CANDIDATE GENE ASSOCIATION MAPPING

IN SPRING WHEAT

by

Jay Robert Kalous

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Plant Science

MONTANA STATE UNIVERSITY
Bozeman, Montana

January 2011

APPROVAL

of a thesis submitted by

Jay Robert Kalous

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency and is ready for submission to the Division of Graduate Education.

Dr. Luther E. Talbert

Approved for the Department of Plant Sciences and Plant Pathology

Dr. John E. Sherwood

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Jay Robert Kalous

January 2011

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

ABSTRACT

Association mapping (AM) is a form of quantitative trait locus (QTL) mapping that utilizes a collection of germplasm rather than a structured mapping population. Marker/trait associations are made through the application of a mixed-effects model that corrects for population stratification. The objective of this study was to evaluate the application of association mapping on a selection of elite spring wheat cultivars. We tested marker/trait associations for known "perfect" markers and markers identified as controlling traits of interest through traditional bi-parental mapping. We also wanted to evaluate the observed linkage disequilibrium (LD) surrounding genes of interest by utilizing closely linked sets of markers in specific regions of the spring wheat genome. Population structure was estimated with fifty-one unlinked microsatellite markers. Two phenotypic datasets were used for evaluation. The first was an unbalanced historical dataset, and the second was a balanced dataset taken from a two year replicated field trial. Marker/trait associations were identified for plant height, stem solidness, heading date, grain protein content, test weight, and seed color. Our analyses identified significant associations between *Rht-D1* and plant height, *Ppd-D1* and heading date, and *Xgwm340* and stem solidness. No associations were identified between *Rht-B1* and plant height, *Ppd-B1* and heading date, nor *Vrn-B1* and heading date. The extent of LD varied depending on breeding history and selection pressure. One LD block was identified around the stem solidness QTL and two blocks were found surrounding a productive tiller QTL. Smaller blocks of LD were observed surrounding the three genes controlling kernel color. No LD was observed surrounding the *Rht-B1* locus.

CHAPTER 1

INTRODUCTION

Molecular markers have been considered useful to the plant breeder for cultivar improvement since the 1980's (Helentjaris et al. 1986). The invention of the molecular marker was hailed as the end to the costly, time-consuming process of conventional breeding based primarily on phenotypic observation. The genetic marker was going to allow plant breeders to identify genes of interest controlling traits such as disease and pest resistance, plant height, seed quality characteristics, and most important of all agronomic traits, yield. The idea behind a molecular marker is that it allows geneticists and plant breeders to mark specific mutations or alleles in an organism's genetic code that give rise to alternative forms of a given trait. Molecular markers useful in wheat include randomly amplified polymorphic DNA (RAPD) markers, simple sequence repeat (SSR) or microsatellite markers, and single nucleotide polymorphism (SNP) markers, to name a few.

The idea of using genetic markers for breeding is referred to as marker assisted selection, and it was supposed to drastically reduce the 7-10 year timeframe typically associated with the development of a new cultivar (Lande 1990). But to this day it still takes roughly 7-10 years to develop a new plant variety or hybrid and crop improvement is still mainly based on phenotypic observations made from evaluating replicated field trials. Marker assisted selection has had some success when applied to a single gene system or when a small number of genes is involved (Kuchel et al. 2007). The

development of linkage maps, from bi-parental populations has allowed for the association of markers with important traits that have later been in used for marker assisted selection (MAS). Markers have been useful in many situations such as parental selection, robust disease control through gene pyramiding, identification of soil toxicity tolerance genes (Navakode et al. 2009), and in back-crossing schemes where gene inheritance can be tracked without the need for phenotypic observations. Many studies have shown success in identifying quantitative trait loci (QTL) controlling important agronomic characteristics (Chapman et al. 2003, Sherman et al. 2010, Tasma et al. 2001).

Thus far in wheat, marker/trait associations have only been made through the application of genetic markers by way of traditional bi-parental mapping in plants. There is however a large gap between identifying important QTL and applying marker assisted selection to breeding material (Xu and Crouch 2008). A new method poised to close this gap is association mapping (AM), also known as association analysis or linkage disequilibrium mapping.

Association mapping was originally developed for the study of animal genetics (Lander and Schork 1994) where convenient mapping populations were infeasible to develop. AM relies on identifying novel QTL present in the currently available germplasm. Instead of creating a population to study by deriving genotypes from a cross between two parents for study, the geneticist develops a population by selecting from previously developed genotypes with varying and often unknown genetic relationships. The principles of linkage disequilibrium are the foundation for QTL discovery in

association mapping. The novel idea of association mapping is identifying novel QTL in the target populations of germplasm.

One advantage of association mapping is that complex, polygenic systems are not disrupted, as is the potential with traditional mapping populations. Also bi-parental populations suffer with regard to the precision of locating novel QTL due to reduced genetic diversity. Only two alleles for any one gene can be evaluated when multiple alleles could potentially exist. Association mapping can exploit the multi-allelic nature that can occur for traits of interest, which allows for more precise characterization of important genes. A third important difference between traditional QTL mapping and association mapping involves the collection of phenotypic data. Bi-parental populations usually don't have phenotypic information and must be analyzed in the greenhouse or field after the populations is created. By making germplasm selections from lines that exist in advanced stages of the breeding process association mapping can potentially utilize historical phenotypic data already collected, reducing costs and time.

A problem with association mapping occurs due to unknown genetic relatedness that likely exists in germplasm from a breeding program causing the power of association mapping to be reduced due to a high type II error rate. The natural relatedness of lines can cause the markers to be associated with a trait due to familial descent, rather than actual genetic linkage. Therefore population stratification must be well characterized and understood before analyzing any data. The relatively new application of association mapping in plant systems further complicates its usage. A clear cut method on the application of AM does not exist and given the subjective nature of the analysis a

straightforward method may never exist. As with traditional QTL mapping, many steps in the AM process are open to interpretation by the researcher and before successful identification of novel QTL can be declared, extensive confirmation steps must be taken.

Association mapping is also statistically intensive requiring mixed-effects modeling and multivariate analyses. Additional understanding of combining experimental data via least squares means estimation or best linear unbiased prediction is necessary when working with unbalanced historical data. But most importantly the researcher needs to obtain or have access to someone with a strong familiarity with the population. This way the subjective leaps that are sometimes taken have a basis in fact. For example, when deciding the proper number of sub-populations to adequately correct the population stratification statistical analyses can reduce this number to a narrow range of options but ultimately the choice is up to the breeder.

The purpose of this study is to assess the feasibility of applying association mapping to a selection of 94 elite spring wheat genotypes evaluated across the state of Montana. Phenotypic data consists of phenotypic means resulting from an unbalanced historical data set from 1997, 2002, and 2007, and means derived from two balanced, replicated, single-row yield trials evaluated in 2009 and 2010 in Bozeman, MT. The questions we are investigating are: Can associations be made between important genes and the markers known to control those genes? Are marker-trait associations consistent when we use historical phenotypic averages derived from an unbalanced, multi-environment dataset compared to phenotypic means from a completely balanced replicated field trial? Are similar results obtained with AM when we screen markers of

interest identified through traditional QTL mapping? And finally, what is the extent of linkage disequilibrium surrounding genes controlling traits of interest under intense selection.

CHAPTER 2

LITERATURE REVIEW

Microsatellites

Simple sequence repeats (SSRs) or microsatellite loci are a common feature found in non-coding regions of DNA and a great source of variation in wheat (Tautz and Renz 1984). SSRs are repeated motifs of genetic sequence, such as $(CA)_n$ or $(GC)_n$ and can vary in length between two individuals of the same species. If the motif differs in size between two individuals the locus is polymorphic and can differentiate the individuals at the microsatellite locus. The genetic variation in length can arise due to unequal crossing over and mispairing during DNA replication and recombination or extension of single-stranded ends of DNA (Tautz et al. 1986). Sequences flanking microsatellites are often conserved, however. Primers are designed from the flanking regions by screening bacterial artificial chromosome (BAC) clones for a specific microsatellite sequence and sequencing the flanking regions of DNA. SSR primers are then used in polymerase chain reactions (PCR) to amplify the microsatellite locus and allow for the possible observation of polymorphisms between two individuals.

Common bread wheat has an inherently low level of polymorphism, compared to other plant species, due to a self-pollinating breeding system (Devos and Gale 1992) and two genetic bottlenecks that occurred during the origination of the species (Haudry et al. 2007). The co-dominant nature, repeatability, and high level of polymorphism relative to

other marker systems make microsatellite markers well suited for individual genetic identification in wheat (Roder et al. 1995). For the time being, microsatellites adequately cover the entire wheat genome and allow for linkage maps to be constructed and lead to the identification of genes controlling phenotypic traits of interest to plant breeders (Roder et al. 1998).

Linkage Mapping

Genetic markers, including SSRs, can be used to construct genetic maps of culturally important crops. Linkage mapping allows scientists to identify genetic markers that are associated with key genes controlling agronomic and grain quality characteristics. Both Mendelian gene systems and quantitative trait loci are evaluated. Wheat microsatellite markers have been used to identify genes controlling stem solidness, plant height, bread quality, and various other useful traits such as disease resistance (http://maswheat.ucdavis.edu). The first step in linkage mapping involves the development of a bi-parental population derived from two individuals showing phenotypic variation for a trait of interest. A population can be developed in several ways each with their own advantages and disadvantages. A common population is called a recombinant inbred line (RIL) population developed from two individuals that vary for a trait of interest.

A RIL population starts with a cross between two unique parents resulting in offspring that share, 50% of their genetic identity with each parent. The offspring are subsequently advanced several generations by self-pollination reducing genetic

heterozygosity. When lines reach an acceptable level of homozygosity the population is evaluated for variation in traits of interest. The population is also screened with genetic markers that are polymorphic between the two original parents. The genotypic and phenotypic data is then analyzed together to identify novel marker-trait associations. Recombination frequency of a marker screened on the population is used to estimate the distance between the marker, a trait of interest, and other markers tested.

Marker Assisted Selection

Traditional plant breeders have relied on phenotypic selection through replicated experimental trials for cultivar enhancement. Phenotypic selection has the potential to be costly and time consuming depending on the trait. Once molecular markers are identified as being linked to traits of interest plant breeders can utilize this information to aid in cultivar development by genotypic selection. Genotypic selection can bypass costly phenotyping methods (Gupta et al. 2010). For example, wheat breeders can pyramid different genes offering unique modes of resistance to a plant disease, such as stripe rust (Castro et al. 2003) and leaf rust (Kloppers and Pretorius 1997). A variety with multiple modes of resistance is more durable (Boskovic and Boskovic 2009) and has the potential to prolong the eventual evolution in the pathogen for the ability to overcome that resistance. Genetic pyramiding is very difficult to accomplish by phenotypic selection unless each gene of interest results in a unique phenotypic reaction as opposed to measuring a resistant or susceptible response.

Linkage Disequilibrium

Linkage disequilibrium (LD) is the co-inheritance of alleles at two different loci at frequencies different than expected under random chance. Linkage disequilibrium is not dependent on genomic location and perhaps a more appropriate term would be gametic disequilibrium. LD can occur between two loci on different chromosomes, such as when epistasis changes the fitness of a specific genotype. Several factors affect the extent and decay of linkage disequilibrium. The extent of LD can be increased by drift (Farnir et al. 2000), selection resulting in a reduction of genetic variability, or a bottleneck/founder event. All three cases reduce the inherent genetic variability previously associated with the population. If fewer alleles are available for two genes the frequency of the remaining alleles increases, increasing the likelihood for co-inheritance of alleles at the two loci. Naturally occurring genetic mutations will also cause an initial increase in linkage disequilibrium because the new allele frequency is not represented by all possible gametic combinations.

Recombination between loci causes LD to decay, by potentially increasing the genetic variability. Breeding system also greatly affects the extent of LD by affecting the rate in which LD decays. An out-crossing species is more likely to have greater LD decay because the entire population's genetic diversity is available every generation, to every individual producing offspring through cross-pollination, in turn increasing the potential effectiveness of recombination between unique loci. A typical self-pollinating species is limited in available genetic material to that contained within each individual reducing the effectiveness of recombination.

Three main measurements of linkage disequilibrium exist which are *D*, *D′*, and r^2. Lewontin and Kojima (1960) first suggested *D* as a measurement of LD where two segregating loci *A* and *B* with four possible alleles *A*, *a*, *B*, and *b*, respectively, *D* is equal to the product of the coupling gametic frequencies minus the repulsion gametic frequencies.

$$D = g_{11}\, g_{00} - g_{10}g_{01}$$

g_{11} , equals the frequency of the gamete *AB*
g_{10} , equals the frequency of the gamete *Ab*
g_{01}, equals the frequency of the gamete *aB*
g_{00}, equals the frequency of the gamete *ab*

In the absence of evolutionary pressure *D* will approach zero after successive generations (Lewontin and Kojima 1960). *D* ranges from -0.25 to 0.25, as would be the case if gametes are present only in repulsion or only in coupling, respectively. *D* is dependent on the allele frequencies of the two loci therefore the maximal values of *D* will vary from locus to locus.

In order to standardize the value of *D* across loci, Lewontin (1964) later proposed the statistic *D′*.

$$|D'| = D/D_{max}$$
If $D < 0$, $D_{max} = min(pApB, qAqB)$
If $D > 0$, $D_{max} = min(pAqB, qApB)$

pA and qA, equal the frequency of allele *A* and *a* respectively, and
pB and qB, equal the frequency of allele *B* and *b* respectively.

$|D'|$ falls within the range of zero to one allowing the researcher to assess the extent of LD between two loci, taking into account the allele frequencies.

The third statistic for estimating the extent of linkage disequilibrium between loci is the correlation coefficient r^2, which is actually a measurement of the correlation between allelic states at two different loci (Hill 1968). r^2 varies from zero to one if the allele frequencies are equal.

$$r^2 = \frac{D^2}{pAqApBqB}$$

pA and qA, equal the frequency of allele *A* and *a* respectively,
pB and qB, equal the frequency of allele *B* and *b* respectively.

The calculations are straight forward for *D*, *D'*, and r^2 when working with bi-allelic markers. The calculations need to be adjusted when the number of alleles is greater than two for a locus. One possible modification is to weight the gametic combinations based on allele frequency. When loci are multi-allelic the values for *D*, *D'*, and r^2 are calculated for every gametic combination and then weighted based on the frequency of each allele within the population (Hedrick 1987).

