Association Mapping of Endosperm Colour in Durum Wheat (*Triticum turgidum* L. var. *durum*).

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Plant Sciences

University of Saskatchewan

Saskatoon, Saskatchewan

By

Sherisse Opal Reimer

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements of a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised this thesis work or, in their absence, by the Head of the Department or the Dean of the College in which this thesis work was done. It is understood that any copying or publication or use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences

51 Campus Drive,

University of Saskatchewan,

Saskatoon, Saskatchewan, Canada,

S7N 5A8

ABSTRACT

Association mapping (AM), based on linkage disequilibrium, is a complementary strategy to traditional quantitative trait loci (QTL) mapping for describing associations between genotypes and phenotypes in crop plants. Yellow endosperm colour, an important quality trait in durum wheat (*Triticum turgidum* L. var. *durum*), was studied to determine the potential of AM to (1) identify previously reported QTL using a genome wide scan and (2) to determine allelic association of the phytoene synthase 1 (*Psy1*) gene using a candidate gene analysis. At present, a number of QTL for endosperm colour have been identified, and phytoene synthase, the initial enzyme of the carotenoid biosynthetic pathway, has been associated with QTL on the group 7 chromosomes which are considered to play a significant role in expression of yellow pigment concentration. CIE 1976 b*, a light reflectance measurement, and water-saturated butanol extracted pigments were assessed on a collection of 93 elite accessions from a variety of geographic origins, and genotyped with 245 markers. Population structure was assessed using genetic distance and Bayesian model based approaches, identifying five sub-populations consistent with breeding origin and pedigree. Association analysis identified significant associations with yellow endosperm colour on all chromosomes, including several previously identified QTL as well as new regions for genomic dissection. Pairwise LD mapping of *Psy1-B1* and *Psy1-A1* located the genes to chromosomes 7B and 7A respectively, to regions which have previously been identified for yellow pigment concentration QTL. The results of this study indicate that AM can be used to complement traditional QTL mapping techniques, and identify novel QTL for further study.

ACKNOWLEDGEMENTS

The financial assistance provided by the Graduate Fellowship of the Canadian Wheat Board, the Canadian Seed Growers Association award, and the Parr Postgraduate Scholarship was greatly appreciated. I would like to thank Dr. Curtis Pozniak (Durum wheat breeder, Crop Development Center) for serving as my supervisor, as well as committee members Dr. Bruce Coulman (Head, Department of Plant Sciences), Dr. Graham Scoles (Acting Dean, College of Agriculture and Bioresources), and Dr. Tom Warkentin (Field pea and Chickpea breeder, Department of Plant Sciences) for their comments and suggestions. Genotyping assistance provided by Dr. Daryl Somers (Agriculture and Agri-Food Canada Cereal Research Centre) and Krystalee Wiebe is greatly appreciated. Germplasm provided by Dr. John Clarke, Agriculture and Agri-Food Canada, Semiarid Prairie Agricultural Research Centre, Swift, Current is gratefully acknowledged. Assistance from the durum wheat field lab, quality lab and wheat molecular lab was appreciated.

THIS THESIS IS DEDICATED TO MY FAMILY FOR THEIR LOVE, SUPPORT AND ENCOURAGEMENT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

In plants, identification of quantitative trait loci (QTL) generally relies on segregating populations upon which polymorphic markers are evaluated, and subsequently associated with a phenotypic trait of interest (Kearsey and Farquhar 1998). This method has proven effective at identifying segregating QTL for traits for which the population was developed, but QTL resolution is generally poor and a limited number of alleles are sampled. In addition, multiple populations may be required for analysis of many traits of interest and this results in repeated expense of creating, phenotyping, and genotyping populations.

Association mapping (AM) is an alternative strategy to identify marker-trait associations and has been used extensively in human and animal genetic experiments (DeWan et al. 2006; Karlsson et al. 2007) where large segregating populations are not available. Association mapping has a number of advantages over other mapping techniques including the potential for increased QTL resolution, and an increased sampling of molecular variation, both factors associated with the use of unrelated populations which is possible with AM (Buckler and Thornsberry 2002; Yu and Buckler 2006). However, the probability of identifying spurious associations is higher in AM studies compared to traditional QTL mapping studies. Spurious associations are often the result of structured relationships within the population and may be reduced by taking population structure into account (Pritchard et al. 2000a; Yu et al. 2006).

Two approaches are commonly applied in association mapping (1) whole genome scans (Rafalski 2002; Kraakman et al. 2004) and (2) a candidate gene approach (Thornsberry et al. 2001; Wilson et al. 2004). Whole genome scans focus on identification of genomic regions on all chromosomes related to the trait of interest. Success and resolution of genome scans is dependent on the extent of linkage disequilibrium (LD). For example, increased LD decay, often represented by plotting LD versus genetic distance, requires a large number of closely linked markers, rendering the use of genome scans more laborious. Where a candidate gene for a trait has been identified, polymorphisms within the gene (for example single nucleotide polymorphisms; SNPs) can be correlated with phenotypic variation (Thornsberry et al.

2001) and are most useful when LD decays rapidly with increasing physical distance. The candidate gene approach has been effective at identifying single nucleotide polymorphisms in *Dwarf8* (Thornsberry et al. 2001) and *Y1* (Palaisa et al. 2003) associated with phenotypic variation in flowering time and β-carotene accumulation, respectively, in maize (*Zea mays* L).

The number of AM studies that have used structured populations of cultivars and breeding lines is increasing in self-pollinated crops, despite LD extending several centimorgans (cM) (Kraakman et al. 2004; Breseghello and Sorrells 2006a; Kraakman et al. 2006; Agrama et al. 2007; Cockram et al. 2008). In common wheat (*Tritcum aestivum* L.), several studies using AM have been reported, (Breseghello and Sorrells 2006a; Roy et al. 2006; Jing et al. 2007; Tommasini et al. 2007; Peng et al. 2008), but most studies have focused on individual chromosomes where QTL have been previously identified, or on traits controlled by one or more major genes (Ravel et al. 2006b; Breseghello and Sorrells 2006a; Tommasini et al. 2007). Common wheat and durum wheat (*Triticum turgidum* L. var. *durum*) are polyploid species with large genomes (16,000 Mb and 13,000 Mbp respectively; Arumuganathan and Earle 1991), and more research is required to validate the potential of AM to identify marker-trait associations using a whole genome scan in these species due to the complexities associated with multiple genomes. Also, to date there are only a few reports examining the potential of using a candidate gene approach for AM in self-pollinated crops.

To validate the usefulness of AM to identify genomic regions associated with a trait of interest, it is best to select trait(s) where an extensive amount of genetic information is already available, including putative genes that have been proposed as candidates for QTL. In durum wheat, the yellow pigment (YP) concentration of grain is a desirable end-use quality trait and elevated pigment concentration has been the target of durum breeding programs worldwide (Troccoli et al. 2000). In durum, the genetic control of YP concentration is multigenic and expression varies with the environment (Clarke et al. 2006), but YP concentration is highly heritable (Braaten et al. 1962; Elouafi et al. 2001; Clarke et al. 2006). QTL for YP have been identified on all chromosomes, except 2B (Parker et al. 1998; Elouafi et al. 2001; Mares and Campbell

2001; Hessler et al. 2002; Cervigni et al. 2005; Atienza et al. 2007; Pozniak et al. 2007; He et al. 2008; Patil et al. 2008; Zhang and Dubcovksy 2008), but the group 7 chromosomes are considered to play an important role in expression of YP in the grain (Elouafi et al. 2001; Pozniak et al. 2007; Patil et al. 2008; Zhang and Dubcovsky 2008). Pozniak et al (2007) identified *Psy1-B1*, a gene coding for a critical enzyme in the carotenoid biosynthethic pathway, as a candidate for the 7B QTL. Given that the genetics of YP concentration in durum wheat are well understood, and a candidate gene for an important QTL has been identified, YP concentration is a good trait to validate the potential of AM in a polyploid species.

2.0 LITERATURE REVIEW

Genes underlying genetic variation of quantitatively inherited traits are responsible for the majority of genetic diversity of interest in plant improvement programs. Quantitative traits are normally controlled by several genes at different loci with varying degrees of genetic contribution and expression is often influenced by the environment (Tanksley 1993) making identification of QTL more challenging than mapping a simple Mendelian trait. Discovery of quantitative trait loci (QTL) is useful for application in marker-assisted selection (MAS) and germplasm enhancement, and to better understand the genetics of complex traits (Asins 2002). In plants, there are two approaches to identify genomic regions influencing expression of quantitative traits. The most common approach is to identify QTL in a segregating population developed from a bi-parental cross of parents contrasting for the trait(s) of interest (Paterson et al. 1988; Kearsey and Farquhar 1998; Asins et al. 2002). Alternatively, a relatively new approach being applied in plants is association mapping, which is based on linkage disequilibrium (LD). In this approach to mapping, diverse populations of unrelated material are used to identify associations between allele frequencies and phenotypic variation. While extensive literature is available on identification of QTL from segregating populations, the use of AM in plants remains preliminary.

2.1 Identification of Quantitative Trait Loci (QTL) in Segregating Populations

Traditional methods to identify QTL in plants involve developing a segregating population from two genotypes varying in phenotypic values for a trait of interest, and via linkage mapping, significant marker-trait associations are identified. A variety of population types may be used, the most common including doubled haploids (DH), recombinant inbred lines (RILs), recombinant substitution lines (RSLs), F_2 and backcross populations. Segregating populations are evaluated with polymorphic markers and marker loci are then associated with variation in phenotypic expression of traits evaluated in multiple environments. Currently, simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) are viewed as the markers of choice for conducting mapping studies because of their abundance, amenability to high through-put analysis, and highly polymorphic nature (Röder et al. 1995; Macdonald et al. 2005). Although this approach to mapping identifies QTLs segregating in the population, large

populations are required to obtain high resolution which may not be feasible for difficult and expensive to measure traits (Buckler and Thornsberry 2002). Poor resolution limits the usefulness of QTL markers for marker-assisted selection, and makes identification of candidate genes coincident with the QTL more challenging. In wheat, the physical to genetic distance is variable along chromosomes, but ranges from 0.02 Mb/cM (Spielmeyer et al. 2000) to 16.7 Mb/cM (Akhunov et al. 2003). Thus a QTL spanning 10 cM could potentially contain hundreds of genes. Such a large number of genes cannot be tested for candidacy, therefore finer mapping in large populations must be pursued. Generally, fine mapping involves "mendelizing" the QTL in a near isogenic pair from which a large population (>1000) is created for fine mapping, often localizing QTL to <1 cM (Chicaiza et al. 2006; Price 2006). This approach has been used to fine map QTL in wheat for grain protein content (Olmos et al. 2003), grain weight (Röder et al. 2008), preharvest sprouting (Torada et al. 2008), and leaf and stripe rust resistance (Spielmeyer et al. 2008). Fine mapping QTL however is often associated with large expense, and is generally most effective for large effect QTL, as assessing phenotypic differences for small effect QTL is limited (Price 2006).

In a bi-parental mapping population, only the alleles present in the two parents are sampled and QTL localization is limited to loci segregating between the two parents. This prevents the identification of novel alleles and QTL outside of the mapping population and limits the ability to define smaller effect QTL and to map unrelated QTL (Jannink et al. 2001; Buckler and Thornsberry, 2002; Buntjer et al. 2005). One approach to overcome this limitation is to create several mapping populations, each of which must undergo genotypic and phenotypic evaluation at multiple environments. For example, in wheat, at least nine mapping populations have been reported for grain protein content and unique QTL have been identified in each of those populations (Blanco et al. 1996; Joppa et al. 1997; Prasad et al. 1999; Groos et al. 2003; Gonzalez-Hernandez et al. 2004; Turner et al. 2004; Blanco et al. 2006; Huang et al. 2006; McCartney et al. 2006). Due to the large number of populations that are required, associated genotyping and phenotyping expenses can be large (Jannink et al. 2001).

2.2 Association Mapping

Association mapping is based on LD, where correlations between alleles in a population occur as a result of non-random segregation at different loci, and though physical linkage may increase LD, LD is not necessarily due to physical linkage. When a mutation initially arises in a population, it is in disequilibrium with many genes throughout the genome. However, over many generations, segregation and recombination will "break" associations with other genes, and the mutation will only be in LD with alleles that are physically linked. Association mapping (AM) is a complementary strategy to QTL mapping to identify associations between genotype and phenotype (Yu and Buckler 2006), and takes advantage of this "historical" LD to identify marker-trait relationships. The basic objective of AM is to detect correlations between genotypes and phenotypes in a sample of unrelated individuals. This technique has been successfully employed in human and animal genetics (DeWan et al. 2006; Karlsson et al. 2007) where creating large populations of segregating individuals is not practical or feasible.

Compared to linkage mapping in traditional biparental populations, AM offers several advantages: increased sampling of allelic variation, increased mapping resolution, and reduced research time (Buckler and Thornsberry 2002; Flint-Garcia et al. 2003; Kraakman et al. 2004; Aranzana et al. 2005).

2.2.1 Sampling of Allelic Variation

Linkage mapping is restricted to sampling only the alleles differing between the two parents. In contrast, AM populations are generally comprised of a diverse collection of accessions and breeding lines, providing a greater number of alleles for sampling (Remington et al. 2001; Kraakman et al. 2004; Breseghello and Sorrells 2006; Stich et al. 2005; Crossa et al. 2007) (Figure 1). For example, in an AM population of common wheat the number of alleles averaged 4.8 per microsatellite locus (Breseghello and Sorrells 2006a).

An attractive feature of AM is that marker-trait associations can be studied in well-phenotyped germplasm pools and breeding populations of locally adapted varieties (Breseghello and Sorrells 2006b). Diverse populations of germplasm such as found in AM offer a greater number of alleles for sampling as a result of more recombination

events present and greater genetic diversity as compared to populations of narrow germplasm. Comparatively, NILs offer greater resolution than either F_2 or RIL mapping populations, however all remain limited by the number of alleles that may be sampled (Figure 1). Association mapping is further advantageous for its application in populations of unrelated individuals, in contrast to related populations studied in QTL mapping (Malosetti et al. 2007). Studying populations of unrelated individuals facilitates increased sampling of meiotic events, and provides the opportunity to identify novel alleles that may be contributing to a trait (Figure 1; Gaut and Long 2003; Tommasini et al. 2007).

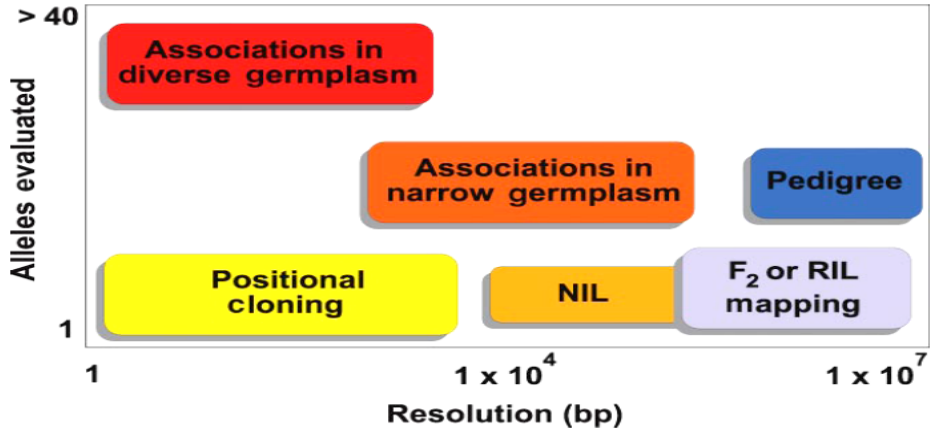

Figure 1. Resolution of QTL discovery methods (Flint-Garcia et al. 2003). Though high resolution mapping has been obtained using current techniques such as F_2, RIL or NIL for mapping, the number of alleles evaluated is limited. Conversely, association mapping provides diverse populations of germplasm from which alleles may be sampled.

Further, large populations increase power by providing the opportunity to identify alleles at higher frequency (Breseghello and Sorrells 2006a; Andersen et al. 2007). Small sample sizes often cause reduced power (Andersen et al. 2007; Chao et al. 2007), with higher levels of LD decay anticipated in smaller populations of low sequence diversity (Hamblin et al. 2004; Stich et al. 2007). Increasing sample size further facilitates increased power that may normally be reduced by interactions between alleles, such as those caused by epistasis, by allowing for interaction terms to be included in models (Wilson et al. 2004).

2.2.2 Resolution of Association Mapping

Association mapping theoretically allows mapping with higher resolution than achieved using bi-parental crosses (Gaut and Long, 2003; Remington et al. 2001; Thornsberry et al. 2001; Morgante and Salamini 2003; Skøt et al. 2007; Tommasini et al. 2007) (Figure 1). The degree of resolution depends on the extent of LD (Remington et al. 2001) and higher resolution is expected when LD declines rapidly with increasing genetic distance (Figure 2).

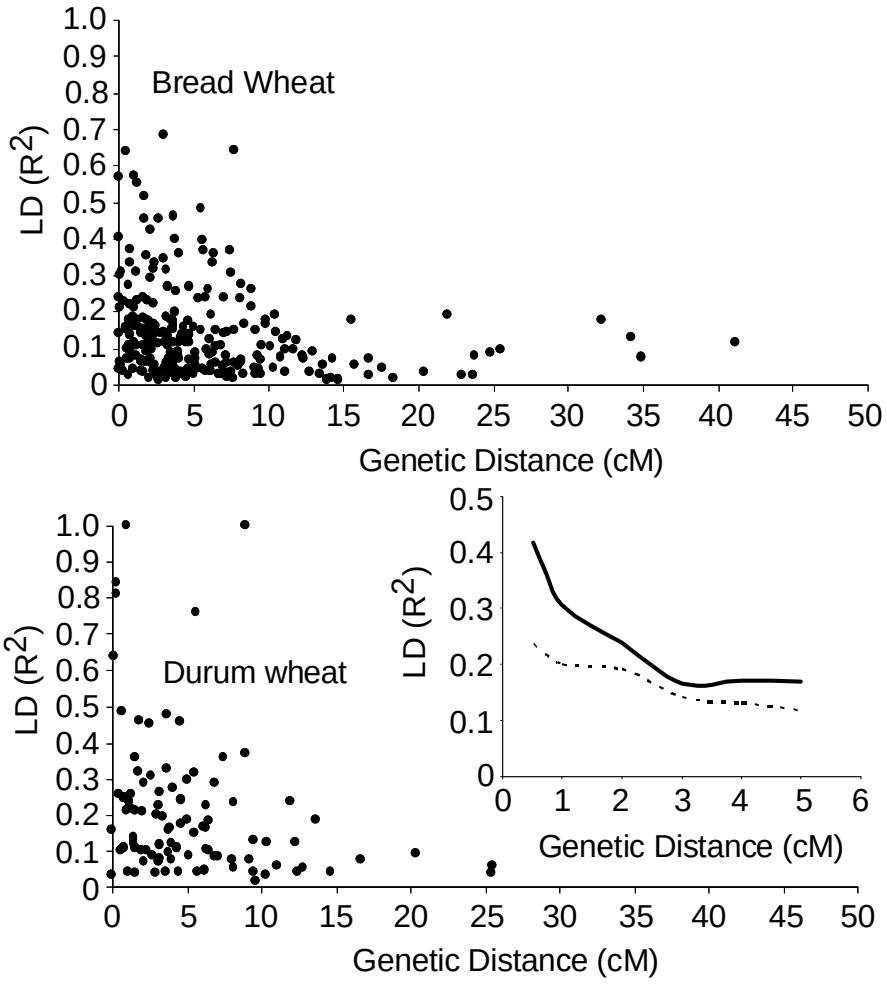

Figure 2. Extent of linkage disequilibrium in bread wheat and durum wheat (Somers et al. 2007). LD decay is visualized by plotting LD (R^2) versus genetic distance (cM). Inset indicates average rate of LD decay for durum wheat (solid line) and bread wheat (dashed line).

