9, the trait associated loci were in high LD with one another (0.75+) whereas others were lower (0.40 or less). It is possible that all QTL were truly trait associated and that selection for WK in OM and OF resulted in the LD between loci. Alternatively, all QTL were not truly trait associated but LD between the genotyped loci and the causative allele resulted in trait association. Further analysis and fine mapping are required to determine if all of these loci are independently associated with WK. More research is necessary to determine if these loci are located in causative genes or if they

are in LD with other regions of the genome which impact oil content, the putative genes suggests that these regions are involved in oil metabolism.

The function of the proteins which might be encoded in these regions could provide clues into the physiological basis of high oil content. Two of the WK-associated loci, M09-021 and M09-022, were similar to the coding sequence of acyl co-A oxidase. This enzyme catalyzes the first step in the break down of fatty acids by beta-oxidation (Kindl, 1993), which is typically associated with seed germination (Olsen and Huang, 1988) and growth under carbohydrate-limited conditions (Hooks et al., 1995). The role of acyl co-A oxidase in seed maturation and oil accumulation is unknown but research in *Arabidopsis thaliana* suggested that functional beta oxidation is required for embryo development (Rylott et al., 2003). Maize lines with high oil concentration tend to accumulate oil at a higher rate for a longer period of time during kernel development than lower oil lines (Curtis et al., 1968), which could be due to either sustained synthesis of oil or to reduced degradation.

The loci with the largest effect on WK, and the only loci associated ENDO, was similar to that of a cellulose synthase-like protein D4 (CSLD4) gene. This locus was also associated with ENDO_WT at FDR<0.05. CSLD4 has no known involvement in oil biosynthesis or degradation, but does have a role in cellulose synthesis, cell wall biogenesis, and plant growth (Hu et al., 2010; Li et al., 2009). CSLD4 was analyzed in rice (Wang et al., 2010) and wheat (Nemeth et al., 2010) grains, but changes in its expression or coding sequence had no impact on grain phenotype. Although these studies indicate that CSLD4 does not have seed-specific function, these results suggest that modifications to CSLD4 could impact maize kernel oil by modulating endosperm oil content and size.

The genetic sequence around EMB-associated M08-080 was similar to that of a glycerol 3-phosphate dehydrogenase (G3PDH) gene. The G3PDH enzyme catalyzes the production of glycerol-3-phosphate, the carbon backbone to which fatty acids are attached in production of triacylglycerols (TAG) (Val et al., 2008). Multiple studies have shown that transgenic modifications to total oil content can be more easily achieved through targeting genes in TAG biosynthesis, such as G3PDH, than by modifications to fatty acid synthesis (Thelen and Ohlrogge, 2002). Transgenic expression of G3PDH in canola increased seed oil content (Vigeolas et al., 2007), and overexpression of DGAT, which catalyzes the final step in TAG biosynthesis, increased whole kernel and embryo oil in maize (Zheng et al., 2008). The identification of a QTL G3PDH genetic sequence highlights the importance TAG synthesis in the determination of final oil content, and supports the theory that oil accumulation is driven by sink strength rather than source availability (Borras et al., 2002).

The number of QTLs identified via association analysis in this population at FDR<0.001 was small but focused the investigation of putative genes to only the most significant loci. In comparison, the number of loci declared trait associated at FDR<0.05 matched the expectation of the quantitative genetic basis of oil content of maize (Berke and Rocheford, 1995; Clark et al., 2006; Dudley and Lambert, 1992 ; Laurie et al., 2004; Song et al., 2004; Yang et al., 2010; Zhang et al., 2008). ENDO_MG, the only trait for which no loci were declared trait-associated at FDR<0.05, had an estimated H^2 equal to 0.79. This suggests that the trait was genetically controlled but that the loci were not detected.

CONCLUSIONS

The analysis of the phenotypic relationship among maize whole kernel, embryo, and endosperm oil contents and the identification of QTL associated with them provided evidence that these traits are under partially independent genetic control. Although it is possible that there were QTLs that impact both tissue types which were not detected here, these results have shown that these traits are quantitative and there is a portion of the loci associated with each tissue which are unique. The focused investigation of potential causative loci on the most significant of the trait associations identified putative genes with known function in oil metabolism. The localization of QTL into putative genes associated with TAG synthesis and fatty acid degradation highlighted the importance of these metabolic pathways in determination of final oil content.

REFERENCES

Alexander D., Collins F., Rodgers R. (1967) Analysis of oil content of maize by wide-line NMR. Journal of the American Oil Chemists' Society 44:555-558.

Allison L.W., William H.B., Jesse Rucker B.M.B. (2008) The genetic architecture of complex traits in teosinte (Zea mays ssp. parviglumis): new evidence from association mapping. Genetics 180:1221-1232.

Alrefai R., Berke T.G., Rocheford T.R. (1995) Quantitative trait locus analysis of fatty acid concentrations in maize. Genome 38:894-901.

Balding D.J. (2006) A tutorial on statistical methods for population association studies. Nature Review Genetics 7:781-791.

Beló A., Zheng P., Luck S., Shen B., Meyer D.J., Li B., Tingey S., Rafalski A. (2008) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Molecular Genetics and Genomics 279:1-10.

Benjamini Y., Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57:289-300.

Berke T.G., Rocheford T.R. (1995) Quantitative trait loci for flowering, plant and ear height, and kernel traits in maize. Crop Science 35:1542-1549.

Bernardo R. (2002) Breeding for quantitative traits in plants Woodbury, Minn.(EUA). Stemma Press.

Borras L., Cura J.A., Otegui M.E. (2002) Maize kernel composition and post-flowering source-sink ratio. Crop Science 42:781-790.

Bradbury P., Zhang Z., Kroon D., Casstevens T., Ramdoss Y., Buckler E. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Breseghello F., Sorrells M.E. (2006) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Science 46:1323-1330.

Buckler E.S., Holland J.B., Bradbury P.J., Acharya C.B., Brown P.J., Browne C., Ersoz E., Flint-Garcia S., Garcia A., Glaubitz J.C., Goodman M.M., Harjes C., Guill K., Kroon D.E., Larsson S., Lepak N.K., Li H., Mitchell S.E., Pressoir G., Peiffer J.A., Rosas M.O., Rocheford T.R., Romay M.C., Romero S., Salvo S., Villeda H.S., Sofia da Silva H., Sun Q., Tian F., Upadyayula N., Ware D., Yates H., Yu J., Zhang Z., Kresovich S., McMullen M.D. (2009) The genetic architecture of maize flowering time. Science 325:714-718.

Ching A., Caldwell K., Jung M., Dolan M., Smith O., Tingey S., Morgante M., Rafalski A. (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:19-33.

Clark D., Dudley J., Rocheford T.R., LeDeaux J. (2006) Genetic analysis of corn kernel chemical composition in the random mated 10 generation of the cross of generations 70 of IHO x ILO. Crop Science 46:807-819.

Cover T., Hart P. (1967) Nearest neighbor pattern classification. Information Theory 13:21-27.

Csaikl U., Bastian H., Brettschneider R., Gauch S., Meir A., Schauerte M., Scholz F., Sperisen C., Vornam B., Ziegenhagen B. (1998) Comparative analysis of different DNA extraction protocols: a fast, universal maxi-preparation of high quality plant

DNA for genetic evaluation and phylogenetic studies. Plant Molecular Biology Reporter 16:69-86.

Curtis P.E., Leng E.R., Hageman R.H. (1968) Development changes in oil and fatty acid content of maize strains varying in oil content. Crop Science 8:689-693.

Doehlert D.C., Lambert R.J. (1991) Metabolic characteristics associated with starch, protein, and oil deposition in developing maize kernels. Crop Science 31:151-157.

Dudley J., Lambert R.J. (1992) Ninety generations for selection for oil and protein in maize. Maydica 37:1-7.

Fan J.B., Oliphant A., Shen R., Kermani B., Garcia F., Gunderson K., Hansen M., Steemers F., Butler S., Deloukas P. (2003) Highly parallel SNP genotyping, Cold Spring Harbor Laboratory Press. pp. 69-78.

Flint-Garcia S.A., Thornsberry J.M., S E., Iv B. (2003) Structure of linkage disequilibrium in plants. Annual review of plant biology 54:357-374.

Goldman I.L., Rocheford T.R., Dudley J.W. (1994) Molecular markers associated with maize kernel oil concentration in an Illinois High Protein x Illinois Low Protein cross. Crop Science 34:908-915.

Halliburton R. (2004) Introduction to population genetics. Pearson/Prentice Hall.

Holland, J.B. (2004) Implementation of molecular markers for quantitative traits in breeding programs- challenges and opportunities. Proc. 4th Intl. Crop Sci. Cong., 26 Sep.−1 Oct. 2004, Brisbane, Australia.

Hooks M.A., Bode K., Couée I. (1995) Regulation of acyl-CoA oxidases in maize seedlings. Phytochemistry 40:657-660.

Hu J., Zhu L., Zeng D., Gao Z., Guo L., Fang Y., Zhang G., Dong G., Yan M., Liu J. (2010) Identification and characterization of NARROW ANDROLLED LEAF 1, a novel gene regulating leaf morphology and plant architecture in rice. Plant Molecular Biology 73:283-292.

Kang H.M., Zaitlen N.A., Wade C.M., Kirby A., Heckerman D., Daly M.J., Eskin E. (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709-1723.

Kempthorne O. (1969) An introduction to statistical genetics, Ames, Iowa: The Iowa State University Press.

Kindl H. (1993) Fatty acid degradation in plant peroxisomes: function and biosynthesis of the enzymes involved. Biochimie 75:225-230.

Krill A.M., Kirst M., Kochian L.V., Buckler E.S., Hoekenga O.A. (2010) Association and linkage analysis of aluminum tolerance genes in maize. PLoS One 5:e9958.

Lambert R. (2001) High-oil corn hybrids, in: A. R. Hallauer (Ed.), Specialty corns, CRC Press, Boca Raton, FL. pp. 131-154.

Lande R., Thompson R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

Laurie C.C., Chasalow S.D., LeDeaux J.R., McCarroll R., Bush D., Hauge B., Lai C., Clark D., Rocheford T.R., Dudley J.W. (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141-2155.

Lawrence C., Dong Q., Polacco M., Seigfried T., Brendel V. (2004) MaizeGDB, the community database for maize genetics and genomics. Nucleic acids research 32:D393-D397.

Lee M., Sharopova N., Beavis W., Grant D., Katt M., Blair D., Hallauer A. (2002) Expanding the genetic map of maize with intermated B73 x Mo17 (IBM) population. Plant Molecular Biology 48:453-461.

Li M., Xiong G., Li R., Cui J., Tang D., Zhang B., Pauly M., Cheng Z., Zhou Y. (2009) Rice cellulose synthase like D4 is essential for normal cell wall biosynthesis and plant growth. The Plant Journal 60:1055-1069.

Lin D., Hu Y., Huang B. (2008) Simple and efficient analysis of disease association with missing genotype data. The American Journal of Human Genetics 82:444-452.

Mangolin C.A., de Souza C.L., Garcia A.A.F., Garcia A.F., Sibov S.T., de Souza A.P. (2004) Mapping QTLs for kernel oil content in a tropical maize population. Euphytica 137:251-259.

McGinnis S., Madden T. (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic acids research 32:W20.

Melchinger A.E., Utz H.F., Schön C.C. (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383-403.

Nemeth C., Freeman J., Jones H.D., Sparks C., Pellny T.K., Wilkinson M.D., Dunwell J., Andersson A.A.M., Åman P., Guillon F., Saulnier L., Mitchell R.A.C., Shewry P.R. (2010) Down-regulation of the CSLF6 gene results in decreased (1,3;1,4)-β-d-glucan in endosperm of wheat. Plant Physiology 152:1209-1218.

Olsen J.A., Huang A.H.C. (1988) Glyoxysomal acyl-CoA synthetase and oxidase from germinating elm, rape and maize seed. Phytochemistry 27:1601-1603.

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

Pritchard J.K., Stephens M., Donnelly P. (2000a) Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pritchard J.K., Stephens M., Rosenberg N.A., Donnelly P. (2000b) Association mapping in structured populations. The American Journal of Human Genetics 67:170-181.

Rafalski J.A. (2010) Association genetics in crop improvement. Current Opinion in Plant Biology 13:174-180.

Rylott E.L., Rogers C.A., Gilday A.D., Edgell T., Larson T.R., Graham I.A. (2003) Arabidopsis mutants in short-and medium-chain acyl-CoA oxidase activities accumulate acyl-CoAs and reveal that fatty acid -oxidation is essential for embryo development. Journal of Biological Chemistry 278:21370-21377.

Saitou N., Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4:406-425.

Song X.F., Song T.M., Dai J.R., Rocheford T., Li J.S. (2004) QTL mapping of kernel oil concentration with high-oil maize by SSR markers. Maydica 49:41-48.

Thelen J.J., Ohlrogge J.B. (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4:12-21.

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler Iv E.S. (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 28:286-289.