LD is dependent on the frequency of alleles in the population being tested, therefore common significance thresholds indicating linkage disequilibrium due to proximity may not be transferable across populations. There is little information in the literature for setting significance thresholds indicating physical linkage when using the statistics that measure linkage disequilibrium. Given the population dependent nature of linkage disequilibrium the best method may be to develop unique thresholds for the specific population in question. Breseghello and Sorrells (2006) set a significance threshold indicating physical linkage, by square root transforming unlinked r^2 values and

taking the parametric 95th percentile of that distribution to be the population-specific significance threshold indicating LD due to physical linkage.

Association Mapping

Association mapping (AM), association analysis, or linkage disequilibrium mapping is a gene mapping analysis based on linkage disequilibrium. The idea is that traits of interest first arise as a mutation in one individual, creating a new allele for a trait. If this new allele improves the individual's fitness it is passed on to subsequent generations. Loci surrounding the advantageous allele are also passed on, creating a block of loci co-inherited non-randomly. Recombination will slowly break down the block of co-inherited loci proportional to the distance from the beneficial allele. Genes that are closest to the mutation will continue to be co-inherited throughout generations. Association mapping seeks to exploit this phenomenon by allowing plant breeders to tag the genetic material surrounding traits of interest with molecular markers. The markers themselves can then be used as genetic flags for cultivar improvement. The goal of association mapping is the same as that of linkage mapping; the methods for achieving this goal however, are different. Several studies in multiple animal species and a few plant species with multiple forms of molecular markers have been completed and reported in the literature (Crossa et al. 2007, Karlsson et al. 2007).

Association mapping populations consist of germplasm selections rather than developing large bi-parental populations, as is done for linkage mapping. The fact that association mapping does not require a bi-parental population may explain why AM

methodology is more widely used in animal genetics (Andersson 2009). Association mapping in animals typically utilizes family-based methods of testing including case-control studies and transmission/disequilibrium tests (Schulze and McMahon 2002). Association mapping in plants typically utilize a linear mixed model approach (Yu et al. 2006). The linear mixed model combines fixed and random effects to help account for population stratification common in typical plant breeding programs. A difficult problem with association mapping in plants revolves around discriminating alleles of interest linked to important traits from alleles in common due to identity by descent. Population substructure can result in spurious marker trait associations reducing the power of association mapping studies, in plants. Two main types of association mapping are done in plants including genome-wide studies and candidate-gene studies. Genome-wide association mapping involves screening a selected population of germplasm with a large set of molecular markers believed to be dispersed throughout the genome. This set of markers is used to estimate the population structure and tested for significant association with a trait of interest. Marker coverage is dependent on the researcher's resources as well as the extent of linkage disequilibrium in the population under study. Candidate-gene association mapping generally consists of two sub-sets of molecular markers. The first set of genetic markers is selected to accurately estimate the population structure with as few markers as necessary. The second subset of markers targets regions of the populations genome thought to be important in the traits of interest.

The general model, for associating a marker and trait, was originally proposed by Henderson (1975) and is depicted below,

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

Where y refers to a vector of observations from the population, β is an unknown vector of fixed effects including the molecular marker and population structure information, u is the random, unknown, additive background genetic effects, X and Z are known incidence matrices. X also called the Q matrix consists of the actual genotypic data and a matrix containing values estimating the population structure. Z also called the K matrix consists of a pair-wise comparison matrix or kinship matrix, and is used to estimate the covariance of each individual's genetic background effect. The term e refers to the residual error in the model. The population's genetic diversity will control whether the full model is best or if a reduced model is more appropriate for the actual analysis.

Several methods exist for estimating underlying population structure, including a Bayesian method implemented in the software STRUCTURE (Pritchard et al. 2000), and two multivariate analyses, principal components analysis (Price et al. 2006), and multi-dimensional scaling (Li et al. 2010). The Bayesian method has the advantage of being model based where initial parameters are set by the researcher; however the rate of calculation is directly proportional to the number of individuals in the population and the number of markers screened. When the number of markers screened is in the thousands the computational time involved can be extensive. The multivariate analyses take much less computational time, dependent only on the number of individuals in the population. However, missing data must first be imputed in order to complete the analysis and the formatting for the marker data is somewhat unconventional when using multi-allelic markers.

The kinship (K) matrix can be developed through a wide variety of methods, (Loiselle et al. 1995, Ritland and Ritland 1996, Queller and Goodnight 1989, Lynch and Ritland 1999, Wang 2002). The software SPAGeDi can calculate up to nine different kinship coefficients. Pedigree information can also be used in estimating the genetic correlation among individuals in the population (Colonna et al. 2009). It should be noted that the kinship matrix is used to estimate the variance/covariance matrix associated with the random effects, which are the effects of the background markers excluding the marker being tested. In this way the model is only testing a single parameter rather than n-1 parameters, which would remove all of the degrees of freedom available for the model.

CHAPTER 3

MATERIALS AND METHODS

Plant Materials

The entries for this study consisted of ninety-four spring wheat genotypes grown in the Montana State University's Spring Wheat advanced yield trial during one or more of the following years 1997, 2002, and 2007. No full-sib genotypes were included in the population.

Lines were evaluated for yield, grain test weight, heading date, plant height, grain protein content, and stem solidness under irrigated and dryland field conditions at the following locations: The Northern Agricultural Research Center (NARC) is located near the city of Havre, MT. Geographical coordinates for the NARC are 48° 11’ 22” North, 114 ° 08' 00" West, with an elevation of 896.3 meters. The soil type for this area is a Joplin clay loam. The Eastern Agricultural Research Center (EARC) is located near the city of Sidney, MT with the following geographical coordinates, 47° 43' 35" North, 104° 03' 01" West. The elevation of the EARC is 650 m and the soil type is a Savage silty clay. The Western Triangle Agricultural Research Center (WTARC), located at 48° 18' 25" North, 111° 55' 34" West, with an elevation of 1233.3 m is near the city Conrad, MT. The soil type for the WTARC is a Scobey clay loam. Two miles west of Moccasin, is the Central Agricultural Research Center (CARC). The geographical coordinates are 47° 03’ 23” North, 109 ° 57' 07" West and the elevation is 1433.3 m. The soil type is a Judith-

Danvers clay loam. The Southern Agricultural Research Center (SARC) is located near the town, Huntley, MT. The geographical coordinates for the SARC are 45° 55' 30" North, 108° 14' 42" West. The elevation of the SARC is 1066.7 m and the soil type is a Fort Collins silt loam. The Northwest Agricultural Research Center (NWARC) can be found just west of Kalispell, MT, near the city Creston, MT. The elevation is 963.3 m and the geographical coordinates are 48° 11' 22" North, 114° 08' 00" West. The soil type is a Creston silt loam. The Arthur H. Post Field Research Farm is located on the west side of Bozeman, MT. 45° 41'N, 111° 00'W are the geographical coordinates and the elevation is 1590.7 m. The Post farm soil type is an Amsterdam silt loam.

In 1997, twenty-nine of the lines were grown and tested across the state at all locations. In 2002, forty-three of the lines were evaluated across the state of Montana at all locations, excluding the Central Agricultural Research Center, in Moccasin In 2007, forty-two lines from the population were tested at all locations

In 2009 and 2010, the all entries were planted in Bozeman, MT at the Arthur H. Post Field Research Farm. The entries were planted with a lattice design with three replications under dryland conditions. The entries were planted in single row plots at a seeding rate of 8-10 grams per plot. In 2009 planting occurred on 14 May and harvest was on 9 September. In 2010, the population was planted on 10 June and harvested on 14 October. The lines were evaluated for yield, grain test weight, heading date, plant height, grain protein content, and stem solidness. Table 1 summarizes the weather information across environments.

Phenotypic Data Collection

Yield

Entries from all environments were tested for yield which was calculated using the following equation:

$$Yield\ \left(\frac{kg}{ha}\right) = \left[\left(\frac{1.6}{Plot\ Length}\right) * Plot\ Grain\ Weight\right] * 67.19$$

Grain Test Weight

Grain test weight was evaluated using a Fairbanks grain weighing scale and adjusted for individual grain densities using the following equation:

$$Test\ Weight\ \left(\frac{lb}{bu}\right) = Test\ Weight\ (g) * 0.4939$$

$$Test\ Weight\ \left(\frac{kg}{m^3}\right) = Test\ Weight\ \left(\frac{lb}{bu}\right) * 12.87$$

Heading Date

Heading date for each entry was recorded using the Julian scale, in the field as the day when 50% of the heads had emerged from the flag leaf sheath.

Plant Height

Plant heights were evaluated by measuring the distance, in centimeters, from the soil surface to the estimated average height of two or three main tillers, excluding the awns. Two measurements per plot were taken at random and averaged together for a final plant height for each plot.

Grain Protein Content

Grain protein content analysis was performed on whole grains using a Foss Infratec 1241 Grain Analyzer (Tecator 1241, Foss Analytical AB, Höganäs, Sweden) in the Montana State University Cereal Quality Lab, Bozeman, MT.

Stem Solidness

Five stems per plot were pulled at random, near crop maturity. A cross section was cut through the center of each internode and a total of five internode ratings were made using a 1-5 scale where 1 designates a hollow stem and 5 designates a completely solid stem. The scores were summed to give each stem a single score, ranging from five to 25. The single stem scores were averaged and recorded as one final stem solidness score for each entry. In 2010, only three stems were used per plot.

Seed Color

A visual observation of each entry was also taken on the seed color. Two phenotypes were observed, either red or white. There are three color genes reported on the group three homologous chromosomes (Sherman et al. 2008). The red allele is dominate to the white allele, therefore an individual must be recessive at all three loci to observe the white phenotype, whereas only one gene needs to be dominate to observe the red phenotype.

Genotyping

The population was genotyped with 103 molecular markers. Forty-two microsatellite markers were selected at one marker per chromosome arm for estimating the genetic population stratification. Genomic regions of interest were then targeted for being previously reported and confirmed as controlling traits of interest. Traits evaluated include plant height (*Rht-B1, Rht-D1,* Ellis et al. 2002), heading date (*Ppd-B1, Ppd-D1,* Blake et al. 2009), vernalization (*Vrn-B1, Vrn-D1,* Blake et al. 2009) kernel hardness (*Pina* , Gautier et al. 1994), and stem solidness (*Qsst.msub-3BL* Cook et al. 2004). The loci controlling grain color (*R-A1, R-B1, R-D1,* McIntosh 2008) were also evaluated. These regions were screened with additional markers and were considered to be candidate genes for the association mapping analysis. Markers identified as being linked to traits of interest in the MSU Conan/Reeder and McNeal/Thatcher bi-parental mapping populations were also screened on the AM entries. The traits linked to the markers of interest identified through traditional QTL analysis included productive tiller number, stem solidness, plant height, heading date, grain test weight, grain protein content, kernel diameter, and kernel hardness. All microsatellite markers were screened using the Li-Cor DNA analysis system, while sequence-tagged-site markers for *Rht-B1, Rht-D1, Ppd-B1, Ppd-D1, Vrn-B1, Vrn-D1*, and *Pina* were analyzed using polyacrylamide gel electrophoresis (Blake et al. 2009, Gautier et al. 1994).

Population Stratification

Forty-two microsatellite markers spanning the entire wheat genome, as well as nine additional markers controlling important agronomic traits were input into the program STRUCTURE, for estimating the subpopulation matrix. The program was set to test the hypothesis for one to six subpopulations (*k*), with admixture and correlated allele frequencies (Falush et al. 2003). All other settings were left at default values. Ten runs were completed at each *k* with a burn-in phase of 10^5 iterations, and a sampling phase of 5×10^4 replicates. To determine the correct number of subpopulations Δk, described by Evanno et al. (2005) was used. The Δk value of three subpopulations was used in creating the final Q matrix.

The raw data output for each of the ten runs at *k=3* were reformatted and input into the program CLUMPP (Jakobsson and Rosenberg 2007). The *Greedy* algorithm with the G' pair-wise matrix similarity statistic were used to find the best average Q matrix computed from across all ten runs. The *Greedy* algorithm with 1000 random input orders was chosen over the *FullSearch* algorithm because of the large computational time involved. G' was used instead of G because G' standardizes the output values to fall within a range of [0, 1], excluding nonsensical negative values. The average Q matrix was then reformatted to be used with the software program Distruct (Rosenberg et al. 2002), which is a software program that creates a visualization of the Q matrix.

The K matrix was created in the program SPAGeDi and based on the same fifty-one markers that were used in developing the Q matrix. The method used to calculate the coefficient of coancestry is described by Loiselle et al. (1995). Negative coefficients

were replaced with zeroes and the entire matrix was then multiplied by a scalar of two because of the inbred and diploid state of all entries.

Statistical Analysis

Least squares entry means were computed for the unbalanced historical dataset across all environments using Proc GLM, in SAS Statistical Analysis Software, version 9.02.01, (© 2002-2008 by SAS Institute Inc., Cary, NC). Least squares entry means for the 2009 and 2010 datasets were computed using Proc LATTICE. Least squares entry means were also computed for each entry across the historical data and the 2009 and 2010 datasets. Proc MIXED with the Type 3 method was used to test the environment by entry interaction between the trials from 2009 and 2010. Correlations among the traits were also computed using the least squares entry means from the 2009, 2010, historical and combined datasets.

Model Testing

The linear mixed model described by Yu et al. (2006) was implemented using the Proc MIXED procedure in SAS to test the variation of each trait explained by each molecular marker. The full model was tested initially.

$$y = X\beta + Qv + Zu + e$$

Where y is a vector of phenotypic observations, the fixed effect X, is a matrix of genotypic data, the fixed effect Q, is the subpopulation matrix, Z is the kinship matrix

estimating the variance/covariance structure of the random effects, and *e* is the residual error. A macro was used to sequentially remove a single locus to be tested against the phenotypic values. Two additional forms of the model, excluding either the Q matrix or K matrix, were also used to test each individual marker against each trait. To control the Type I error rate the significance threshold was adjusted correcting for multiple comparisons using the Bonferroni correction method (Cheverud 2001).

Two sets of phenotypic data were used to test the marker associations with the traits of interest. The first set included the least squares entry means from the historical dataset. The second phenotypic dataset tested, included the least square entry mean observations from 2009 and 2010.

Linkage Disequilibrium

The amount of linkage disequilibrium in specific areas of interest throughout the genome was also evaluated. The areas were selected depending on whether we believed them to be under selection or not in the breeding program. These regions included the *Rht-B1* locus controlling plant height, the *R-A1*, *R-B1*, and *R-D1* loci controlling grain color, the QTL *Qsst.msub-3BL*, controlling stem solidness and a new potential QTL controlling productive tiller number discovered in the Conan/Reeder and McNeal/Thatcher MSU mapping population.