Linkage disequilibrium is a pairwise measure of allele frequencies at two polymorphic sites, where alleles that are highly correlated are said to be in linkage disequilibrium (Gaut and Long 2003; Nordborg and Tavare 2002). LD is commonly measured using either D' or r^2 (Lewontin 1964; Weir 1996; Hamblin et al. 2004; Gupta et al. 2005; Zhao et al. 2005; Somers et al. 2007), and summarized by plotting LD versus distance, allowing for visualization of the rate of LD decay over distance (Figure 2). As is indicated in figure 2, on average genome-wide LD decay in durum and bread wheat extends approximately 2 to 3 cM (r^2 <0.2). Small sample sizes however have a large effect on D' (Flint-Garcia et al. 2003). Therefore, for AM, the r^2 statistic is favored (Breseghello and Sorrells 2006; Rostoks et al. 2006; Agrama et al. 2007; Andersen et al. 2007; Crossa et al. 2007; Malosetti et al. 2007; Tommasini et al. 2007; Weber et al. 2007; Jun et al. 2008) for determining correlation of markers with the QTL of interest and resolution, as a result of its ability to account for mutational as well as recombinational history (Flint-Garcia et al. 2003).

Several factors influence LD among unlinked markers. LD is higher in autogamous species, due to lower effective recombination as individuals in a population are homozygous at any given locus (Nordborg et al. 2002; Morrell et al. 2005; Rostoks et al. 2006; Skøt et al. 2007). For example, LD in maize (an allogamous species) decays rapidly with increasing physical distance and extends approximately 100-2,000 bp (Remington et al 2001; Tenaillon et al. 2001). In contrast, LD is extensive in *Arabidopsis*, extending to 1 cM (250 kb) in a global sampling of ecotypes to as high as 50-100 cM (bp) when examining local ecotypes (r^2<0.2); Nordborg et al. 2002). Similarly, LD extends between 90 to 574 kb (r^2<0.4, Hyten et al. 2007) to 2.0–2.5 cM (1.0–1.5 Mbp) in soybean (*Glycine max* L. Merr.) (r^2 <0.1, Zhu et al. 2003), and 10 cM (bp) in barley (*Hordeum vulgare* L.) (r^2<0.2, Kraakman et al. 2004), both self-pollinating species. Genome wide LD was estimated to be less than 10 cM (bp) among a collection of 43 US wheat cultivars (r^2<0.2, Chao et al. 2007). LD estimates in durum are reported ranging from approximately 2-3 cM (r^2<0.2, Somers et al. 2007) to 20 cM (r^2<0.2, Maccaferri et al. 2005).

Admixture, or gene flow between individuals of genetically distinct populations, results in the introduction of alleles of different ancestry and will influence allele frequencies. This results in a buildup of LD along chromosomes and among unlinked genes. However, random mating will quickly result in LD breakdown, but this is not the case in self-pollinated crops for the reasons indicated above. Like admixture, LD blocks can occur in populations that have recently experienced a genetic bottleneck due to domestication. Morrell et al. (2005) report LD in wild barley (*Hordeum vulgare* ssp. *spontaneum*) as low as 300 bp, whereas LD in elite cultivars ranged from approximately 212 kb (Caldwell et al. 2006) to 10 cM (Kraakman et al. 2004). In hexaploid wheat, large LD blocks exist on the D genome (Somers et al. 2007), compatible with the genetic bottleneck created by recent polyploidization and addition of the D genome to form hexaploid wheat (Caldwell et al. 2004). Soybean has undergone several genetic bottlenecks, having lost 81% of its rare alleles when compared to its wild progenitors (Hyten et al. 2006), resulting in increased blocks of LD identified in three different genomic regions (Hyten et al. 2007).

As expected, selection pressure (locus specific bottleneck; Flint-Garcia et al. 2003) increases LD and is usually measured by comparing populations of elite cultivars which tend to have higher LD relative to distantly related populations (Andersen et al. 2007). In rice (*Oryza sativa* L.), Zhu et al. (2007) found higher LD in cultivated rice compared to populations of unselected wild species confirming simulation studies in maize (Stich et al. 2007). A study of elite breeding lines of maize indicated LD in phytoene synthase 2 (*Psy2)* extends 100-200 bp in comparison to phytoene synthase 1 (*Y1)* where LD extends 2000 bp as the result of selecting for increased endosperm colour associated with positive alleles at *Y1* (Palaisa et al. 2003). In wheat, selection at the *Yr17* locus has resulted in increased LD at that locus (Rhoné et al. 2007).

The extent of LD (and thus resolution) is not uniform across genomes and populations (Nordborg and Tavare 2002; Mather et al. 2007). In maize, LD has been reported to extend from 100 – 200 bp (Tenaillon et al. 2001; Palaisa et al. 2003) to as far as 1500 to 2000 bp (Remington et al. 2001; Palaisa et al. 2003). In rice, LD estimates were also found to be variable, relative to selection, outcrossing and recombination rates,

and greatest among five 500 kb regions in temperate *japonica* (>500kb), tropical *japonica* (150kb), *indica* (75 kb) and *O. rufipogon* (<40kb) (Mather et al. 2007), contrasting the genome wide LD extending 20 - 30 cM between SSR loci (Agrama et al. 2007). In a study of the *xa5* locus of rice, a race-specific bacterial blight resistance gene, LD extended only 100kb (Garris et al. 2003), comparatively larger than that of maize *Y1* (Palaisa et al. 2003), yet lower than the genome average suggesting possible effects of selection. Variability in LD exists between chromosomes in bread wheat, extending less than 1 cM on chromosome 2D, to approximately 5 cM on 5A (Breseghello and Sorrells 2006a). In common and durum wheat, LD ranges from 2 to 3 cM (Figure 2) but several genomic regions with LD less than 0.5 cM to greater than 30 cM were observed (Somers et al. 2007). The differences in LD along the wheat chromosomes could be due to recombination "hot spots" within the genome (Faris et al. 2000). In wheat, recombination preferentially occurs in the distal regions of chromosomes, where high density "gene-rich" regions are commonly found (Sandhu and Gill 2002). Though LD is variable in wheat, high recombination rates in distal chromosome regions correspond to low levels of LD observed on several wheat chromosomes, including the distal end of the group 7 chromosomes (Somers et al. 2007).

Understanding the extent of LD in the genome is required prior to conducting AM studies as the extent of genotyping required increases with rapid LD decay (Garris et al. 2003). Marker availability may be a limiting factor, particularly if LD is low (Rostoks et al. 2006). The best genotyping method must be chosen on the basis of the specific requirements of the envisioned genotyping project, and the resources available. Single nucleotide polymorphisms (SNPs) are preferred for genotyping as a result of their abundance, providing high marker densities for mapping (Ching et al. 2002). Technologies currently exist with the ability to genotype thousands of sites simultaneously (for example, Perlegen Sciences Inc. genotyping arrays, Affymetrix Inc. GeneChip arrays, and Illumina Inc. BeadArray technology coupled with the GoldenGate genotyping assay), however they are not necessarily cost effective for genotyping large panels with a modest number of SNPs (Macdonald et al. 2005). The majority of studies have found that simple sequence repeats (SSRs) or SNPs are the markers of choice when performing association studies, as a result of their ability to detect genetic variability

(Eujayl et al. 2001; Stich et al. 2006a). The high level of polymorphism that SSRs provide increases the power to detect LD and facilitates higher resolution mapping (Stich et al 2006a).

2.2.3 Approaches for Association Mapping

Recently, several AM studies have been published on a variety of crops including common wheat (Breseghello and Sorrells 2006a; Ravel et al. 2006b; Roy et al. 2006; Crossa et al. 2007; Jing et al. 2007; Tommasini et al. 2007; Peng et al. 2008), barley (Kraakman et al. 2004; Kraakman et al. 2006; Rostoks et al. 2006; Cockram et al. 2008), potato (*Solanum tuberosum* L.) (Malosetti et al. 2007), maize (Remington et al. 2001; Wilson et al. 2004; Weber et al. 2007), rice (Agrama et al. 2007), and recently the first report for durum wheat (Sanguineti et al. 2007). Kraakman et al. (2006) provided support for the potential of AM in barley with a number of the associations identified in their study in regions of QTL previously identified through linkage analysis.

Two approaches are used for AM studies: whole genome scans and candidate gene analysis (Thornsberry et al. 2001; Rafalski 2002; Kraakman et al. 2004; Rostoks et al. 2006). Whole genome scans are accomplished by saturating the genome with adequate marker coverage (as determined by the extent of LD; see section 2.2.2), in order to identify associations between markers and phenotypes of interest (Rafalski 2002). This approach is favored in situations where marker availability is a limiting factor or when LD extends for large distances, allowing for potential candidate regions associated with a trait of interest to be identified for further study (Remington et al. 2001). For example, Kraakman et al. (2004) estimated LD to extend approximately 10 cM in a collection of barley cultivars, comparatively greater than other inbreeding crop species (Breseghello and Sorrells 2006; Somers et al. 2007). The high level of LD in their study was not conducive to fine resolution mapping, but was useful for identifying regions which may be the subject of further fine mapping experiments (Kraakman et al. 2004). If LD decays too rapidly, the number of markers required to conduct genome-wide AM analysis increases significantly, resulting in AM focused on a candidate gene as an alternative approach for attaining high resolution.

Candidate genes that have been shown or are suspected to have a functional role in expression of a phenotype of interest can be used in AM studies where allelic variants are associated with phenotypic variation. In cases where LD among single nucleotide polymorphisms (SNPs) within the gene decays rapidly, AM could be used to identify the causal molecular polymorphism(s) responsible for trait differences. Using a candidate gene approach, SNPs in *dwarf8* were evaluated for association with flowering time and plant height in 92 maize inbred lines (Thornsberry et al. 2001). Nine polymorphisms, including a miniature transposable element (MITE) insertion in the promoter were associated with flowering time (Thornsberry et al. 2001) and this gene has since been validated as a causal factor influencing flowering time in maize (Andersen et al. 2005). In maize, molecular differences at *Y1* were associated with phenotypic variation in grain carotenoid concentration (Palaisa et al. 2003) and this gene has since been identified as the causal factor for elevated carotenoids in maize. However, association of SNPs with a trait still requires verification, as the SNP could be in disequilibrium with the causal factor, particularly if LD is high in the genomic region surrounding the gene. Thus candidate gene approaches are generally utilized to eliminate putative candidates for detailed functional studies. For example, a candidate gene approach was useful at eliminating three of eight candidates in a 70 kb region conferring resistance to *Xanthomonas oryzae* (Blair et al. 2003).

2.2.4 Population Structure

Population structure results from selection and high levels of admixture (individual accession membership proportion found in multiple sub-populations) in a population and results in increased LD between unlinked markers. (Nordborg and Tavare 2002; Cardon and Palmer 2003; Farnir et al. 2000; Rostoks et al. 2006). Population structure is often used in genetic studies to summarize relationships between individuals within and among populations, and can provide insight into evolutionary relationships. The probability of a Type I error increases in AM studies if population structure is not accounted for (Flint-Garcia et al. 2003; Gupta et al. 2005). The most well known example of population structure-related Type I error is a study of Type 2 diabetes among the Native Americans of the Pima and Papago tribes, where associations were initially made between the *Gm* haplotype and Type 2 diabetes (Knowler et al.

1988). However, the haplotype was later associated with Caucasian admixture (Knowler et al. 1988), where the frequency of *Gm* was higher in individuals with Caucasian alleles. When accounting for population structure, the *Gm* haplotype was no longer associated with Type 2 diabetes (Knowler et al. 1988). Likewise, Zhao et al. (2007a), found a decrease in the number of associations identified when population structure was modeled and incorporated into analysis, as opposed to when structure was unaccounted for in a germplasm collection of *Brassica rapa* L.

Several methods have been proposed for estimating population structure and modeling population structure in AM studies, including distance- and model-based methods (Pritchard et al. 2000a; Ahmad 2002; Lu et al. 2005; Yu et al. 2006; Camus-Kulandaivelu et al. 2007; Peleg et al. 2008). Distance-based estimates of population structure are generally based on clustering of individuals based on pairwise genetic distance estimates between individuals (Nei 1972; Rogers 1972; Nei 1978). Although visually appealing, distance-based methods are not suitable for statistical inference (Pritchard et al. 2000a). In contrast, model-based methods assign individuals probabilistically to one or more sub-populations (Pritchard et al. 2000a). The most common model-based approach is Bayesian modeling where allele frequencies are used to estimate the likelihood of an individual belonging to a particular subpopulation. This approach allows assignment of individuals to respective populations that can be integrated into statistical models to account for population structure in AM studies (Pritchard et al. 2000a).

With Baysian modeling, the number of sub-populations is usually estimated *a priori.* Often, known relationships (pedigree, origin of the individual) and/or genetic distance methods are used to estimate a realistic number of sub-populations for calculation of model-based assignments (Liu et al. 2003; Lu et al. 2005; Agrama et al. 2007; Chao et al. 2007; Hai et al. 2007). Hai et al. (2007) found that when assigning population structure among 69 bread wheat accessions, both methods led to similar assignment of individuals to sub-populations. In contrast, distance- and model-based methods were conflicting in a collection of US wheat cultivars and breeding lines, with

model-based assignments detecting population structure missed by distance-based analysis (Chao et al. 2007).

The software STRUCTURE (Pritchard et al. 2000a) has been developed to account for population structure (Pritchard et al. 2000a) and has been implemented in AM studies in a number of crop species including barley (Rostoks et al. 2006), wheat (Breseghello and Sorrells 2006; Crossa et al. 2007; Tommasini et al. 2007), and rice (Agrama et al. 2007). *Structure* utilizes a Bayesian modeling approach to assign individuals to sub-populations to minimize the LD among unlinked markers among subpopulations. Yu et al. (2006) proposed a unified mixed-model method to determine relatedness of samples in populations, resulting in a reduction in both type I and type II error, by combining population structure (Q) with relative kinship (K), accounting for multiple levels of relatedness. In an association study by Zhao et al. (2007b), when used alone or when combined with estimates of population structure, inclusion of the kinship matrix resulted in a reduced false-positive rate. Pedigree information has been proposed as a means to estimate K, but factors such as missing/incorrect pedigree information, selection and genetic drift can make interpretation of pedigree information difficult (Liu et al. 2003).

Rare alleles (commonly defined as occurring at frequencies <5 - 10%) (Tenaillon et al. 2001; Barnaud et al. 2006; Caldwell et al. 2006; Ravel et al. 2006a; Chao et al. 2007; Rhoné et al. 2007), inflate estimates of LD, reducing statistical power in AM studies (Wilson et al. 2004; Somers et al. 2007; Crossa et al. 2007). The presence of rare alleles can also increase LD between unlinked markers and increase the Type I error rate in AM studies. Removal of rare alleles, or subsequently pooling rare alleles into their own class (Pritchard et al. 2000a; Pritchard et al. 2000b; Maccaferri et al. 2005; Somers et al. 2007) is a common practice prior to conducting AM studies. Tightly linked markers may result in increased LD among unlinked markers and are best avoided when assessing structure (Falush et al. 2003).

Closely related individuals are more easily assigned to related populations, which may result in overestimating the number of sub-populations as a result of background LD (Pritchard et al. 2000a; Falush et al. 2003) which reduces the statistical power of AM

studies (Yu et al. 2006). In addition, Camus-Kulandaivelu et al. (2007) noted that assignment of individuals to sub-populations was variable when closely related individuals were present in the AM population. However, removal of highly related individuals to estimate the number of sub-populations has been suggested as an approach to minimize overestimation of the number of subpopulations (Liu et al. 2003; Breseghello and Sorrells 2006; Camus-Kulandaivelu et al. 2007).

2.3 Association Mapping of Yellow Pigment Concentration

Recently, a number of AM studies have been conducted in wheat (Ravel et al. 2006b; Breseghello and Sorrells 2006; Roy et al. 2006; Crossa et al. 2007; Sanguineti et al. 2007; Tommasini et al. 2007; Jing et al. 2007; Peng et al. 2008). However, several of these studies have focused on individual chromosomes where QTL have been previously identified (Breseghello and Sorrells 2006) or on traits controlled by one or more major genes (Ravel et al. 2006b Tommasini et al. 2007). A single study in durum wheat has been reported for root and shoot traits (Sanguineti et al. 2007) and research is required to validate the potential of AM to identify marker-trait associations using a whole genome scan in durum. Also, to date there are no reports in durum wheat examining the potential of using a candidate gene approach for AM. To validate AM studies, it is best to use traits where extensive genetic information is already available, such as for yellow pigment (YP) concentration in durum wheat.

2.3.1 Yellow Pigment in Durum Wheat

Yellow pigment (YP) concentration is a desirable end-use quality trait in durum wheat both visually and nutritionally, and elevated pigment concentration has been the target of durum breeding programs worldwide (Troccoli et al. 2000). Pasta yellowness is determined by several factors, including carotenoid pigments (Hentschel et al. 2002; Panfili et al. 2004), semolina extraction rate (Matsuo and Dexter 1980), processing conditions (Borrelli et al. 1999), and oxidative degradation by lipoxygenases (Borrelli et al. 1999).