Val D.L., Schwartz S.H., Kerns M.R., Deikman J. (2008) Development of a High Oil Trait for Maize, in: A. L. Kriz and B. A. Larkins (Eds.), Molecular genetic approaches to maize improvement, Springer-Verlag, Berlin. pp. 303-323.

Veldboom L., Lee M., Woodman W. (1994) Molecular marker-facilitated studies in an elite maize population: I. linkage analysis and determination of QTL for morphological traits. Theoretical and Applied Genetics 88:7-16.

Vigeolas H., Waldeck P., Zank T., Geigenberger P. (2007) Increasing seed oil content in oil seed rape (Brassica napus L.) by over expression of a yeast glycerol 3 phosphate dehydrogenase under the control of a seed specific promoter. Plant Biotechnology Journal 5:431-441.

Visscher P.M., Hill W.G., Wray N.R. (2008) Heritability in the genomics era - concepts and misconceptions. Nature Review Genetics 9:255-266.

Wang L., Guo K., Li Y., Tu Y., Hu H., Wang B., Cui X., Peng L. (2010) Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. BMC Plant Biology 10:282.

Wassom J.J., Mikkelineni V., Bohn M.O., Rocheford T.R. (2008a) QTL for fatty acid composition of maize kernel oil in Illinois High Oil x B73 backcross-derived lines. Crop Science 48:69-78.

Wassom J.J., Wong J.C., Martinez E., King J.J., DeBaene J., Hotchkiss J.R., Mikkilineni V., Bohn M.O., Rocheford T.R. (2008b) QTL associated with maize kernel oil, protein,

and starch concentrations; kernel mass; and grain yield in Illinois High Oil x B73 backcross-derived lines. Crop Science 48:243-252.

Watson S. (1987) Structure and composition, in: S. Watson and P. Ramstad (Eds.), Corn Chemistry and Technology, American Association of Cereal Chemists, Inc., St. Paul, MN. pp. 53-82.

Weber E. (1987) Lipids of the kernel, in: S. Watson and P. Ramstad (Eds.), Corn Chemistry and Technology, American Association of Cereal Chemists, Inc., St Paul, MN. pp. 311-349.

Willmot D.B., Dudley J., Rocheford T.R., Bari A. (2006) Effect of random mating on marker-QTL associations for grain quality traits in the cross of Illinois High Oil x Illinois Low Oil. Maydica 51:187-199.

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

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

Yan J., Kandianis C.B., Harjes C.E., Bai L., Kim E.H., Yang X., Skinner D.J., Fu Z., Mitchell S., Li Q. (2010) Rare genetic variation at Zea mays crtRB1 increases [beta]-carotene in maize grain. Nature Genetics 42:322-327.

Yan X., Liao H., Beebe S.E., Blair M.W., Lynch J.P. (2004) QTL mapping of root hair and acid exudation traits and their relationship to phosphorus uptake in common bean. Plant and Soil 265:17-29.

Yang X., Guo Y., Yan J., Zhang J., Song T., Rocheford T., Li J.S. (2010) Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. Theoretical and Applied Genetics 120:665-678.

Yu J., Pressoir G., Briggs W.H., Bi I.V., Yamasaki M., Doebley J.F., McMullen M.D., Gaut B.S., Nielsen D.M., Holland J.B. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203-208.

Zhang J., X.Q. L., X.F. S., Yan J.B., Song T.M., Dai J.R., Rocheford T.R., Li J.S. (2008) Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. Euphytica 162:335-344.

Zhao K., Aranzana M.J., Kim S., Lister C., Shindo C., Tang C., Toomajian C., Zheng H., Dean C., Marjoram P. (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4.

Zheng P., Allen W.B., Roesler K., Williams M.E., Zhang S., Li J., Glassman K., Ranch J., Nubel D., Solawetz W., Bhattramakki D., Llaca V., Deschamps S., Zhong G.-Y., Tarczynski M.C., Shen B. (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nature Genetics 40:367-372.

Zhu C., Gore M., Buckler E.S., Yu J. (2008) Status and prospects of association mapping in plants. The Plant Genome 1:5-20.

Table 1. Description of traits analyzed in association mapping of maize kernel oil content and calculations used to generate reported values.

Trait[a]	Tissue	Units[b]	Calculation[c]
WK	whole kernel	mg g^{-1}	NIT percent oil *1000
EMB	embryo	mg g^{-1}	NMR percent oil *1000
EMB_MG	embryo	mg	embryo weight (mg)*NMR percent oil
EMB_WT	embryo	mg	-
ENDO	endosperm	mg g^{-1}	NMR percent oil*1000
ENDO_MG	endosperm	mg	endosperm weight (mg)*NMR percent oil
ENDO_WT	endosperm	mg	-

[a] WK = whole kernel oil in mg g^{-1}, EMB = embryo oil in mg g^{-1}, EMB_MG = embryo oil in mg, EMB_WT = embryo weight in mg, ENDO = endosperm oil in mg g^{-1}, ENDO_MG = endosperm oil in mg, ENDO_WT = endosperm weight in mg,

[b] mg=milligrams, g=grams

[c] NIT=Near Infrared Transmittance, NMR=Nuclear Magnetic Resonance,

Table 2. Genomic coverage for 1311 loci analyzed in the association mapping of maize kernel oil content. The average distance between markers is in megabases.

Chromosome	Count of Loci	Average Distance between Loci
1	202	2.05
2	156	1.93
3	140	2.31
4	145	2.04
5	160	1.77
6	123	1.82
7	121	1.77
8	85	2.73
9	99	1.87
10	80	2.22

.

Table 3. Summary of missing data and PIC values for markers analyzed in the association mapping of maize kernel oil content. Values were calculated for the whole population and for the sub-groups identified by breeding history and utility in the breeding program.

		Missing			PIC[b]		
Sub-group[a]	**Number of individuals**	**Minimum**	**Maximum**	**Average**	**Minimum**	**Maximum**	**Average**
All	208	0%	10%	4%	0.10	0.50	0.36
CF	11	0%	25%	1%	0.00	0.50	0.22
CM	33	0%	25%	1%	0.00	0.50	0.26
OF	41	0%	35%	10%	0.00	0.50	0.34
OM	123	0%	13%	3%	0.00	0.50	0.30

[a] ALL = all inbreds of association mapping population, CF = conventional female, CM = conventional male, OF = oil female, OM = oil male

[b] PIC = polymorphism information content

Table 4. Phenotype means, ranges, variances, and broad sense heritabilities estimated from 208 maize inbred lines of association mapping study. Trait means and ranges were based upon the across environment means.

Trait[a]	Mean[a]	Phenotypic Range[a]	$\hat{\sigma}_e^{2}$ [b]	$\hat{\sigma}_g^{2}$ [b]	$\hat{\sigma}_{ge}^{2}$ [b]	H^2 [b]	H^2 95% CI [b]
WK	84 mg g^{-1}	25-175 mg g^{-1}	12.3	3492	35.5	0.98	0.97, 0.99
EMB	360 mg g^{-1}	186-475 mg g^{-1}	656	32466	324	0.96	0.95, 0.97
EMB_MG	11.5 mg	3.8-21.3 mg	1.5	83	1.5	0.91	0.89, 0.93
EMB_WT	31.5 mg	20 mg -50 mg	4	189	19	0.76	0.69, 0.80
ENDO	25 mg g^{-1}	7-66 mg g^{-1}	18.6	892	41	0.89	0.87, 0.91
ENDO_MG	4.0 mg	1.3-10.8 mg	0.9	14.9	1.2	0.79	0.75, 0.83
ENDO_WT	183 mg	104 mg – 293 mg	201	9440	523	0.92	0.90, 0.94

[a] WK = whole kernel oil in mg g^{-1}, EMB = embryo oil in mg g^{-1}, EMB_MG = embryo oil in mg, EMB_WT = embryo weight in mg, ENDO = endosperm oil in mg g^{-1}, ENDO_MG = endosperm oil in mg, ENDO_WT = endosperm weight in mg, mg=milligrams, g=gram

[b] $\hat{\sigma}_e^2$ = estimated variance component for environment, $\hat{\sigma}_g^2$ = estimated variance component for genotype, $\hat{\sigma}_{ge}^2$ = estimated variance component for genotype by environment interaction, H^2=broad sense heritability, CI = confidence interval

Table 5. Phenotypic correlations between traits measured in maize inbreds of association mapping study. Correlations were calculated within the whole population or within sub-group CF, CM, OF, or OM. *, ** significant at $p<0.05$ and $p<0.01$, respectively.

Trait1-Trait2[a]	ALL[b]	CF[b]	CM[b]	OF[b]	OM[b]
EMB_MG-EMB_WT	0.85**	0.77**	0.74**	0.89**	0.85**
EMB_MG-ENDO	0.41**	-0.02	0.37**	0.15	-0.10
EMB_MG-ENDO_MG	0.54**	0.29	0.47**	0.54**	0.24**
EMB_MG-ENDO_WT	-0.24**	0.42	0.15	0.32*	0.53**
EMB_MG-WK	0.65**	0.73**	0.77**	0.43**	0.13
EMB_WT-ENDO	0.23**	0.17	0.11	-0.04	-0.07
EMB_WT-ENDO_MG	0.47**	0.44	0.39**	0.42**	0.31**
EMB_WT-ENDO_WT	0.06	0.52*	0.43**	0.48**	0.56**
EMB_WT-WK	0.41**	0.67**	0.44**	0.25*	0.05
EMB-EMB_MG	0.82**	0.77**	0.75**	0.69**	0.55**
EMB-EMB_WT	0.45**	0.29	0.23	0.35*	0.13
EMB-ENDO	0.54**	-0.13	0.55**	0.50**	-0.05
EMB-ENDO_MG	0.46**	0.10	0.39**	0.52**	-0.07
EMB-ENDO_WT	-0.58**	-0.03	-0.21	-0.22	0.03
EMB-WK	0.79**	0.62*	0.76**	0.67**	0.26**
ENDO_MG-ENDO_WT	-0.25**	0.60*	0.48**	-0.02	0.12
ENDO_MG-WK	0.65**	0.68**	0.46**	0.68**	0.38**
ENDO_WT-WK	-0.76**	0.32	-0.17	-0.56**	-0.59**
ENDO-ENDO_MG	0.82**	0.86**	0.72**	0.76**	0.73**
ENDO-ENDO_WT	-0.61**	0.54*	0.00	-0.50**	-0.41**
ENDO-WK	0.83**	0.52*	0.52**	0.76**	0.70**

[a] WK = whole kernel oil mg g^{-1}, EMB = embryo oil mg g^{-1}, EMB_MG = embryo oil in mg, EMB_WT = embryo weight in mg, ENDO = endosperm oil mg g^{-1}, ENDO_MG = endosperm oil in mg, ENDO_WT = endosperm weight in mg, mg = milligrams, g = grams

[b] ALL = all inbreds, CF = conventional female, CM = conventional male, OF = oil female, OM = oil male

Table 6. Number of loci associated with WK, EMB, EMB_MG, EMB_WT, ENDO, and ENDO_WT at a False Discovery Rate <0.05 in an association mapping of 208 maize inbred lines. The number located at the intersection of two traits represents the number of common loci between the traits.

Trait[a]	Number of Loci	EMB[a]	EMB_MG[a]	EMB_WT[a]	ENDO[a]	ENDO_WT[a]
WK	60	15	8	0	3	2
EMB	58	-	32	1	0	0
EMB_MG	37	-	-	3	0	0
EMB_WT	3	-	-	-	0	0
ENDO	6	-	-	-	-	2
ENDO_WT	9	-	-	-	-	-

[a] WK = whole kernel oil mg g^{-1}, EMB = embryo oil mg g^{-1}, EMB_MG = embryo oil in mg, EMB_WT = embryo weight in mg, ENDO = endosperm oil mg g^{-1}, ENDO_WT = endosperm weight in mg, mg = milligrams, g = grams.

Table 7. Loci identified as associated to maize kernel oil traits at False Discovery Rate ≤ 0.001 from association mapping of 208 inbred lines.