Molecular markers, referred to as primary markers and believed to be linked to the genes controlling the traits were screened. Three to nine additional markers per loci, reported on the website GrainGenes (http://wheat.pw.usda.gov/GG2/index.shtml) as

ranging in proximity to the primary markers, were also screened. Estimated location of the markers in centiMorgans (cM) was also taken from GrainGenes (http://wheat.pw.usda.gov/GG2/index.shtml). The software program TASSEL (www.maizegenetics.net) was used to compute the r^2 estimates among all loci and the comparison-wise significance was computed by 1000 permutations. The value, 0.2292 was used as the critical r^2 estimate indicating linkage between two adjacent markers. To estimate the population specific critical value of r^2 for significant linkage, estimates of r^2 from 42 unlinked markers were square root transformed and the parametric 95th percentile of that distribution was used (Breseghello and Sorrells 2006), as indicated in the following equation:

$$Significant\ r^2\ Threshold = \bar{y} + 1.645 * \sqrt{\sigma^2}$$

Table 1. Weather summary by location, year, and averaged across years for the seven Montana Ag Experiment Stations.

Location	Year	Weather Measurement	APR.	MAY	JUNE	JULY	AUG.	SEP.	OCT.	Total	Average
Havre, MT	Average	Precipitation (cm)	2.46	4.47	6.48	3.63	3.05	2.92	1.68	24.69	-
	1916-2007	Temperature (°C)	6.44	12.28	16.61	20.67	19.61	13.39	7.72	-	13.82
Conrad, MT	Average	Precipitation (cm)	2.44	4.42	7.47	3.35	3.28	2.97	1.55	25.48	-
	1985-2007	Temperature (°C)	6.50	11.39	15.50	19.56	19.00	13.83	7.33	-	13.30
Moccasin, MT	Average	Precipitation (cm)	3.02	6.45	8.03	4.32	4.06	3.56	2.26	31.70	-
	1909-2007	Temperature (°C)	4.94	10.11	14.39	18.78	18.22	12.61	7.11	-	12.31
Sidney, MT	Average	Precipitation (cm)	2.84	5.16	7.24	5.23	3.56	3.23	2.31	29.57	-
	1949-2007	Temperature (°C)	7.00	13.39	18.06	21.11	20.39	14.22	7.72	-	14.56
Kalispell, MT	Average	Precipitation (cm)	4.65	6.22	8.08	4.09	2.87	4.24	3.30	33.45	-
	1980-2007	Temperature (°C)	6.33	10.89	14.28	17.94	17.44	11.94	5.67	-	12.07
Huntley, MT	Average	Precipitation (cm)	3.45	5.36	6.05	2.90	2.31	3.28	2.64	25.98	-
	1911-2007	Temperature (°C)	7.49	12.73	17.39	21.42	20.34	14.33	8.28	-	14.57
Bozeman, MT	2009	Precipitation (cm)	7.16	4.09	6.65	7.09	3.84	1.30	4.60	34.72	-
		Temperature (°C)	5.06	12.11	14.06	18.56	18.39	17.61	3.72	-	12.79
	2010	Precipitation (cm)	3.78	8.59	11.91	1.02	4.52	4.04	3.38	37.24	-
		Temperature (°C)	6.22	8.28	13.83	18.50	17.94	14.00	6.35	-	12.16
	Average	Precipitation (cm)	4.14	6.68	7.04	3.43	3.23	3.68	3.58	31.78	-
	1958-2010	Temperature (°C)	5.72	10.67	14.83	18.72	18.11	13.06	7.33	-	12.63

CHAPTER 4

RESULTS

Phenotypic Data

Phenotypic data were obtained from two sources. The historical dataset consisted of least squares entry means calculated from yield trial data collected in 1997, 2002, and 2007. The population was evaluated across the state of Montana for heading date, stem solidness, plant height, yield, test weight, and grain protein content. Data were completely balanced within years and unbalanced between years with only five entries common across 1997, 2003, and 2007. A second dataset consisted of least squares entry means estimated from field trials grown in Bozeman in 2009 and 2010. The 2009/2010 least squares entry means were based on completely balanced and triple replicated field trials. The phenotypic information is depicted in Appendix A.

The 2010 field trial was planted about one month later than typical for the Bozeman area. The reason for the late planting was to shift the grain filling period to a time of year with higher temperatures creating a high-heat stress environment. Two additional trials were planted, one under dryland conditions and one under irrigated conditions, in Bozeman at normal planting times. However the Bozeman area suffered a catastrophic weather event and the two early planted trials were destroyed. The late planted trial was still very immature and suffered no measurable damage.

The late planting as well as other environmental differences between 2009 and 2010 could be problematic. However, correlation (*r*) analysis between the six traits for 2009 and 2010 indicated significant and highly positive correlations between matching traits for the two environments (Table 2).

Analysis of variance (ANOVA) was used to test the effects of variation of environment within year on genotype for 2009 and 2010. The F-values and their significance levels are reported in Table 3 for the variation due to environment, genotype, and the environment by genotype interaction. The interaction between environment and genotype was significant in all traits, but the variance associated with the interaction was small in comparison to the variance associated with the environment or the genotype alone. The complete analysis of variance is reported in Appendix B.

A second correlation analysis was conducted on the least squares entry means between the 2009/2010 dataset and the historical dataset, described in Table 4. Matching traits had significant and positive correlations between the two datasets. The correlation for heading date, stem solidness, plant height, yield, test weight, and grain protein content was 0.85, 0.87, 0.95, 0.45, 0.59, and 0.55, respectively.

Molecular Marker Analysis

Ninety-four polymorphic SSR markers were screened on the AM entries, of which ninety-nine polymorphic loci were scored. Observed residual heterozygosity was roughly 0.3%. The small number of heterozygotes was replaced with missing data values, removing twenty-nine data points out of 10,181. The average number of alleles

per locus was 5.46. Table 5 summarizes the results. The SSR markers were classified by observed alleles per locus. The number of loci with the given number of alleles per locus was summed in the second column of Table 5. Alleles that had a minor frequency of less than 10% in the population are indicated in the third column totaling 279 minor alleles out of 541 total alleles. Of the 279 minor alleles, 101 were accession specific meaning a single individual had a unique allele polymorphism compared to the rest of the population.

Nine additional bi-allelic, sequence tagged site (STS) polymorphic markers were screened on the population, as well. There were no minor or accession specific alleles associated with these markers. These markers characterized the population for photoperiod sensitivity, vernalization, plant height, and kernel hardness. The *Ppd-D1* locus had seventy-nine individuals characterized as photoperiod sensitive and fifteen as photoperiod insensitive. The *Vrn-B1* locus had sixty-nine individuals scored as the spring type and twenty-five scored as the winter type. The *Vrn-D1* locus had fourteen individuals scored as the spring type and eighty individuals scored as the winter type. Sixty-five individuals were semi-dwarf in stature and twenty-nine were tall in stature Seventy-two individuals were wild type at the *Rht-B1* locus and twenty-two were mutant at the *Rht-B1* locus. Fifty-one individuals were wild type and forty-two were mutant at the *Rht-D1* locus. Forty-one individuals had the wild type allele and fifty-three had null alleles at the *Pina* locus.

Population Substructure

Estimation of population structure is critical for interpretation of marker-trait relationships inferred from AM experiments. Diversity measurements were obtained from the fifty-one microsatellite markers and used to estimate underlying population substructure. The diversity measurements were created in the program STRUCTURE. The number of subpopulations (K) ranged from one to six. ΔK, described by Evanno et al. (2005) was calculated to estimate the correct number of subpopulations, which resulted in $K=3$. At $K=3$, the value for ΔK peaks and approaches zero for $K>3$. A Q matrix was created by averaging across the ten runs of $K=3$, from STRUCTURE, and visualized using the program Distruct, shown in Figure 1. The Q matrix groups individuals together based on the proportion of their genome belonging to each of the three sub-populations. The three sub-populations are represented by different colors (Figure 1). Individuals are organized chronologically based on when the genotypes first entered the MSU Spring Wheat Advanced Yield Trial. Named varieties and experimental lines indicated by "MT", "MTHW" and "BZ" make up the population. Actual model analysis utilized a matrix with numeric values describing genotypes, which summed to one for each individual.

The genetic constitution of the spring wheat breeding program changed over time, as can be seen in Figure 1. Over time there is a shift in the number of individuals with the blue color towards more gray. The red color seems to have a consistent presence over the years. Most individuals break down into one or two colors; interestingly the line "Thatcher" has proportions of all three colors, which makes sense given that that this line

was a founding member of the MSU Spring Wheat project and used extensively in crossing strategies. Lines classified with "MTHW" are hard white spring wheat lines and these individuals are comprised of mostly the blue color. "Amidon" and "Ernest" are lines from North Dakota germplasm and share mostly the red color.

Pair wise estimates of relatedness between individuals calculated in SPAGeDi made up the additive relationship, or K matrix. The K matrix was incorporated into the mixed model as an estimation of the variance associated with the unique genetic background contained within each individual.

Linkage Disequilibrium Analysis

Linkage disequilibrium is affected by founding/bottleneck events, drift, and selection causing the extent of LD to be population-specific; because of this we calculated a population-specific threshold for r^2. Appendix C shows a matrix of forty-two unlinked microsatellite markers, reported to be on different chromosome arms (http://wheat.pw.usda.gov/GG2/index.shtml). Below the diagonal are the estimated p-values, calculated in TASSEL, between each pair of markers indicating significant linkage disequilibrium. Several markers were observed to be in significant LD at the 0.01 level. Above the diagonal are pair-wise r^2 estimates between markers, also calculated in TASSEL. Based on these values the population specific r^2 significance threshold was found to be 0.2292, indicating linkage disequilibrium due to physical linkage.

We tested several markers suspected to be in LD due to physical linkage for regions of the genome under varying degrees of selection. Figure 2 is a matrix depicting the markers found to be physically linked in our regions of interest. The markers within the figure are grouped by the targeted genomic regions. Significance values between markers are listed below the diagonal and r^2 values between markers are listed above the diagonal.

The first set of markers and their locations are reported on the group three chromosomes (http://wheat.pw.usda.gov/GG2/index.shtml). These three genomic regions control seed color in hard wheat. On chromosome 3A two markers were found to be physically linked, *Xcfa2170.2* and *Xgwm155*. These two markers are an estimated 5.7 cM apart. On 3B *Xcfa2170.1* and *Xbarc84* were linked and reported as 1.0 cM apart. The 3A markers were found to be in significant LD not due to linkage with the 3B markers. *Xgpw5235* and *Xgwm4306* were linked on chromosome 3D, with an estimated distance of 4.5 cM between the markers. These markers were in significant LD not due to linkage with *Xcfa2170.1* on 3B.

A second region on chromosome 3B was targeted for controlling stem solidness. This region is suspected of being under intense selection. We found five markers to be physically linked, *Xwmc632*, *Xgwm181*, *Xwmc274*, *Xgwm547*, and *Xgwm340*. These markers were spread across a distance of 8.2 cM.

We tested five markers surrounding the *Rht-B1* locus at distances between 1 cM and 17 cM away. No marker tested on chromosome 4B, was found to be physically

linked to *Rht-B1*. However, the locus *Rht-D1*, on chromosome 4D had significant r^2 and LD values with the *Rht-B1* marker.

The final region of interest surrounded a QTL controlling productive tiller number in spring wheat on chromosome 6B. We found seven markers spread across 14 cM to be in significant LD. *Xwmc397*, *Xgwm58*, *Xgwm70*, and *Xgwm193* were physically linked. *Xwmc487*, *Xwmc182.1* and *Xbarc354* were found to be physically linked in this region, as well.

Association Mapping Analysis

Seed Color

Three markers linked to the genes controlling seed color in spring wheat have been reported on the group three homologous chromosomes (Sherman et al. 2008). Red seed color is dominant to white seed color, therefore requiring all three loci to be recessive for the white phenotype. We measured seed color qualitatively rather than quantitatively, creating a categorical trait that is not distributed across a range of values such as yield, because of this testing in multiple environments is unnecessary. Ten of the ninety-four individuals in this AM population displayed the white phenotype.

Association analysis was conducted in SAS using the Proc MIXED function. Association analysis results are presented in Table 6. Marker/trait associations were tested with three models. The first model was a general linear model (Q), testing seed color against the genetic marker and Q matrix which were included in the model as fixed effects. The second model was a mixed linear model (K) and included the marker as a

fixed effect and the K matrix in the random effects term. The third model tested included the marker and Q matrix as fixed effects and the K matrix as a random effect in a mixed linear model (QK).

A total of fourteen SSR markers were found to be significantly associated with seed color after the Bonferroni correction was applied to correct for multiple testing issues. *Xgwm155*, *Xgwm4010*, and *Xgwm4306* were previously reported as being the most closely linked markers on chromosomes 3A, 3B, and 3D respectively. *Xgwm4010* and *Xgwm4306* both associated with seed color in at least one form of the model in our analysis. *Xgwm155*, reported as being near the 3A gene, did not have any significant associations with seed color. Markers associated with seed color were not localized to the group three chromosomes either. For some tests of associations between marker and seed color the Proc MIXED procedure in the SAS software failed to converge causing an error to occur. These errors are indicated by the "N/A" designation and only affected models that included a random effects term.

Heading Date

Blake et al. (2009) reported that the *Ppd-D1* gene had a significant effect on heading date in two bi-parental mapping populations. Our results reported in Table 7 agreed with that report when testing both sets of phenotypic data and across all three models. However, the *Ppd-B1* and *Vrn-B1* genes were also reported to effect heading date whereas we found no associations in this study. Two additional markers, *Xgdm132* on chromosome 6D and *Xgwm980* on chromosome 3B, associated with heading date. These markers were only significant in the general linear model. *Xgdm132* only

associated with the historical dataset and *Xgwm980* only associated in the 2009/2010 dataset.

Stem Solidness

Our study found nine markers associated with stem solidness. *Xgwm340*, *Xgwm547*, and *Xwmc274* all reported on chromosome 3B were associated, at the 0.01 alpha level, with stem solidness in both datasets and across all models. Additional markers reported on 3B (*Xgwm980, Xgwm181, Xwmc78),* 5B and 5D (*Xwmc160),* and 6B (*Xgwm88, Xgwm193)* were found associated with stem solidness at the 0.05 alpha level in one of the possible models and datasets.

Plant Height

In our study *Rht-D1* was associated with plant height across all models and both phenotypic data sets at the 0.01 alpha level. *Rht-B1* was only found associated with plant height in the general linear model. Three additional markers on 3D (*Xgwm4306*), 5B (*Xwmc810*), and 6A (*Xgwm427*) were associated with plant height in at least one phenotypic dataset and at least one model at the 0.05 to 0.01 alpha level.