2.3.2 Components of Yellow Pigment

The yellow pigment in durum grain is comprised primarily of the xanthophylls lutein and zeaxanthin and small concentrations of β-cryptoxanthin and β-carotene

(Hentschel et al 2002; Panfili et al. 2004). However, lutein is the predominant component, accounting for over 80% of the total yellow pigment concentration (Youngs 1988; Panfili et al. 2004). Lutein is evenly distributed throughout the kernel (Hentschel et al. 2002; Panfili et al. 2004) whereas β-carotene and zeaxanthin are concentrated near the outer layers of the kernel (Hentschel et al. 2002).

In plants, carotenoids are synthesized via the isoprenoid biosynthetic pathway. Carotenoid biosynthesis occurs in the plastids and enzymes involved in biosynthesis are associated with, or integrated in the membrane, and are nuclear encoded (Cunningham and Gantt 1998; van den Berg et al 2000; Gallagher et al. 2004). Phytoene synthase (*Psy*) is the first critical enzyme of carotenoid biosynthesis, catalyzing the formation of phytoene from two molecules of geranyl-geranyl pyrophosphate (GGPP; C20) (Cunningham and Gantt 1998). Phytoene, a colourless product, undergoes four desaturation reactions catalyzed by the enzymes phytoene desaturase (*Pds*) and ζ-carotene desaturase (*Zds*) to produce lycopene (Cunningham and Gantt 1998; van den Berg et al 2000; Fraser and Bramley 2004). From lycopene, cyclization reactions occur, yielding carotenoid products. Zeaxanthin is produced from β-carotene by the oxidation of the number three carbon on each ring (Cunningham and Gantt 1998). Lutein, the xanthophyll primarily responsible for yellow endosperm colour in durum, contains two rings, one β ring and one ε ring, each of which are oxygenated on the number three carbon (Cunningham and Gantt 1998).

Phytoene synthase is the rate-limiting enzyme of carotenoid biosynthesis and has been the target of study to improve levels of grain carotenoids in durum (Pozniak et al. 2007; Zhang and Dubcovsky, 2008), maize (Buckner et al. 1996; Palaisa et al. 2003; Wong et al. 2004), and rice (Wurtzel et al. 2001). In rice, maize, and sorghum (*Sorghum bicolor* L. Moench), at least three copies of *Psy* exist and have been designated as *Psy1*, *Psy2*, and *Psy3* (Gallagher et al. 2004; Li et al. 2008; Welsch et al. 2008). In maize, only *Psy1* has been associated with carotenoid accumulation in the grain. *Psy2* is only expressed in leaf tissue, and may be associated with the production of carotenoids responsible for protecting the photosynthetic apparatus from photo-oxidative degradation (Palaisa et al. 2003; Gallagher et al. 2004). *Psy3* has only recently been

identified and is involved in root carotenogensis. In addition, this gene is upregulated in response to abiotic stresses, drought, salt, and abscisic acid and has been suggested as a target gene to improve tolerance to abiotic stress in cereal crops (Li et al. 2008; Welsch et al. 2008). In wheat and durum, only *Psy1* and *Psy2* have been identified (Cenci et al. 2004; Atienza et al. 2007; Pozniak et al. 2007; Zhang and Dubcovsky 2008). Since durum wheat is an allotetraploid, two copies of *Psy1* and *Psy2* are present and map to the group 7 A and B, and 5 A and B chromosomes, respectively (Pozniak et al. 2007). Like maize, only *Psy1* is involved in expression of endosperm colour in durum (Pozniak et al. 2007).

2.3.3. Determination of Yellow Pigment Concentration

Several methods have been developed to determine yellow endosperm colour and pigment concentration in durum wheat including CIE 1976 b*, water-saturated butanol extraction (WSB) (method 14-50; AACC 2000), high-performance liquid chromatography (HPLC), and near-infrared reflectance (NIR) spectrophotometry (Johnston et al. 1980; McCaig et al. 1992; Oliver et al. 1992; Hentschel et al. 2002; Panfili et al. 2004; Fratianni et al. 2005; McCaig et al. 2006). Commission Internationale de L'Éclairage (CIE) 1976 b* is a light reflectance measurement that assesses the relative degree of yellowness of a sample, and can be used to estimate differences in colour without the use of chemical extraction of pigments (Commission Internationale de L'Éclairage 1986; Oliver et al. 1992). NIR measures the near-infrared and visible range, and similar to CIE 1976 b* measurements, is advantageous as it does not require the use of chemicals (McCaig et al. 1992). High correlations have been observed between chemical methods and NIR (McCaig et al. 1992). Reflectance measurements are particularly beneficial for breeding programs, facilitating rapid screening of yellow colour among accessions, without the need for chemical extraction. Extraction of pigments using WSB and assessment of extract using a spectrophotometer at 435 nm is the standard method for determination of yellow pigment concentration (AACC 2000). Although WSB is used to estimate total yellow pigment concentration extracted from grain, HPLC is required to estimate relative carotenoid composition of grain kernels (Panfili et al. 2004; Fratianni et al. 2005). However, given that the majority of carotenoids in durum grain are lutein (Panfili et al. 2004), the correlations

between HPLC estimates of carotenoid components and WSB extraction are in the range of 0.94-0.98 (Fratianni et al. 2005; Abdel-Aal et al. 2007).

2.3.4 Genetics of Yellow Pigment

Yellow pigment concentration is a quantitative trait controlled by additive gene action (Johnston et al. 1983; Elouafi et al. 2001; Mares and Campbell 2001) and is highly heritable (Parker et al. 1998; Elouafi et al. 2001; Clarke et al. 2006; Patil et al. 2008). Several QTL for yellow endosperm colour have been identified in both durum and bread (hexaploid) wheat on chromosomes 1A (Patil et al. 2008), 1B (Cervigni et al. 2005; He et al. 2008), 3A (Parker et al. 1998), 3B (Mares and Campbell 2001; Patil et al. 2008), 4A and 5A (Hessler et al. 2002), 2A, 4B and 6B (Pozniak et al. 2007), 5B (Patil et al. 2008), 6A (Cervigni et al. 2005) and the group 7 chromosomes (Parker et al. 1998; Elouafi et al. 2001; Cervigni et al. 2005; Atienza et al. 2007; Pozniak et al. 2007; He et al. 2008; Patil et al. 2008; Zhang and Dubcovsky 2008). Two minor QTL on 4A and 5A were associated with flour colour by Hessler et al. (2002). Genes affecting expression of yellow pigment in durum have been associated with a number of chromosomes including caroteoid biosynthetic enzymes phytoene synthase (group 7 chromosomes; Elouafi et al. 2001; Atienza et al. 2007; Pozniak et al. 2007; Zhang and Dubcovsky 2008), phytoene desaturase (group 4 chromosomes; Cenci et al. 2004), zeta-carotene desaturase (group 2 chromosomes; Cenci et al. 2004), as well as lipoxygenases, which are known to result in degradation of yellow pigment (group 4 chromosomes; Carrera et al. 2007). Thousand-seed weight was significantly associated with the locus on 5A for flour colour, suggesting that larger kernels potentially result in a dilution of yellow colour (Hessler et al. 2002). Clarke et al. (2006) have also suggested a relationship with kernel weight and pigment concentration.

The majority of mapping studies are in agreement that QTL on the group 7 chromosomes largely influence the expression of grain pigment concentration in wheat and durum. Elouafi et al. (2001) identified two minor QTL located on chromosome 7A and a third major QTL on chromosome 7BL, believed to be homoeologous to one of the QTL on 7AL. Together, these three QTL accounted for 62% of the total phenotypic variation for yellow pigment concentration, the QTL on 7B alone explaining 53% of the variation (Elouafi et al. 2001). In a recent study published in durum, Patil et al. (2008)

identify a QTL for yellow pigment concentration on chromosome 7A accounting for up to 55% of the variation of the trait. Mares and Campbell (2001) identified a QTL on 7A in two common wheat populations having a large effect on flour yellowness, corresponding to regions associated with both CIE 1976 b* and xanthophyll. In common wheat, Kuchel et al. (2006) identified a QTL for flour colour on chromosome 7BL centered around *gwm344*.

Further to this work, Pozniak et al. (2007) located a QTL on chromosome 7BL, in the same region previously identified by Elouafi et al. (2001) contributing significant variation to yellow pigment concentration. The same authors showed that a *Psy1-B1* was associated with that QTL and a likely candidate for functional analysis. Zhang and Dubcovsky (2008) and Atienza et al. (2007) confirmed *Psy-B1* as a candidate gene. Recently, *Psy1-A1* (a homoeologue of *Psy1-B1*) has been localized to the yellow pigment QTL on 7AL in durum (C. Pozniak, personal communication).

2.4 Hypotheses and Objectives

To validate AM it is best to select trait(s) with genetic information already available, including putative genes that have been proposed as candidates for QTL. Yellow pigment (YP) concentration is a desirable end-use quality trait in durum wheat and elevated pigment concentration has been the target of durum breeding programs worldwide (Troccoli et al. 2000). Pasta yellowness is determined by several factors, including carotenoid pigments (Hentschel et al. 2002; Panfili et al. 2004), semolina extraction rate (Matsuo and Dexter 1980), processing conditions (Borrelli et al. 1999), and oxidative degradation by lipoxygenases (Borrelli et al. 1999). The genetics of this trait have been well studied (Parker et al. 1998; Elouafi et al. 2001; Mares and Campbell 2001; Hessler et al. 2002; Cervigni et al. 2005; Atienza et al. 2007; Pozniak et al. 2007; He et al. 2008; Patil et al. 2008; Zhang and Dubcovsky 2008), and a candidate gene for one of the QTL have been reported (Pozniak et al. 2007; Zhang and Dubcvosky 2008). Yellow pigment concentration is an attractive trait for evaluation of AM in durum, given its high heritability and the localization of several QTL regions associated with yellow endosperm colour using traditional QTL mapping techniques. The hypothesis of this research is that association mapping (AM) will identify genomic regions influencing

expression of grain yellow pigment concentration in durum wheat. If this hypothesis is true, AM should be able to detect the same QTL previously identified using bi-parental mapping populations. The objectives of this study were to evaluate the effectiveness of genome scans to identify previously reported QTL for YP in bi-parental wheat mapping populations using a diverse collection of durum wheat varieties and breeding lines collected from major durum wheat breeding programs worldwide. A second objective was to determine the association of allelic variation in two homoeologous *Psy1* genes with variation in yellow pigment concentration using AM (candidate gene approach).

3.0 MATERIALS AND METHODS

3.1 Plant Materials and Field Testing

A collection of 93 diverse accessions obtained from breeding programs across global durum growing regions were used in this study and will collectively be referred to as the AM population (Appendix 1). This collection was composed of germplasm from breeding programs in Canada (Agriculture and Agri-Food Canada (AAFC) – n=20; University of Saskatchewan Crop Development Centre (CDC) – n=5), Argentina (n=5), Australia (n=8), France (n=3), Italy (n=17), Germany (n=2), Mexico (n=3), Morrocco (n=3), United States (n=11), New Zealand (n=2), Russia (n=1), Iran (n=4), and Spain (n=9). In 2005 and 2006, the AM population was planted in an alpha-lattice experimental design (Patterson and Williams 1976) with two replications grown in replicated field trials conducted at Swift Current and Saskatoon, Saskatchewan, Canada. At Saskatoon, trials were grown in 5 row plots seeded 1.22 m wide x 2.44 m long. Swift Current trials were grown in 4 row plots 0.92 m wide x 3 m long using the same seeding rate as Saskatoon. The seeding rate was approximately 250 seeds m^{-2} at both locations. Once all plots had reached physiological maturity, they were harvested with a small plot combine, dried to a constant moisture and stored for lab analysis.

3.2 Trait Analysis

For assessment of endosperm colour, grain from each plot was ground in a UDY Cyclone Sample Mill (UDY Corporation, Fort Collins, Colorado) with a 1-mm screen. The sample was placed in a flat-bottomed glass dish and a Hunter Lab Miniscan XE Plus (Hunter Lab Associates Inc) was used to measure CIE 1976 b*, a measure of the "yellowness" of meal samples (Commission Internationale de L'Éclairage 1986). Yellow pigment (YP) concentration (mg kg-1) was measured following AACC Method 14-50 (AACC 2000). Briefly, 40 mL water-saturated butyl alcohol was added to 8 g of whole meal, shaken for one minute (270 rpm) and extracted for 16 hours. Extract was then filtered through Whatman No. 1 filter paper and absorbance measured at 435nm using a Microplate Reader (BioRad, CA, USA). Three individual measurements per extracted sample were recorded and absorbance values were averaged and converted to YP concentration (mg kg^{-1}) using the extinction coefficient for β-carotene (AACC

2000). Percent moisture was determined using AACC method 44-15A (AACC 2000) and YP concentrations were converted to a 14% moisture basis prior to data analysis. The weight of one thousand kernels (g) was measured on a plot basis following counting of seeds with an electronic seed counter.

3.3 Phenotypic Data Analysis

Data for yellow pigment concentration (mg kg^{-1}), CIE 1976 b*, and seed thousand kernel weight (KWT; g) were analyzed separately for each environment using PROC MIXED in SAS (Littell et al. 1996) with accessions as fixed effects and replications and blocks considered as random effects. Error variances were homogenous across environments, so data were combined and subject to ANOVA using PROC MIXED with blocks (b), replications (r), locations (l), years (y) considered as random factors. All interactions with random factors were also considered as random effects in the mixed linear model. Pearson's correlation coefficients among least square (LS) means for each environment were estimated using PROC CORR in SAS. Heritability (h^2) estimates were calculated as ratios of genetic and phenotypic variances (σ^2_g/σ^2_p), with accessions considered random. For individual environments, the phenotypic variance (σ^2_p) was estimated as $[\sigma^2_g + \sigma^2_e/r]$ where σ^2_g is the genetic variance, σ^2_e is the residual variance and r is the number of replications. For the combined analysis, phenotypic variances for each trait were estimated as the sum of the genetic variance and average variance estimates for genotype by location (σ^2_{g*l}), genotype by year (σ^2_{g*y}), genotype by location by year (σ^2_{g*l*y}) and residual variances (σ^2_e), such that $\sigma^2_p = (\sigma^2_g + \sigma^2_{g*l}/l + \sigma^2_{g*y}/y + \sigma^2_{g*l*y}/ly + \sigma^2_e/lyr)$ (Clarke et al. 2006).

3.4 Microsatellite Analysis

Microsatellite analysis of the 93 accessions was performed on a single plant of each accession (Somers et al. 2007). Given that most cultivars were expected to be heterogeneous at some regions in the genome, a representative plant was identified for the accession chosen based on genotyping of ten individual plants with twelve unlinked, genome specific microsatellite markers (Somers et al. 2007). A total of 245 microsatellite markers (Appendix 2) were selected based on a) good genome coverage (Somers et al. 2004), b) high polymorphism information content (PIC, Botstein et al.

1980), c) robust amplification and d) locus specific amplification. In addition, microsatellite markers previously associated with phenotypic variation in YP (Pozniak et al. 2007; Zhang and Dubcovsky, 2008) were also included. Polymerase chain reaction (PCR) was performed in a 25 μL reaction containing: 60 ng DNA, 1X buffer (GenScript), 200 μM dNTP, 0.08 μM forward primer, 0.32 μM reverse primer and 0.15 μM M13 primer 1.25 U Taq DNA Polymerase (GenScript, Cat No. E00007). Genotyping was performed on an ABI 3100 capillary electrophoresis (Applied Biosystems, Foster City, California) using M13-labeled primers (Schuelke 2000). Number of alleles at each locus, and polymorphism information content (PIC) (Botstein et al. 1980) were determined using the program PowerMarker v 3.25 (Liu and Muse 2005).

3.5 Population Structure

Two approaches were used to estimate population structure within the AM population. A pairwise genetic similarity matrix was calculated based on Rogers' Euclidean distance (Rogers 1972), and was utilized to construct a dendrogram within the program PowerMarker (Liu and Muse 2005). Only those microsatellite markers that produced polymorphic fragments were used to calculate the similarity matrix. To test the reliability of the relationships, a bootstrap analysis with 100 replications was performed and a majority-rule consensus tree was reconstructed using UPGMA and displayed using "TreeView" software (Page 1996). A Bayesian clustering approach was also used to infer the number of sub-populations (K) and to assign individuals to sub-populations based on membership proportion in each sub-population (Q-matrix) with the software STRUCTURE V.2 (Pritchard et al. 2000a). For Bayesian analysis, 28 unlinked microsatellites (one for each chromosome arm) were selected based on high PIC values (Appendix 2). Markers with a high frequency of rare alleles were avoided because rare alleles inflate LD (Gaut and Long 2003; Somers et al. 2007). Three independent STRUCTURE runs were conducted for each K=3 to 7 and an average log likelihood value across runs was estimated for each K. Posterior probability of the data was calculated from the average log likelihood and the highest likelihood of the data was observed for K=5. Therefore, the Q matrix, which is defined as the assignment of individuals to a subpopulation based on membership proportion, was estimated as the

average of three runs for K=5. STRUCTURE parameter settings included admixture model, allele frequencies correlated, an initial run length (burn-in) of 10000, and 10000 repetitions.

3.6 Marker-Trait Associations

Marker-trait associations were tested with a linear mixed-effects model within the program TASSEL v. 2.0.1 (Yu et al. 2006) using LS means from each environment. The Q matrix estimated for K=5 provided covariates. Prior to analysis, rare alleles (frequency <5%) were either a) combined into one genotypic class if their combined frequency was greater than 5% or b) scored as missing data. Associations were declared significant if the p-value of the F test was $\leq$0.05. Significance levels were corrected to adjust for multiple comparisons by performing 10,000 permutations.