Trait[a]	Locus	FDR[b]	Estimated effect	Public bin	Putative gene	E[c]
EMB	M04-054	0.00085	19 mg g^{-1}	4.06	-	-
EMB	M08-080	0.00039	21 mg g^{-1}	8.08	Glycerol-3-phosphate dehydrogenase (G3PDH)-like protein	0
EMB_MG	M03-066	0.00049	2.0 mg	3.06	60S ribosomal protein L5-1	1e-153
EMB_MG	M08-006	0.00026	1.4 mg	8.01	Putative HGA6	7e-39
EMB_MG	M08-020	0.00026	1.5 mg	8.03	-	-
EMB_MG	M08-023	0.00048	1.3 mg	8.03	-	-
ENDO	M09-017	0.00039	5 mg g^{-1}	9.03	Cellulose synthase-like protein D4 (CSLD4)	1e-145
WK	M09-012	0.00007	10 mg g^{-1}	9.02	Multidrug-resistance associated protein 3 (MRP3)	0
WK	M09-017	1.44E-11	12 mg g^{-1}	9.03	Cellulose synthase-like protein D4 (CSLD4)	1e-145
WK	M09-021	0.00017	9 mg g^{-1}	9.03	Putative acyl-CoA oxidase	2e-59
WK	M09-022	0.00039	8 mg g^{-1}	9.03	Putative acyl-CoA oxidase	2e-59
WK	M09-027	0.00026	8 mg g^{-1}	9.03	-	-

[a] WK = whole kernel oil mg g^{-1}, EMB = embryo oil mg g^{-1}, EMB_MG = embryo oil in mg, ENDO = endosperm oil mg g^{-1}, ENDO_MG = endosperm oil in mg, mg = milligrams, g = grams

[b] FDR = False discovery rate

[c] = Refers to results of BLAST analysis

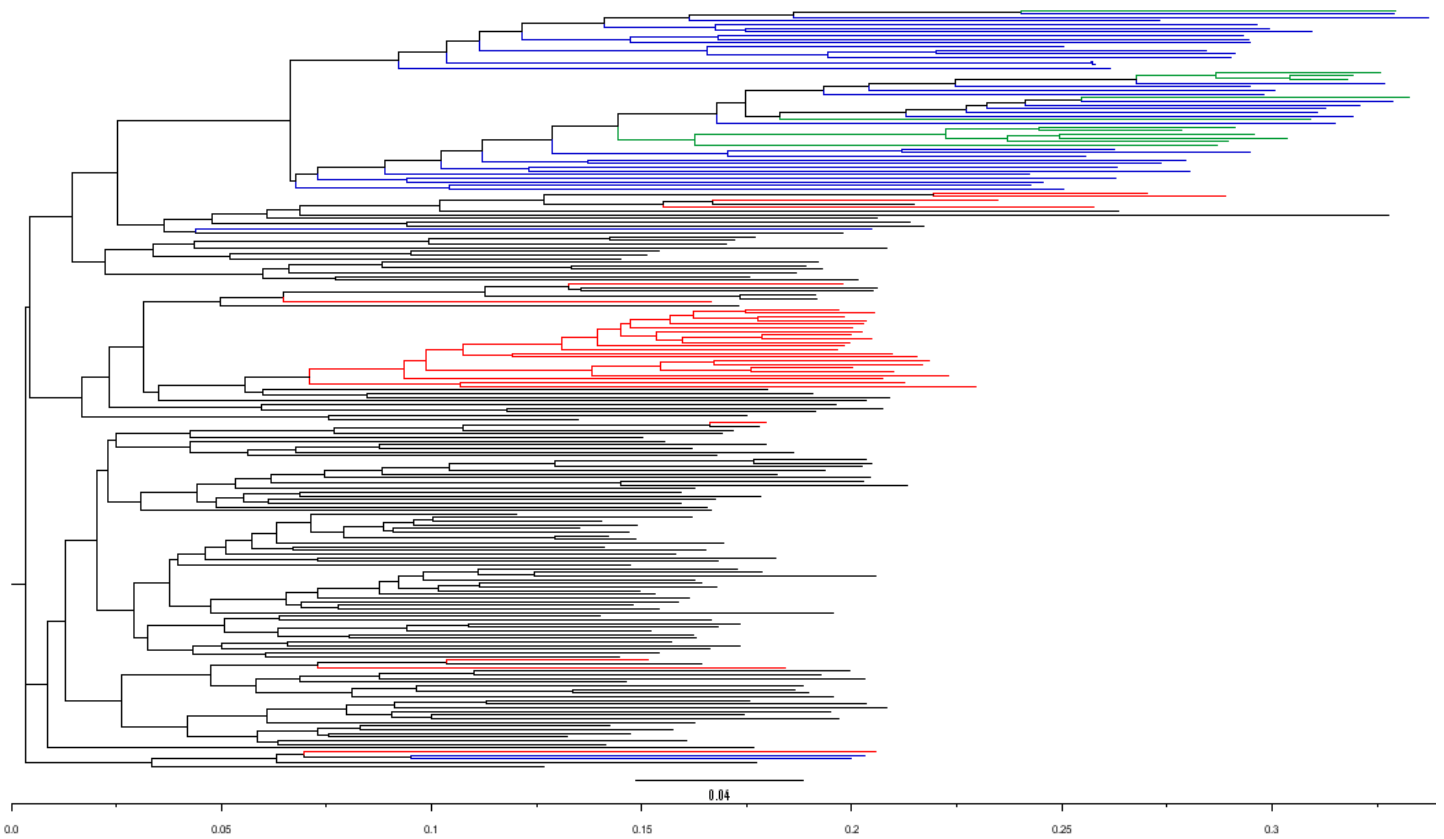

Figure 1 Cladogram generated from genotypic data from 1311 loci in 208 inbred lines via neighbor-joining algorithm using the simple parsimony substitution model. Nodes were color coded to differentiate CF (green), CM (red), OF (blue), and OM (black). CF = conventional female, CM = conventional male, OF = oil female, OM = oil male

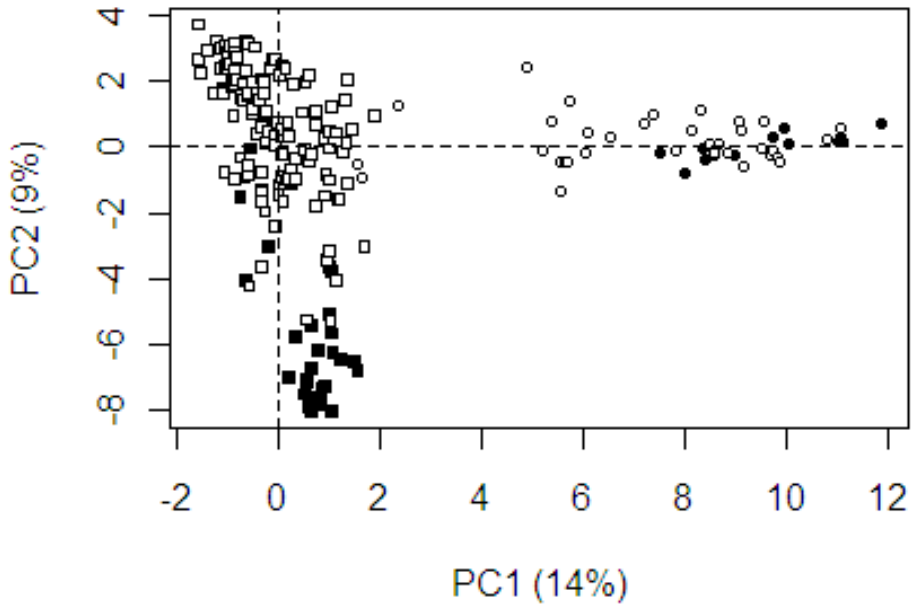

Figure 2. Plot of principal components 1 and 2 calculated from 500 randomly selected loci genotyped in 208 inbreds of the association mapping population. The percent variation explained by the principal component is listed in parentheses.

PC1 = principal component 1, PC2 = principal component 2 ■ =conventional males, □ = oil males, ● = conventional females, ○ = oil females

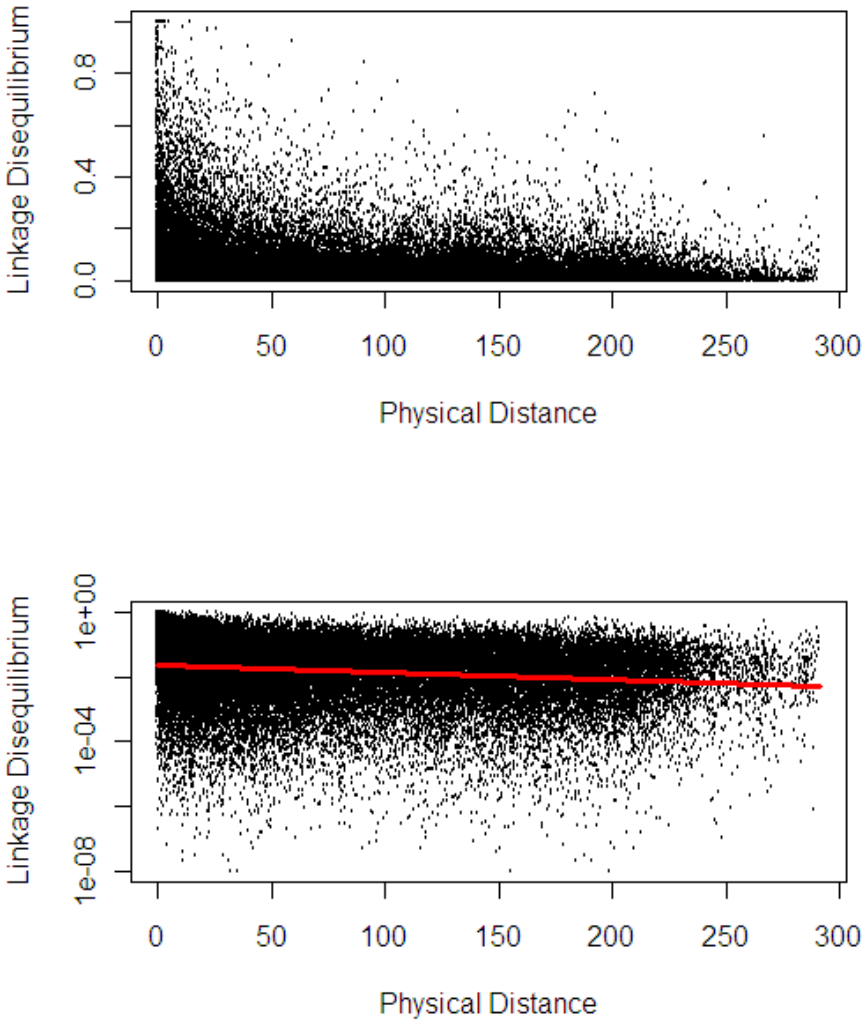

Figure 3. Linkage disequilibrium, measured as r^2, plotted against physical distance between markers in megabases. The upper plot is r^2 by physical distance; the lower plot is the same data plotted with the y-axis on $\log_{10}$ scale and contains a best fit line. Linkage disequilibrium was calculated only for pairs of markers on the same chromosome.

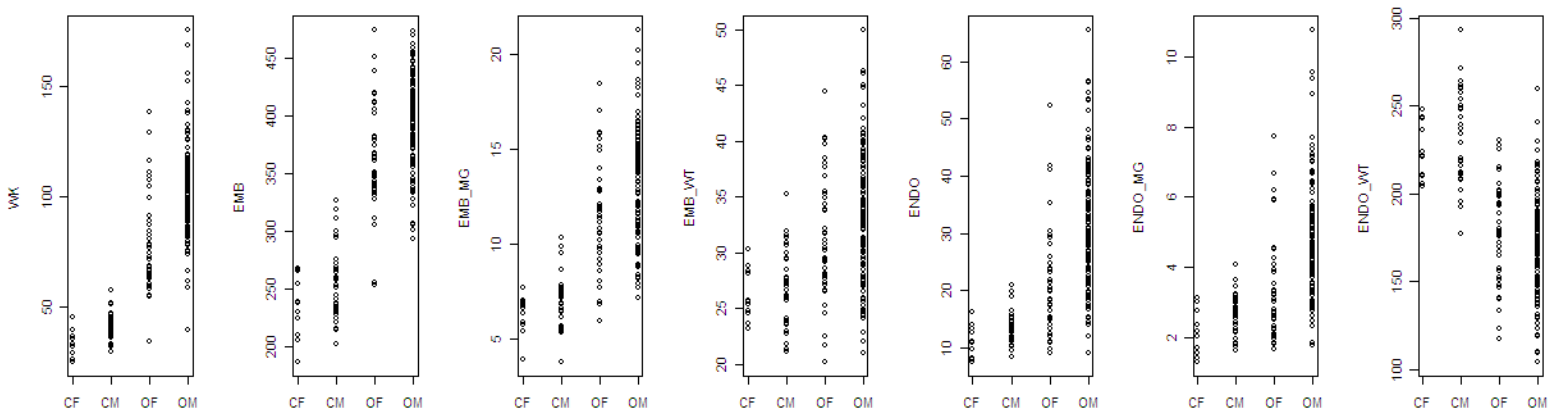

Figure 4. Inbred least square mean phenotypic values of WK, EMB, EMB_MG, EMB_WT, ENDO, ENDO_MG, and ENDO_WT within sub-groups of CF, CM, OF, and OM. Each datapoint on the plot represents a single inbred.