Yield

In our study three markers associated with yield, *Xgwm161, Xwmc274, Xwmc413*, but only in the historical dataset and only with the general linear model. Coincidentally, *Xwmc274* associated with both stem solidness and yield. There was a negative association with yield and a positive association with stem solidness.

Test Weight

Four markers were observed to be associated with test weight on chromosomes 2D, 3D, 3A, 4B, 5B, and 7B. No known associations between these markers and test weight have been reported in the literature nor were they identified in the QTL studies.

Grain Protein Content

Xbarc119 mapped to the group one chromosomes and associated with grain protein content for all three models in the historical dataset only. This marker is reported as being 7.2 cM away from Glu-3, a low molecular weight glutenin gene and 15.5 cM away from Gli-A3, a gliadin gene (http://wheat.pw.usda.gov/GG2/index.shtml).

Table 2. Correlations (r) between like traits for the 2009 and 2010 least squares entry means.

Agronomic Trait	r
Heading Date	0.86**
Stem Solidness	0.90**
Plant Height	0.85**
Yield	0.50**
Test Weight	0.78**
G. P. C.	0.87**

[a] Significance Level: *,** = $P < 0.05$, $P < 0.01$, respectively

[b] G.P.C. = Grain Protein Content

Table 3. 2009 and 2010 F-values reported from the experimental analysis of variance.

	Heading Date	Stem Solidness	Plant Height	Yield	Test Weight	Grain Protein Content
Environment	48176.6**	8.62*	265.84**	161.67**	23.17**	25.22**
Genotype	26.83**	24.45**	20.44**	7.07**	13.31**	56.32**
G x E	2.58**	1.45**	2.03**	2.33**	1.88**	4.15**

[a] Significance Level: *,** = $P < 0.05$, $P < 0.01$, respectively

Table 4. Correlations (r) between like traits for the 2009/2010 combined least squares entry means and historical least squares entry means.

Agronomic Traits	r
Heading Date	0.85**
Stem Solidness	0.87**
Plant Height	0.95**
Yield	0.45**
Test Weight	0.59**
G. P. C.	0.55**

[a] Significance Level: *,** = $P < 0.05$, $P < 0.01$, respectively

[b] G.P.C. = Grain Protein Content

Table 5. Number of alleles per locus for 99 polymorphic loci observed with 94 SSR markers

Alleles/Locus	No. of Loci	No. Alleles ≤10% Frequency	No. Accession Specific Alleles
14	1	9	6
10	4	27	11
9	5	30	12
8	10	55	16
7	13	52	17
6	12	37	16
5	16	37	14
4	15	19	7
3	13	10	2
2	10	3	0
Total	99	279	101

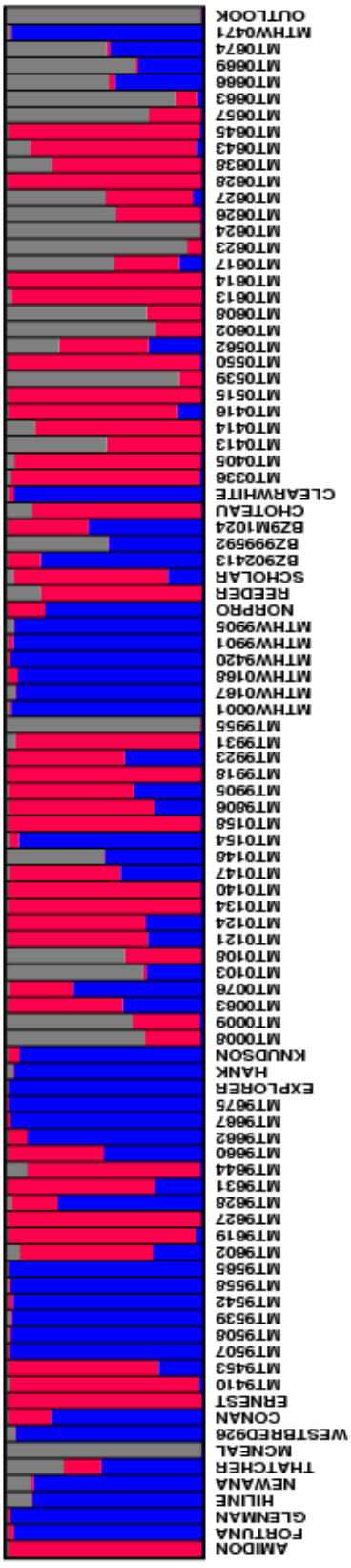

Figure 1. A representation of the population stratification for the association mapping panel. Individuals are divided among three sub-populations as indicated by the red, blue and gray colors. Lines are organized based on when that individual was first tested in the MSU Spring Wheat Advanced Yield Trial.

	Distance in cM	5.7		1		4.5		8.2							14							
		3A		3B		3D		3B					4B	4D	6B							
		Xcfa2170.2	*Xgwm155*	*Xcfa2170.1*	*Xbarc84*	*Xgpw5235*	*Xgwm4306*	*Xwmc632b*	*Xgwm181*	*Xwmc274*	*Xgwm547*	*Xgwm340*	*Rht-B1b*	*Rht-D1b*	*Xwmc397*	*Xgwm58*	*Xgwm70*	*Xgwm193*	*Xbarc354*	*Xwmc182.1*	*Xwmc487*	
3A	*Xcfa2170.2*		0.28	0.17	0.11	0.01	0.02	?	0.02	0.03	0.03	0.02	0.02	0.08	0.01	0.01	0.02	0.02	0.02	0.04	0.02	
	Xgwm155	0.00		0.13	0.04	0.01	0.02	0.07	0.02	0.04	0.04	0.03	0.03	0.14	0.00	0.00	0.00	0.00	0.02	0.02	0.06	
3B	*Xcfa2170.1*	0.00	0.00		0.24	0.09	0.10	?	0.05	0.06	0.06	0.06	0.05	0.17	0.01	0.01	0.01	0.00	0.02	0.03	0.01	
	Xbarc84	0.00	0.00	0.00		0.04	0.09	0.06	0.04	0.05	0.06	0.05	0.00	0.07	0.04	0.04	0.03	0.04	0.03	0.03	0.02	
3D	*Xgpw5235*	0.31	0.05	0.00	0.10		0.59	0.01	0.00	0.00	0.00	0.02	0.01	0.02	0.00	0.00	0.00	0.03	0.00	0.00	0.01	
	Xgwm4306	0.43	0.01	0.01	0.13	0.00		0.04	0.04	0.04	0.05	0.05	0.01	0.06	0.01	0.01	?	0.01	0.01	0.00	0.03	
3B	*Xwmc632*	?	0.00	?	0.00	0.12	0.00		0.23	0.28	?	0.23	0.01	0.12	0.05	0.04	0.04	0.03	0.02	0.01	?	
	Xgwm181	0.30	0.19	0.00	0.26	0.44	0.00	0.00		0.55	0.55	0.48	0.00	0.14	0.08	0.08	0.08	0.07	0.02	0.01	?	
	Xwmc274	0.08	0.05	0.00	0.06	0.94	0.00	0.00	0.00		0.69	0.52	0.00	0.14	0.06	0.07	0.07	0.08	0.02	0.01	0.02	
	Xgwm547	0.16	0.05	0.02	0.12	0.85	0.00	?	0.00	0.00		0.54	0.00	0.14	0.09	0.10	0.10	0.08	0.03	0.01	0.02	
	Xgwm340	0.03	0.24	0.00	0.27	0.01	0.00	0.00	0.00	0.00	0.00		0.00	0.17	0.05	0.05	0.05	0.04	0.01	0.01	?	
4B	*Rht-B1*	0.03	0.32	0.05	0.92	0.43	0.57	0.13	0.96	0.75	0.40	0.93		0.26	0.04	0.07	0.11	0.12	0.03	0.01	0.02	
4D	*Rht-D1*	0.00	0.00	0.00	0.03	0.26	0.04	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.05	
6B	*Xwmc397*	0.50	0.49	0.39	0.11	0.61	0.07	0.00	0.00	0.00	0.00	0.01	0.09	0.64		0.63	0.56	0.49	0.09	0.04	?	
	Xgwm58	0.11	0.24	0.64	0.05	0.45	0.23	0.00	0.01	0.04	0.00	0.03	0.03	0.80	0.00		0.59	0.53	0.09	0.01	0.01	
	Xgwm70	0.42	0.36	0.53	0.23	0.22	?	0.00	0.00	0.01	0.00	0.02	0.00	1.00	0.00	0.00		0.72	0.09	0.02	?	
	Xgwm193	0.49	0.10	0.95	0.15	0.40	0.13	0.01	0.01	0.01	0.00	0.22	0.00	1.00	0.00	0.00	0.00		0.09	0.03	0.01	
	Xbarc354	0.03	0.03	0.01	0.01	0.59	0.10	0.00	0.00	0.01	0.02	0.06	0.00	0.61	0.00	0.00	0.00	0.00		0.32	0.12	
	Xwmc182.1	0.00	0.01	0.01	0.02	0.81	0.55	0.00	0.00	0.00	0.17	0.00	0.49	0.95	0.00	0.04	0.01	0.21	0.00		0.25	
	Xwmc487	0.04	0.00	0.45	0.01	0.64	0.04	?	?	0.16	0.06	?	0.26	0.01	?	0.15	?	0.06	0.00	0.00		

Figure 2. Linkage disequilibrium between markers of interest. Below the diagonal are P-values and above are r^2 estimates. The P-values indicate estimates of general LD. r^2 values greater that 0.2292 indicate physical linkage. The top row indicates estimated distances in cM between each group of markers, taken from GrainGenes (http://wheat.pw.usda.gov/GG2/index.shtml).

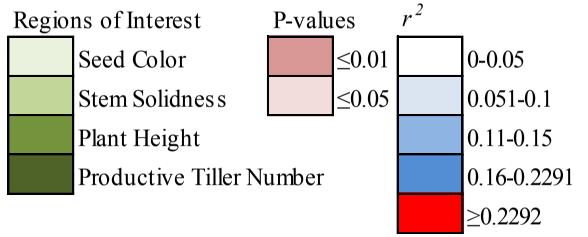

Table 6. Association analysis results for seed color. The relevant significance values are shown.

Trait	Locus	Chromosome	Q	K	QK
Seed Color	*Xwmc59*	1A, 6A	0.011	N/A	NS
	Xwmc177	2A	6.13E-05	0.014	N/A
	Xgwm349	2D	0.001	0.001	0.001
	Xgpw2109	3A	0.015	N/A	N/A
	Xwmc215	3A, 5A, 5D	6.45E-05	1.59E-04	0.003
	Xcfa2170.1	3B	4.83E-06	NS	0.004
	Xgwm340	3B	0.012	NS	NS
	Xgwm4010	3B	0.002	0.012	NS
	Xwmc505	3B	0.045	N/A	N/A
	Xgpw5235	3D	7.12E-05	N/A	N/A
	Xgwm4306	3D	6.13E-05	2.14E-04	0.004
	Xwmc238	4B	NS	0.008	N/A
	Xwmc160	5B, 5D	1.24E-07	2.52E-04	0.001
	Xgwm427	6A	NS	0.012	N/A

[a] NS = Not Significant

[b] N/A = Not Applicable

[c] Q indicates the general linear model testing the trait against the marker and Q matrix.

[d] K indicates the mixed linear model testing the trait against the marker and K matrix.

[e] QK indicates the mixed linear model testing the trait against the marker, Q matrix, and K matrix.

Table 7. Association analysis results for the least squares entry means calculated from the historical and 2009/2010 data sets. Specific significance values are given in the table.

Trait	Locus	Chromosome	Historical Data			2009/2010 Data		
			Q	K	QK	Q	K	QK
Heading Date	*Xgdm132*	6D	0.015	NS	NS	NS	NS	NS
	Ppd-D1	2D	0.001	0.003	0.002	0.043	0.029	0.039
	Xgwm980	3B	NS	NS	NS	0.040	NS	NS
	Ppd-B1	2B	NS	NS	NS	NS	NS	NS
	Vrn-B1	5B	NS	NS	NS	NS	NS	NS
Stem Solidness	*Xgwm980*	3B	0.009	N/A	0.020	0.044	N/A	0.019
	Xgwm181	3B	NS	NS	NS	0.014	NS	0.025
	Xgwm193	6B	NS	NS	NS	0.040	N/A	NS
	Xgwm340	3B	0.002	0.008	0.007	1.91E-06	3.74E-05	3.45E-05
	Xgwm547	3B	2.96E-06	2.48E-06	3.67E-05	3.66E-15	4.00E-13	4.61E-12
	Xgwm88	6B	NS	NS	NS	0.022	NS	NS
	Xwmc160	5B, 5D	NS	NS	NS	NS	0.001	0.001
	Xwmc274	3B	2.60E-06	2.91E-06	3.11E-05	1.28E-15	3.02E-13	3.13E-12
	Xwmc78	3B	NS	NS	NS	0.048	NS	NS
Plant Height	*Xgwm427*	6A	NS	N/A	N/A	0.025	0.009	0.006
	Xgwm4306	3D	NS	0.005	0.007	NS	0.003	0.003
	Rht-B1a	4B	0.002	NS	NS	0.003	NS	NS
	Rht-B1b	4B	0.001	NS	NS	0.002	NS	NS
	Rht-D1b	4D	0.002	3.50E-04	0.001	0.003	0.001	0.002
	Xwmc810	5B	0.009	N/A	N/A	0.004	0.004	0.002

Table 7. Continued

Trait	Locus	Chromosome	Historical Data			2009/2010 Data		
			Q	K	QK	Q	K	QK
Yield	*Xgwm161*	4A, 5D	0.023	NS	NS	NS	NS	NS
	Xwmc274	3B	0.015	NS	NS	NS	NS	NS
	Xwmc413	4B	0.018	NS	NS	NS	NS	NS
Test Weight	*Xgpw294.1*	2D, 3D	0.009	0.016	0.046	NS	NS	NS
	Xgwm513	4B, 5B, 7B	NS	NS	NS	0.004	0.032	0.006
	Xwmc559	3A	0.042	0.029	0.035	NS	NS	NS
	Xwmc710	7B	NS	NS	NS	0.005	NS	0.007
G. P. C.	*Xbarc119*	1A, 1B, 1D	0.012	0.013	0.012	NS	NS	NS

[a] NS = Not Significant

[b] N/A = Not Applicable

[c] Q indicates the general linear model testing the trait against the marker and Q matrix.