3.7 Allelic Variation at Phytoene Synthase 1 (*Psy1*) and LD Mapping of *Psy1*

Two *Psy1* loci have been reported in durum wheat on chromosomes 7AL and 7BL (Pozniak et al. 2007). Partial sequences of the *Psy1* genes were obtained for a subset of 39 lines from the AM population to identify allelic variation at each of the two *Psy1* loci (*Psy1-A1* and *Psy1-B1)* for LD mapping. Lines were selected to include the majority of Canadian accessions, plus a subset of high and low pigment concentration lines from the remaining sub-populations (Appendix 1). Partial sequences (approximately 1125bp) of *Psy1-A1* and *Psy1-B1* were PCR amplified and cloned from the 5' region using primers Psy1F5 (5'-GCG AGG AGT ATG CCA AGA CCT-3') and Psy1R5 (5'-AAG GCC GAC AAA CGA AAC AAT-3') as this primer set is known to amplify both loci (Pozniak et al. 2007). PCR reactions were performed in 25 µL volumes containing 100 ng genomic DNA, 1X PCR Buffer, 1.5 mM $MgCl_2$, 200 µM dNTP, 0.5 µM of each primer, 5% (v/v) dimethyl sulfoxide (DMSO), and 2.5 units of *Taq* DNA polymerase (Invitrogen). PCR cycle conditions were: 95˚C for 5 min, 42 cycles of 94˚C for 1 min, 57˚C for 1 min, 72˚C for 1 min 20 sec, final extension of 72˚C for 30 min. Following PCR, 4 µL of fresh PCR product was cloned into TOPO TA Cloning® Kit for Sequencing (Invitrogen) as per the manufacturer's instructions. Ligation reactions were incubated at room temperature for 30 min and transformed into One Shot® TOP10 Competent Cells. Approximately 75 µL of each transformation was

spread on a pre-warmed selective plate containing 100 μg mL^{-1} ampicillin. Though the One Shot® TOP 10 competent cells contain a *lac*Zα-*ccd*B lethal gene, which when disrupted in theory only allows growth of positive colonies, each plate was spread with 40 μL Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) (20 mg mL^{-1}) to facilitate identification of positive colonies using blue/white colony screening, as a result of the large number of false positive colonies present during an initial screening. Plates were incubated for 18 hours at 37˚C and recombinant clones transferred into 3 mL Luria Broth Base (LB) media (Becton, Dickinson and Company) containing 100 ug mL^{-1} ampicillin and incubated at 37˚C overnight, while shaking at 200 rpm. To confirm insertion of the PCR amplicon, PCR analysis was performed as described above using 1 uL of colony growth as DNA template. Amplicons were digested with *Hpa*II (Invitrogen) in a 10 μL reaction containing 2.5 uL PCR Product, 0.5 uL *Hpa*II (Invitrogen), and 1 uL 10X React 8 Buffer at 37˚C for 1hour, followed by 90˚C for 15 min. Digested fragments were separated on a 1.5% (w/v) agarose gel. *Psy1-B1* cuts with a major identifying band of 857 bp, while *Psy1-A1* was indicated by a band of 759 bp (Pozniak et al. 2007). Two Langdon disomic substitution lines (Joppa and Williams 1988) were included to validate chromosome location. The Langdon7D(7A) substitution is a line where chromosome 7D from hexaploid wheat cultivar 'Chinese Spring' is substituted for 7A. Likewise, The Langdon7D(7B) is missing chromosome 7B but has in place 7D. A minimum of two clones each of *Psy1-B1* and *Psy1-A1* obtained from independent PCR reactions were sequenced from each of the 39 accessions at the Plant Biotechnology Institute, National Research Council, Saskatoon, Saskatchewan, Canada. Raw sequences were edited to remove vector sequence and aligned to reported sequences of *Psy1-B1* and *Psy1-A1* (Genebank accession no. DQ642440 and DQ642439; DQ642443 respectively) using Vector NTI Advance 10 (Invitrogen). Singleton variants were confirmed by re-sequencing (both directions) of clones derived from additional PCR products to eliminate PCR artifacts.

Partial sequence analysis of *Psy1-A1* and *Psy1-B1* revealed the presence of two alleles at each locus. The two alleles correspond to previously identified *Psy1-B1* alleles (Pozniak et al. 2007) and a co-dominant marker distinguishing these alleles (Zhang and Dubcovsky 2008) was scored on the AM population. Briefly, primers PSY1BF3

(GTGGAACTTGCATGCTATACA) and PSY1BR2 (GAACCTCAGGTTCACATTCC) (Zhang and Dubcovsky 2008) were used to amplify genomic DNA using the following PCR conditions: 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 2.5 μM of each primer, 1 U *Taq* polymerase (Invitrogen), and 100 ng of genomic DNA. Cycling conditions were: 94 ˚C for 1 min, 60 ˚C for 1 min, and 72 ˚C for 1 min. The PCR was completed with a 10 min extension at 72 ˚C. The *Psy1-B1a* allele was characterized by an approx. 217 bp fragment, whereas the *Psy1-B1b* allele was approx. 200 bp as the result of a 17 bp deletion in the second intron (Pozniak et al. 2007; Zhang and Dubcovsky 2008). For *Psy1-A1,* 10 μL aliquots of Psy1F5/R5 PCR product were subjected to restriction digestion with 1.5 U of *Tse*I (New England BioLabs Inc.) at 65°C for 1 hr. Following digestion, fragments were separated on a 1% agarose gel and all accessions scored for the presence of the 81/1053 bp fragments indicative of the *Psy1-A1a* haplotype. Lack of the restriction site (1148 bp) was indicative of *Psy1-A1b*.

The full length hexaploid wheat sequence for *Psy1-A1* has been reported (He et al. 2008) and a third allele, *Psy1-A1c*, has been reported in durum wheat (Zhang and Dubcovsky 2008). Evaluation of the AM population for this allele (*Psy1-A1c*) was performed using primers Psy1_AF1 (ATGTACTGCTACTATGTAGCC) and Psy1_R3 (GACTCCTTTGACGATGTCTTC) (Zhang and Dubcovsky 2008) using the following PCR conditions: 1X PCR Buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 mM of each primer, 1 U *Taq* polymerase (Invitrogen), and 100 ng of genomic DNA. Cycle conditions were: 95 ˚C for 5 min, followed by 35 cycles of 95 ˚C for 30 sec, 56 ˚C for 30 sec 72 ˚C for 1 min, and final extension of 72 ˚C for 10 min. The *Psy1-A1a* allele was characterized by an amplicon 359 bp in length and *Psy1-Alc* by an amplicon of 1047 bp.

Pairwise LD of *Psy1-B1* and *Psy1-A1* marker data was calculated with SSR marker data as the squared allele frequency correlation (r^2, Weir 1996) within the program TASSEL (http://www.maizegenetics.net). The distribution of r^2 for unlinked markers was calculated and the parametric 95th percentile of that distribution was taken as the critical value of r^2 to declare significant (Breseghello and Sorrells 2006a).

4.0 RESULTS

4.1 Phenotypic Data

Yellow pigment (YP) concentration and CIE 1976 b* were assessed on 93 lines of the AM population on a plot basis from trials conducted at Swift Current and Saskatoon, Saskatchewan in 2005 and 2006. Analysis of variance (ANOVA) of YP for individual environments indicated a significant effect due to blocks for the majority of environments (exception of Swift Current 2005; Table 1). For the combined analysis, the variance estimate for Accession (A)*Location (L)*Year (Y) interaction was significantly greater than zero ($p<0.01$) but small (0.062) and similar to the residual variance estimate (0.087) (Table 2). The variance estimate for A*L was larger (0.154), than the A*Y interaction, indicating that the location had a greater effect on expression of YP than the year. The F-tests revealed significant differences between accessions ($P<0.001$) at each location and for the combined analysis (Table 2). Correlations among LS means ranged from 0.88 ($P<0.001$) to 0.96 ($P<0.001$), indicating relative consistency among accessions between environments (Table 3), suggesting that the accession x environment interaction was non-crossover type.

For all environments, the least square (LS) means for each accession are presented in Table 4. Averaged over all environments, YP ranged from 4.09 mg kg^{-1} to 12.13 mg kg^{-1} in the field grown samples (Table 4). The greatest range in YP was 9.55 mg kg^{-1} (Table 4) at Swift Current in 2006. The mean YP was similar among environments, except at Swift Current 2006, where the average YP was higher (8.59 mg kg^{-1}). Heritability estimates for YP were high for individual environments, with estimates ranging from 0.97 at Saskatoon in 2005 to 0.99 at Swift Current and Saskatoon 2006 (Table 4), confirming earlier reports that YP is highly heritable in durum wheat (Elouafi et al. 2001; Clarke et al. 2006).

A relationship between YP expression and geographic origin was evident. All Canadian accessions exhibited intermediate to high pigment concentration with Kyle (Townley-Smith et al. 1987) and AC Melita (Clarke et al. 2005a) having the lowest YP of the Canadian accessions when averaged across all four environments.

Table 1. Variance estimates for random effects and F-values for fixed effects from analysis of variance (ANOVA) of yellow pigment (YP) concentration, CIE 1976 b* and thousand kernel weight (KWT). Data was analyzed for individual environments at Swift Current (SC) and Saskatoon (SK) 2005 and 2006 using PROC MIXED in SAS (Littell et al. 1996).

	YP				CIE 1976 b*				KWT			
	2005		2006		2005		2006		2005		2006	
Effect	**SC**	**SK**	**SC**	**SK**	**SC**	**SK**	**SC**	**SK**	**SC**	**SK**	**SC**	**SK**
					RANDOM EFFECTS VARIANCE ESTIMATES							
Rep	0.086	0.011	0.047	0	0.001	0.033	0.102	0	0	0.021	3.001	0
Block(Rep)	0.095	0.227**	0.057**	0.023*	0.027*	0	0.010	0.042*	1.083*	0.150	6.627**	2.519**
Residual	0.120***	0.112***	0.052***	0.064***	0.058***	0.058***	0.156***	0.084***	2.646***	2.109***	2.615***	2.005***
					FIXED EFFECT F-VALUES							
Accession	50.02***	33.75***	149.53***	84.08***	57.28***	30.46***	21.10***	32.08***	9.16***	10.85***	5.35***	9.00***

*P<0.05, **P<0.01, ***P<0.001

Of the Canadian accessions, DT709 consistently expressed the highest YP across all four environments (Table 4). AC Navigator (Clarke et al. 2000a), Commander (Clarke et al. 2005a), and Strongfield (Clarke et al. 2005b) had higher pigment concentration than AC Morse (Clarke et al. 2005a), Kyle (Townley-Smith et al. 1987), AC Avonlea (Clarke et al. 1998) and AC Napoleon, consistent with previous reports (Clarke et al. 2005a, Clarke et al. 2005b).

Table 2. Variance estimates for random effects and F values for fixed effects from combined analysis of variance (ANOVA) of yellow pigment (YP) concentration, CIE 1976 b* and thousand kernel weight (KWT) using PROC MIXED.

	YP	CIE 1976 b*	KWT
	2005-2006	2005-2006	2005-2006
Random Effects	Combined	Combined	Combined
Year	0.154	0	0
Location	0.339	0.835	6.259
Location*Year	0	0.336	7.097
Rep (location*year)	0.031	0.033	0.708
Block (location*year*rep)	0.074***	0.020***	2.619***
Year*Accession	0.094***	0.020	0.843*
Location*Accession	0.154***	0.058***	1.488***
Location*Year*Accession	0.062***	0.059***	1.506***
Residual	0.087***	0.090***	2.363***
Fixed Effects			
Accession	18.85***	20.44***	4.02***

*P<0.05, **P<0.01, ***P<0.001

Table 3. Pearson correlations based on least square means of yellow pigment (YP) concentration analysis performed on association mapping field trials conducted at Swift Current (SC) and Saskatoon (SK), 2005 and 2006.

	YP		
	2005 SC	2005 SK	2006 SC
2005 SK YP	0.91**		
2006 SC YP	0.96**	0.88**	
2006 SK YP	0.93**	0.91**	0.94**

With the exception of Langdon, a low-pigment concentration line, the U.S.-derived accessions also expressed intermediate to high YP similar to the Canadian breeding material. The high YP of Kofa, a US durum variety, is consistent with previous reports,

and has been used as a parent in two independent mapping populations to identify QTL for YP (Pozniak et al. 2007; Zhang and Dubcovsky 2008). The majority of Australian accessions evaluated in this study had statistically higher YP than Commander and AC Navigator, the two Canadian varieties valued for their high YP (Clarke et al. 2005a; Clarke et al. 2000a).

Table 4. Least square means, heritability estimates, and summary of yellow pigment concentrations (mg kg^{-1}) for the association mapping population evaluated at Swift Current (SC) and Saskatoon (SK) in 2005 and 2006.

		2005		2006		2005-2006
Accession	Origin	SC	SK	SC	SK	Combined
Bonaerance Inta Cumenay	Argentina	5.07	6.51	5.46	5.61	5.68
Bonaerance Quilaco	Argentina	5.25	5.64	5.53	5.07	5.37
Bonaerance Valverde	Argentina	6.37	6.12	6.51	5.63	6.17
Buck Ambar	Argentina	5.92	5.33	6.66	5.58	5.92
Buck Topacio	Argentina	11.35	8.63	11.91	10.37	10.57
920334	Australia	11.83	11.03	12.79	11.96	11.90
940030	Australia	11.31	10.10	12.85	11.82	11.46
940435	Australia	12.54	10.72	13.56	11.58	12.13
950329	Australia	9.86	9.12	10.54	9.84	9.86
950844	Australia	10.29	9.41	10.90	10.48	10.27
940955	Australia	11.77	10.55	13.34	11.75	11.86
Tamaroi	Australia	8.17	8.30	9.28	8.04	8.42
Wollaroi	Australia	9.85	8.99	11.30	11.11	10.33
AC Morse	Canada	8.63	7.10	9.89	7.89	8.33
AC Napoleon	Canada	9.41	7.14	10.90	8.40	8.98
9661-AF1D	Canada	8.91	7.33	9.53	8.00	8.41
9661-CA5E	Canada	8.77	7.20	9.63	7.93	8.39
AC Avonlea	Canada	9.12	7.83	9.84	7.88	8.66
AC Melita	Canada	8.46	6.90	9.43	7.11	7.97
AC Navigator	Canada	9.49	8.89	10.34	9.29	9.48
AC Pathfinder	Canada	8.49	7.30	9.63	7.79	8.29
DT691	Canada	9.10	7.80	10.20	8.07	8.76
DT695	Canada	8.32	6.97	9.23	7.94	8.08
DT696	Canada	8.98	8.29	9.53	8.27	8.72
Kyle	Canada	7.90	6.59	8.52	6.71	7.44
Commander	Canada	9.73	8.63	10.30	9.49	9.54
DT704	Canada	8.66	7.89	9.44	8.17	8.54
DT705	Canada	9.12	8.40	10.18	8.44	9.04
DT707	Canada	8.96	8.19	9.73	8.58	8.85
DT709	Canada	10.10	10.03	11.93	10.35	10.61
DT710	Canada	9.23	8.31	9.84	8.41	8.98
DT711	Canada	9.24	7.86	9.60	8.61	8.85
Strongfield	Canada	9.42	8.65	10.37	9.01	9.35
D24-1773	Canada	9.01	7.93	10.35	8.73	9.06
DT513	Canada	8.10	7.48	9.35	8.12	8.27

		2005		2006		2005-2006
Accession	Origin	SC	SK	SC	SK	Combined
DT536	Canada	8.20	7.02	9.00	8.70	8.28
DT540	Canada	10.34	8.59	11.08	9.57	9.85
Carioca	France	6.12	5.73	6.86	6.48	6.32
RABD 93.40	France	9.70	8.04	10.57	9.02	9.29
Tetradur	France	9.61	8.53	8.43	8.25	8.72
Durabon	Germany	8.79	8.74	9.00	8.42	8.75
Durafit	Germany	6.29	6.28	7.18	6.20	6.52
44616	Iran	6.23	7.97	6.78	6.63	6.89
44721	Iran	6.90	5.59	6.01	5.46	6.00
CRDW17	Iran	8.15	8.82	8.79	8.46	8.54
D-73-15	Iran	5.36	5.41	5.94	5.15	5.48
Simeto	Italy	7.01	6.32	8.46	8.19	7.47
Colosseo	Italy	5.48	6.11	5.75	5.32	5.66
Duilio	Italy	6.86	6.36	8.09	7.42	7.17
Grazia	Italy	6.74	6.23	6.66	6.55	6.54
Fortore	Italy	-[a]	7.73	7.35	7.55	7.49
Lesina	Italy	6.20	5.61	7.45	6.55	6.48
Varano	Italy	5.60	5.04	6.46	6.28	5.88
Bronte	Italy	5.21	5.10	5.66	5.58	5.37
Ciccio	Italy	6.88	7.12	7.35	7.26	7.17
Demetra	Italy	4.63	4.86	5.57	4.96	5.00
Gianna	Italy	4.43	4.30	5.04	4.70	4.59
Iride	Italy	6.31	6.35	7.29	7.15	6.78
Medda	Italy	5.50	6.08	5.85	5.52	5.75
Mongibello	Italy	7.58	7.25	8.75	7.72	7.83
Parsifal	Italy	5.84	5.41	6.35	5.55	5.78
Svevo	Italy	8.89	8.25	9.70	9.06	9.00
Tresor	Italy	7.08	7.01	7.01	6.87	7.03
Green 27	Mexico	7.90	6.50	6.16	6.11	6.67
Green 34	Mexico	7.24	6.44	7.26	7.38	7.08
Nacori 97	Mexico	7.78	6.47	8.03	6.76	7.23
DHTON 1	Morrocco	8.43	7.13	8.28	8.15	7.99
Gidara 17a	Morrocco	4.42	5.54	4.81	4.97	4.93
Marjak	Morrocco	4.15	4.31	4.01	3.84	4.09
Arrivato	New Zealand	5.69	5.94	6.39	5.53	5.91
CFR5001	New Zealand	7.55	7.43	7.86	7.64	7.62
K-39099	Russia	6.66	5.01	7.79	5.41	6.20
Agridur	Spain	7.47	6.15	7.73	7.45	7.17
Altar-Aos	Spain	8.14	7.46	8.97	8.02	8.19
Arcobelano	Spain	7.95	6.43	8.42	6.93	7.41
Ariesol	Spain	7.62	6.48	7.65	6.90	7.15
Borli	Spain	7.25	6.79	7.14	7.00	7.03
Camacho	Spain	5.40	5.18	5.44	4.97	5.22
Gallareta	Spain	8.15	7.91	8.54	7.40	7.96
Mexa	Spain	7.21	6.08	8.18	7.57	7.26
Vitron	Spain	6.23	5.88	6.66	6.64	6.41
Ocotillo	U.S.	7.68	6.59	7.62	6.26	7.01
D940027	U.S.	7.44	6.62	8.86	7.34	7.58

		2005		2006		2005-2006
Accession	Origin	SC	SK	SC	SK	Combined
D940098	U.S.	7.45	6.79	9.27	7.85	7.87
D941038	U.S.	9.20	7.50	10.75	8.52	9.01
D95580	U.S.	8.29	7.10	9.25	7.73	8.08
Plaza	U.S.	9.53	7.11	10.33	8.46	8.85
Durex	U.S.	8.99	7.31	9.34	8.49	8.55
Langdon	U.S.	5.52	5.15	6.41	5.35	5.61
Westbred881.	U.S.	8.55	7.61	9.34	8.60	8.52
Kofa	U.S.	9.71	8.01	10.90	9.51	9.53
Kronos	U.S.	9.74	7.90	10.71	9.86	9.52
Ave		7.93	7.19	8.59	7.66	7.84
Min		4.15	4.30	4.01	3.84	4.09
Max		12.54	11.03	13.56	11.96	12.13
$LSD_{0.05}$		0.71	0.74	0.50	0.53	0.98
Heritability		0.98	0.97	0.99	0.99	0.95

[a]sample was lost due to a seeding error

Tamaroi and 950329 had the lowest YP of Australian accessions, and were not statistically different from Strongfield (Table 4). In general, South American, Mexican and European accessions had low-intermediate pigment concentration, with the exception of Buck Topacio (Argentina), RABD 93.40 (France), and Svevo (Italy) which showed statistically similar ($P<0.05$) YP to Australian (Buck Topacio) and the high pigment Canadian accessions Commander and Strongfield (RABD 93.40 and Svevo). Tetradur (France) and Durabon (Germany) displayed YP similar to AC Avonlea, but both lines have North American parentage, tracing back to the founder variety "Edmore". Accessions from Argentina had low YP, with the exception of Buck Topacio, which had nearly two times higher YP than other lines originating from Argentina. DHTON 1 had significantly higher YP then either Gidara 17a or Marjak, all lines originating in Morrocco, with DHTON 1 having YP similar to Kyle and AC Melita. Two Iranian breeding lines (44721 and D-73-15) and the Russian cultivar K-39099 had significantly less YP than Kyle and AC Melita.