WK = whole kernel oil mg g-1, EMB = embryo oil mg g-1, EMB_MG = embryo oil in mg, EMB_WT = embryo weight in mg, ENDO = endosperm oil mg g-1, ENDO_MG = endosperm oil in mg, ENDO_WT = endosperm weight in milligrams, CF = conventional female, CM = conventional male, OF = oil female, OM =oil male, mg = milligrams, g = grams

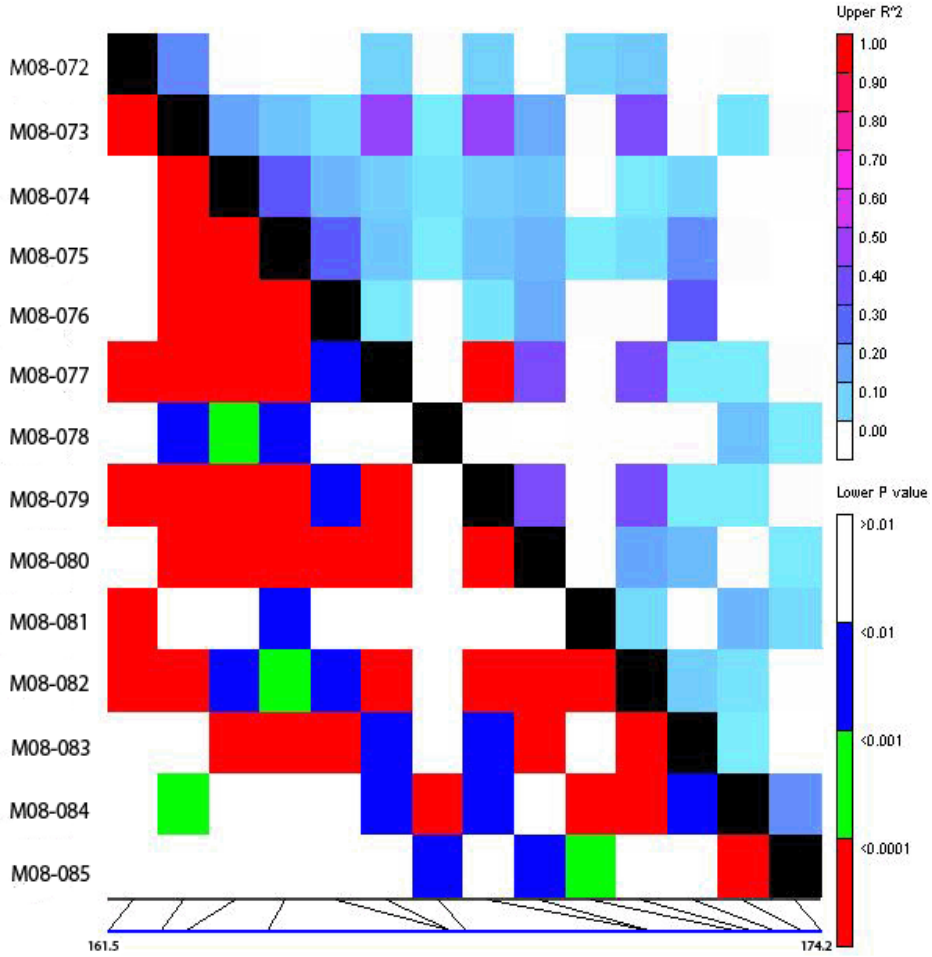

Figure 5. Linkage disequilibrium (r^2) between pairs of polymorphic sites in the interval between M08-072 and M08-085 was calculated (upper right), and significance was determined by Fisher's exact test (lower left). Shading indicates the magnitude of r^2 and significance level. The physical positions of the loci, in megabases, are indicated on the scale at the bottom of the figure. This interval spanned 12.7 megabases.

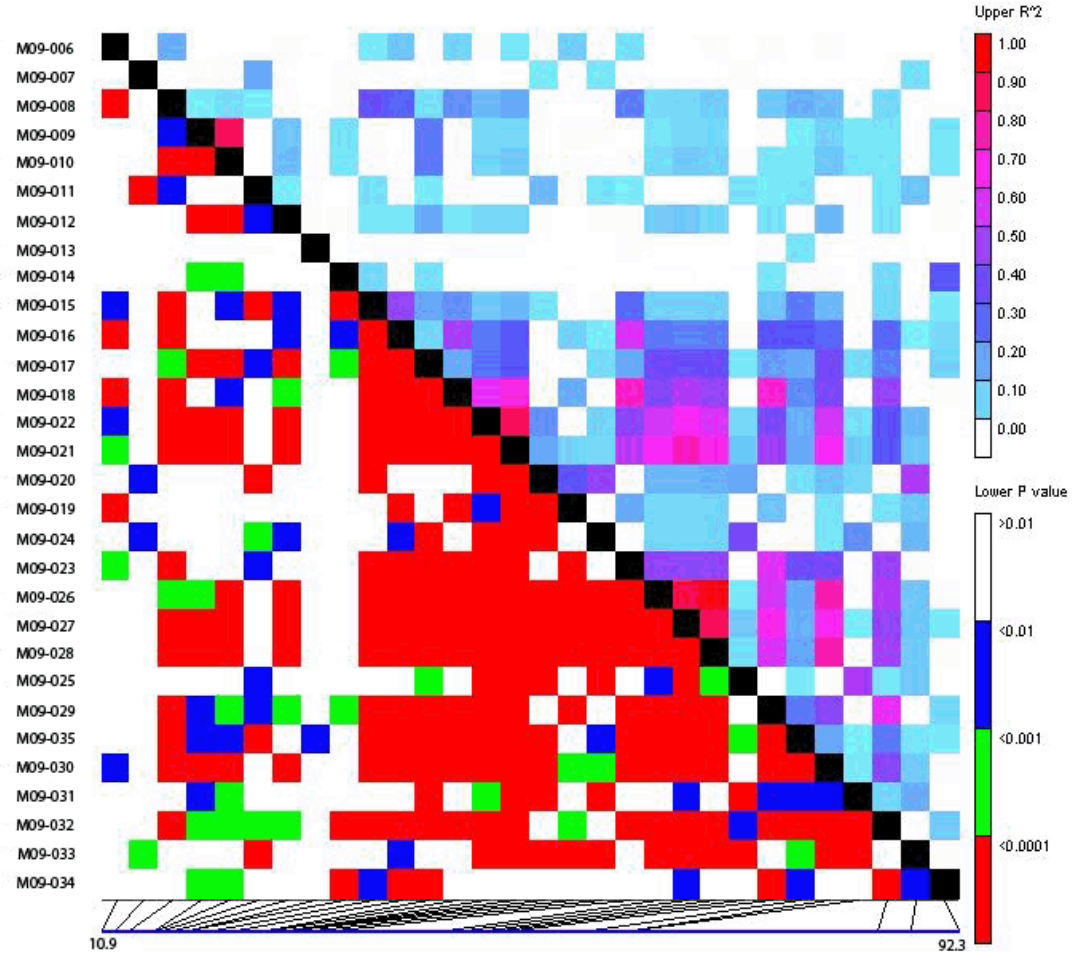

Figure 6. Linkage disequilibrium (r^2) between pairs of polymorphic sites in the interval between M09-006 and M09-035 was calculated (upper right), and significance was determined by Fisher's exact test (lower left). Shading indicates the magnitude of r^2 and significance level. The physical positions of the loci, in megabases, are indicated by the scale at the bottom of the figure. This interval spanned 81.4 megabases.

Supplemental Material

Locus	EMB_WT[a]	EMB[a]	EMB_MG[a]	ENDO_WT[a]	ENDO[a]	WK[a]
M01-019		0.0379				0.0036
M01-020		0.0301				0.0320
M01-024			0.0368			
M01-025		0.0421	0.0359			
M01-027		0.0042	0.0057			
M01-036						0.0244
M01-042		0.0301				0.0013
M01-047		0.0382				
M01-058		0.0302				
M01-064			0.0487			
M01-076		0.0375				
M01-103						0.0386
M01-134				0.0053		
M01-135				0.0203		
M01-170			0.0346			
M01-196						0.0197
M01-200						0.0197
M02-039		0.0211				
M02-069						0.0367
M02-072						0.0445
M02-089			0.0368			
M02-107						0.0367
M02-110		0.0119				
M02-112		0.0119				
M02-125		0.0382				
M02-132		0.0301				
M03-001					0.0110	
M03-019						0.0036
M03-023			0.0487			
M03-026						0.0073
M03-034						0.0073
M03-042						0.0256
M03-050						0.0388
M03-065		0.0301				
M03-066		0.0013	0.0005			0.0452

Supplemental Material, continued.

M03-072		0.0241				0.0253
M03-118						0.0388
M03-119						0.0342
M04-005		0.0445				
M04-033		0.0452				
M04-034				0.0269		
M04-035			0.0191			
M04-038				0.0269		
M04-046		0.0042				
M04-049		0.0382				
M04-052		0.0319				
M04-054		0.0009	0.0487			0.0226
M04-056		0.0189	0.0496			
M04-058						0.0452
M04-105		0.0211	0.0217			
M04-144		0.0109	0.0201			0.0459
M05-014						0.0459
M05-015						0.0197
M05-027						0.0266
M05-032				0.0446		
M05-034			0.0368			
M05-057					0.0012	
M05-090						0.0051
M05-106						0.0298
M05-109		0.0195				
M05-116						0.0197
M05-131			0.0257			
M05-135		0.0013				
M06-042			0.0201			
M06-045		0.0119				
M06-047		0.0119				
M06-050			0.0201			
M06-054		0.0151				
M06-055		0.0061	0.0359			
M06-065		0.0151				0.0197
M06-068		0.0452				

Supplemental Material, continued.

M06-083		0.0119	0.0238			
M06-086		0.0211				0.0363
M07-001						0.0297
M07-043		0.0421				
M07-060		0.0379				
M07-075			0.0098			
M07-081					0.0089	
M07-086		0.0301	0.0191			0.0197
M07-087		0.0306	0.0191			0.0197
M07-102		0.0151				0.0244
M07-112			0.0068			0.0298
M07-119		0.0420				
M08-004		0.0109				
M08-005		0.0042	0.0217			
M08-006	0.0131	0.0012	0.0003			
M08-011			0.0129			
M08-017			0.0331			
M08-018			0.0201			
M08-020	0.0013		0.0003			
M08-023	0.0251		0.0005			
M08-025		0.0379	0.0238			0.0229
M08-026			0.0068			0.0036
M08-032		0.0119				
M08-035		0.0401				
M08-036		0.0300				
M08-048		0.0421	0.0068			
M08-049		0.0252	0.0277			0.0239
M08-057		0.0211	0.0138			
M08-061				0.0446		
M08-062				0.0269		
M08-078		0.0445				
M08-079		0.0445				
M08-080		0.0004				0.0097
M09-012					0.0087	0.0001
M09-013		0.0253				
M09-014						0.0298

Supplemental Material, continued.

M09-017				0.0083	0.0004	0.0000
M09-018						0.0253
M09-021						0.0002
M09-022						0.0004
M09-025						0.0097
M09-026						0.0142
M09-027						0.0003
M09-028						0.0121
M09-029						0.0388
M09-030						0.0016
M09-031						0.0010
M09-034						0.0069
M09-054						0.0262
M09-058		0.0211				
M09-059		0.0088	0.0487			
M09-065						0.0197
M10-010						0.0072
M10-015				0.0446	0.0395	0.0253
M10-042						0.0367
M10-068		0.0223				
M10-069						0.0405
M10-073			0.0496			0.0010

Supplemental Material. Loci identified as associated to maize kernel oil traits at FDR ≤ 0.05 from association mapping of 208 inbreds. FDR is listed at the intersection of a trait column and a locus row.

[a] WK = whole kernel oil mg g^{-1}, EMB = embryo oil mg g^{-1}, EMB_MG = embryo oil in mg, EMB_WT = embryo weight in mg, ENDO = endosperm oil mg g^{-1}, ENDO_WT = endosperm weight in mg, mg = milligrams, g = grams.

CHAPTER 3: GENETIC MAPPING OF QUANTITATIVE TRAIT LOCI AFFECTING MAIZE WHOLE KERNEL, EMBRYO, AND ENDOSPERM OIL CONTENT USING LINKAGE ANALYSIS

A paper to be submitted to Euphytica

Karen Grote and Michael Lee

ABSTRACT

Maize hybrids with elevated oil content have value for human consumption, livestock feed, and energy production. Selection for increased oil content has typically been conducted on the level of the whole kernel. Selection for embryo and endosperm oil content could be included during the screening process and could enable the production of inbreds and hybrids with oil targeted to a specific tissue type. Eighty-five F2:3 lines were used to analyze the phenotypic relationship among whole kernel oil in milligrams oil per gram of kernel (WK), embryo milligrams oil per gram of embryo tissue (EMB), embryo milligrams oil per embryo (EMB_MG), endosperm milligrams oil per gram of endosperm tissue (ENDO), and endosperm milligrams of oil per endosperm (ENDO_MG) and to map quantitative trait loci associated with the traits. All traits were significantly correlated with one another but were low when calculated between embryo and endosperm traits. Six QTL were identified for WK, three for EMB, three for EMB_MG, and one for ENDO. QTL associated with WK and EMB_MG also coincided in their genomic position on chromosome 3. WK, EMB, and EMB_MG had QTL located in a single region on chromosome 6.

INTRODUCTION

Maize hybrids with elevated oil content have value for human consumption, livestock feed, and energy production (Lambert, 2001). Whole kernel oil is typically 40 mg g^{-1} (Watson, 1987), but with selection can be as much as 200 mg g^{-1} (Dudley and Lambert, 1992). Whole kernel oil content can be increased thorough oil concentration in the embryo and endosperm or by altering the ratio of embryo to endosperm tissue (Curtis et al., 1968). Modifications to the embryo-endosperm ratio have been associated with reduced endosperm and kernel size and lower grain yield (Ottaviano and Camussi, 1981; Wassom et al., 2008b), which is unfavorable for commercial grain production. Screening for embryo and/or endosperm oil content, in addition to whole kernel oil, could be part of the selection criteria used in the development of high whole kernel oil maize lines. Additionally, the ability to screen and select for embryo or endosperm oil content could enable production of inbreds and hybrids with tissue-specific oil accumulation.