[d] K indicates the mixed linear model testing the trait against the marker and K matrix.

[e] QK indicates the mixed linear model testing the trait against the marker, Q matrix, and K matrix.

[f] G. P. C. = Grain Protein Content

CHAPTER 5

DISCUSSION

Phenotypic Data

The historical dataset was estimated from unbalanced data across 1997, 2002, and 2007. This may have caused the least squares means to be measured with less precision. However, the ANOVA table in Appendix B calculated between the 2009 and 2010 field trials showed little variation due to the interaction between genotype and the environment. This result gives justification for combining the historical dataset across years and locations. Also, by combining data across environments we are able to adjust for the genetic by environment component that can drastically affect some traits such as yield, test weight, and grain protein content.

Observed Linkage Disequilibrium

Qsst.msub-3BL Stem Solidness QTL

The stem solidness trait has been under intense selection since approximately 1950 (Hayat et al. 1995), because of the traits ability to control the wheat stem sawfly. The solid stem phenotype can be traced back to a single genotype (Hayat et al. 1995), which may be the reason for the large observed LD block surrounding this QTL. It is likely that genetic material surrounding this QTL has been co-inherited with little

recombination to break down the linkage disequilibrium in this area. Given the self-fertilized nature of spring wheat recombination would be further reduced in effectiveness.

Productive Tiller Number

The large LD block observed around the productive tiller number QTL could be due to a variety of reasons. It is difficult to assess the population variability for this trait since it was not actually measured in the population. But one could postulate that high tiller number is a relatively new trait that is suspected to have been incorporated into the breeding program within the last twelve years through the North Dakota genotype "Reeder" and there has not been enough time for LD to decay through recombination. You could also attribute the appearance of a large block of 14 cM on the reduced precision of bi-parental mapping which has a resolving power of roughly 10 to 20 cM (Doerge 2002). The reduced precision can be attributed to the fact that traditional QTL mapping is able to take advantage of a limited number of recombination events that occurred during the population construction whereas association mapping can exploit recombination events on potentially an evolutionary or historical level.

Seed Color

The LD observed within and between chromosomes for the color genes could be due to a number of factors. You would expect to find markers physically linked on the same chromosome and in LD between the chromosomes because the breeder is actively selecting the recessive allele, at all three loci, at the same time in the lines exhibiting the white phenotype. The breeder would simultaneously be selecting for the surrounding

regions of these loci as well. A small number of lines were used to bring in the white phenotype with recessive alleles at each locus. This founding effect should have caused a reduction in the diversity of these regions. The red and white phenotypes are maintained in separate populations within the breeding program. The separation of genotypes reduces the potential for recombination with diverse material and maintains levels of linkage disequilibrium. Our estimation of the extent of LD in regions controlling seed color is potentially skewed, given the unequal proportion of red and white phenotypes and given the fact that there is a direct separation of the two gene pools. As far as LD between chromosomes this could be due to the simultaneous selection for the white allele at all three loci in the white seeded genotypes.

Reduced Height, Rht-B1 Locus

Rht-B1 and *Rht-D1* were observed to be in linkage disequilibrium in this study. Researchers know that these two genes are on separate chromosomes (Ellis 2002). The observed LD estimates are due to the fact that the dwarf genotypic class, which is mutant at both loci, is missing from the population and because *Rht-B1* and *Rht-D1* are linked in repulsion in the semi-dwarf genotypes. LD estimates are automatically inflated when one gametic combination is missing in the population. In our population, only three of the four possible combinations of genotypes are present, wild type, semi-dwarf due to a mutant allele at *Rht-B1* and a wild type allele at *Rht-D1*, or semi-dwarf due to a wild type allele at *Rht-B1* and a mutant allele at *Rht-D1*. The dwarf genotype which would have the mutant allele at both loci is absent causing inflated LD estimates. The fact that *Rht-*

B1 and *Rht-D1* are linked in repulsion, in semi-dwarf genotypes also gives the appearance of linkage between the two loci.

As far as the low levels of observed LD for markers surrounding the *Rht-B1* locus this could be due to a number of reasons. One explanation could be that the *Rht-B1* locus happens to be in a recombination hot spot. A second cause for the low observed LD could be due to a large amount of natural diversity in the genome surrounding the *Rht-B1* locus. It is also interesting to note the unequal distribution of mutant to wildtype alleles for *Rht-B1* in this population. Only twenty-two lines have the mutant allele for *Rht-B1*, the other seventy-two lines have the wildtype allele. The allelic distribution could potentially cause greater amounts of diversity for the *Rht-B1* locus since the introduction of the mutant allele would also cause a founder effect for this region.

Association Mapping

Candidate Gene Associations and False Positive Associations

We were able to discover several of the candidate genes we initially expected to find with this association mapping study, including *Ppd-D1*, *Rht-B1*, *Rht-D1*, *Xgwm340*, *R-B1*, and *R-D1*. We failed to discover all of the expected candidate genes, however. Why were we unable to detect associations between heading date and *Ppd-B1*, *Vrn-B1* and *Vrn-D1*? One reason may be due to the fact that *Ppd-D1* accounted for most if not all of the observed variation with heading date leaving little or no variation to detect the remaining genes. Alternatively, the reason could be due to epistasis where the combined effect of *Ppd-B1*, *Vrn-B1* and *Vrn-D1* is more important than each gene alone. A

possible solution to these two problems would be to test the interactions between *Ppd-D1* and *Ppd-B1*, *Vrn-B1* and *Vrn-D1*.

Also it is interesting to note that *Rht-D1* was found to be significant regardless of model or phenotypic dataset, while *Rht-B1* was only significant with the general linear model. Discrepancies in allele frequency may be the cause of this result. In the population seventy-two individuals had the wild type allele at the *Rht-B1* locus leaving twenty-two individuals with the mutant allele. *Rht-D1* on the other hand had fifty-one individuals with the wild type allele, forty-two individuals with the mutant allele, and one individual with missing data. The *Rht-B1* allele frequency is skewed toward the wild type allele whereas the *Rht-D1* allele frequency is relatively even.

It would appear that there is still more population stratification not yet corrected for causing false positive associations. The false positive associations are likely when considering plant height. The genes controlling plant height are well characterized yet three additional markers on 3D, 5B, and 6A were significantly associated. False positive associations are also likely to be included in with the seed color trait. This isn't as straight forward however since these genes are not as well described. One reason that fourteen seemingly dispersed markers were associated with seed color may be due to the fact that the red phenotype can be observed from multiple combinations of alleles. Furthermore the trait was essentially categorical; lines were either red or white even though variation in the intensity of red and white is present. Perhaps a better way would be to estimate seed color based on a more quantitative measurement accounting for subtle differences in hue or tinting, and then it might be able to account for variation caused by

each locus. The phenotypic distribution for seed color in the population may also be the reason for the excessive associations, given that roughly 10% of the population has the white phenotype with the remaining 90% with the red phenotype.

Differences between Phenotypic Datasets

There were differences in the association analysis between the two phenotypic datasets specifically for the traits that are largely influenced by environment such as, yield, test weight, and grain protein content. For these traits any given significant marker/trait association was only observed in one of the two datasets. *Xbarc119*, for example associated with grain protein content only in the historical dataset. Referring back to the correlation analysis in Table 4 the fact that yield, test weight, and grain protein content, appear to show a greater dependence on environment may make sense. There was an observed reduction in the correlation coefficients between the 2009/2010 and historical datasets for these traits compared to heading date, stem solidness, and plant height, indicating that yield, test weight, and plant height may in fact be more dependent on environment. This reinforces the utility of marker assisted selection with markers identified from historical phenotypic datasets.

The variation from environment to environment may cause a "back and forth" selection when based on phenotype alone. The breeder selects lines thus selecting alleles that perform well one year. The next growing season those same lines perform poorly causing the breeder to select against those associated alleles perpetuating the cycle. The results described also suggest that association mapping may be better suited for quantitative traits that are influenced by environment versus traditional bi-parental

mapping with limited environmental diversity. Being able to analyze historical data with association mapping, we can potentially increase the number of environments that our genotypes are tested in and hopefully identify genes that perform well regardless of the environment that they are used in. Multiple-environment testing is not impossible with bi-parental mapping but it is costly especially considering the limited number of traits that can be observed with each population and given the fact that a bi-parental mapping population is ultimately, not the target population.

Finally, the model and trait together had an impact on the discovery of marker/trait associations. It is interesting to note that in regards to yield only significant associations were found in the historical dataset but more importantly only significant associations were found using the general linear model that included the marker and the Q matrix. When we include the K matrix in the model essentially we are trying to estimate the effects that genetic background has on the marker under testing. The fact that we were unable to detect any associations when we included the genetic background effect may speak to the polygenic nature of yield indicating that no one locus accounts for a large enough proportion of variation alone.

CHAPTER 6

CONCLUSION

For this experiment, we applied candidate gene association mapping to a collection of ninety-four advanced spring wheat genotypes grown in Montana in one or more of the years 1997, 2002, and 2007. The lines were genotyped with ninety-three microsatellite markers and nine STS markers. A subset was used to determine relatedness and population structure. Alleles for marker loci, controlling candidate genes and previously identified QTL were determined for the population. Phenotypic data included data from previous breeding trials containing overlapping subsets of the ninety-four lines as well as a yield trial in 2009 and 2010 containing all ninety-four lines.

Did we identify associations between important genes and the markers known to control those genes? We identified marker/trait associations for major genes controlling heading date, plant height and a major QTL for stem solidness as predicted from previous genetic studies. We failed to identify associations between the markers, *Ppd-B1* and *Vrn-B1* and the trait, heading date. Did our results vary depending on the source of phenotypic information? Traits heavily influenced by environment appeared to have greater associations with markers when the phenotypic information was derived from a larger number of environments. Also, when we estimated the interaction between environment and genotype we observed a small effect. This would indicate that historical datasets may have great potential for being incorporated into association mapping studies.

Did we get similar results with AM when we tested markers first identified via traditional QTL mapping? We did not find any associations with AM between markers and traits that were initially identified through QTL mapping. Finally, what was the extent of LD for traits of interest under selection? Blocks of linkage disequilibrium were observed for the markers surrounding the *Qsst.msub-3BL* QTL and the productive tiller number QTL. Smaller blocks of LD were observed around the *R-A1*, *R-B1* and *R-D1* locus. However, a significant LD block was not observed around the *Rht-B1* locus.

Two drawbacks to this study centered on the small population size and use of microsatellite markers. The small number of individuals limited the genetic diversity needed for finding associations. Microsatellite markers presented complications due to the locus ambiguity in amplification. The multi-allelic nature of SSRs may lead future studies to be based on SNP markers instead. Differences in results were observed depending on the phenotypic dataset tested and model tested. Therefore careful considerations about the population and traits tested within the population must be taken into account before proceeding with an association mapping study.

REFERENCES CITED

Andersson L. (2009) Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. Genetica 136:341-349. DOI: 10.1007/s10709-008-9312-4.

Blake N.K., Lanning S.P., Martin J.M., Doyle M., Sherman J.D., Naruoka Y., Talbert L.E. (2009) Effect of Variation for Major Growth Habit Genes on Maturity and Yield in Five Spring Wheat Populations. Crop Science 49:1211-1220. DOI: 10.2135/cropsci2008.08.0505.

Boskovic J., Boskovic M. (2009) DURABLE RESISTANCE TO PUCCINIA TRITICINA BY ACCUMULATION OF RESISTANCE GENES. Genetika-Belgrade 41:353-378. DOI: 10.2298/gensr0903355b.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635. DOI: 10.1093/bioinformatics/btm308.

Breseghello F., Sorrells M.E. (2006) Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172:1165-1177. DOI: 10.1534/genetics.105.044586.

Castro A.J., Capettini F., Corey A.E., Filichkina T., Hayes P.M., Kleinhofs A., Kudrna D., Richardson K., Sandoval-Islas S., Rossi C., Vivar H. (2003) Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theoretical and Applied Genetics 107:922-930. DOI: 10.1007/s00122-003-1329-6.

Chapman A., Pantalone V.R., Ustun A., Allen F.L., Landau-Ellis D., Trigiano R.N., Gresshoff P.M. (2003) Quantitative trait loci for agronomic and seed quality traits in an F-2 and F-4 : 6 soybean population. Euphytica 129:387-393.

Cheverud J.M. (2001) A simple correction for multiple comparisons in interval mapping genome scans. Heredity 87:52-58.

Colonna V., Nutile T., Ferrucci R.R., Fardella G., Aversano M., Barbujani G., Ciullo M. (2009) Comparing population structure as inferred from genealogical versus genetic information. European Journal of Human Genetics 17:1635-1641. DOI: 10.1038/ejhg.2009.97.

Cook J.P., Wichman D.M., Martin J.M., Bruckner P.L., Talbert L.E. (2004) Identification of microsatellite markers associated with a stem solidness locus in wheat. Crop Science 44:1397-1402.

Crossa J., Burgueno J., Dreisigacker S., Vargas M., Herrera-Foessel S.A., Lillemo M., Singh R.P., Trethowan R., Warburton M., Franco J., Reynolds M., Crouch J.H., Ortiz R. (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889-1913. DOI: 10.1534/genetics.107.078659.

Devos K.M., Gale M.D. (1992) THE USE OF RANDOM AMPLIFIED POLYMORPHIC DNA MARKERS IN WHEAT. Theoretical and Applied Genetics 84:567-572.

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

Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620. DOI: 10.1111/j.1365-294X.2005.02553.x.

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

Farnir F., Coppieters W., Arranz J.J., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Mni M., Nezer C., Simon P., Vanmanshoven P., Wagenaar D., Georges M. (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Research 10:220-227.

Gautier M.F., Aleman M.E., Guirao A., Marion D., Joudrier P. (1994) TRITICUM-AESTIVUM PUROINDOLINES, 2 BASIC CYSTINE-RICH SEED PROTEINS - CDNA SEQUENCE-ANALYSIS AND DEVELOPMENTAL GENE-EXPRESSION. Plant Molecular Biology 25:43-57.

Gupta P.K., Langridge P., Mir R.R. (2010) Marker-assisted wheat breeding: present status and future possibilities. Molecular Breeding 26:145-161. DOI: 10.1007/s11032-009-9359-7.

Hardy O.J., Vekemans X. (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2:618-620.

Haudry A., Cenci A., Ravel C., Bataillon T., Brunel D., Poncet C., Hochu I., Poirier S., Santoni S., Glemin S., David J. (2007) Grinding up wheat: A massive loss of nucleotide diversity since domestication. Molecular Biology and Evolution 24:1506-1517. DOI: 10.1093/molbev/msm077.