Meal CIE 1976 b*, a measure of ground meal "yellowness" was also assessed as a measure of endosperm colour. For each environment, the ANOVA F-test indicated significant differences ($P<0.001$) among accessions for CIE 1976 b* (Table 1). The combined ANOVA across environments revealed that the variance estimates for the random factors Block(L*Y*R), A*L and A*L*Y (0.020, 0.058 and 0.059 respectively),

were all lower than the variance estimate for the residual (0.090) (Table 2). Combined analysis of CIE 1976 b* revealed significant differences among accessions (F=20.44; p<0.01, Table 2) and a heritability of 0.95 (Table 5), similar to the heritability for YP (Table 4). The range in CIE 1976 b* values was similar across all environments, with the exception of SC 2005, which had a greater range in CIE1976 b* expression (range = 6.7; Table 5). Averaged over all environments, the range in CIE 1976 b* was 5.8.

Table 5. Summary of least square means of CIE 1976 b* analysis for endosperm colour of the AM population grown at Swift Current (SC) and Saskatoon (SK), 2005 and 2006. CIE 1976 b* was measured on a plot basis on a dry sample of ground durum whole meal.

	CIE 1976 b*				
	2005		2006		2005-2006
	SC	SK	SC	SK	Combined
Ave	18.9	16.9	17.8	17.0	17.6
Min	15.5	14.6	14.0	13.8	14.6
Max	22.2	19.4	20.7	20.0	20.4
$LSD_{0.05}$	0.51	0.47	0.80	0.61	0.71
Heritability	0.98	0.97	0.95	0.97	0.95

Whole meal CIE 1976 b* values were highly correlated with YP measurements at all environments, with correlation coefficients ranging from 0.94 in Saskatoon 2005 (P<0.0001) to 0.97 in Swift Current 2005 and 2006 (P<0.0001) (Table 6), suggesting that both methods provided similar estimates of yellow pigment concentration. Given the high correlation between CIE 1976 b* and YP, the remainder of this thesis will focus on presentation, discussion and association mapping of only the YP results. The LS means for CIE 1976 b* of accessions at each environment are presented in Appendix 3.

Table 6. Pearson correlation coefficients of least square means of data analysis of phenotypic traits from field trials conducted at Swift Current and Saskatoon, 2005 and 2006 between yellow pigment (YP) concentration and CIE 1976 b* and thousand kernel weight (KWT).

	YP				
	2005		2006		2005-2006
	SC	SK	SC	SK	Combined
CIE 1976 b*	0.97**	0.94**	0.97**	0.95**	0.98**
KWT	-0.08ns	-0.11ns	-0.23*	-0.15ns	-0.11ns

*P<0.05, **P<0.0001; ns non-significant

In several mapping populations, variation in YP concentration has been associated with differences in seed size, likely due to dilution of YP by increased starch content in the seed (Hessler et al. 2002; Clarke et al. 2006). To assesses if this was the case in the AM population, seed size (as assessed using 1000-kernel weight) was assessed at all four environments. Variance estimates for kernel weight indicated significant variation among field blocks (2.619) and to a lesser extent A*L (1.488) and A*L*Y (1.506) (Table 2). Averaged over all environments, significant differences ($P<0.001$) in 1000-kernel weight were evident among accessions. The least square means for kernel weight at each environment are presented in Appendix 4. The largest range in kernel weight occurred in Saskatoon 2005 (32.6 mg kg^{-1} to 51.2 mg kg^{-1}; Table 7). Kernel weight was highly heritable at individual environments, ranging from 0.83 at Swift Current 2006 to 0.91 in Saskatoon 2006 (Table 7), and was moderately to highly heritable (0.76) when all environments were considered, similar to previous heritability estimates for seed size (Sharma and Knott 1964; Belay et al. 1993). Correlations between kernel weight and YP were non-significant, with the exception of Swift Current 2006, where these traits were weakly negatively correlated ($r= -0.23$; $P<0.05$, Table 6). These results suggest that variation in YP observed in the AM population was due to genetic differences in expression of YP, and not due to the pleiotropic effects of seed size. Several Argentinian lines demonstrated consistently high thousand kernel weight, and a number of Spanish lines had low thousand kernel weight, however no distinct relationship between 1000-kernel weight and geographic origin was apparent.

Table 7. Summary of least square means for 1000-kernel weight (g) of the AM population grown at Swift Current and Saskatoon, 2005 and 2006.

	2005		2006		2005-2006
	SC	SK	SC	SK	Combined
Ave	41.5	42.0	37.7	46.0	41.8
Min	30.4	32.6	30.8	37.8	33.7
Max	48.2	51.2	44.7	55.0	48.1
Ave LSD0.05	3.5	2.9	3.5	3.1	3.8
Heritability	0.90	0.91	0.83	0.90	0.76

4.2 Microsatellite Analysis

A total of 245 microsatellite markers were scored on the AM population, amplifying 251 loci (Table 8). Based on the hexaploid wheat consensus map (Somers et al. 2004), the average distance between adjoining markers was 6.9 cM. On average, 18 loci were scored for each chromosome with a minimum of 9 loci on chromosome 6A and a maximum of 23 loci on chromosomes 1A, 1B, 5A and 7A (Table 8). The number of alleles detected ranged from 1 to 15 alleles. Fifteen markers amplified only a single allele. Across all marker loci, 1184 alleles were amplified, with a mean of 5 alleles per locus. The average polymorphism information content (PIC) values ranged from 0.28 on chromosome 2B to 0.50 on chromosome 6B, with an average of 0.40 (Table 8). The A genome, contained a greater number of alleles than did the B genome, however more loci were amplified in the A genome than the B genome. The average PIC values and the average number of alleles per locus were similar for the A and B genomes (approximately 5 alleles per locus).

Rare alleles can inflate linkage disequilibrium estimates (Remington et al. 2001; Flint-Garcia et al. 2003; Mather et al. 2007; Somers et al. 2007) and could result in identification of spurious associations when identifying marker-trait associations (Thornsberry et al. 2001). Of the microsatellite markers analyzed,193 loci produced rare alleles, defined as having a frequency of less than <5%, (Tenaillon et al. 2001; Lu et al. 2005; Breseghello and Sorrells 2006a; Chao et al. 2007; Hai et al. 2007; Rhoné et al. 2007; Tommasini et al. 2007). The number of rare alleles amplified ranged from 7 on chromosomes 6A to 19 on chromosome 7A (Table 8). The number of loci amplifying rare alleles was similar for both genomes (A=97; B=96).

4.3 Genetic Diversity and Population Structure

Two approaches were used to quantify population structure in the AM population, genetic distance- and model-based methods. Pair-wise genetic-distances were estimated on Rogers' genetic coefficient of similarity (Rogers 1972). Roger's coefficient of similarity was selected over other similarity coefficients as it is suitable to evaluate genetic similarity among germplasm derived from breeding programs where selection has been practiced (Reif et al. 2005). A dendrogram was constructed and

revealed three major groups of accessions consistent with their geographic origin (Figure 3). The first cluster consisted primarily of North American derived accessions. Within the North American accessions, two clusters were evident (Figure 3), and were generally consistent with known pedigrees. The first cluster consisted of related accessions resulting from common breeding material used in Canadian cultivar development. This includes US desert durum varieties Kofa, Durex and Westbred881. AC Pathfinder (Clarke et al. 2000b) and DT711, both Westbred881 derivatives, grouped together. The close relationship between Commander and one of its parents AC Pathfinder (Clarke et al. 2005a) is evident (Figure 3). The sister lines, DT704, DT705, DT707, and Strongfield are grouped with their common parent AC Avonlea (Clarke et al. 1998), and progenitor Kyle (Townley-Smith et al. 1987), along with sister lines 9661-AF1D and 9661-CA5E (Clarke et al. 2003). The second North American cluster included accessions from Canada and the North Dakota State University (NDSU) breeding programs.

Table 8. Summary of SSR data. Number of loci per chromosome, number of alleles amplified, number of loci containing rare alleles, mean number of alleles amplified per chromosome, and PIC values.

Chromosome	No. of Loci	No. of Allele	No. of Loci with Rare Alleles	Mean Allele No. (range)	Mean PIC value (range)
1A	23	102	15	4 (1-12)	0.37 (0 - 0.76)
1B	23	116	18	5 (1-11)	0.45 (0 - 0.72)
2A	18	101	16	6 (1-12)	0.47 (0 - 0.77)
2B	18	74	16	4 (1-10)	0.28 (0- 0.68)
3A	14	64	10	5 (1-9)	0.38 (0 - 0.62)
3B	21	98	18	5 (1-13)	0.40 (0 - 0.72)
4A	17	72	14	4 (2-7)	0.38 (0.02 - 0.67)
4B	14	63	8	5 (1-10)	0.42 (0 - 0.74)
5A	23	105	16	5 (2-13)	0.40 (0.06 - 0.83)
5B	22	87	14	4 (1-9)	0.35 (0 - 0.73)
6A	9	40	7	4 (2-7)	0.41 (0.02 - 0.57)
6B	10	50	10	5 (3-7)	0.50 (0.22 - 0.76)
7A	23	119	19	5 (2-13)	0.37 (0.02 - 0.64)
7B	16	94	12	6 (1-15)	0.47 (0 - 0.86)
Total	251	1185	193		
Mean	18	85	14	5	0.40

Pedigree analysis indicated that the majority of these lines derived from the cultivar Edmore, including French cultivar Tetradur, Spanish cultivar Agridur, and the German cultivar Durabon (Appendix 1).

The second major cluster grouped accessions from Australian, CIMMYT (Mexico) and European breeding programs (Figure 3). Within this cluster, the majority of Australian accessions grouped together. Italian accessions Duilio, Simeto, Vitron, Gianna and Bronte were genetically similar to CIMMYT sister accessions Green27 and Green34. DT710, a Canadian breeding accession was most similar to Green27, its ancestor (Figure 3; Appendix 1). Duilio, Vitron, and Bronte clustered together and are all derived from the founder Anhinga (Maccaferri et al. 2003). Arcobelano, Gallareta, Iride, Nacori 97, Svevo and Tamaroi all grouped with Altar-Aos. Altar84 is a founder of Arcobelano, Iride and Nacori 97, and is a sister accession of Gallareta (Maccaferri et al. 2007) and all are found to cluster in the same region of the dendrogram. Altar 84 and Tamaroi, an Australian accession, share common ancestors including Ruff, Flamingo, Shearwater and Mexicali-75 (Maccaferri et al. 2007; Appendix 1), accounting for the close relationship of Tamaroi to the Spanish/CIMMYT accessions versus Australian.

A third cluster comprised Italian varieties derived from the founder accessions Valforte and Cappelli (Figure 3). Bonaerance Valverde and Bonaerance Cumenay, both cultivars developed in Argentina, clustered with Lesina and Fortore, suggesting common ancestry. Two accessions, 44616 (from Iran) and K-39099 (from Russia), showed little similarity to other accessions and were separated from the clusters present.

Although cluster analysis can be used to visually assess similarity of accessions, it is difficult to draw statistical inference as to the level of admixture present in accessions within the population (Pritchard et al. 2000a). A more useful approach would be to identify actual subpopulations and to assign individuals to these sub-populations probabilistically. Pritchard et al. (2000a) developed a Bayesian model that attempts to identify genetically distinct subpopulations based on patterns of allele frequencies based on multi-locus data collected from unlinked markers. Bayesian analysis was performed using 28 unlinked markers (Appendix 2) with the number of *a priori* assigned subpopulations (K) ranging from 3 to 7.

Figure 3. A consensus UPGMA dendrogram constructed using Rogers' genetic distance (Rogers 1972) of 93 cultivars of the association mapping population. The bootstrap

values >50% are shown in the tree. Coloring represents the three major clusters identified blue (North American), green (Australian, CIMMYT, European) and pink (Italian). Symbols to the right of accession names indicate model-based subpopulation assignments as inferred by STRUCTURE (Pritchard et al. 2000a) for K=5 subpopulations and are defined as Subpopulation 1 (■), subpopulation 2 (◆), subpopulation 3 (□), subpopulation 4 (◇), and subpopulation 5 (▼). Accessions carrying the *Psy1-B1b* allele are indicated with (1), *Psy1-A1b* allele by (2) and *Psy1-A1c* by (3).

Averaged over three independent runs, the lowest natural log probability was assigned to K=5 (Table 9) and the posterior probability of the data indicated the highest likelihood for K=5 (Table 9). Using K=5, the average membership proportion of accessions in each sub-population (Q-matrix) was estimated (Appendix 5) as the average of three runs for K=5.

Table 9. Estimated Ln likelihood as assigned by *Structure* v.2 (Pritchard et al. 2000a) for *K* subpopulations and posterior probabilities of the data for each K.

	K				
Run	3	4	5	6	7
1	-2698	-2597	-2551	-3547	-2514
2	-2672	-2587	-2564	-2518	-2522
3	-2669	-2592	-2553	-2498	-2703
Ave Ln Likelihood	-2680	-2592	-2556	-2854	-2580
Posterior Probability	1.42×10^{-54}	2.33×10^{-16}	0.98	3.9×10^{-130}	3.78×10^{-11}

Using the Q matrix, the majority of individuals could be discretely assigned to 1 of 5 sub-populations (designated as different symbols in Figure 3) based on membership proportion greater than 50% (Appendix 5). Five accessions, DT691, DT696, Bronte, Tamaroi and Carioca displayed high levels of admixture and all shared near equal membership in two subpopulations (Figure 3; Appendix 5), indicating admixture of alleles from separate sub-populations. Sub-population 4 had the fewest accessions (n=13), with a similar number of accessions (range 19-22) found in the remaining sub-populations (Table 10).

Generally, the five sub-populations were in agreement with the distance-based groupings (Figure 3), but several differences were observed. Using genetic distance estimates, Bonaerance Quilaco grouped with the North American accessions, but model-based analysis grouped this accession with accessions from Australia and CIMMYT. Demetra and Marjak also grouped with Australian and CIMMYT material. RABD 93.40 was associated with the Canadian-derived material using genetic distance, but model-based analysis grouped RABD 93.40 with European accessions. Among these accessions, pedigree information is lacking with the exception of Demetra, whose pedigree indicates common parents with accessions present in the distance-based clustering. Pedigree information available indicates a large degree of shared germplasm common among European accessions, and similarly among North American accessions, with some cross-over of germplasm between groups which may account for potential differences in sub-population assignment.

Table 10. Summary of average yellow pigment concentration (mg kg-1) and range for each environment, Swift Current (SC) and Saskatoon (SK), according to sub-population as assigned by STRUCTURE using K=5 sub-populations. Accessions were classified into one of five sub-populations (as indicated by symbols corresponding to accessions in Figure 3). The intraclass correlation coefficient (ICC) was determined for each environment as the proportion of phenotypic variation explained by sub-populations.

	2005		2006		2005-2006
Sub-population	SC	SK	SC	SK	Combined
Sub-pop 1 (■)	9.04±0.34	7.94±0.29	9.87±0.39	8.49±0.36	8.83±0.33
(n=22)	(7.68-10.34)	(6.59-10.03)	(7.62-11.93)	(6.36-10.34)	(7.05-10.61)
Sub-pop 2 (◆)	8.34±0.36	7.14±0.30	9.20±0.40	7.89±0.38	8.15±0.35
(n=20)	(5.52-11.35)	(5.01-8.99)	(6.41-11.91)	(5.36-11.10)	(5.62-10.57)
Sub-pop 3 (□)	8.10±0.37	7.49±0.31	8.55±0.41	7.87±0.39	8.00±0.36
(n=19)	(4.15-12.54)	(4.31-11.03)	(4.01-13.56)	(3.84-11.96)	(4.10-12.18)
Sub-pop 4 (◇)	6.93±0.44	6.51±0.37	7.64±0.50	7.03±0.47	7.03±0.43
(n=13)	(4.43-8.17)	(4.30-8.30)	(5.04-9.28)	(4.75-8.15)	(4.60-8.40)
Sub-pop 5 (▼)	6.65±0.38	6.55±0.31	7.16±0.41	6.80±0.38	6.80±0.36
(n=19)	(5.07-9.74)	(5.04-8.82)	(5.44-10.71)	(4.99-12.04)	(5.22-10.10)
ICC[a]	0.25	0.13	0.26	0.11	0.20

[a] Intraclass correlation coefficient $ICC = \sigma^2_p / (\sigma^2_p + \sigma^2)$ (Neter et al. 1996).

The greatest range in YP occurred in subpopulation 3 (Table 10), which contained both the highest (940435) and lowest (Marjak) pigment concentration accessions (Figure 3).

Among the 5 subpopulations, subpopulation 1 (containing the majority of Canadian breeding material and cultivars) consistently displayed high yellow pigment concentration, while subpopulation 5 displayed low yellow pigment concentration. However dependent upon environment YP concentration of one subpopulation was not always significantly different from other subpopulations, for example subpopulation 1 was not found to be significantly different from subpopulations 2 and 3 at all environments (Table 10). Variation for YP concentration in subpopulation 1 however was significantly different from subpopulations 4 and 5 regardless of environment. The intraclass correlation coefficient (Neter et al. 1996) indicated that population structure accounted for approximately 20% of phenotypic variation in YP, ranging from 11% at Saskatoon 2006 to 26% at Swift Current 2006 (Table 10).