Independent analyses of the oil contents of maize embryo and endosperm tissues requires manual dissection of tissues; a slow, expensive, and destructive process. The difficulty in obtaining phenotypic data necessitates improvements be made in screening for these traits, and could include development of new methods such as identification of trait associated loci and marker-assisted breeding. The genetic basis of embryo and endosperm oil has received little attention in comparison to whole kernel oil, which has been analyzed in many populations and environments (Laurie et al., 2004; Mangolin et al., 2004; Song et al., 2004; Wassom et al., 2008b; Willmot et al., 2006; Yang et al., 2010; Zhang et al., 2008). Several studies have been conducted in which loci with alleles of large estimated genetic effects were associated with oil or fatty acid content of the embryo. Zheng et al. (2008)

identified a region on chromosome 6 which increased embryo oil concentration by approximately 18% in a backcross population and Belo et al. (2008) determined that a fatty acid desaturase, *fad2*, on chromosome 4 affected oleic acid concentration in embryo tissue in an association analysis of over 500 elite inbreds. Over-expression of a transcription factor, *ZmWRI1*, resulted in a 30% increase in embryo oil content in transgenic maize inbreds relative to the non-transgenic lines, but no increases in endosperm oil content (Shen et al., 2010). No investigations of the genetic basis of endosperm oil content were found.

The utility of screening and selecting for embryo or endosperm specific oil content requires further investigation into the relationship among these traits and whole kernel oil content and into their genetic control. Genetic mapping could identify quantitative trait loci (QTL) and help elucidate the genetic basis of these traits, which would have an impact on selection methodology and breeding strategies. The first objective in this study was to characterize whole kernel, embryo, and endosperm oil in a bi-parental F2-derived population and estimate broad-sense heritability and phenotypic correlations among the traits. The second objective of this study was to identify QTL associated with these traits and estimate their genetic effects within this population.

MATERIALS AND METHODS

Population Development

Selection of parents for the genetic mapping population was based upon divergence for whole kernel oil content in milligrams oil per gram of kernel mass and inbred relevance to the Monsanto Maize Breeding program. The high oil inbred (HO) contained 85 mg g^{-1} whole kernel oil and the low oil inbred (LO) contained 41 mg g^{-1}. An F1 cross between HO

and LO was created in Huxley, IA in 2008, which was planted and self-pollinated in Hawaii in the 2008/2009 season to generate F2 seed. F2 kernels were planted in Huxley, IA in 2009 for self-pollination and the generation of 85 F2:3 ears. The number of F2:3 ears returned was lower than expected and was partly due to weather conditions at the time of anthesis. The F2:3 ears were planted in three field replications in Hawaii in the November 2009 to March 2010 growing season and self-pollinated to generate the samples for phenotypic analysis.

Genotypic Data and Analysis

Fifteen kernels of each F2:3 ear were bulked together for DNA extraction from which the genotypes of the F2 plants were inferred. DNA was extracted by a standard procedure (Dellaporta et al., 1983) and samples were genotyped at a density of 137 single nucleotide polymorphisms (SNP) known to be polymorphic between the parents of the population. The SNPs assayed in this experiment were developed by Monsanto. The primer and probe sequences for the Taqman genotyping assays are available on the request of K. E. Grote at Monsanto. Access requires that the sequences will be used only for noncommercial research and will not be transferred to a third party.

A chi-square test was conducted to identify loci which deviated from the expected 1:2:1 segregation ratio. Experiment-wise type I error rate was controlled by modifying the alpha significance level according to the Bonferroni procedure (Kutner et al., 2005). For 137 independent chi-square tests, the alpha level for any individual test was 0.05/137 or 0.000365. Loci for which the genotypic ratio deviated significantly from expected were further investigated and removed from the set prior to construction of the genetic map. Creation of the genetic map from 135 genotyped loci was done using the Kosambi mapping

function in R/qtl (Broman et al., 2003). Loci from the IBM2 Neighbors maize genetic map (Lee et al., 2002) were included on the Monsanto genetic map, which enabled the identification of public markers within 0.5 cM of the Monsanto SNPs.

The genetic composition of each F2:3 line was calculated by summing the loci which were scored within a genotypic class and dividing that by the total number of loci which had genotypic data (Veldboom et al., 1994).

Phenotypic Data and Analysis

Three replications of the F2:3 lines were grown in single row plots in Hawaii during the 2009/2010 season. The genetic material was blocked by replications and randomized within block. Plants were spaced approximately 15 centimeters apart within a row and the rows were separated by 0.76 meters. At least 5 plants were self- pollinated in each row, harvested at maturity, and dried to approximately 10% moisture.

From each field plot, a balanced bulk of 24 F2:4 kernels was created from self-pollinated ears of F2:3 plants to form the samples for kernel dissection and embryo and endosperm analysis. Kernels were imbibed in water to facilitate removal of the pericarp and separation of embryo and endosperm. Following separation, embryo and endosperm tissues were air-dried overnight at room temperature and weighed. The embryo and endosperm tissues were analyzed for oil percentage via nuclear magnetic resonance using a Maran Ultra-20 from Oxford Instruments. On whole kernel samples, oil percentage was measured by near infrared transmittance (NIT) with a Foss Infratec 1221 NIT Grain Analyzer instrument.

The five traits measured on each sample were whole kernel oil in milligrams oil per gram of kernel (WK), embryo milligrams oil per gram of embryo tissue (EMB), embryo milligrams oil per embryo (EMB_MG), endosperm milligrams oil per gram of endosperm

tissue (ENDO), and endosperm milligrams of oil per endosperm (ENDO_MG). Descriptions of trait name, tissue type, units of measurement, and calculations are listed in Table 1.

Analysis of variance conducted for each trait using R (Ihaka and Gentleman, 1996) and estimates of genetic and phenotypic variance and broad-sense heritabilities were calculated from the observed mean squares. Residuals were analyzed (data not shown) to verify that the assumptions of the analysis of variance were met for each trait. A mean trait value for each F2-derived line was calculated as the average of the three field replications.

From the analysis of variance, genetic variance, $\hat{\sigma}_g^2$, was estimated as (MSG-MSE)/R, where R was equal to the number of replications, MSG was the mean squares for genotypes and MSE for error. Phenotypic variance, $\hat{\sigma}_p^2$, was estimated as MSG/R. From the genotypic and phenotypic variance estimates, broad-sense heritability was calculated as $H^2 = \hat{\sigma}_g^2/\hat{\sigma}_p^2$ and was equivalent to repeatability (Kempthorne, 1969). The 95% confidence intervals for heritabilities were calculated following the protocol of Knapp and Ross (1985). Spearman rank correlations and significance were calculated for all pairs of traits as described by Kutner et al. (2005).

QTL Mapping

Trait means, genetic map positions, and composite interval mapping (CIM) (Zeng, 1994) were used to identify QTL which affected WK, EMB, EMB_MG, ENDO, and ENDO_MG. CIM was conducted using Model 6 of QTL Cartographer Version 2.5009, (Wang et al., 2007) to determine the most likely position of QTLs and their additive and dominance genetic effects. Intervals of 1 cM were scanned and window size was set at 10cM to eliminate the impact of markers within 10cM of the interval under consideration. Markers for background control were determined by forward-backward regression with a probability

of 0.1 for entering or exiting the model. To control for experiment-wise type I error rate, a critical likelihood ratio (LR) value was set for each trait (alpha =0.05) by 1000 random permutations (Doerge and Churchill, 1996). The LR corresponding to $\alpha = 0.05$ level of significance for each trait was WK: 16.62; EMB: 16.48; EMB_MG: 17.27; ENDO: 16.87; ENDO_MG: 16.85.

The presence of a QTL was declared when the LR for the test of the presence of additivity and dominance exceeded the critical LR as determined by the permutation test. QTL location was determined by peak LR. Additive (a) and dominance (d) genetic effects for the QTL within this F2 population and the percentage of the phenotypic variation explained were estimated and reported. Levels of dominance (LD) were calculated as $|d|/|a|$ which was used to determine gene action using the criteria of Bohn et al. (1996). The criteria were: additive (A) $0.0 \leq LD \leq 0.20$; partial dominance (PD) $0.21 \leq LD \leq 0.80$; dominance (D) $0.81 \leq LD \leq 1.2$; over-dominance (OD) $LD > 1.20$. Since only F2:4 24 kernels were phenotyped for each plot, it is possible that genetic sampling did not accurately represent the heterozygous genotypic class from the F2 plant and this could affect estimates of genetic effects. Since the effects of the QTLs were not determined in an independent sample, the effect of the QTL could be biased upward (Melchinger et al., 1998).

RESULTS

Genotypic Data

Two of the 137 genotyped loci had segregation ratios which were significantly different from the expected 1:2:1 for F2 pedigrees. The loci were located on different chromosomes and further analysis of the scoring of the SNP markers indicated that

genotypes were difficult to call. As a result, the accuracy of the genotypic calls reported was questionable and these loci were removed from the data set prior to genetic map creation. The genetic map created from the 135 SNP markers spanned 10 chromosomes and had a total genetic distance of 1569.7 cM (Table 2). The number of markers per chromosome ranged from 9 on chromosome 9 to 20 on chromosome 1. The genetic distance between adjacent markers ranged from 1.4 cM to 41.5 cM with an average spacing of 12.6 cM across the genome.

Within this set of F2-derived lines, the average genetic composition was 24.2% homozygous HO, 24.4% homozygous LO, and 51.3% heterozygous (Figure 1). The standard deviations were 8.6%, 8.3%, and 10.2%, respectively. These estimates of the genetic class proportions meet the expectations for randomly generated F2-derived lines and suggest that there was no unintentional selection which favored the genotype of one parent over the other.

Phenotypic Data

The ranges of the phenotypic values were much greater in WK, and EMB, than ENDO (Table 3). The difference between the smallest and largest values for WK and EMB values were 31 mg g-1 and 186 mg g^{-1}, respectively, whereas for ENDO the range was only 7 mg g^{-1}. Similarly, the phenotypic range for ENDO_MG was much smaller than that of EMB_MG. Embryo tissue contained much more oil than did the endosperm both in mg g^{-1} and in mg. On average, EMB was 32 times greater than ENDO and EMB_MG was 5 times greater than ENDO_MG.

A continuous distribution was observed for all traits (Figure 2), which suggested multi-genic control of phenotypes. All traits had significant differences among progenies ($p<0.01$) and all H^2 were significantly greater than zero (Table 3). The H^2 estimate for WK,

the only trait not dependent upon kernel dissection, was the highest (0.92) and ENDO was the lowest (0.23). Variation in H^2 estimates among traits could be due to differences in underlying genetic factors and field environmental sensitivity, but could have been the result of variation in the precision of dissection and contamination of one tissue type with the other. Phenotyping errors would increase the impact of non-genetic factors on line means and would reduce H^2. The phenotypic ranges for ENDO and ENDO_MG, which were much smaller than those for WK, EMB, and EMB_MG, could be indicative of low genotypic variation which would influence H^2 estimates.

All Spearman rank correlations were positive and statistically greater than 0 (Table 4). The highest correlations with WK were EMB (0.69) and EMB_MG (0.74); in comparison, correlations between WK and the endosperm traits was lower. These estimates are in agreement with the general assumption that maize whole kernel oil content is primarily determined by the embryo and that the endosperm has a relatively small effect on WK (Zheng et al., 2008). Correlations calculated between the two tissues types - EMB-ENDO and EMB_MG - ENDO_MG - were also lower than WK-EMB or WK-EMB_MG and correlations calculated within a tissue type. The highest correlation among all pairs of traits was 0.86 between ENDO and ENDO_MG.

QTL Mapping

Sixteen QTL were identified for four of the five traits, and each QTL explained between 8.8% and 34.8% of the phenotypic variation. The numbers of QTL detected were three for EMB and EMB_MG, one for ENDO, and six for WK (Table 5). No QTL were associated with ENDO_MG at α=0.05 level of significance. Two regions, located on chromosomes 3 and 6, were associated with more than one trait. The region on chromosome

6 was associated with WK, EMB, and EMB_MG and explained 23.5%, 34.8%, and 17.9% of the phenotypic variation, respectively. Of the QTLs associated with WK and EMB, the QTL on chromosome 6 explained the greatest amount of the phenotypic variation for each trait. The favorable allele, which corresponded with the allele conferring higher oil either in mg g^{-1} or mg, originated from HO in 12 of the 13 QTLs identified. The favorable allele originated from LO for *emb-05*.

All types of gene action were detected within this population (Table 5). WK was the only trait which had a QTL effect estimated as OD. The QTL with OD gene action, *wk-03,* co-located with *emb_mg-03* which had D gene action for EMB_MG. The gene action for the QTL on chromosome 6 was A for WK and EMB, but D for EMB_MG. All QTL associated with EMB_MG had D gene action. Across all traits, 31% of QTL were PD, 31% were A, 31% were D, and 7% were OD.