Hayat M.A., Martin J.M., Lanning S.P., McGuire C.F., Talbert L.E. (1995) Variation for stem solidness and its association with agronomic traits in spring wheat. Canadian Journal of Plant Science 75:775-780.

Hedrick P.W. (1987) GAMETIC DISEQUILIBRIUM MEASURES - PROCEED WITH CAUTION. Genetics 117:331-341.

Helentjaris T., Slocum M., Wright S., Schaefer A., Nienhuis J. (1986) CONSTRUCTION OF GENETIC-LINKAGE MAPS IN MAIZE AND TOMATO USING RESTRICTION-FRAGMENT-LENGTH-POLYMORPHISMS. Theoretical and Applied Genetics 72:761-769.

Henderson C.R. (1975) BEST LINEAR UNBIASED ESTIMATION AND PREDICTION UNDER A SELECTION MODEL. Biometrics 31:423-447.

Hill W.G., Robertson A. (1968) LINKAGE DISEQUILIBRIUM IN FINITE POPULATIONS. Theoretical and Applied Genetics 38:226-231.

Jakobsson M., Rosenberg N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806. DOI: 10.1093/bioinformatics/btm233.

Karlsson E.K., Baranowska I., Wade C.M., Salmon Hillbertz N.H.C., Zody M.C., Anderson N., Biagi T.M., Patterson N., Pielberg G.R., Kulbokas E.J., Comstock K.E., Keller E.T., Mesirov J.P., von Euler H., Kampe O., Hedhammar A., Lander E.S., Andersson R., Andersson L., Lindblad-Toh K. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. Nature Genetics 39:1321-1328. DOI: 10.1038/ng.2007.10.

Kloppers F.J., Pretorius Z.A. (1997) Effects of combinations amongst genes Lr13, Lr34 and Lr37 on components of resistance in wheat to leaf rust. Plant Pathology 46:737-750.

Kuchel H., Fox R., Reinheimer J., Mosionek L., Willey N., Bariana H., Jefferies S. (2007) The successful application of a marker-assisted wheat breeding strategy. Molecular Breeding 20:295-308. DOI: 10.1007/s11032-007-9092-z.

Lande R., Thompson R. (1990) EFFICIENCY OF MARKER-ASSISTED SELECTION IN THE IMPROVEMENT OF QUANTITATIVE TRAITS. Genetics 124:743-756.

Lander E.S., Schork N.J. (1994) GENETIC DISSECTION OF COMPLEX TRAITS. Science 265:2037-2048.

Lewontin R.C. (1964) INTERACTION OF SELECTION + LINKAGE .I. GENERAL CONSIDERATIONS - HETEROTIC MODELS. Genetics 49:49-&.

Lewontin R.C., Kojima K. (1960) THE EVOLUTIONARY DYNAMICS OF COMPLEX POLYMORPHISMS. Evolution 14:458-472.

Li M., Reilly M.P., Rader D.J., Wang L.-S. (2010) Correcting population stratification in genetic association studies using a phylogenetic approach. Bioinformatics (Oxford) 26:798-806. DOI: 10.1093/bioinformatics/btq025.

Loiselle B.A., Sork V.L., Nason J., Graham C. (1995) SPATIAL GENETIC-STRUCTURE OF A TROPICAL UNDERSTORY SHRUB, PSYCHOTRIA OFFICINALIS (RUBIACEAE). American Journal of Botany 82:1420-1425.

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

Navakode S., Weidner A., Lohwasser U., Roder M.S., Borner A. (2009) Molecular mapping of quantitative trait loci (QTLs) controlling aluminium tolerance in bread wheat. Euphytica 166:283-290. DOI: 10.1007/s10681-008-9845-8.

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. DOI: 10.1038/ng1847.

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

Queller D.C., Goodnight K.F. (1989) ESTIMATING RELATEDNESS USING GENETIC-MARKERS. Evolution 43:258-275.

Ritland K., Ritland C. (1996) Inferences about quantitative inheritance based on natural population structure in the yellow monkeyflower, Mimulus guttatus. Evolution 50:1074-1082.

Roder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M.H., Leroy P., Ganal M.W. (1998) A microsatellite map of wheat. Genetics 149:2007-2023.

Roder M.S., Plaschke J., Konig S.U., Borner A., Sorrells M.E., Tanksley S.D., Ganal M.W. (1995) ABUNDANCE, VARIABILITY AND CHROMOSOMAL LOCATION OF MICROSATELLITES IN WHEAT. Molecular & General Genetics 246:327-333.

Rosenberg N.A., Pritchard J.K., Weber J.L., Cann H.M., Kidd K.K., Zhivotovsky L.A., Feldman M.W. (2002) Genetic structure of human populations. Science 298:2381-2385.

Schulze T.G., McMahon F.J. (2002) Genetic association mapping at the crossroads: Which test and why? Overview and practical guidelines. American Journal of Medical Genetics 114:1-11.

Sherman J.D., Souza E., See D., Talbert L.E. (2008) Microsatellite markers for kernel color genes in wheat. Crop Science 48:1419-1424. DOI: 10.2135/cropsci2007.10.0561.

Sherman J.D., Weaver D.K., Hofland M.L., Sing S.E., Buteler M., Lanning S.P., Naruoka Y., Crutcher F., Blake N.K., Martin J.M., Lamb P.F., Carlson G.R., Talbert L.E. (2010) Identification of Novel QTL for Sawfly Resistance in Wheat. Crop Science 50:73-86. DOI: 10.2135/cropsci2009.03.0145.

Tasma I.M., Lorenzen L.L., Green D.E., Shoemaker R.C. (2001) Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. Molecular Breeding 8:25-35.

Tautz D., Renz M. (1984) SIMPLE SEQUENCES ARE UBIQUITOUS REPETITIVE COMPONENTS OF EUKARYOTIC GENOMES. Nucleic Acids Research 12:4127-4138.

Tautz D., Trick M., Dover G.A. (1986) CRYPTIC SIMPLICITY IN DNA IS A MAJOR SOURCE OF GENETIC-VARIATION. Nature 322:652-656.

Wang J.L. (2002) An estimator for pairwise relatedness using molecular markers. Genetics 160:1203-1215.

Xu Y.B., Crouch J.H. (2008) Marker-assisted selection in plant breeding: From publications to practice. Crop Science 48:391-407. DOI: 10.2135/cropsci2007.04.0191.

Yu J.M., 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., Kresovich S., Buckler E.S. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203-208. DOI: 10.1038/ng1702.

APPENDICES

APPENDIX A

PHENOTYPIC LEAST SQUARES ENTRY MEANS

Genotype	Heading Date	Solid Stem Score	Plant Height (cm)	Yield (kg/ha)	Test Weight (kg/m^3)	Protein %
GLENMAN	180.41	16.49	77.67	3447.8	744.57	13.90
	204.02	17.59	75.73	4313.6	767.52	13.18
MT9410	176.29	17.68	86.53	3653.9	781.71	14.46
	199.92	17.90	80.44	3725.9	789.13	15.06
MT9453	179.67	16.91	86.50	3539.0	766.35	14.88
	202.34	15.73	81.57	4370.6	763.75	15.24
MT9507	176.16	16.00	78.32	3213.4	776.95	15.64
	199.15	19.07	74.36	3466.2	779.88	15.05
MT9508	176.53	19.09	75.12	3358.6	768.84	15.52
	198.46	20.59	70.91	3452.2	772.22	14.94
MT9539	179.18	15.80	75.88	3637.0	772.17	14.48
	201.60	17.06	70.84	3851.9	788.99	13.82
MT9542	180.28	18.54	76.13	3791.1	757.34	13.40
	203.74	20.18	72.29	4442.0	776.77	13.85
MT9558	180.58	18.86	78.22	3332.9	756.67	14.28
	202.18	19.38	74.74	4294.9	783.46	14.28
MT9565	176.69	7.16	83.76	3366.5	783.98	14.74
	200.40	8.69	76.56	3581.7	799.56	14.37
MT9602	177.20	20.20	84.64	3120.7	774.73	15.72
	199.10	22.10	79.45	3254.5	773.06	15.74
MT9619	179.75	17.13	86.42	3477.2	773.47	14.70
	203.07	15.87	81.30	4040.8	779.34	14.75
MT9627	179.49	13.17	75.55	3813.0	779.35	13.90
	202.97	13.87	71.25	4122.2	793.80	13.68
MT9628	176.92	5.32	78.17	3712.2	773.30	15.64
	199.50	7.67	74.27	3678.4	787.38	14.84
MT9631	180.65	17.07	91.05	3536.6	780.46	14.66
	203.81	17.32	83.93	4321.8	775.39	15.01
MT9644	179.33	12.85	90.29	3739.2	776.10	14.42
	201.14	14.43	80.67	3741.0	777.76	14.83
MT9660	179.33	3.89	79.46	3588.1	758.96	14.98
	203.76	7.55	75.36	4090.0	768.39	15.20
MT9662	181.34	5.85	72.38	3839.4	758.00	13.30
	204.59	7.44	68.75	5071.1	772.65	13.60
MT9667	178.96	14.16	87.22	3126.9	763.96	14.78
	200.69	13.60	81.16	3664.7	780.55	15.18
MT9675	181.91	7.93	79.22	3937.5	739.27	13.30
	203.98	8.48	73.55	4558.6	760.55	14.31
AMIDON	179.55	10.74	91.86	3665.9	760.34	14.15
	202.64	17.28	84.30	4141.7	754.75	15.33

Genotype	Heading Date	Solid Stem Score	Plant Height (cm)	Yield (kg/ha)	Test Weight (kg/m^3)	Protein %
HILINE	177.06	6.62	72.75	3729.7	764.80	14.83
	201.36	7.56	69.44	4193.0	791.61	14.98
NEWANA	181.41	6.76	74.71	3764.3	763.39	13.69
	205.23	8.73	71.75	4370.0	766.92	14.63
WESTBRED926	175.86	7.72	72.26	3685.2	755.65	15.17
	198.35	9.93	71.60	3816.1	772.73	14.88
SCHOLAR	180.75	11.23	88.25	3680.2	772.99	14.63
	203.60	14.40	81.19	4015.9	772.18	15.31
BZ902413	176.59	9.94	74.75	3849.9	781.04	14.56
	199.41	11.98	71.27	3802.5	774.86	14.64
BZ999592	179.39	7.34	76.69	3889.1	765.86	14.26
	202.99	8.59	72.78	4656.5	769.08	14.92
BZ9M1024	178.25	6.99	73.34	4044.4	770.19	13.99
	202.09	9.28	65.67	3910.5	790.73	13.70
CLEARWHITE	175.64	8.40	65.69	3793.9	763.81	13.79
	201.80	8.33	67.95	3818.3	782.58	12.92
MT0336	178.95	20.14	78.49	3994.8	778.48	14.29
	202.21	22.04	74.60	5166.4	773.16	14.12
MT0405	177.62	6.86	75.61	3881.8	773.82	14.59
	201.33	8.14	71.68	4403.3	773.26	14.83
MT0413	176.39	16.18	76.09	3861.0	775.19	14.52
	199.70	19.18	70.44	3691.2	787.16	14.37
MT0414	178.04	7.01	83.36	4102.5	778.70	14.72
	203.38	6.87	75.08	4663.5	781.23	15.53
MT0416	178.12	16.00	76.26	4139.7	776.11	13.76
	202.11	15.96	71.94	4272.1	775.38	14.60
MT0515	179.45	15.71	78.80	4209.3	789.85	14.49
	203.66	16.87	73.94	5199.3	771.71	14.42
MT0539	178.63	16.19	75.26	4033.1	756.83	14.06
	201.42	17.99	70.92	4030.4	773.11	15.56
MT0550	175.92	10.44	78.52	3635.3	789.51	14.09
	198.24	14.81	72.67	3475.0	785.58	14.86
MT0562	179.89	16.75	78.17	3749.3	764.68	14.22
	203.80	18.46	73.15	3864.4	780.41	14.62
MT0602	179.15	6.17	82.04	3891.0	760.82	14.26
	202.68	6.72	75.58	3944.1	779.43	14.55
MT0608	177.17	9.32	83.92	3674.7	777.94	14.39
	201.86	13.11	75.88	3590.7	779.42	15.08
MT0613	178.96	11.39	74.73	4059.1	781.39	14.56
	203.07	14.87	69.11	4219.0	787.84	14.69

Genotype	Heading Date	Solid Stem Score	Plant Height (cm)	Yield (kg/ha)	Test Weight (kg/m^3)	Protein %
MT0614	178.48	20.53	82.31	3789.6	772.88	14.59
	202.34	19.90	77.27	3679.4	759.53	15.70
MT0617	176.00	7.80	72.76	4024.3	759.14	14.12
	199.73	8.88	72.04	4271.8	757.97	14.66
MT0623	180.88	7.30	78.01	3844.7	751.66	14.49
	205.90	8.11	74.20	4638.5	753.96	14.19
MT0624	179.69	5.60	83.12	3815.9	775.39	13.56
	203.23	7.70	75.61	4111.3	773.26	15.08
MT0626	177.88	8.63	78.48	3956.8	771.61	13.72
	202.82	16.76	72.88	4609.8	767.26	15.08
MT0627	178.77	6.57	78.32	3767.7	718.32	14.29
	201.87	9.04	73.91	3952.7	748.02	15.66
MT0628	178.10	17.22	80.27	3942.8	752.93	14.59
	201.71	20.95	74.43	4294.8	777.92	15.09
MT0638	177.28	7.23	78.74	3991.2	783.16	14.59
	200.33	8.97	71.97	3820.1	779.24	15.72
MT0643	176.79	21.67	73.92	3794.2	779.97	14.52
	200.26	21.43	71.11	3363.5	779.32	15.96
MT0645	180.36	20.29	75.53	3686.9	765.43	14.56
	202.40	20.12	72.79	4384.2	770.89	14.47
MT0657	178.43	5.46	73.71	4025.9	766.06	13.82
	202.99	7.11	69.46	4096.0	790.31	14.09
MT0663	178.32	6.29	76.65	4131.0	782.29	13.69
	203.18	7.11	69.93	4660.5	797.43	13.99
MT0666	179.83	5.74	78.25	4018.5	755.62	13.79
	204.91	8.95	72.52	4439.8	777.80	13.36
MT0669	180.15	6.92	77.81	3959.9	780.70	14.52
	205.14	6.70	74.07	4162.5	776.50	14.48
MT0674	176.20	12.10	78.42	3826.3	761.75	14.56
	198.99	17.05	75.21	3626.6	759.82	15.13
MTHW0471	181.04	18.21	88.54	3778.7	781.87	14.86
	204.84	19.66	80.60	4495.1	786.30	13.66
HANK	176.82	8.86	73.93	3969.5	748.46	14.40
	198.64	8.74	69.58	4184.5	770.15	14.07
NORPRO	179.06	7.36	71.55	3976.2	765.84	13.92
	200.82	7.61	67.53	4168.3	771.82	14.40
REEDER	178.33	7.08	79.01	4010.5	777.60	14.67
	204.09	7.58	73.07	4549.7	771.80	14.88
CHOTEAU	178.42	22.64	74.25	3917.9	770.83	14.67
	201.89	22.50	69.77	4387.9	776.11	15.32