4.4 Phytoene Synthase Gene

Psy1-B1 has been suggested as a candidate gene for the YP QTL on chromosome 7B in durum wheat (Pozniak et al. 2007). *Psy1-A1* has also been determined to be associated with a QTL for YP on 7A in hexaploid wheat (He et al. 2008) and in durum wheat (C. Pozniak, unpublished results). To evaluate the potential of AM to detect associations with allelic variation at *Psy1-A1* and *Psy1-B1* and variation for YP, a portion of the two *Psy1* genes was sequenced. Amplification using primers Psy1F5 and Psy1R5 produced a single amplicon of approximately 1125 bp (Figure 4A). These fragments were subjected to restriction digestion with *HpaII* (5'-C|CGG-3') as this enzyme targets a known SNP that differentiates sequences from the A (*Psy1-A1*) and B (*Psy1-B1*) genomes (Pozniak et al. 2007). Digestion with *HpaII* resulted in two fragments (857 and 759 bp) from all lines evaluated, including 'Langdon' (Figure 4B). The 759 bp fragment was absent in Langdon7D(7A) confirming that the 759 bp fragment was derived from chromosome 7A (Figure 4B). Likewise, the 857 bp fragment was absent in Langdon7D(7B), indicating that this fragment was from 7B (Figure 4B). Thus the Psy1F5 and Psy1R5 primers amplified partial sequences of *Psy1* from both the A and B genomes, and it was possible to differentiate these clones using *HpaII* restriction mapping prior to sequencing.

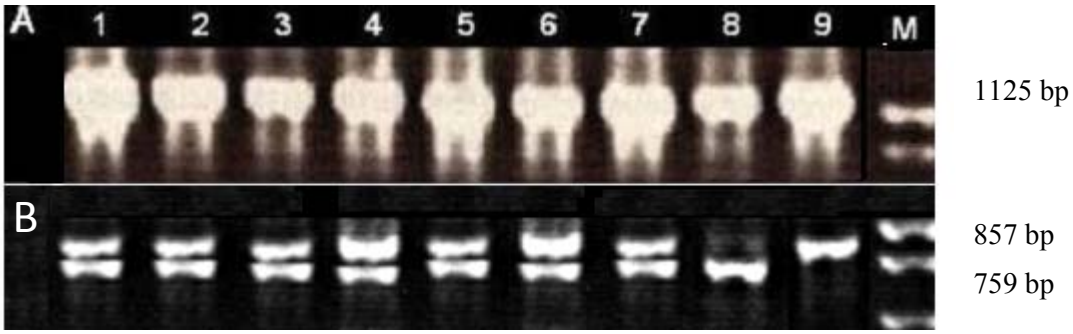

Figure 4. PCR amplification and restriction analysis with *HpaII* of *Psy1* from durum wheat accessions from the association mapping population. (A) PCR amplification of approximately 1125 bp fragment from genomic DNA using primers Psy1F5 and Psy1R5. Accessions are AC Morse (lane 1), AC Napoleon (lane 2), AC Avonlea (lane 3), AC Melita (lane 4), AC Navigator (lane5), AC Pathfinder (lane 6), Langdon (lane 7), Langdon7D(7B) (lane 8), and Langdon7D(7A) (lane 9). (B) Restriction digestion of the 1125 bp fragment with *HpaII* resulted in two fragments 857 and 759 bp in length.

Alignment of sequence results for *Psy1-B1* sequences with reported *Psy1-B1a* and *Psy1-B1b* revealed that only these two alleles were present in the 39 lines sequenced from the AM population. The two alleles could be easily differentiated based on a 17 bp insertion/deletion (Indel) site (Figure 5). Subsequent analysis of the remaining lines of the AM population with a co-dominant marker designed to detect the Indel (Zhang and

```
Psy1-B1a    CTAGTTCTGGATGCGAATAATGGCAAATAGAAACATCGTGGAACTTGCATGCTATACATT
Psy1-B1b    CTAGTTCTGGATGCGAATAATGGCAAATAGAAACATCGTGGAACTTGCATGCTATATATT

Psy1-B1a    TATATAGATACTCCTATATAGTATAGTCAGTGAAGAATAAAGGCCTCACATAACACTTTT
Psy1-B1b    TATATAGATACTCCTATATAGTATAGTCAGTGAAGAATAAAGGCCTCACATAACACTTTT

Psy1-B1a    TTTATATGCCATTATGTGTGGAAGAATCAAATTAGGCTTTTTGTTGGCTAAATGGCCTCA
Psy1-B1b    TTTATATGCCATTATGTGTGGAAGAATCAAATTAGGCTTTTTGTTGGCTAAATGGCCTCA

Psy1-B1a    ATAGGATCAAAGTACATGAGAAAAGTTTGCAAGAACATATTCCTCACTACTTAAGGAATG
Psy1-B1b    ATAGGATCAAAGTACATGAGAA----------------TTCCTCACTACTTAAGGAATG

Psy1-B1a    TGAACCTGAGGTTCTGTCAGTTCTAAATGAGATATACTCTAGGCATCGAACACTTTCAGA
Psy1-B1b    TGAACCTGAGGTTCTGTCAGTTCTAAATGAGATATACTCTAGGCATCGAACACTTTCAGA
```

Figure 5. Partial sequence alignment of two *Psy1-B1* alleles identified from 39 lines sequenced from the association mapping population. *Psy1-B1a* (Genbank accession no. DQ642440) and *Psy1-B1b* (Genbank accession no. DQ642439) have both been reported previously (Pozniak et al. 2007; Zhang and Dubcovsky 2008). Primers used to detect the insertion/deletion differentiating the two alleles (indicated by a box) are underlined.

Dubcovsky 2008; Figure 6) identified 17 of 93 lines in the AM population carrying the *Psy1-B1b* allele (Figure 3; Appendix 1). The *Psy1-B1b* allele was predominant in US desert durum cultivars including Kofa, Kronos, Durex and Westbred 881 (Figure 3; Table 11; Appendix 1). Westbred 881 is in the pedigrees of Canadian accessions Commander, AC Navigator, DT709 and DT711 and these lines also carried *Psy1-B1b* (Figure 3; Appendix 1).

Sequence alignment of *Psy1-A1* clones revealed two *Psy1-A1* haplotypes, designated *Psy1-A1a* and *Psy1-A1b* (Figure 7) which could be differentiated based on numerous SNPs and Indels (Figure 7). The *Psy1-A1b* allele, identified in this study, has yet to be reported in the literature. The *Psy1-A1a* allele contains a unique *Tse*I restriction digest site (Figure 7) corresponding to band sizes of 1043/82 bp (Figure 8). Scoring the *TseI* cleaved amplified polymorphic sequence (CAPS) marker on the AM population identified nine lines in which the *Tse*I restriction site was absent, corresponding to the *Psy1-A1b* allele (Figure 3; Appendix 1).

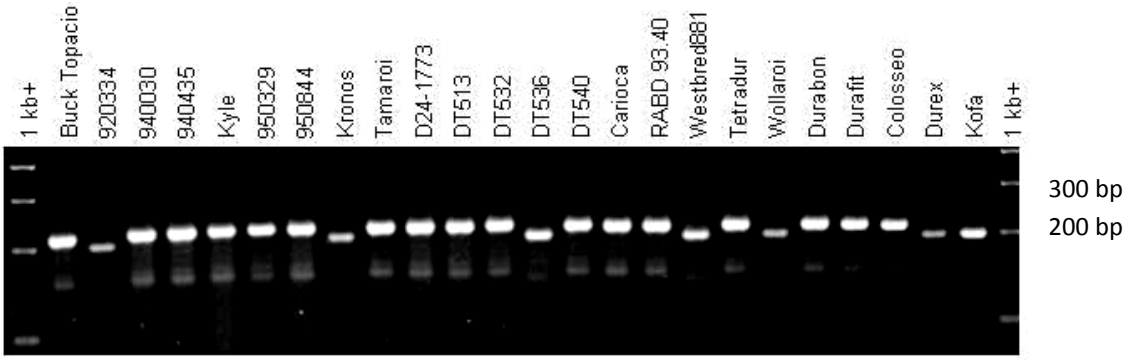

Figure 6. A co-dominant marker differentiating the two reported alleles of *Psy1-B1* (Based on a 17 bp insertion deletion (see Figure 5). The 217 bp fragment was designated as *Psy1-B1a* and the 200 bp fragment as *Psy1-B1b*. 920334 (lane 2), Kronos (lane 8), DT536 (lane 13), Westbred881 (lane 17), Wollaroi (lane 19), Durex (Lane 23), and Kofa (lane 24) all carry *Psy1-B1b*. Lines possessing each of the *Psy1-B1* alleles are summarized in Figure 3.

During the writing of this thesis, the full length hexaploid wheat sequence for *Psy1-A1* was reported (He et al. 2008). Alignment of the hexaploid wheat *Psy1-A1* sequence with the near full length sequence of *Psy1-A1a* reported by Pozniak et al. (2007) indicated a 688 bp discrepancy (Figure 9) at the 3' end of the sequence, outside

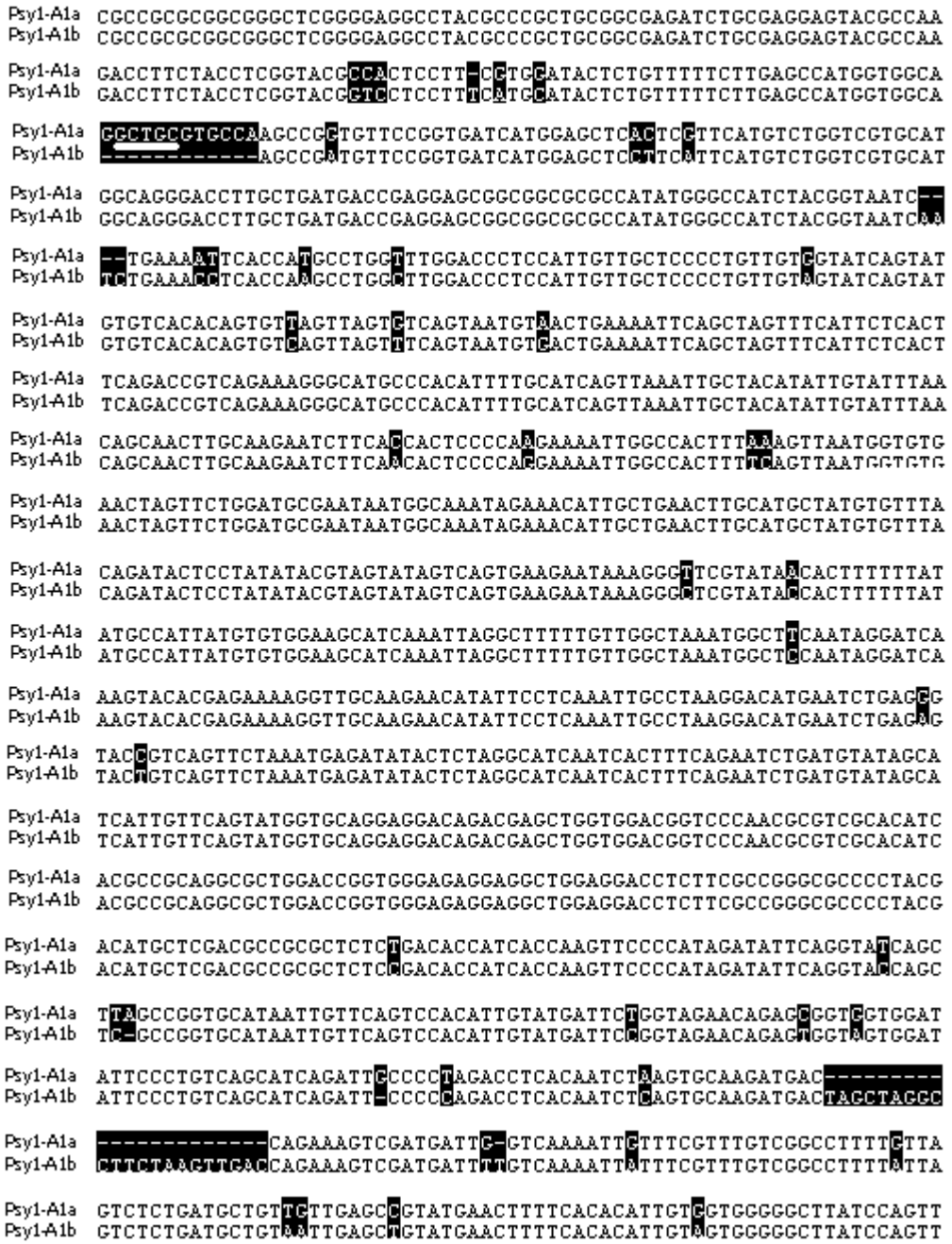

Figure 7. Partial sequence alignment of two *Psy1-A1* alleles identified. *Psy1-A1a* has been reported previously (Pozniak et al. 2007; Genbank accession no. DQ642443) and *Psy1-A1b* was identified in this study. The underlined sequence in white represents a unique *TseI* digestion site which was used to develop a CAPS marker to differentiate the two alleles (see Figure 8).

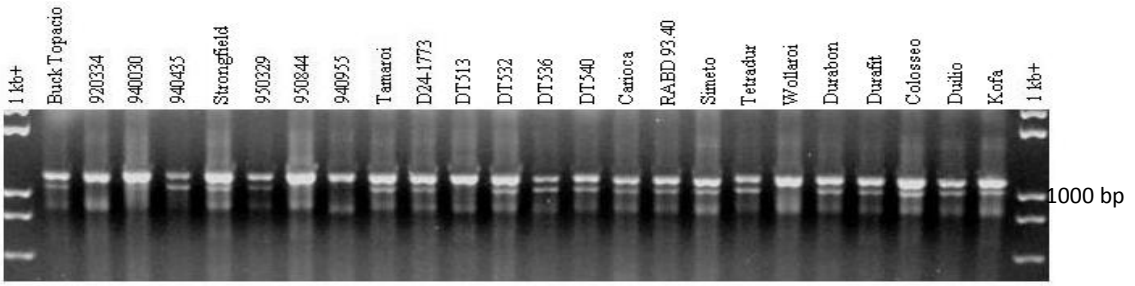

Figure 8. A CAPS marker differentiating two alleles of *Psy1-A1*. The lack of the *Tse1* restriction site (singlet) was indicative of lines carrying *Psy1-A1b* (see Figure 7), whereas the presence of the doublet was indicative of *Psy1-A1a*.

```
T.aestivum Psyl-Ala      CATGTACTGCTACTATGTTGCCGGCACCGTGGGGTTGATGAGCGTCCCGGTGATGGGCAT
T.turgidum Psyl-Ala      CATGTACTGCTACTATGTTGCCGGCACCGTGGGGTTGATGAGCGTCCCGGTGATGGGCAT

T.aestivum Psyl-Ala      TGCCCCCGACTCCAAGGCGACAGCTGAGACCGTCTATGGCGCCGCTCTGGCTCTCGGGCT
T.turgidum Psyl-Ala      TGCMCCCGACTCCAAGGCGACAGCTGAGACCGTCTATGGCGCCGCTCTGGCTCTCGGGCT

T.aestivum Psyl-Ala      CGCGAACCAGCTCACCAACATACTCAGGGATGTCGGAGAAGAGTAAGCCACTCACTCACT
T.turgidum Psyl-Ala      CGCGAACCAGCTCACCAACATACTCAGGGATGTCGGAGAAGAGTAAGCCACTCACTCACT

T.aestivum Psyl-Ala      ACCAATACAATGCAATAGTTTTCCCTTGTAAAATCATTTTTTTTAGAAAAGGAGCATGAC
T.turgidum Psyl-Ala      ACCAATACAATGCAATGGTTTTCCCTTGTAAAATCAT----------------------

T.aestivum Psyl-Ala      CCCCGGCCTCTGCATCTGAGAGATGCATACGGCCACTTTATTGATTATTCTCAGGACCTT
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      ACAAAGTATTACAACAATGAGCCTGAATCCACCATCTTGACAACACATGCCGCTACTCCT
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      ATCCAAAATGATGAAGGGGTGCTAGCTGGGCCACTACCCAAACCACTCACCAAAGCCTAA
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      CATCAAAAGCCGGAAACCGAAACATATTCGGAAGCCCCAGCCGAGCCACATACCGGGTCT
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      GGGCACAATCCGGTCAGACGCACTCGTGTGTCGTCGCCGCCATCTTCCACAGGTCCGTCT
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      TCAGATCATATTGAGGCTTCTACCTTGTCTGGCCACTCTACCATCGACGTCACCATGACG
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      CCAAACAGCAACCTCCTCCTGCGCGAGTCCATCTCCGTGCATCGGGCGGCGAGCCTCCGC
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      AGCGCCATGCCGCCGATCTTCGCCGCCATCAATGAGTGAGATGAAGTACCGCTCCACCAC
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      GGCATGTACAAGGTGACGAAGGGCGAGGTCCCCATCGGAGACACGGGCGGAAGAGAAGCA
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      CCGCAGCCCCGAGACACTGCCCGGAGTTGCGACGCAGTAGATCAGGCGGGCCGTCACCAG
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      GAACCAGACAAGCACGCCATGCACCCCAGCATCCCCATCCCCATGCCCATCCCTTGTAAA
T.turgidum Psyl-Ala      ------------------------------------------------------------

T.aestivum Psyl-Ala      ATCATGATATGGACATTTCGAGATAGCACTATCCTGAATTTTTTGGTGTTTCGAATAATT
T.turgidum Psyl-Ala      -----GATATGGACATTTTGAGATAGCACTATCCTGAATTTTTGGTGTTTCGAATAATT

T.aestivum Psyl-Ala      TCCCAGTGCAAGAAGAGGAAGGATATATTTGCCGCAAGACGAGCTTGCGGAGGCAGGGCT
T.turgidum Psyl-Ala      TCCCAGTGCAAGAAGAGGAAGGATATATTTGCCGCAAGACGAGCTTGCGGAGGCAGGGCT

T.aestivum Psyl-Ala      CTCCGATGAAGACATCTTCAAAGGAGTCGTCACCGACAAGTGGAGAAAATTCATGAAGAG
T.turgidum Psyl-Ala      CTCCGATGAAGACATCTTCAAAGGAGTCGTCACCGACAAGTGGAGAAAATTCATGAAGAG
```

Figure 9. Partial sequence alignment of *Psy1-A1a* sequence from durum wheat (Genbank accession no. DQ642443) and hexaploid wheat (EF600063). Sequence alignment indicates an Indel present in the fourth intron corresponding to the *Psy1-A1c* allele.

of the region sequenced in this study. Recently, Zhang and Dubcovsky (2008) developed a molecular marker for this Indel (Figure 9), and found that 12% of the 48 durum wheat accessions characterized in their study carried the hexaploid wheat allele (*Psy1-A1c*). The *Psy1-A1c* molecular marker was evaluated in the AM population (Figure 10) and 7 lines were found to carry *Psy1-A1c* (Figure 3; Table 11). Vitron carries the *Psy1-A1c* allele, confirming earlier results (Zhang and Dubcovsky 2008) and the majority of accessions carrying the *Psy1-Alc* allele fell within either the third or fourth subpopulations (Figure 3). Buck Ambar, an accession originating in Argentina carries the *Psy1-A1c* allele and did not cluster with the other Argentinian accessions, but did cluster with accessions carrying the same allele (Figure 3). Marjak, identified as the accession with the lowest yellow pigment concentration, contained the *Psy1-A1c* allele (Table 4; Figure 3; Appendix 1).