DISCUSSION

Quality of Phenotypic Data and Potential Impact on QTL Detection

Phenotypic data quality and the importance of the genotype in determination of the phenotype are factors which impact identification of trait associated loci and ascertainment of the genetic basis of traits. Within this population, WK H^2, 0.92, fell within the range reported for F2:3 families (Goldman et al., 1994; Mangolin et al., 2004). The H^2 estimates for EMB, EMB_MG, ENDO, and ENDO_MG were all lower than that of WK, and there are no known published examples to which the values obtained in this experiment can be compared. EMB and EMB_MG H^2 were higher than 0.75 and relatively close to WK H^2. In comparison, ENDO and ENDO_MG H^2 were statistically lower than WK and EMB. Several

factors could have contributed to the lower H^2 values for these traits, including differences in underlying genetic factors, trait environmental sensitivity, and phenotyping errors. Given the high H^2 of WK in the population, and the dependency of WK upon oil contents of the embryo and endosperm fractions, phenotyping errors during the dissection process likely contributed to environmental variation and lower H^2 values. If phenotyping errors substantially increased environmental variation, these H^2 values are underestimates of the true effect of genotype on phenotype for these traits. The amount of genetic variation for a trait within a population also impacts H^2 estimates, particularly if genetic variation is low (Gallais and Hirel, 2004; Price and Schluter, 1991) relative to environmental variation.

Conversely, it is possible that the H^2 reported here could be biased upward since all biological replications were grown in the same environment. The use of a common environment causes variation associated with the genotype by environment interaction to be indistinguishable from genetic variance in the analysis of variance, and the effect of environment cannot be tested (Bernardo, 2002). Maize WK has been shown to be sensitive to genotype by environment interactions, but the variance associated with genotype is typically much larger than that associated with the genotype by environment (Clark et al., 2006; Laurie et al., 2004; Yang et al., 2010). Although the interaction between genotype and this specific environment could have contributed to the differences detected between lines, it is unlikely that the phenotypic values were heavily biased by the interaction since oil content of maize kernels is primarily genetically determined. The results of this experiment are reported with the acknowledgement that the specific environment in which the progeny were grown could have influenced the phenotypic values associated with the genotypes, the correlations among traits, and the QTL identified by CIM. Further analysis conducted in

additional environments and potentially utilizing additional bi-parental segregating populations can be conducted and compared to these results to determine the influence of this specific environment and population on the QTL identified within this experiment.

The power to detect QTL for traits with high H^2 estimates should be greater than traits with lower H^2 (Beavis, 1994; Holland, 2004). WK, which had the highest H^2 within this population, also had the greatest number of QTL associated with the trait. In comparison, two QTL were identified for ENDOt, which had the lowest H^2 estimate in this population and no QTL were associated with ENDO_MGt. QTL can be detected for low H^2 traits if the loci explain a large proportion of the genetic variance (Lande and Thompson, 1990), which suggests that the QTL associated with ENDO had major effects. The lack of QTL associated with ENDO_MG also suggests that small effect QTL, which were not detectable in this population with this estimated H^2 for the trait, comprised the genetic basis of ENDO_MG. The number of individuals genotyped and phenotyped and the number of phenotypic replications also may not have been sufficiently large to detect small effect QTL (Lynch and Walsh, 1998; Melchinger et al., 1998), which could have impacted all traits regardless of their H^2 values. Increasing sample size and replications would have a minimal impact on the power of QTL detection for ENDO and ENDO_MG if imprecise kernel dissections have a large impact on their phenotypic values.

QTL Detected for WK, EMB, EMB_MG, and ENDO

The genetic control of oil content of the maize embryo and endosperm could have an impact both on the ability to select high whole kernel oil lines without modification of the embryo-endosperm ratio and the ability to develop lines which accumulate oil preferentially in one tissue. There are few previous studies in which the oil content of the embryo and

endosperm were analyzed independently of whole kernel oil. Consequently, most of the QTL associated with EMB, EMB_MG, and ENDO were unique to this experiment. The one exception was *emb-06* which co-located with a region associated with increased oil concentration of the embryo identified by Zheng et al (2008). Several of the QTL identified for EMB, EMB_MG and ENDO within this population were located in regions of the genome previously associated with WK. Although WK could be considered similar or related to the embryo and endosperm oil traits, comparisons should be made with caution since the traits are not completely equivalent and does not distinguish oil content from the correlated and undesirable changes to embryo-endosperm ratio.

Many of the QTL associated with WK in this population were in similar genomic regions to QTL which had been previously identified in other populations and environments (Alrefai et al., 1995; Mangolin et al., 2004; Yang et al., 2010; Zheng et al., 2008). The only region identified in this study which did not co-locate with at least one previously reported known QTL was *wk-03*. The QTL similarities between this study and previous experiments are particularly notable on chromosomes 6 and 9. The QTL located in the interval defined by M06-03 and M06-04 had the largest effect on WK within this population, explaining almost 24% of the phenotypic variation, and was also associated with EMB and EMB_MG in this experiment. This region had been associated with WK in multiple other studies (Alrefai et al., 1995; Mangolin et al., 2004; Séne et al., 2001; Song et al., 2004; Willmot et al., 2006; Yang et al., 2010; Zheng et al., 2008). Zheng et al. (2008) identified a DGAT locus as the causative gene which impacted both whole kernel and embryo oil concentration in this region. DGAT and other genes which encode enzymes involved in TAG biosynthesis (Thelen and Ohlrogge, 2002; Vigeolas et al., 2007) have been shown to have major effects on

kernel or seed oil content. The QTL associated with WK on chromosome 9 were in the same region as several which had been associated with WK in the BHO population (Song et al., 2004; Yang et al., 2010). This region includes the gene encoding *shrunken1* (*sh1*). Although *sh1* has never been shown to be directly involved in oil biosynthesis or degradation, it does impact kernel starch content (Wilson et al., 2004), and starch content has been shown to be negatively correlated with maize kernel oil content (Séne et al., 2001; Wassom et al., 2008b; Zhang et al., 2008).

The number of QTL declared significant in this population was similar to the number identified in some populations (Wassom et al., 2008b; Yang et al., 2010; Zhang et al., 2008) but much less than others (Alrefai et al., 1995; Berke and Rocheford, 1995; Clark et al., 2006; Laurie et al., 2004; Willmot et al., 2006). Differences in the QTL identified in this study and others could be due to genetic materials, population size, statistical analysis, and marker density. The high oil parent in this experiment had been selected for agronomic characteristics in addition to high WK so it is possible that loci negatively correlated with agronomic traits such as yield (Wassom et al., 2008b) had been selected against and were not present within this population. The size of the segregating population analyzed also may not have provided the statistical power to detect small effect loci and consequently only the regions with large effect were identified (Melchinger et al., 1998).

Phenotypic correlations and common QTL

Phenotypic correlations can be attributed to pleiotropy, genetic linkage, and common environmental effects (Aastveit and Aastveit, 1993), all of which could have contributed to the significant correlations observed among traits in this study. EMB and EMB_MG had the highest correlation with WK, which was expected since the embryo contains much more oil

than does the endosperm, and embryo oil content is considered the primary determinant of WK (Curtis et al., 1968; Watson, 1987; Zheng et al., 2008). In comparison, the correlation between WK and ENDO or ENDO_MG was relatively low.

Typically, traits which have high positive phenotypic correlations tend to share more common QTL than do traits with lower phenotypic correlations (Veldboom et al., 1994; Wang et al., 2008; Wassom et al., 2008a; Yang et al., 2010). Comparisons of QTL identified across traits can provide insight to their shared or unique genetic base, which can have a significant impact on breeding strategies. In this population, the pair of traits with the highest correlation, WK and EMB_MG, also shared the largest number of QTL. Overall, there was relatively little overlap in the QTL regions for each trait with the exception of the QTL on chromosome six, which was declared associated with three traits. The identification of a causative gene associated with whole kernel and embryo oil content in this region by Zheng et al (2008) suggests that a single locus is pleiotropic for the traits. To differentiate between pleiotropy and tightly linked loci in this population, additional fine mapping experiments would have to be conducted in which more loci are genotyped, more progeny evaluated, and additional lines with a high frequency of genetic recombination in this region developed. The unique QTL associated with each trait suggests that at least a portion of the underlying genetic base is independent and that selection for embryo or endosperm-specific oil content could occur.

CONCLUSIONS

The analysis of the phenotypic relationship among maize whole kernel, embryo, and endosperm oil contents and the identification of QTL associated with them in this F2-derived

population provide support that these traits are under partially independent genetic control and that selection for increased embryo and/or endosperm oil content could be used in conjunction with selection for WK in the development of high oil lines. Phenotypic correlations and shared QTL suggest that embryo oil content was the primary determinant of WK, but that there was some variability in endosperm oil content which could potentially be selected upon. Although it is possible that there were QTLs that impact both tissue types which were not detected in this experiment, these results have shown that there is a portion of the loci associated with each tissue which are independent. Overall, the amount of phenotypic variation explained by each QTL was relatively large, which suggests that these regions could be utilized in marker assisted breeding after further investigation in additional populations and environments.

REFERENCES

Aastveit A., Aastveit K. (1993) Effects of genotype-environment interactions on genetic correlations. Theoretical and Applied Genetics 86:1007-1013.

Alrefai R., Berke T.G., Rocheford T.R. (1995) Quantitative trait locus analysis of fatty acid concentrations in maize. Genome 38:894-901.

Beavis, W.D. (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. *In:* D.B. Wilkinson(ed.) 49th Ann Corn Sorghum Res Conf. Am Seed Trade Assoc., Chicago, IL. Pp.250-266.

Berke T.G., Rocheford T.R. (1995) Quantitative trait loci for flowering, plant and ear height, and kernel traits in maize. Crop Science 35:1542-1549.

Bernardo R. (2002) Breeding for quantitative traits in plants Woodbury, Minn.(EUA). Stemma Press.

Bohn M., Khairallah M., González-de-León D., Hoisington D., Utz H., Deutsch J., Jewell D., Mihm J., Melchinger A. (1996) QTL mapping in tropical maize: I. Genomic regions affecting leaf feeding resistance to sugarcane borer and other traits. Crop Science 36:1352-1361.

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

Clark D., Dudley J., Rocheford T.R., LeDeaux J. (2006) Genetic analysis of corn kernel chemical composition in the random mated 10 generation of the cross of generations 70 of IHO x ILO. Crop Science 46:807-819.

Curtis P.E., Leng E.R., Hageman R.H. (1968) Development changes in oil and fatty acid content of maize strains varying in oil content. Crop Science 8:689-693.

Dellaporta S.L., Wood J., Hicks J.B. (1983) A plant DNA minipreparation: version II. Plant Molecular Biology Reporter 1:19-21.

Doerge R.W., Churchill G.A. (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics 142:285-294.

Dudley J., Lambert R.J. (1992) Ninety generations for selection for oil and protein in maize. Maydica 37:1-7.

Gallais A., Hirel B. (2004) An approach to the genetics of nitrogen use efficiency in maize. Journal of Experimental Botany 55:295-306.

Goldman I.L., Rocheford T.R., Dudley J.W. (1994) Molecular markers associated with maize kernel oil concentration in an Illinois High Protein x Illinois Low Protein cross. Crop Science 34:908-915.

Holland, J.B. (2004) Implementation of molecular markers for quantitative traits in breeding programs- challenges and opportunities. Proc. 4th Intl. Crop Sci. Cong., 26 Sep.–1 Oct. 2004, Brisbane, Australia.

Ihaka R., Gentleman R. (1996) R: A language for data analysis and graphics. Journal of computational and graphical statistics 5:299-314.

Kempthorne O. (1969) An introduction to statistical genetics, Ames, Iowa: The Iowa State University Press.

Knapp S., Ross W. (1985) Exact confidence intervals for heritability on a progeny mean basis. Crop Science 25:192-194.

Kutner M., Nachtsheim C., Neter J., Li W. (2005) Applied linear statistical models.

Lambert R. (2001) High-oil corn hybrids, in: A. R. Hallauer (Ed.), Specialty corns, CRC Press, Boca Raton, FL. pp. 131-154.

Lande R., Thompson R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

Laurie C.C., Chasalow S.D., LeDeaux J.R., McCarroll R., Bush D., Hauge B., Lai C., Clark D., Rocheford T.R., Dudley J.W. (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141-2155.

Lee M., N S., WD B., D G., M K., D B., A. H. (2002) Expanding the genetic map of maize with intermated B73 x Mo17 (IBM) population. Plant Molecular Biology 48:453-461.

Lynch M., Walsh B. (1998) Genetics and analysis of quantitative traits. Sinauer Sunderland, MA.

Mangolin C.A., de Souza C.L., Garcia A.A.F., Garcia A.F., Sibov S.T., de Souza A.P. (2004) Mapping QTLs for kernel oil content in a tropical maize population. Euphytica 137:251-259.

Melchinger A.E., Utz H.F., Schön C.C. (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383-403.