Genotype	Heading Date	Solid Stem Score	Plant Height (cm)	Yield (kg/ha)	Test Weight (kg/m^3)	Protein %
OUTLOOK	180.79	7.75	78.03	3970.1	752.63	14.05
	204.82	9.36	75.85	4309.8	749.19	15.31
EXPLORER	176.49	9.33	73.88	3426.7	754.58	14.21
	198.55	13.75	70.99	3668.7	768.07	14.55
KNUDSON	179.24	7.61	75.73	3846.2	770.99	14.21
	202.15	7.43	70.93	3772.4	775.34	14.23
MT0008	177.18	10.42	84.59	3462.2	763.94	15.18
	199.11	14.01	77.36	3820.5	771.70	15.68
MT0009	177.51	11.03	73.46	3871.8	779.87	14.11
	199.50	16.17	67.98	3584.9	778.95	14.89
MT0063	177.54	9.36	82.63	3709.7	756.44	14.25
	199.77	13.48	77.45	3708.5	762.34	14.92
MT0076	177.21	11.09	83.14	3585.4	766.96	15.25
	198.96	19.24	77.41	3878.4	774.66	15.84
MT0103	180.45	10.29	78.20	3868.7	766.49	14.41
	205.04	17.98	73.86	3813.1	762.86	14.14
MT0108	178.89	9.19	89.14	3311.8	756.41	15.08
	203.19	14.58	79.73	3755.8	754.26	15.77
MT0121	177.96	11.88	84.58	3244.5	790.03	15.35
	200.62	19.51	75.44	3620.7	789.76	15.87
MT0124	179.12	10.91	86.88	3570.9	764.53	14.95
	200.44	19.05	81.90	3748.3	783.53	14.89
MT0134	179.88	17.07	88.31	3639.1	764.26	14.01
	202.98	17.92	80.94	4354.0	773.45	14.99
MT0140	180.70	10.29	85.96	3240.6	784.44	14.38
	204.55	16.02	81.49	4316.9	784.30	15.72
MT0147	176.52	13.18	75.62	3798.5	793.12	14.71
	200.09	19.17	71.86	4162.0	791.93	15.37
MT0148	179.57	7.78	79.86	3755.4	794.53	14.45
	202.63	9.50	74.38	3710.5	788.11	15.06
MT0154	181.17	8.40	78.28	3603.2	781.00	14.38
	206.35	11.83	75.70	4406.8	756.20	15.00
MT0158	179.50	12.81	91.13	3741.2	773.28	14.55
	202.01	20.19	83.43	4185.2	774.42	15.09
MT9806	179.04	8.43	78.46	3877.1	767.73	15.08
	201.68	9.37	74.37	4732.5	765.17	15.18
MT9905	180.51	14.78	80.89	3645.8	790.87	14.85
	204.76	20.07	78.43	5238.9	785.09	14.82
MT9918	177.28	12.09	86.86	4137.9	755.64	13.95
	201.07	20.59	81.34	4259.7	766.49	14.39

Genotype	Heading Date	Solid Stem Score	Plant Height (cm)	Yield (kg/ha)	Test Weight (kg/m^3)	Protein %
MT9923	179.88	8.77	83.52	3438.1	781.89	14.78
	203.75	16.83	79.14	4715.5	796.92	15.32
MT9931	178.85	11.88	77.41	3794.7	783.81	14.51
	201.02	18.19	71.42	3866.3	789.99	15.41
MT9955	179.87	7.57	74.54	3966.6	765.70	14.15
	203.59	8.57	70.57	4092.7	771.88	14.05
MTHW0001	177.94	9.69	79.95	3582.0	775.63	13.85
	200.22	15.69	75.23	4582.9	790.95	14.38
MTHW0167	177.41	9.37	77.27	3443.9	767.43	13.71
	199.89	9.98	72.62	3671.4	785.43	13.65
MTHW0168	180.48	7.64	81.07	3857.6	772.64	13.61
	203.13	9.07	77.46	4437.0	774.09	14.87
MTHW9420	178.12	7.91	73.29	3469.0	753.83	13.68
	199.87	10.09	69.61	3987.6	778.78	13.70
MTHW9901	179.25	9.09	87.68	3577.0	776.57	14.11
	201.88	16.13	80.35	4468.2	788.93	14.41
MTHW9905	177.70	8.45	78.93	3697.4	763.79	13.95
	201.39	9.39	73.74	3913.5	787.38	13.66
FORTUNA	178.67	15.18	91.97	3352.4	768.63	14.66
	202.00	20.22	86.95	3751.3	786.51	15.06
THATCHER	181.41	7.13	96.29	3152.0	741.56	14.93
	204.14	8.97	90.68	3174.3	763.75	15.02
MCNEAL	180.25	7.04	79.02	3897.7	760.53	14.56
	203.60	7.84	73.57	3795.8	764.49	15.20
CONAN	178.22	9.88	72.54	3520.7	771.61	14.77
	200.95	11.81	68.94	3513.4	758.97	14.90
ERNEST	179.18	12.91	90.48	3614.1	773.32	15.05
	201.75	16.31	80.97	3691.7	777.71	15.17

[a] Values are color coded by the year the genotype was tested between 1997 and 2007. Light Blue = 1997, Tan = 2002, Pink = 2007, Dark Blue = 1997 and 2002, Dark Red = 2002 and 2007, and Black = All Years

[b] Historical values are listed above 2009/2010 values.

APPENDIX B

TYPE 3 ANALYSIS OF VARIANCE

Trait	Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Heading Date	ENV	1	62710	62710	48176.6	<.0001
	ENTRY	99	2446.66	24.713737	26.83	<.0001
	ENV*ENTRY	99	235.406667	2.377845	2.58	<.0001
	REP(ENV)	4	5.206667	1.301667	1.41	0.2289
	Residual	396	364.793333	0.921195	.	.
Stem Solidness	ENV	1	1226.892334	1226.892334	8.62	0.0425
	ENTRY	99	15056	152.082621	24.45	<.0001
	ENV*ENTRY	99	892.115305	9.011266	1.45	0.0072
	REP(ENV)	4	569.23658	142.309145	22.88	<.0001
	Residual	396	2462.787309	6.21916	.	.
Plant Height	ENV	1	12056	12056	265.84	<.0001
	ENTRY	99	13584	137.208127	20.44	<.0001
	ENV*ENTRY	99	1348.524583	13.62146	2.03	<.0001
	REP(ENV)	4	181.398333	45.349583	6.76	<.0001
	Residual	396	2658.268333	6.712799	.	.
Yield	ENV	1	103820	103820	161.67	0.0002
	ENTRY	99	30900	312.121854	7.07	<.0001
	ENV*ENTRY	99	10156	102.58362	2.33	<.0001
	REP(ENV)	4	2568.662907	642.165727	14.56	<.0001
	Residual	396	17470	44.116237	.	.
Test Weight	ENV	1	393.92391	393.92391	23.17	0.0086
	ENTRY	99	517.139794	5.223634	13.31	<.0001
	ENV*ENTRY	99	73.213387	0.739529	1.88	<.0001
	REP(ENV)	4	68.010701	17.002675	43.32	<.0001
	Residual	396	155.434087	0.39251	.	.
Grain Protein Content	ENV	1	16.006667	16.006667	25.22	0.0074
	ENTRY	99	248.095	2.50601	56.32	<.0001
	ENV*ENTRY	99	18.263333	0.184478	4.15	<.0001
	REP(ENV)	4	2.538933	0.634733	14.26	<.0001
	Residual	396	17.621067	0.044498	.	.

APPENDIX C

LINKAGE DISEQUILIBRIUM ESTIMATES BETWEEN FORTY-TWO RANDOM UNLINKED MICROSATELLITE MARKERS

	Xbarc83	Xwmc716	Xbarc187	Xwmc44	Xcfd63	Xwmc432	Xgwm312	Xwmc177
Xbarc83		0.02	0.04	0.01	0.02	0.04	0.01	0.02
Xwmc716	0.42		0.02	0.01	0.01	0.01	0.01	0.01
Xbarc187	0.41	0.04		0.02	0.01	?	?	0.01
Xwmc44	0.02	0.34	0.22		0.02	0.01	?	0.01
Xcfd63	0.32	0.01	0.75	0.01		0.02	0.03	0.01
Xwmc432	0.00	0.07	?	0.02	0.36		0.02	0.01
Xgwm312	0.27	0.53	?	?	0.01	0.07		0.03
Xwmc177	0.05	0.49	0.39	0.06	0.11	0.05	0.00	
Xwmc154	0.01	0.04	?	0.10	0.11	0.05	0.02	0.07
Xwmc361	0.89	0.06	0.01	0.78	0.14	0.52	0.09	0.38
Xgwm261	0.80	0.00	0.17	0.83	0.14	0.02	?	0.08
Xgwm349	0.67	0.64	?	0.23	0.08	0.58	?	0.07
Xwmc532	0.03	0.00	?	0.02	0.06	0.06	0.02	0.02
Xwmc559	0.53	0.00	?	0.04	0.43	0.00	0.42	0.35
Xbarc84	0.09	0.01	0.01	0.63	0.03	0.84	0.63	0.05
Xwmc78	0.96	0.67	0.25	0.16	0.03	0.00	0.05	0.00
Xgwm161	0.85	0.33	0.36	0.01	0.20	0.47	0.78	0.03
Xwmc552	0.78	0.10	?	?	0.02	?	0.35	0.80
Xbarc343	0.56	0.89	?	0.04	0.47	0.61	0.19	0.10
Xgwm601	0.00	0.00	0.47	0.03	0.00	0.01	0.02	0.02
Xwmc413	0.21	0.98	0.10	0.01	0.09	0.81	?	0.76
Xwmc710	0.35	0.00	?	0.33	0.02	0.00	?	0.01
Xcfd84	0.83	0.15	0.10	0.40	0.49	0.82	0.99	0.37
Xwmc285	0.63	0.41	?	0.12	0.01	0.62	0.00	0.01
Xbarc319	0.90	0.69	?	0.00	0.56	0.21	0.42	0.18
Xwmc713	0.01	0.04	?	?	0.02	0.00	?	0.00
Xgwm408	0.02	0.06	0.15	0.01	0.51	0.38	?	0.15
Xwmc376	0.60	0.47	0.76	0.26	0.08	0.02	0.04	0.00
Xcfd189	0.58	0.55	?	0.02	0.25	0.16	?	0.13
Xcfd29	0.73	0.10	0.65	?	0.40	0.76	?	0.09
Xgwm334	?	0.13	0.75	0.50	0.04	0.06	?	0.04
Xgwm427	0.18	0.77	0.01	0.08	0.60	0.03	0.02	0.13
Xgwm219	0.28	0.01	0.11	0.59	0.00	?	0.01	0.03
Xwmc487	0.07	0.05	?	?	0.01	0.12	?	0.00
Xbarc175	0.06	0.43	0.03	0.83	0.19	0.47	?	0.68
Xgdm132	0.02	0.04	0.67	0.81	?	0.04	0.46	0.16
Xgwm260	0.68	0.00	?	0.12	0.01	0.19	0.00	0.02
Xgwm332	0.02	0.00	0.02	0.01	0.09	0.09	?	0.02
Xwmc311	0.22	0.19	0.73	0.96	0.00	0.64	0.14	0.17
Xwmc426	0.13	0.03	0.00	0.02	0.02	0.81	?	0.01
Xgwm428	0.99	0.39	0.00	0.69	0.28	0.27	0.12	0.52
Xwmc506	0.66	0.00	?	0.06	0.04	0.16	0.00	0.01

	Xwmc154	Xwmc361	Xgwm261	Xgwm349	Xwmc532	Xwmc559	Xbarc84	Xwmc78
Xbarc83	0.03	0.01	0.03	0.01	0.04	0.01	0.03	0.00
Xwmc716	0.02	0.01	0.05	0.01	0.02	0.01	0.01	0.01
Xbarc187	?	0.02	0.02	?	?	?	0.02	0.01
Xwmc44	0.01	0.08	0.00	0.02	0.02	0.03	0.01	0.02
Xcfd63	0.02	0.04	0.07	0.01	0.01	0.01	0.04	0.04
Xwmc432	0.02	0.00	0.02	0.01	0.01	0.04	0.00	0.02
Xgwm312	0.01	0.02	?	?	0.02	0.01	0.01	0.01
Xwmc177	0.01	0.03	0.03	0.01	0.02	0.02	0.01	0.03
Xwmc154		0.02	0.08	0.01	0.03	0.01	0.04	0.00
Xwmc361	0.00		0.01	0.01	0.01	0.03	0.03	0.02
Xgwm261	0.06	0.34		0.01	0.04	0.04	0.06	0.01
Xgwm349	0.01	0.28	0.24		0.01	0.01	0.07	0.01
Xwmc532	0.00	0.04	0.01	0.20		0.03	0.04	0.03
Xwmc559	0.09	0.00	0.00	0.25	0.00		0.01	0.02
Xbarc84	0.07	0.02	0.21	0.46	0.00	0.17		0.02
Xwmc78	0.61	0.02	0.76	0.75	0.00	0.12	0.08	
Xgwm161	0.06	0.00	0.21	0.00	0.00	0.00	0.00	0.09
Xwmc552	0.02	0.12	?	?	0.53	0.00	0.37	0.33
Xbarc343	0.04	0.21	0.84	0.15	0.00	0.01	0.94	0.04
Xgwm601	0.00	0.01	0.00	?	0.00	0.01	0.02	0.04
Xwmc413	0.04	0.07	0.65	0.61	0.00	0.03	0.29	0.12
Xwmc710	0.01	0.00	0.00	0.00	0.25	0.00	0.44	0.00
Xcfd84	0.27	0.00	?	0.89	0.31	0.13	0.18	0.29
Xwmc285	0.02	0.56	0.69	0.00	0.01	0.11	0.72	0.26
Xbarc319	0.19	0.99	0.85	0.14	0.10	0.36	0.44	0.47
Xwmc713	?	0.04	0.06	?	?	?	0.01	0.11
Xgwm408	0.14	0.38	0.02	0.67	0.02	0.01	0.05	0.06
Xwmc376	0.01	0.02	0.11	0.07	0.13	?	0.39	0.00
Xcfd189	?	?	0.00	0.60	0.69	0.00	0.72	0.77
Xcfd29	0.09	?	0.02	0.20	0.03	0.47	0.43	0.36
Xgwm334	0.00	?	0.46	0.00	?	0.02	0.00	0.27
Xgwm427	0.50	0.01	0.18	0.12	0.10	0.03	0.36	0.18
Xgwm219	0.00	0.01	0.00	0.03	0.00	0.19	0.05	0.13
Xwmc487	?	0.00	?	?	?	?	0.01	0.00
Xbarc175	0.38	0.20	0.75	0.09	0.18	0.61	0.13	0.55
Xgdm132	0.00	0.01	0.13	0.10	0.07	0.12	0.42	0.12
Xgwm260	0.00	0.00	0.00	0.37	?	0.00	0.14	0.06
Xgwm332	0.01	0.01	0.10	0.00	0.00	0.01	0.07	0.25
Xwmc311	0.36	0.79	0.19	0.22	0.23	0.38	0.01	0.01
Xwmc426	0.04	0.00	0.95	0.45	0.01	0.51	0.13	0.15
Xgwm428	0.24	0.00	0.37	0.21	0.04	0.01	0.23	0.15
Xwmc506	0.04	0.05	0.02	0.16	0.01	0.01	0.01	0.00