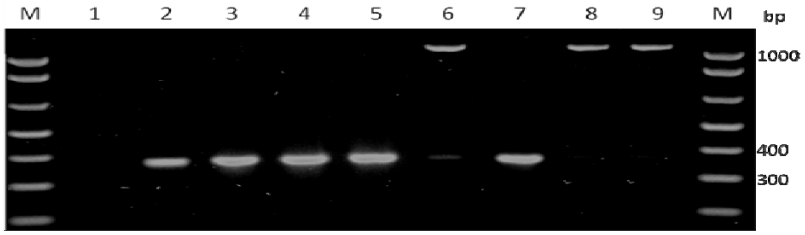

Figure 10. Evaluation of a co-dominant marker differentiating *Psy1-A1c* and *Psy1-A1a* alleles of on a sampling of 10 lines from the association mapping population. The 1050 bp fragment is indicative of the *Psy1-A1c* allele, and the 360 bp fragment indicative of *Psy1-A1a*. Commander (lane 1) carries the *Psy1-A1b* allele (Figure 3) which was not amplified with this marker. Accessions are Strongfield (lane 2), DT540 (lane 3), Demetra (lane 4), Kofa (lane 5), Vitron (lane 6), Mongibello (lane 7), D-73-15 (lane 8), and Marjak (lane 9).

4.5 Marker-trait Associations

In total, 84 loci were identified as significantly ($p \leq 0.05$) associated with YP, with 56 loci significant across all four environments (Figure 11; Appendix 6). On chromosome 1A, six markers were significantly associated with YP in all four test environments (Figure 11). Three SSR markers on 1B were identified and are within a region previously identified to be associated with YP (Cervigni et al. 2005). On 2A,

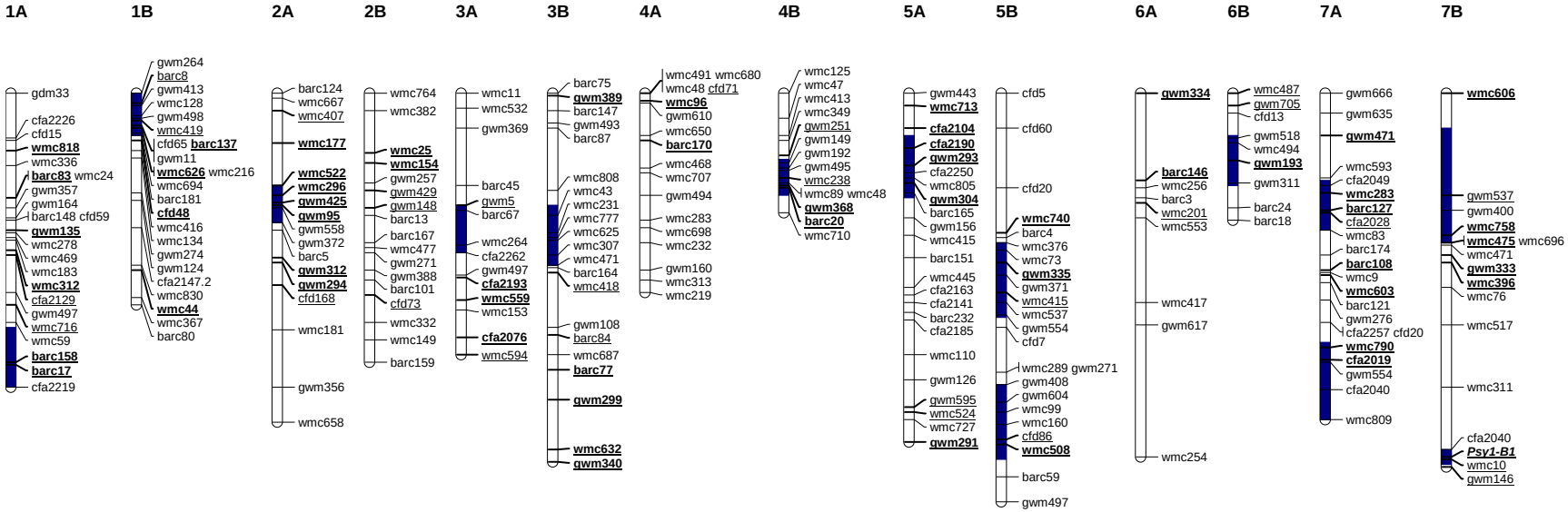

Figure 11. Significant marker-trait associations as identified by TASSEL ($P \leq 0.05$). Blue regions of chromosomes indicate previously identified QTL for yellow pigment concentration. Bold and underlined markers indicate areas identified as significant for yellow pigment concentration at all locations. Underlined markers indicate markers significantly associated at one to three locations. Marker distances and location based on wheat consensus map (Somers et al. 2004).

Xgwm425 and linked markers *Xwmc522*, *Xwmc296*, and *Xgwm95* were significant. X*gwm425* was previously associated with YP in a doubled haploid mapping population designed for QTL localization of YP (Pozniak et al. 2007). A second region on 2A distal to *Xgwm425* was also identified spanning *Xgwm312-Xgwm294-Xcfd168*, in a region where QTL have not previously been identified (Figure 11). Homoeologous regions on the distal end of the group 3 chromosomes were identified as being significant in all four testing environments. Marker associations were identified on 5A at the interval *gwm595-wmc524-wmc727-gwm291*, although only *Xgwm291* was significant in all four environments. On 5A, *Xcfa2104*, *Xcfa2190*, *Xgwm293*, and *Xgwm304* were all associated with YP, where Hessler et al. (2002) previously identified a locus significant for endosperm colour. In addition, *Xgwm193* on 6B, and a region on 7A flanked by *wmc790* and *wmc809* were also significant at all four testing environments, and both regions have been associated with YP previously in bi-parental mapping populations (Pozniak et al. 2007). *Psy1-A1*, which has recently been mapped 2 cM from *wmc809* and *cfa2257* (C. Pozniak, unpublished results) was also significantly associated with YP in the AM population. Two genomic regions associated with YP were identified on 7B. The region on 7BL including *Psy1-B1-Xwmc10-Xgwm146* was significant ($P<0.05$), but only *Psy1-B1* was significant in all environments (Figure 11). A second region on the 7BS centered at *Xgwm475* was also associated with YP. *Xgwm537*, *Xwmc758* and *Xwmc475*, all significant in the AM population are located within a second QTL interval on 7B recently associated with YP in durum wheat (Patil et al. 2008). In contrast to the above results which incorporate the Q matrix as defined for the presence of five subpopulations, the number of associations increased when population structure was left unaccounted for. When ignoring the presence of population structure in the AM population, a total of 134 significant marker trait associations were identified, distributed across the genome, and though a number of these associations correspond to those identified by STRUCTURE, it is likely that a large number of the associations are spurious.

4.6 LD mapping of Psy1

Both *Psy1-A1* and *Psy1-B1* have been associated with variation in yellow pigment concentration in independent studies in durum wheat and the potential of using

LD mapping to place these genes was assessed using the AM population. Pairwise LD analysis of the *Psy1* genes was evaluated. The r^2 revealed that *Psy1-B1* was in strong LD with *Xgwm146* and *cfa2040* (Figure 12) on chromosome 7B, and *Xgwm146* was significantly associated with variation in YP (Figure 11). These results are consistent with results from two mapping populations where *Psy1-B1* mapped approximately 2-4 cM from *Xgwm146* and 5cM from cfa2040 (Pozniak et al. 2007; Zhang and Dubcovsky 2008). *Psy1-B1* was also found to be in LD with *barc148* on chromosome 1A (Figure 12). Using data from the CAPS markers, *Psy1-A1* was in significant LD with *cfa2257* and *cfd20* on 7A (Figure 12). *Psy1-A1* maps 2 cM from *cfa2257* in three independent mapping populations (C. Pozniak, personal communication).

Table 11. Summary of number and frequency of *Psy1-B1* and *Psy1-A1* alleles. Yellow pigment concentration (mg kg^{-1}) least square means (± SD) were determined for each environment, Swift Current (SC) and Saskatoon (SK) in 2005 and 2006, based on accessions carrying the corresponding allele.

			2005		2006		2005-2006
Allele		Frequency (%)	SC	SK	SC	SK	Combined
Psy1-B1	*A*	76 (82)	7.70 ± 0.20	7.02 ± 0.16	8.31 ± 0.22	7.34 ± 0.18	7.59 ± 0.19
	B	17 (18)	8.99 ± 0.44	7.96 ± 0.33	9.85 ± 0.48	9.10 ± 0.39	8.96 ± 0.39
Psy1-A1	*A*	77 (83)	7.91 ± 0.18	7.17 ± 0.14	8.56 ± 0.20	7.60 ± 0.16	7.81 ± 0.16
	B	9 (10)	9.96 ± 0.52	8.95 ± 0.40	11.08 ± 0.58	10.03 ± 0.48	10.00 ± 0.47
	C	7 (7)	5.46 ± 0.59	5.13 ± 0.45	5.71 ± 0.65	5.28 ± 0.54	5.41 ± 0.54

Since LD mapping placed these genes in disequilibrium with microsatellite markers associated with YP, the effects of these genes were assessed in the AM population. At all environments, accessions containing the *Psy1-Alb* allele had significantly higher YP than accessions carrying any other allele (Table 11). In contrast, accessions carrying the *Psy1-A1c* allele contained significantly less YP. The *Psy1-B1b* allele also contributed positively to YP, with accessions carrying this allele having significantly higher pigment concentration than those carrying *Psy1-B1a*. These results are consistent with genetic mapping experiments that showed an average increase in YP of approx. 0.20 mg kg^{-1} in lines carrying the *Psy1-B1b* allele (Pozniak et al. 2007).

Several accessions in the AM population contained both *Psy1-A1b* and *Psy1-B1b* haplotypes including Commander, Wollaroi, 940955 and 920334 (Figure 3; Appendix

1). Averaged over all environments, accessions carrying both *Psy1-A1b* and *Psy1-B1b* had higher YP values (10.91 ± 0.70 mg kg^{-1}) compared to lines carrying any other combination of alleles.

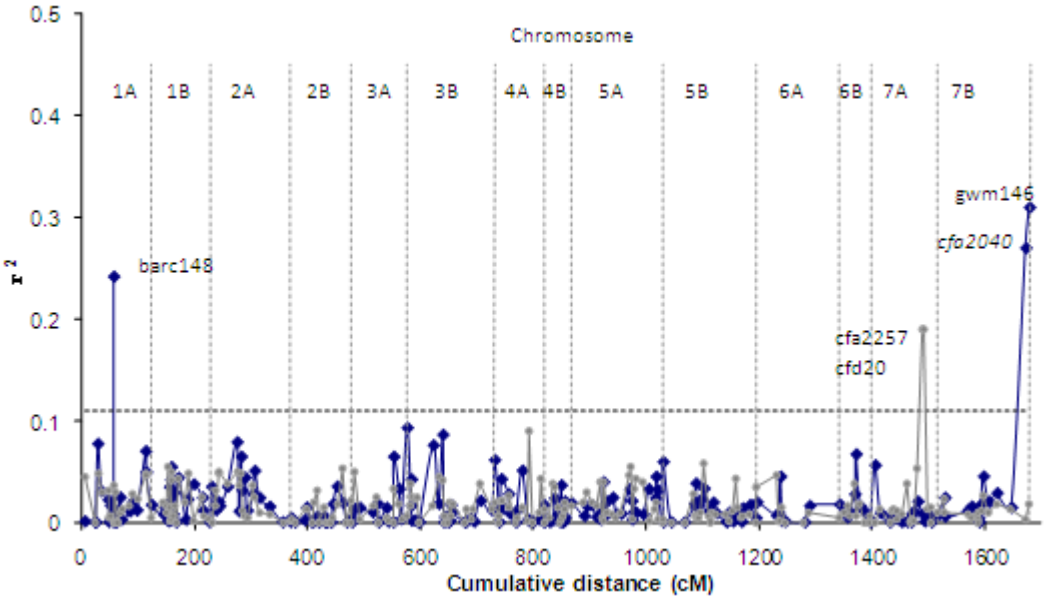

Figure 12. LD mapping of *Psy1-B1* and *Psy1-A1*. Pairwise LD (r^2) between *Psy1-B1* (blue) and *Psy1-A1* (grey) and all scored loci was calculated and plotted against the linkage map order. The r^2 peak on chromosome 7B indicates the putative location of the *Psy1-B1* using LD mapping. The dashed horizontal line represents the r^2 value where the genome wide relative cumulative frequency of r^2 was = 0.95.

5.0 DISCUSSION

5.1 Genetic Diversity and Population Structure

The aim of this research was to validate the potential of AM using a global collection of cultivars and breeding lines sampled from many durum wheat breeding programs. For AM studies, it is best to select lines to maximize genetic diversity as this minimizes pairwise LD between loci and contributes to greater power for association mapping (Yu et al. 2006). The experimental material used in this study was representative of the current elite cultivars and breeding lines of global durum wheat breeding programs. The level of polymorphism detected in this study was high, with as many as 15 alleles detected per microsatellite locus and a mean of five alleles per locus. These results were similar to previous reports of six alleles per locus among 40 elite European hexaploid wheat cultivars (Plaschke et al. 1995), 5.4 alleles among 68 CIMMYT hexaploid wheat breeding lines (Dreisigacker et al. 2004), and 4.8 alleles among 95 eastern U.S. soft winter wheats (Breseghello and Sorrells 2006a). However, the number of alleles was lower than the 7.2 alleles among 43 US wheat lines (Chao et al. 2007) and the 6.9 alleles detected among 134 elite durums sampled from wide geographic origins (Maccaferri et al. 2005). The lower polymorphism observed in this study compared to the previous durum study is likely due to a reduced sample size (93 vs. 134) as the average number of alleles for a 93 cultivar subset of Maccaferri et al.'s (2005) durum population was 5.5, similar to this study. These results suggest that larger population sizes than used here are justifiable as it would allow a greater sampling of alleles leading to decreased pairwise LD between unlinked loci and enhanced QTL resolution (Thornsberry et al. 2001; Gaut and Long 2003; Hamblin et al. 2004; Wilson et al. 2004).

Population structure contributes to LD between unlinked loci and must be accounted for to reduce type I error rates in AM studies. For AM studies, model-based estimates of population structure have been suggested as being superior to distance-based estimates (Pritchard et al. 2000a) as individual membership can be probabilistically assigned and allows modeling of admixture and assignment of individuals to more than one subpopulation. For example, Canadian accessions DT696

and DT691 share near equal admixture between North American subpopulations, a result of shared breeding material (Appendix 1; Appendix 5). Since germplasm exchange is common among global breeding programs, admixture among individuals would be expected in a global collection of lines like those used in this study.

The results from genetic distance-based and model-based approaches were similar, and generally grouped accessions based on geographic origin and known pedigrees. However, several differences were noted between distance and model-based estimates of population structure, which may in part be due to bias from the selection of 28 markers from which STRUCTURE was analyzed. Bayesian modeling did separate the North American material into two sub-populations, despite their close relationship in distance-based analysis (Figure 3). The first sub-population comprised accessions from Canadian wheat breeding programs, many progeny of the Canadian variety Kyle and US variety Westbred881. The second sub-population consisted of descendants from the North Dakota founder Edmore. Distance based analysis grouped Wollaroi, an Australian accession with the majority of Australian accessions, however model based analysis grouped the accession among the North American lines deriving from Edmore (Figure 3). Though Wollaroi does not derive from Edmore, Kamilaroi, present in the pedigree of Wollaroi, contains in its pedigree common progenitors (Capdur, Cando, Lloyd) to those of accessions such as Tetradur and AC Melita, present in the second North American subpopulation (http://www.jic.ac.uk/GERMPLAS/bbsrc_ce/Pedw.txt; Maccaferri et al. 2007). Buck Topacio a line from Argentina also clusters with the accessions from Australia and CIMMYT according to distance based analysis, but with the North American Edmore derived lines based on Bayesian modeling (Figure 3). However, pedigree information for Buck Topacio was unknown. Based on this information, Bayesian modeling is able to identify relationships between accessions that were undetected using genetic distance.

Pedigree information can be useful to assess coefficients of parentage as a means to classify relationship among lines. Lu et al (2005) in a study of 115 U.S. rice cultivars found that distance-based results correspond better to pedigree data than model-based. In contrast, population assignment with model-based methods in this study was largely

consistent with known pedigree data, similar to the results of Chao et al. (2007) who found that model-based estimates were similar to pedigree information when assigning population structure in a collection of 43 US wheat cultivars. Modeling of pedigree information in AM studies has been suggested as a means to improve statistical power and reduce Type II error rates (Breseghello and Sorrells 2006a; Yu et al. 2006). However, the effects of selection in breeding programs can make interpretation difficult (Kim and Ward 1997; Liu et al. 2003; Chao et al. 2007). The potential use of molecular coefficients of parentage (Bered et al. 2002; van Becelaere et al. 2005; Dreisigacker et al. 2005) should be the subject of further assessment as a means to improve power and reduce Type II error rates.

5.2 Phenotypic Expression of Yellow Pigment Concentration

To validate the use of AM in durum wheat, this study focused on analysis of YP concentration as the genetic control of YP has been well documented in the literature (Joppa and Williams 1988; Parker et al. 1998; Elouafi et al. 2001; Mares and Campbell 2001; Hessler et al. 2002; Cervigni et al. 2005; Atienza et al. 2007; Pozniak et al. 2007; He et al. 2008; Patil et al. 2008; Zhang and Dubcovsky 2008). Two methods were used to assess the yellow colour of durum endosperm, including total pigment extraction by water-saturated butanol and colorimetric assessment using CIE 1976 b*. High correlations between water-saturated butanol extractable pigments and CIE 1976 b* were evident (Table 6) and are consistent with earlier reports where significant correlations between CIE b* and butanol extracted pigments were identified (r=0.81, Carrera et al. 2007; r=0.83, Johnston et al. 1980). These results suggest that either method is an acceptable means of endosperm colour determination. For this thesis, the AM analysis was limited to water-saturated butanol extractable pigments as the majority of genetic studies published previously used this method for pigment concentration assessment (Elouafi et al 2001; Pozniak et al. 2007; Zhang and Dubcovsky 2008).