Ottaviano E., Camussi A. (1981) Phenotypic and genetic relationships between yield components in maize. Euphytica 30:601-609.

Price T., Schluter D. (1991) On the low heritability of life-history traits. Evolution:853-861.

Séne M., Thévenot C., Hoffmann D., Bénétrix F., Causse M., Prioul J.L. (2001) QTLs for grain dry milling properties, composition and vitreousness in maize recombinant inbred lines. Theoretical and Applied Genetics 102:591-599.

Shen B., Allen W.B., Zheng P., Li C., Glassman K., Ranch J., Nubel D., Tarczynski M.C. (2010) Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. Plant Physiology 153:980-987.

Song X.F., Song T.M., Dai J.R., Rocheford T., Li J.S. (2004) QTL mapping of kernel oil concentration with high-oil maize by SSR markers. Maydica 49:41-48.

Thelen J.J., Ohlrogge J.B. (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4:12-21.

Veldboom L., Lee M., Woodman W. (1994) Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits. Theoretical and Applied Genetics 88:7-16.

Vigeolas H., Waldeck P., Zank T., Geigenberger P. (2007) Increasing seed oil content in oil seed rape (Brassica napus L.) by over expression of a yeast glycerol 3 phosphate dehydrogenase under the control of a seed specific promoter. Plant Biotechnology Journal 5:431-441.

Wang C., Cheng F., Sun Z., Tang J., Wu L., Ku L., Chen Y. (2008) Genetic analysis of photoperiod sensitivity in a tropical by temperate maize recombinant inbred population using molecular markers. Theoretical and Applied Genetics 117:1129-1139.

Wang S., Basten C., Zeng Z. (2007) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC.

Wassom J.J., Mikkelineni V., Bohn M.O., Rocheford T.R. (2008a) QTL for fatty acid composition of maize kernel oil in Illinois High Oil x B73 backcross-derived lines. Crop Science 48:69-78.

Wassom J.J., Wong J.C., Martinez E., King J.J., DeBaene J., Hotchkiss J.R., Mikkilineni V., Bohn M.O., Rocheford T.R. (2008b) QTL associated with maize kernel oil, protein, and starch concentrations; kernel mass; and grain yield in Illinois High Oil x B73 backcross-derived lines. Crop Science 48:243-252.

Watson S. (1987) Structure and composition, in: S. Watson and P. Ramstad (Eds.), Corn Chemistry and Technology, American Association of Cereal Chemists, Inc., St. Paul, MN. pp. 53-82.

Willmot D.B., Dudley J., Rocheford T.R., Bari A. (2006) Effect of random mating on marker-QTL associations for grain quality traits in the cross of Illinois High Oil x Illinois Low Oil. Maydica 51:187-199.

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

Yang X., Guo Y., Yan J., Zhang J., Song T., Rocheford T., Li J.S. (2010) Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. Theoretical and Applied Genetics 120:665-678.

Zeng Z.B. (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

Zhang J., X.Q. L., X.F. S., Yan J.B., Song T.M., Dai J.R., Rocheford T.R., Li J.S. (2008) Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. Euphytica 162:335-344.

Zheng P., Allen W.B., Roesler K., Williams M.E., Zhang S., Li J., Glassman K., Ranch J., Nubel D., Solawetz W., Bhattramakki D., Llaca V., Deschamps S., Zhong G.-Y.,

Tarczynski M.C., Shen B. (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nature Genetics 40:367-372.

Table 1. Description of traits analyzed of maize kernel oil content and calculations used to generate reported phenotype values in F2-derived progeny from a cross of HO and LO. HO = high oil inbred, LO = low oil inbred

Trait[a]	**Tissue**	**Units**[b]	**Calculation**[b]
WK	whole kernel	mg g^{-1}	NIT percent oil *1000
EMB	embryo	mg g^{-1}	NMR percent oil *1000
EMB_MG	embryo	mg	embryo weight (mg)*NMR percent oil
ENDO	endosperm	mg g^{-1}	NMR percent oil*1000
ENDO_MG	endosperm	mg	endosperm weight (mg)*NMR percent oil

[a] WK = whole kernel oil in mg g^{-1}, EMB = embryo oil in mg g^{-1}, EMB_MG = embryo oil in mg, ENDO = endosperm oil in mg g^{-1}, ENDO_MG = endosperm oil in mg

[b] NIT=Near Infrared Transmittance, NMR=Nuclear Magnetic Resonance, mg=milligrams, g=grams

Table 2. Summary of genetic map produced with 135 segregating loci genotyped in 85 F2:3 progeny from a cross between HO and LO. HO = high oil inbred, LO = low oil inbred

Chromosome	Number of Marker Loci	Chromosome Length (cM)[a]	Average Spacing (cM)[a]	Maximum Spacing (cM)[a]
1	20	216.2	11.4	28.5
2	13	149.6	12.5	37.3
3	15	193.7	13.8	27.6
4	14	150.9	11.6	26.8
5	18	157.3	9.3	24.0
6	9	167.3	20.9	39.2
7	17	149.3	9.3	21.4
8	14	151.9	11.7	26.6
9	7	91.7	15.3	35.9
10	8	141.7	20.2	41.5

[a] cM=centiMorgans

Table 3. Estimated trait phenotype means, ranges, and heritabilities of maize kernel oil phenotypes in the F2-derived lines from a cross between LO and HO. HO = high oil inbred, LO = low oil inbred

Trait[a]	Mean[b]	Range[b]	Least Significant Difference	H^2 (95% confidence interval)[c]
WK	64 mg g^{-1}	47-78 mg g^{-1}	5.41	0.92 (0.88-0.94)
EMB	391 mg g^{-1}	304-490 mg g^{-1}	48.5	0.84 (0.77-0.89)
EMB_MG	16 mg	10.8-20.4 mg	2.75	0.79 (0.70-0.85)
ENDO	12 mg g^{-1}	9-16 mg g^{-1}	10.1	0.23 (-0.12, 0.46)
ENDO_MG	3.3 mg	2.1-5.3 mg	1.21	0.40 (0.11, 0.58)

[a] WK = whole kernel oil in mg g^{-1}, EMB = embryo oil in mg g^{-1}, EMB_MG = embryo oil in mg, ENDO = endosperm oil in mg g^{-1}, ENDO_MG = endosperm oil in mg

[b] mg = milligram, g = gram

[c] H2 = broad-sense heritability

Table 4. Spearman rank correlations between trait phenotypic values with significance levels of 0.05 (*), 0.01 (**), and 0.001 (***) in F2-derived lines from a cross between LO and HO. HO = high oil inbred, LO = low oil inbred

Trait[a]	EMB	EMB_MG	ENDO	ENDO_MG
WK	0.69***	0.74***	0.33**	0.18
EMB		0.59***	0.34**	0.26*
EMB_MG			0.22*	0.32**
ENDO				0.86***

[a] WK = whole kernel oil in mg g^{-1}, EMB = embryo oil in mg g^{-1}, EMB_MG = embryo oil in mg, ENDO = endosperm oil in mg g^{-1}, ENDO_MG = endosperm oil in mg, mg = milligram, g = gram

Table 5. QTL for maize kernel oil content detected by composite interval mapping in segregating F2:4 progeny lines from a cross between HO and LO.

Trait	QTL code[a]	Chromo-some	Flanking loci, Monsanto[b]	Flanking loci, public[b]	**Estimated position of QTL (cM)[c]**	**Additive genetic effect[d]**	**Dominance genetic effect[d]**	**Gene action and LD estimate[e]**	**Origin of favorable allele[f]**	R^2_p [g]
WK	*wk-01*	1	M01-03 – M01-04	csu1190-IDP2552	35.3	3.7	-1.2	PD, 0.32	HO	10.8%
	wk-03	3	M03-08 – M03-09	IDP1974-plt1	79.2	2.9	-3.9	OD, 1.34	HO	20.8%
	wk-06	6	M06-03 – M06-04	IDP1966-AI737983	72.5	4.9	0.25	A, 0.05	HO	23.5%
	wk-07	7	M07-13 – M07-14	npi263-bnl14.34	114.4	3.1	-0.1	A. 0.03	HO	10.1%
	wk-09-1	9	M09-01 – M09-02	IDP1969-chr113	2.1	2.5	2.6	D, 1.04	HO	10.0%
	wk-09-2	9	M09-02 – M09-03	chr113- les8	10.2	3.2	1.7	PD, 0.53	HO	10.9%
EMB	*emb-05*	5	M05-12 – M05-13	IDP8535-TIDP3037	99.5	14.0	3.8	PD, 0.27	LO	8.8%
	emb-06	6	M06-03 – M06-04	IDP1966-AI737983	72.5	29.5	2.5	A, 0.08	HO	34.8%
	emb-08	8	M08-03 – M08-04	IDP4220-umc1807	43.0	21.2	-9.1	PD, 0.43	HO	20.8%
EMB_MG	*emb_mg-03*	3	M03-08 – M03-09	IDP1974-plt1	77.9	1.16	-1.03	D, 0.89	HO	18.5%
	emb_mg-06	6	M06-03 – M06-04	IDP1966-AI737983	75.5	1.06	1.04	D, 0.98	HO	17.9%
	emb_mg-07	7	M07-08 – M07-09	his2b1-mmc0411	68.8	1.11	1.02	D, 0.92	HO	19.8%
ENDO	*endo-07*	7	M07-13 – M07-14	npi263-bnl14.34	106.7	0.08	0.006	A, 0.08	HO	20.10%

Table 5, continued.

[a] QTL were designated using the trait name followed by a number indicating the chromosome. If more than one QTL for a trait was identified on a chromosome, a sequential number was added.

[b] Markers flanking the QTL. The public markers were included on the Monsanto reference genetic map and are located within 0.5 cM of the Monsanto marker

[c] Position estimated by the peak LR associated with the QTL

[d] Genetic effects estimated by QTL Cartographer

[e] Gene action estimated as |d|/|a| and LD (level of dominance) determined using the thresholds set by Bohn et al. (1996)

[f] HO = high oil parent, LO = low oil parent

[g] Proportion of phenotypic variation explained by the QTL

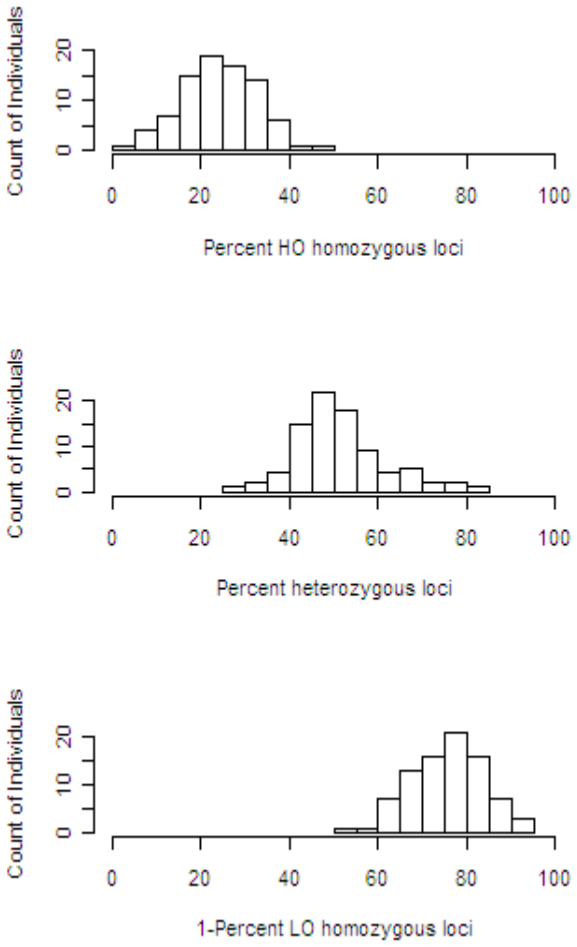

Figure 1. Genetic composition of F2:3 progeny generated from a cross between HO and LO. HO = high oil inbred, LO = low oil inbred

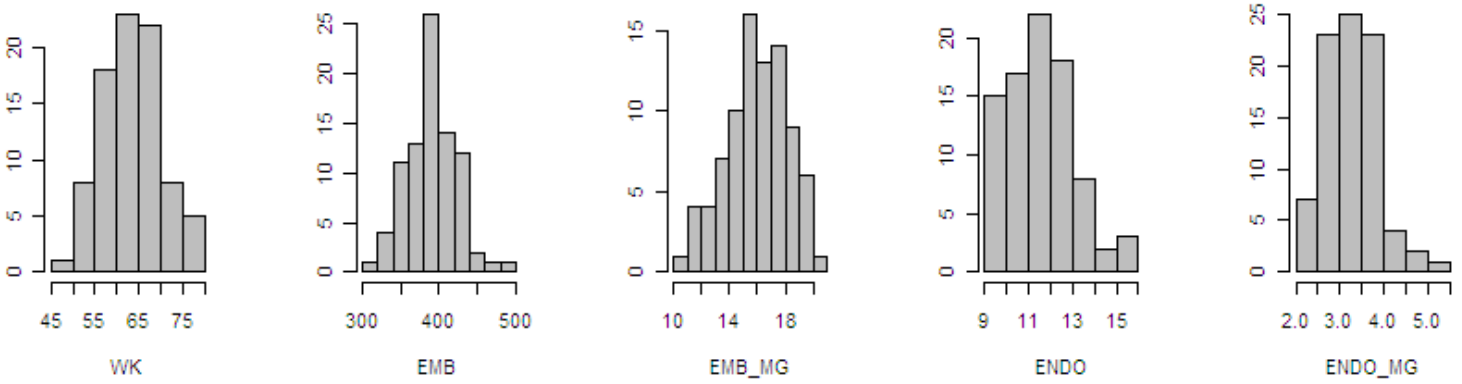

Figure 2. Distribution of phenotypic values in F2:3 progenies generated from a cross between HO and LO evaluated in F2:4 generation. ENDO and ENDO_MG distributions were created using the untransformed data. HO = high oil inbred, LO = low oil inbred, WK = whole kernel oil in mg g^{-1}, EMB = embryo oil in mg g^{-1}, EMB_MG = embryo oil in mg, ENDO = endosperm oil in mg g^{-1}, ENDO_MG = endosperm oil in mg, mg = milligram, g = gram