	Xgwm161	Xwmc552	Xbarc343	Xgwm601	Xwmc413	Xwmc710	Xcfd84	Xwmc285
Xbarc83	0.01	0.01	0.02	0.05	0.04	0.01	0.01	0.01
Xwmc716	0.01	0.01	0.01	0.04	0.00	0.02	0.02	0.03
Xbarc187	0.02	?	?	0.02	0.03	?	0.01	?
Xwmc44	0.04	?	0.05	0.02	0.02	0.02	0.04	0.01
Xcfd63	0.02	0.03	0.00	0.03	0.05	0.03	0.01	0.02
Xwmc432	0.01	?	0.03	0.02	0.01	0.02	0.01	0.01
Xgwm312	0.01	0.01	0.03	0.02	?	?	0.00	0.05
Xwmc177	0.02	0.01	0.01	0.02	0.00	0.02	0.01	0.03
Xwmc154	0.02	0.02	0.03	0.06	0.04	0.02	0.01	0.03
Xwmc361	0.10	0.04	0.07	0.02	0.03	0.02	0.03	0.00
Xgwm261	0.01	?	0.00	0.06	0.01	0.03	?	0.00
Xgwm349	0.03	?	0.01	?	0.00	0.02	0.01	0.08
Xwmc532	0.02	0.01	0.03	0.03	0.04	0.01	0.01	0.01
Xwmc559	0.02	0.03	0.06	0.02	0.02	0.02	0.02	0.01
Xbarc84	0.03	0.01	0.00	0.03	0.00	0.01	0.00	0.03
Xwmc78	0.02	0.03	0.01	0.02	0.04	0.02	0.02	0.03
Xgwm161		0.04	0.03	0.04	0.02	0.01	0.03	0.02
Xwmc552	0.00		0.01	0.01	?	?	?	0.02
Xbarc343	0.12	0.32		0.01	0.00	0.01	0.04	0.01
Xgwm601	0.00	0.41	0.42		0.01	0.02	0.01	0.02
Xwmc413	0.13	?	0.53	0.18		0.03	0.01	0.01
Xwmc710	0.17	?	0.55	0.00	0.07		0.01	0.02
Xcfd84	0.08	?	0.25	0.73	0.48	0.20		0.00
Xwmc285	0.07	0.33	0.01	0.05	0.86	0.04	0.93	
Xbarc319	0.35	0.36	0.13	0.14	0.02	0.05	0.54	0.28
Xwmc713	0.03	?	0.78	0.00	0.16	?	0.12	0.14
Xgwm408	0.00	?	0.31	0.00	0.01	0.04	0.39	1.00
Xwmc376	0.00	0.03	0.03	0.03	0.00	0.00	0.07	0.01
Xcfd189	?	0.03	?	0.23	0.24	0.40	0.74	1.00
Xcfd29	0.08	?	0.11	0.00	0.08	?	0.81	0.24
Xgwm334	0.11	?	0.11	0.11	0.01	?	?	0.15
Xgwm427	0.33	0.29	0.81	0.06	0.11	0.00	0.44	0.04
Xgwm219	0.62	0.37	0.18	0.04	0.19	0.05	0.55	0.05
Xwmc487	?	0.11	?	0.00	0.02	?	0.21	?
Xbarc175	0.16	?	1.00	?	0.31	0.65	0.19	0.40
Xgdm132	0.07	0.02	0.13	0.02	0.39	0.02	0.23	0.85
Xgwm260	0.04	0.00	0.37	0.00	0.01	?	?	0.34
Xgwm332	0.40	?	0.03	?	0.37	0.00	?	0.04
Xwmc311	0.02	0.70	0.94	0.87	0.62	0.57	0.01	0.68
Xwmc426	?	?	0.10	?	0.75	0.02	0.38	0.26
Xgwm428	0.01	0.04	0.00	0.25	0.01	0.03	0.62	0.38
Xwmc506	0.01	0.26	0.12	0.00	0.02	0.03	0.00	0.05

	Xbarc319	Xwmc713	Xgwm408	Xwmc376	Xcfd189	Xcfd29	Xgwm334	Xgwm427
Xbarc83	0.01	0.04	0.06	0.02	0.00	0.03	?	0.02
Xwmc716	0.02	0.02	0.03	0.02	0.01	0.02	0.02	0.01
Xbarc187	?	?	0.04	0.01	?	0.01	0.04	0.01
Xwmc44	0.03	?	0.05	0.03	0.01	?	0.01	0.02
Xcfd63	0.01	0.03	0.02	0.04	0.01	0.01	0.01	0.01
Xwmc432	0.01	0.04	0.01	0.02	0.01	0.01	0.02	0.01
Xgwm312	0.00	?	?	0.01	?	?	?	0.03
Xwmc177	0.01	0.03	0.02	0.06	0.07	0.01	0.03	0.02
Xwmc154	0.02	?	0.03	0.03	?	0.03	0.02	0.01
Xwmc361	0.02	0.01	0.07	0.03	?	?	?	0.02
Xgwm261	0.00	0.05	0.04	0.02	0.13	0.05	0.01	0.01
Xgwm349	0.01	?	0.02	0.01	0.02	0.01	0.04	0.01
Xwmc532	0.02	?	0.03	0.02	0.01	0.03	?	0.02
Xwmc559	0.02	?	0.03	?	0.03	0.01	0.02	0.03
Xbarc84	0.02	0.01	0.04	0.01	0.01	0.01	0.06	0.05
Xwmc78	0.01	0.02	0.05	0.05	0.00	0.01	0.02	0.01
Xgwm161	0.01	0.02	0.09	0.04	?	0.01	0.02	0.02
Xwmc552	0.02	?	?	0.03	0.07	?	?	0.01
Xbarc343	0.02	0.01	0.01	0.01	?	0.02	0.02	0.01
Xgwm601	0.01	0.05	0.08	0.02	0.02	0.07	0.02	0.02
Xwmc413	0.07	0.05	0.05	0.11	0.02	0.03	0.03	0.01
Xwmc710	0.02	?	0.04	0.03	0.01	?	?	0.02
Xcfd84	0.01	0.01	0.03	0.01	0.00	0.02	?	0.01
Xwmc285	0.00	0.04	0.03	0.02	0.01	0.01	0.04	0.01
Xbarc319		0.02	0.01	0.03	?	0.04	0.03	0.01
Xwmc713	0.40		0.07	?	?	?	?	0.01
Xgwm408	0.26	0.08		0.02	0.01	0.02	0.02	0.03
Xwmc376	0.21	?	0.28		0.04	?	0.02	0.01
Xcfd189	?	?	0.57	0.04		?	0.00	?
Xcfd29	0.00	?	0.11	?	?		?	0.02
Xgwm334	0.00	?	0.20	0.03	0.86	?		0.01
Xgwm427	0.03	0.03	0.07	0.28	?	0.39	0.62	
Xgwm219	0.88	0.11	0.16	0.07	0.05	0.02	0.06	0.34
Xwmc487	0.24	?	0.51	?	0.75	0.09	0.02	0.90
Xbarc175	1.00	0.44	0.36	0.65	1.00	0.47	?	0.51
Xgdm132	0.64	?	0.22	0.05	0.71	?	0.00	0.90
Xgwm260	0.94	?	0.11	0.00	?	?	0.28	0.29
Xgwm332	0.16	?	0.26	0.15	?	?	?	0.03
Xwmc311	0.92	0.48	0.61	0.14	0.42	0.82	0.00	0.51
Xwmc426	?	?	0.17	0.67	0.95	0.09	?	?
Xgwm428	0.71	0.31	0.08	0.23	0.43	0.94	0.22	0.69
Xwmc506	0.03	?	0.08	0.01	?	?	0.00	0.41

	Xgwm219	Xwmc487	Xbarc175	Xgdm132	Xgwm260	Xgwm332	Xwmc311	Xwmc426
Xbarc83	0.07	0.02	0.00	0.02	0.01	0.05	0.01	0.02
Xwmc716	0.03	0.01	0.01	0.02	0.02	0.02	0.00	0.01
Xbarc187	0.02	?	0.02	0.01	?	0.02	0.01	0.06
Xwmc44	0.03	?	0.02	0.01	0.01	0.03	0.01	0.03
Xcfd63	0.07	0.02	0.01	?	0.03	0.02	0.04	0.01
Xwmc432	?	0.01	0.03	0.01	0.01	0.01	0.01	0.03
Xgwm312	0.02	?	?	0.01	0.02	?	0.00	?
Xwmc177	0.01	0.07	0.00	0.01	0.02	0.03	0.03	0.01
Xwmc154	0.06	?	0.04	0.03	0.02	0.03	0.01	0.02
Xwmc361	0.02	0.06	0.01	0.02	0.03	0.01	0.01	0.11
Xgwm261	0.06	?	0.02	0.03	0.04	0.02	0.00	0.02
Xgwm349	0.02	?	0.03	0.01	0.01	0.02	0.01	0.01
Xwmc532	0.03	?	0.02	0.01	?	0.02	0.02	0.02
Xwmc559	0.02	?	0.01	0.01	0.01	0.01	0.02	0.01
Xbarc84	0.02	0.02	0.00	0.01	0.01	0.02	0.05	0.01
Xwmc78	0.01	0.06	0.01	0.02	0.02	0.01	0.04	0.01
Xgwm161	0.01	?	0.03	0.01	0.02	0.01	0.01	?
Xwmc552	0.02	0.04	?	0.02	0.02	?	0.01	?
Xbarc343	0.01	?	0.00	0.03	0.01	0.04	0.00	0.02
Xgwm601	0.02	0.05	?	0.02	0.05	?	0.00	?
Xwmc413	0.01	0.08	0.02	0.01	0.03	0.01	0.01	0.00
Xwmc710	0.03	?	0.02	0.02	?	0.02	0.01	0.02
Xcfd84	0.01	0.04	0.03	0.02	?	?	0.09	0.03
Xwmc285	0.01	?	0.01	0.01	0.01	0.02	0.00	0.01
Xbarc319	0.03	0.01	0.00	0.01	0.01	0.02	0.01	?
Xwmc713	0.02	?	0.00	?	?	?	0.01	?
Xgwm408	0.01	0.04	0.03	0.01	0.03	0.03	0.02	0.04
Xwmc376	0.01	?	0.02	0.02	0.03	0.02	0.03	0.01
Xcfd189	0.01	0.01	0.00	0.01	?	?	0.02	0.00
Xcfd29	0.03	0.02	0.02	?	?	?	0.01	0.00
Xgwm334	0.01	0.04	?	0.02	0.02	?	0.06	?
Xgwm427	0.01	0.01	0.04	0.01	0.01	0.02	0.01	?
Xgwm219		?	0.03	0.03	0.01	0.03	0.03	0.02
Xwmc487	?		0.06	0.05	?	?	0.02	0.01
Xbarc175	0.40	0.05		0.03	0.02	0.01	0.07	0.01
Xgdm132	0.15	0.01	0.01		0.02	0.03	0.02	0.01
Xgwm260	0.36	?	0.08	0.00		0.02	0.01	0.01
Xgwm332	0.00	?	0.13	0.01	0.08		0.02	?
Xwmc311	0.15	0.30	0.02	0.02	0.41	0.34		0.00
Xwmc426	0.68	0.06	0.01	0.93	0.35	?	0.09	
Xgwm428	0.20	0.05	0.04	0.00	0.13	0.05	0.09	0.13
Xwmc506	0.02	?	0.01	0.04	0.17	0.10	0.09	?

	Xgwm428	Xwmc506
Xbarc83	0.00	0.00
Xwmc716	0.02	0.03
Xbarc187	0.01	?
Xwmc44	0.04	0.02
Xcfd63	0.02	0.02
Xwmc432	0.01	0.01
Xgwm312	0.02	0.01
Xwmc177	0.00	0.02
Xwmc154	0.03	0.01
Xwmc361	0.09	0.01
Xgwm261	0.00	0.08
Xgwm349	0.02	0.02
Xwmc532	0.01	0.01
Xwmc559	0.04	0.02
Xbarc84	0.01	0.04
Xwmc78	0.02	0.04
Xgwm161	0.05	0.02
Xwmc552	0.03	0.01
Xbarc343	0.14	0.01
Xgwm601	0.02	0.04
Xwmc413	0.05	0.07
Xwmc710	0.02	0.03
Xcfd84	0.03	0.03
Xwmc285	0.00	0.01
Xbarc319	0.05	0.02
Xwmc713	0.03	?
Xgwm408	0.05	0.06
Xwmc376	0.04	0.03
Xcfd189	0.02	?
Xcfd29	0.00	?
Xgwm334	0.04	0.06
Xgwm427	0.00	0.01
Xgwm219	0.01	0.01
Xwmc487	0.08	?
Xbarc175	0.07	0.04
Xgdm132	0.03	0.02
Xgwm260	0.02	0.02
Xgwm332	0.02	0.01
Xwmc311	0.05	0.06
Xwmc426	0.00	?
Xgwm428		0.01
Xwmc506	0.90	

P-values		r^2	
	≤.01		0-.05
	≤.05		.051-.1
			.11-.15
			.16-.2291
			≥.2292