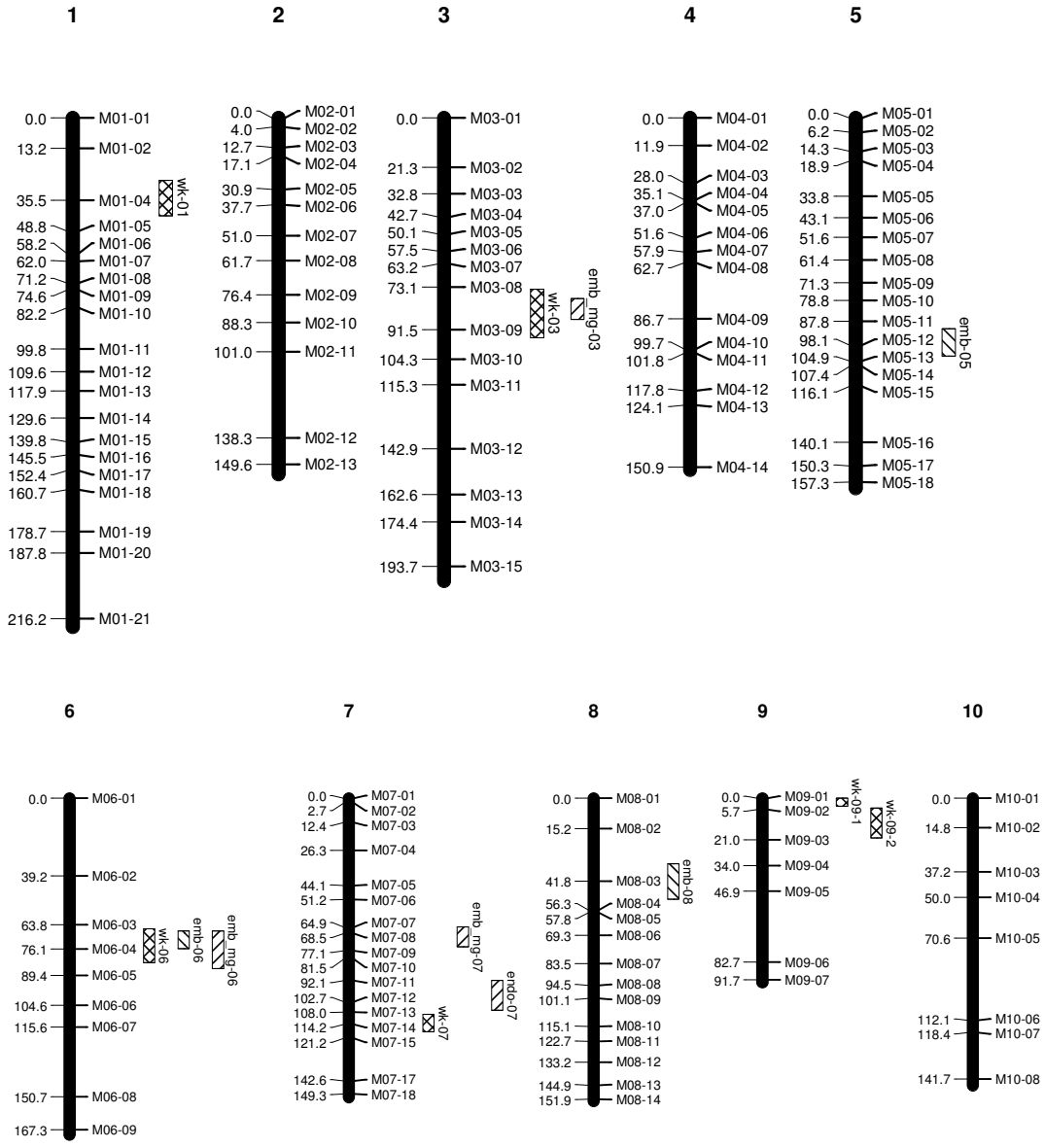

Figure 3. Genetic map of ten maize chromosomes constructed with 135 loci and location of QTL identified in 85 F2-derived progeny of a cross between LO and HO. Genetic map position in centiMorgans (cM) is listed to the left and marker name to the right on the map. The 1-LOD confidence intervals for the 13 QTLs are indicated by the boxes to the right of the chromosomes. QTLs were named according to the trait to which the region was associated (wk, emb, emb_mg, and endo), followed by the chromosome number, and a sequential identifying number if more than a single QTL was identified for a trait on a single chromosome. LO = low oil inbred, HO = high oil inbred

CHAPTER 4

GENERAL CONCLUSIONS

Quantitative trait loci for maize kernel oil traits were mapped in a panel of commercially relevant inbred lines via association mapping and by linkage mapping in F2-derived progeny from a cross between a high whole kernel oil content inbred and a low whole kernel oil content inbred. Phenotypic relationships, trait repeatability, and impact of genotype by environment interaction were also assessed.

Analysis of the association mapping inbred lines across multiple environments indicated that environment, genotype, and genotype by environment had significant effects on whole kernel, embryo, and endosperm oil traits. The variance associated with genotype was at least 10 times larger than either environment or the interaction between genotype and environment. Repeatability, or H^2, was greater than 0.75 for all traits in the association mapping population. Since the effect of the genotype was much greater than either the effect of the environment or the interaction between genotype and environment for all traits in the inbred lines, power of QTL detection via association mapping should have been relatively high. In the F2:4 progeny, H^2 estimates could have been biased upward by genotype by environment variance.

Most maize kernel oil traits were significantly correlated, but the magnitude and direction of the correlation were trait and germplasm dependent. The low correlation between EMB and ENDO and EMB_MG and ENDO_MG suggested that the oil contents of the tissues were independently controlled. The EMB-ENDO correlation was not statistically greater than zero in OM In F2-derived progeny, EMB-ENDO and EMB_MG-ENDO_MG were 0.34 and 0.32, respectively. The relationship between WK and embryo and endosperm

oil were also germplasm dependent. Within OM, EMB-WK and EMB_MG-WK were lower and ENDO-WK was higher than in CM.

A total of 173 associations were made between 129 loci and at least one of six traits via association mapping in inbred lines. Between six (ENDO) and 60 (WK) loci were detected. No loci were associated with ENDO_MG. There was no overlap in the QTL associated with EMB or EMB_MG and ENDO which provided evidence of partially independent genetic control of oil content. Investigation of the genomic sequence at the loci associated with traits at FDR<0.001 identified several putative genes with known or possible function in oil metabolism. The function of the putative loci suggested that metabolic processes which occur after fatty acid synthesis are important in determination of final oil content.

Linkage mapping in F2-derived progeny identified between two and six regions of the genome associated with the traits WK, EMB, EMB_MG, ENDOt. No regions of the genome were linked to ENDO_MGt. WK, EMB, EMB_MG, and ENDOt all had QTL which were located in a 10cM region on chromosome six. Each trait was also associated with at least one QTL which did not co-locate with QTL associated with a different trait.

The loci which were used in map construction and QTL identification in the bi-parental population were not coded equivalent to those utilized for association mapping. The loci used in the two studies could be placed on the Monsanto genetic map and several comparisons could be made. The region on chromosome 6 which was associated with WK, EMB, EMB_MG, and ENDOt in the bi-parental population contained several associations with EMB and EMB_MG in the association mapping analysis.

Because these studies were conducted using commercially relevant germplasm, the results obtained here should have application to commercial genetic material. The combined results of the phenotypic evaluation and QTL detection suggest that embryo and endosperm oil contents have partially independent genetic control, which control could be important for the development of new inbreds and hybrids with tissue specific oil accumulation. The QTL detected herein could be used to augment phenotypic selection for whole kernel oil content, or could be used replace expensive and time consuming phenotypic evaluation of embryo and endosperm oil content.

APPENDIX: FORMULAE USED IN STATISTICAL ANALYSIS

Linear Model for Analysis of Phenotypic Data

The linear model for analysis of phenotypic data in the association mapping population was:

$$Y_{ijk} = \mu + E_i + G_j + GE_{ij} + \varepsilon_{ijk}$$

where μ is the mean for the trait, E_i is the random effect of the environment, G_j is the fixed effect of the genotype, and GE_{ij} is the random effect of the interaction between genotype j and environment i, and ε_{ijk} is the residual variance.

The linear model for the analysis of the phenotypic data in the linkage mapping population was:

$$Y_{ij} = \mu + G_i + \varepsilon_{ij}$$

where μ is the mean for the trait, G_i is the fixed effect of the genotype, and ε_{ij} is the residual variance.

Broad-Sense Heritability

Broad-sense heritability was equivalent to repeatability and the intraclass correlation and as the proportion of genetic variation relative to the sum of genetic and environmental variation:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

Broad-sense heritability on a progeny mean basis was estimated using the methodology outlined by Knapp et al. (1985) which used the following formula:

$$H^2 = 1 - M_2/M_1$$

where M_1 is the mean square associated with genotype in the analysis of variance. M_2 is the mean square associated with environment in the analysis of variance of the phenotypic data in the linkage mapping population. In the analysis of variance of the phenotypic data

obtained from the association mapping population, M_2 is the mean square associated with the genotype by environment interaction. The confidence intervals for H^2 estimates were calculated according to Knapp et al. (1985).

Pearson and Spearman Phenotypic Correlations

The Pearson correlation between two phenotyped traits *i* and *j* was estimated as:

$$\hat{r}_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j}$$

where $\hat{\sigma}_{ij}$ is the phenotypic covariance between traits *i* and *j*, $\hat{\sigma}_i$ is the square root of the phenotypic variance for trait *i*, and $\hat{\sigma}_j$ is the square root of the phenotypic variance for trait *j*.

The Spearman rank correlation is a nonparametric measure of association based upon the ranks of the data for two phenotyped traits. The formula is:

$$\hat{\theta} = \frac{\sum_i (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_i (R_i - \bar{R})^2 (S_i - \bar{S})^2}}$$

where R_i is the rank of x_i, S_i is the rank of y_i, $\bar{R}$ is the mean of the R_i values, and $\bar{S}$ is the mean of the S_i values.

Statistical differences between estimated $\hat{r}_{ij}$ or $\hat{\theta}$ and zero were calculated according to Kutner et al. (2005). For each estimated correlation, a t-statistic was calculated as

$$\mathrm{t} = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$$

and degrees of freedom were equal to n-2.

Chi-square Tests for Deviation from Expected Frequencies

For each locus the expected genotypes in the F2 population are AA (HO homozygote), Aa (heterozygote), and aa (LO homozygote) with an expected ratio of 1:2:1. Deviation from this expected ratio was tested with a chi-squared test:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where O_i was the observed count of individuals which had a genotypic class i and E_i was the expected count of individuals which had a genotypic class i. The calculated chi-square statistic, X^2, was compared to the central chi-squared distribution with two degrees of freedom to estimate a p-value.

Experiment-wise type I error rate was controlled by modifying the alpha significance level according to the Bonferroni procedure. For 137 independent chi-square tests, the alpha level for any individual test was 0.05/137 or 0.000365.

ACKNOWLEDGEMENTS

I would first like to thank my advisor Dr. Michael Lee for his support and guidance during my Ph.D. program. His influence has particularly affected how I approach scientific problems and communication of both the scientific question and results.

I also need to thank Monsanto for granting me the opportunity to complete my Ph.D. while working in a part/full-time status. There are people too numerous to mention all who I have gained knowledge from, both in regard to science and how business should be conducted. Additionally, the ability to devote some of my working time to my Ph.D. has been instrumental in completion of this degree.

My Program of Study committee deserves many thanks. Through interactions during this Ph.D. program, in classes, and in discussions regarding the research, this group has helped me immensely during this course of study. I would like to particularly thank Maria Salas-Fernandez for her input on the association mapping analysis and interpretation.

Finally, I would like to thank my family and friends for supporting me through this period of time. My husband, Jesse, always reminded me of the importance of life outside of school and work and my son, George who arrived midway through the course of studies, who doesn't yet care about agronomy, genetic mapping, or plant breeding but brought much joy.