For genetic studies, a large range in phenotypic expression is desirable to increase the potential of identifying associations with polymorphic markers and QTL controlling the trait. In this study, a large range in expression of YP was observed in the AM population (average range of 8.04 mg kg^{-1}), regardless of environment. On average,

sub-populations accounted for approximately 20% of the phenotypic variation observed (Table 10), with the largest range in YP observed in sub-population 3 (composed predominantly of Australian and CIMMYT derived accessions). The large differences in phenotypic expression observed among sub-populations stresses the need to account for population-structure in AM studies (Breseghello and Sorrells 2006b). This was evidenced by the decrease in the number of associations identified when population structure was incorporated as covariates as opposed to when left unaccounted for. In this study, the accession x environment interaction was small and the heritability estimates were high (Table 2 and Table 4) and are consistent with previous reports (Elouafi et al. 2001; Clarke et al. 2006).

Hessler et al. (2002) reported that in some mapping populations, QTL for kernel weight were associated with variation in yellow pigment concentration, likely due to a pigment dilution effect with increasing seed size. In this study, a weak correlation ($r=-0.23$; $p<0.05$; Table 6) between seed size and YP concentration was detected only at Swift Current in 2006, despite a large range in seed size at all environments (Table 7). These results suggest that the variation in YP observed in the AM population were not due to differences in seed size.

5.3 Marker-trait Associations

The genetics of yellow pigment concentration is known to be complex, with QTL being reported on all chromosomes (Joppa and Williams 1988; Parker et al. 1998; Elouafi et al. 2001; Mares and Campbell 2001; Hessler et al. 2002; Cervigni et al. 2005; Atienza et al. 2007; Pozniak et al. 2007; He et al. 2008; Patil et al. 2008; Zhang and Dubcovsky 2008). After accounting for population structure, many of the marker-YP associations identified through AM were in genomic regions previously identified for yellow pigment concentration using conventional QTL mapping strategies (Figure 12). On chromosome 1A, *barc158* and *barc17* were localized to a QTL region reported previously (Patil et al. 2008). On chromosome 1B, Cervigni et al. (2005) reported QTL for semolina color in durum, which we confirmed with a clustering of markers near *wmc419* (Figure 12). Pozniak et al. (2007) reported a yellow pigment concentration QTL on chromosome 2A centered at *Xgwm425,* and that locus and its flanking markers

were declared significant in this study. Three yellow pigment concentration QTL were reported on the group 5 chromosomes (Mares and Campbell 2000; Hessler et al. 2002; Patil et al. 2008), and markers associated with all three regions were identified with AM (Figure 12). On chromosome 4B, a QTL was reported centered at *gwm495* (Pozniak et al. 2007), and the same region was identified in the AM population (Figure 12). In addition, a QTL on 6B centered at *Xgwm193* was reported in the same population, and this marker was significantly associated with YP in the AM population ($P<0.01$) (Figure 12). The marker interval X*wmc790-cfa2019* as well as *Xwmc283* and *barc127* on 7A fell within known QTL regions (Elouafi et al. 2001; Mares and Campbell 2001; Patil et al. 2008). On 7B, Elouafi et al. (2001) reported and Pozniak et al. (2007) and Zhang and Dubcovsky (2008) validated a major QTL in durum wheat near *gwm146*, which was detected with AM. Another region on the short arm of 7B around X*gwm537*, X*wmc758* and X*wmc475*, that was significantly associated with YP (Figure 12), is located within a second YP QTL on 7BS in durum wheat (Patil et al. 2008). Therefore, AM identified the majority of QTL for YP reported in durum wheat, while also identifying regions which have not been previously reported. Heritability of YP was high in the AM population (Table 4), and may explain, at least in part, why AM was able to successfully identify a number of markers located within known QTL locations associated with yellow endosperm colour. Further research is required to determine if AM would be as effective at detecting marker-trait associations for those traits with a low heritability.

An attractive feature of AM over conventional QTL analysis is the ability to detect additional QTL that were not segregating in published mapping populations, and identification of previously unreported YP-marker associations. Several markers on chromosomes 1A and 1B were identified in all environments in regions where QTL have not yet been reported. Homoeologous regions on the distal end of the group 3 chromosomes were identified as being significant in all four test environments (Figure 12) and are distinct from YP QTL reported on those chromosomes (Mares and Campbell 2001; Patil et al. 2008). On 2A, two regions were identified by AM in all four environments, but the region associated with linked markers *Xgwm312*, *Xgwm294*, and *cfd168* (Figure 12) has yet to be reported in the literature. Joppa et al. (1988) concluded that there were major genes for yellow colour on 2B, and it is reasonable to hypothesize

that the genomic region identified in this study may be the potential site for a QTL for YP on that chromosome. Although the possibility that these associations are spurious cannot be ruled out, the ability of AM to detect the majority of known QTL for YP in this population and that the majority of these markers were consistently detected in each of the four environments suggests that these novel associations should be the target of confirmation studies. A simple process of genotyping F2 plants and subsequent phenotyping of F2:3 families could be conducted to confirm the effect associated with the novel marker loci identified here (Breseghello and Sorrells 2006a). Some of the accessions in the AM population were selected as they are parents of current breeding populations, and thus populations that have already been extensively phenotyped are readily available for further genetic dissection of these novel QTL.

The number of molecular markers required for association mapping studies is largely based on the extent of LD in the population being evaluated (Flint-Garcia et al. 2003; Yu and Buckler 2006). A comparison of the available LD studies in wheat indicates that LD varies widely among different populations, both within and among species (Maccaferri et al. 2005; Breseghello and Sorrells 2006a). Somers et al. (2007) reported that LD in the AM population used in this thesis extends approximately 2-3 cM on average, suggesting that association mapping should localize genomic regions with resolution to that level. This was true for several regions in this study. The QTL reported on 6B centered at *Xgwm193* spanned 13.2 cM (Pozniak et al. 2007), but with association mapping, only *Xgwm193* was identified as being significant. Somers et al. (2007) reported low pairwise LD estimates (r^2<0.1) on 6B in this AM population, and this could explain the enhanced QTL resolution noted in this study. The QTL reported on 1A by Patil et al. (2008) was larger than the interval identified by AM in this study (Figure 11), despite high levels of LD reported in that region (Somers et al. 2007). The 2A interval *Xwmc522-Xgwm95* identified by AM spanned a similar interval previously reported in the bi-parental mapping population reported by Pozniak et al. (2007), and LD in that region was similar to the genome average (r^2= 0.2). Somers et al. (2007; Table 10) reported high levels of LD on 5AS (r^2>0.9) near the region where significant marker-YP associations were identified in this study. Because of the high LD present in

this region, the possibility that this association is spurious cannot be ruled out, although this region has a reported YP QTL (Hessler et al. 2002; Figure 11).

5.4 The Role of *Psy1* in Yellow Pigment Expression

The group 7 chromosomes appear to contain genes most critical to yellow pigment concentration (Parker et al. 1998; Elouafi et al. 2001; Mares and Campbell 2001; Pozniak et al. 2007; He et al. 2008; Patil et al. 2008; Zhang and Dubcovsky 2008). Elouafi et al. (2001) identified QTL on the group 7 chromosomes which together explained 62% of the total phenotypic variation observed for that trait. The 7B QTL alone explained 53% of the phenotypic variation. AM confirmed the role of the group 7 chromosomes on YP expression in wheat. In previous mapping experiments, a phytoene synthase gene *Psy1-B1* was linked to *gwm146* on 7B and was suggested as a potential candidate for the yellow pigment concentration QTL localized to 7B (Pozniak et al. 2007). In this study, *Psy1-B1* was in strong LD with *gwm146* (Figure 12) and confirmed that LD mapping can be used as a strategy to localize genes in the durum wheat genome. The majority of US desert durum varieties carried *Psy1-B1b*, including Westbred 881. Westbred 881 is in the pedigree of all Canadian lines carrying the *Psy1-B1b* allele, with the exception of DT536. Zhang and Dubcovsky (2008) confirmed that Cappelli *ph1c*, is the source of *Psy1-B1b* allele in modern durum wheat cultivars. Italian cultivars Fortore, Simeto, and Varano all carried *Psy1-B1b*, and all lines trace back to the founder Cappelli and its progeny Capeiti 8. It is thus likely that Cappelli *ph1c* is in the pedigree of these Italian cultivars, and not Cappelli, as reported (Maccaferri et al. 2003; Sanguineti et al. 2007; Appendix 1).

Although the AM results presented here suggest that *Psy1-B1* is associated with variation in YP in the AM population, Zhang and Dubcovsky (2008) hypothesized that an additional gene proximal to *Psy-B1* may also be influencing expression of grain pigment concentration in durum wheat. Although LD is low among markers in the distal regions of chromosome 7B in the AM population (Somers et al. 2007), AM could not resolve if two independent QTL for YP exist in that region. High resolution mapping will be required to confirm this hypothesis. Interestingly, pairwise LD estimates indicate that *Psy1-B1* was also in LD with *barc148* on 1A (Figure 12), suggesting these

two loci have been inherited together, perhaps as a result of selection efforts to elevate yellow pigment concentration in durum breeding programs. This could suggest an epistatic interaction between these two loci, where full expression of YP due to *Psy1-B1* is contingent on the allelic state at or near *barc148*. Interestingly, however when LD mapping was restricted to Canadian accessions, *barc148* was no longer in LD with *Psy1-B1* suggesting that among Canadian breeding programs selection may have occurred against or independent of this interaction. Conversely, it cannot be ruled out that the LD present between *barc148* and *Psy1-B1* may be the result of background LD.

Several studies have reported QTL for YP on the distal end of 7AL in wheat (Patil et al. 2007; Zhang and Dobcovsky 2008) and this region was detected using AM (Figure 11). In bread wheat, *Psy1-A1* is linked to *wmc809* on 7AL (He et al. 2008), a marker associated with YP in the AM population. Three alleles for *Psy1-A1* were detected in the AM population, and using allelic frequencies, this gene was found to be in LD with cfa2257 and cdf20, both on 7AL (Figure 12), and was associated with variation in YP (Figure 11) These results confirm current results where *Psy1-A1* mapped 2 cM from *cfa2257* (C. Pozniak, personal communication) and was associated with variation in YP in two independent mapping populations. Currently, the source of either *Psy1-A1b or Psy1-A1c* alleles is unknown (Appendix 1).

Results from AM studies confirmed the association of two *Psy1* genes and suggest that both of these genes could be targets for marker assisted selection and the markers used here for LD mapping (Figure 11 and Figure 12) could easily be implemented in breeding programs. In the AM population, the average effects on YP at the *Psy1-A1* where *Psy1-A1b*>*Psy1-A1a*>*Psy1-A1c* (Table 11), similar to that observed in bi-parental mapping populations (C. Pozniak, personal communication). Likewise, at *Psy1-B1*, lines carrying the *Psy1-B1b* allele had 1.37 mg kg^{-1} more YP, which is consistent with the additive effect of 0.6-0.8 mg kg-1 associated with that allele in two independent mapping populations (Pozniak et al. 2007; Zhang and Dubcovsky 2008). It does appear that combining *Psy1-A1b* and *Psy1-B1b* results in an additive effect, as on average, lines carrying both the *Psy1-A1b* and *Psy1-B1b* alleles had higher YP than lines carrying any other combination of alleles (Table 4; Figure 3). However, exceptions

were observed that contradict this hypothesis. For example, 940435, the cultivar with the highest yellow pigment concentration (Table 4) does not possess either the *Psy1-B1b* or *Psy1-A1b* alleles (Figure 3). In contrast, Mexa carries both alleles, and only expressed intermediate grain YP. However, given the complexity of the genetics of YP, it is not surprising that factors other than *Psy1* are influencing expression of YP. Perhaps incorporation of *Psy1-A1b* and *Psy1-B1b* into 940435 would result in even greater YP expression.

6.0 CONCLUSIONS AND FUTURE WORK

The overall objective of this study was to determine if association mapping could be used as a strategy to identify genomic regions associated with phenotypic variation for quantitative traits. The following conclusions were made based on the data collected:

- AM using cultivars and breeding lines was effective at detecting many of the QTL previously reported for YP concentration in bi-parental populations. These results suggest that AM can be an effective tool to identify genomic regions for multiple traits of interest commonly assessed in breeding programs where extensive phenotypic data sets are generated each year.

- AM confirmed the association of two *Psy1* loci with phenotypic variation in yellow pigment concentration. Taken together with published genetic mapping results, these loci could be the target of marker assisted selection to elevate YP in durum grain.

- Several genomic regions not associated with previously identified QTL were identified in the AM population.

- The resolution using AM was similar to that for reported QTL. These results suggest that AM is a complementary strategy to traditional QTL mapping where a relatively small set of core microsatellite markers can be used successfully to identify putative regions with AM for validation and finer genetic dissection in relevant segregating populations.

- The level of LD in this study is adequate for identification and selection of genes associated with traits of interest among durum breeding lines.

Future Work

Several gaps of 20 cM (based on the wheat consensus map, Somers et al. 2004) were present on chromosomes (ie. 6A), resulting in a number of regions being untested for significant associations. Markers in these regions may also be significantly associated

with YP, therefore it would be beneficial to increase marker saturation in regions lacking markers. One option is DArT® technology (Wenzl et al. 2004; Peleg et al. 2008), which has previously been successfully applied in wheat, or a second option focused on identification of SNPs may be used.

- Understanding population structure is critical to successful identification of associations. This study found that both genetic distance and Bayesian model based methods assigned population structure consistent with geographic origin and pedigree. Assignment of population structure is currently an area of research and further consideration of the use of pedigree data and coefficients of parentage for population assignment are potential areas of focus.

- Yellow pigment concentration was highly heritable in this study and the potential for AM in low to moderately heritable traits remains to be determined.

- Previous association of *Psy1* with yellow pigment concentration was confirmed in this study. The current QTL interval for *Psy1-B1* spans approximately 5 cM, therefore fine mapping the QTL region would determine if *Psy1* is the cause of elevated YP concentration or is linked to the causal factor. The presence of Indels and SNPs in regions of introns suggests potential linkage to a causal factor yet to be identified. Furthermore, in this study elevated pigment concentration was noted among lines carrying both *Psy1-A1b* and *Psy1-B1b* alleles, however, several lines lacking one or both alleles but with high pigment concentration were also observed. Therefore, study of the alleles in a similar background is required to determine whether the positive effect of *Psy1* on YP is the result of an additive effect or complementary effect.

- Novel QTL regions which were discovered as being significantly associated with yellow pigment concentration via association mapping need to be confirmed in segregating populations or NILs.

7.0 REFERENCES

AACC (American Association for Cereal Chemistry). 2000. Approved methods of the AACC-method, 10 edn. St. Paul Minnesota.

Abdel-Aal, E.-S. M., Young, J. C., Rabalski, I., Hucl, P., and Fregeau-Reid, J. 2007. Identification and quantification of seed carotenoids in selected wheat species. Journal of Agricultural and Food Chemistry **55**, 787-794.

Agrama, H. A., Eizenga, G. C., and Yan, W. 2007. Association mapping of yield and its components in rice cultivars. Molecular Breeding **19**, 341-356.

Ahmad, M. 2002. Assessment of genomic diversity among wheat genotypes as determined by simple sequence repeats. Genome **45**, 646-651.

Akhunov, E. D., Goodyear, A. W., Geng, S., Qi, L.-L., Echalier, B., Gill, B. S., Miftahudin, Gustafson, J. P., Lazo, G., Chao, S., Anderson, O. D., Linkiewicz, A. M., Dubcovsky, J., La Rota, M., Sorrells, M. E., Zhang, D., Nguyen, H. T., Kalavacharla, V., Hossain, K., Kianian, S. F., Peng, J., Lapitan, N. L. V., Gonzalez-Hernandez, J. L., Anderson, J. A., Choi, D.-W., Close, T. J., Dilbirligi, M., Gill, K. S., Walker-Simmons, M. K., Steber, C., McGuire, P. E., Qualset, C. O., and Dvorak, J. 2003. The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Research **13**, 753-763.

Andersen, J. R., Schrag, T., Melchinger, A. E., Zein, I., and Lübberstedt, T. 2005. Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). Theor Appl Genet **111**, 206-217.

Andersen, J. R., Zein, I., Wenzel, G., Krützfeldt, B., Eder, J., Ouzunova, M., and Lübberstedt, T. 2007. High levels of linkage disequilibrium and associations with forage quality at a *Phenylalanine Ammonia-Lyase* locus in European maize (*Zea mays* L.) inbreds. Theor Appl Genet **114**, 307-319.

Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P., and Nordborg, M. 2005. Genome-Wide Association Mapping in *Arabidopsis* Identifies Previously Known Flowering Time and Pathogen Resistance Genes. PLOS Genetics **1**, 531-539.

Arumuganathan, K., and Earle, E.D. 1991. Estimation of nuclear DNA content of plants by flow cytometry. Plant Mol Biol Rep **9**, 229-233.

Asins, M. 2002. Present and future of quantitative trait locus analysis in plant breeding. Plant Breeding **121**, 281-291.

Atienza, S. G., Avila, C. M., and Martin, A. 2007. The development of a PCR-based marker for *PSY1* from *Hordeum chilense*, a candidate gene for carotenoid content accumulatoin in tritordeum seeds. Australian Journal of Agricultural Research **58**, 767-773.

Barnaud, A., Lacombe, T., and Doligez, A. 2006. Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. Theor Appl Genet **112**, 708-716.

Belay, G., Tesemma, T., Becker, H. C., and Merker, A. 1993. Variation and interrelationships of agronomic traits in Ethiopian tetraploid wheat landraces. Euphytica **71**, 181-188.

Blair, M. W., Garris, A. J., Iyer, A. S., Chapman, B., Kresovich, S., and McCouch, S. R. 2003. High resolution genetic mapping and candidate gene identification at the *xa5* lous for bacterial blight resistance in rice (*Oryza sativa* L.). Theor Appl Genet **107**, 62-73.

Blanco, A., de Giovanni, C., Laddomada, B., Sciancalepore, A., Simeone, R., Devos, K. M., and Gale, M. D. 1996. Quantitative trait loci influencing grain protein content in tetraploid wheats. Plant Breeding **115**, 310-316.

Blanco, A., Simeone, R., and Gadaleta, A. 2006. Detection of QTLs for grain protein content in durum wheat. Theor Appl Genet **112**, 1195-1204.

Borrelli, G. M., Troccoli, A., Di Fonzo, N., and Fares, C. 1999. Durum Wheat Lipoxygenase Activity and Other Quality Parameters that Affect Pasta Color. Cereal Chemistry **76**, 335-340.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. 1980. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. Am.J.Hum.Genet **32**, 314-331.

Braaten, M. O., Lebsock, K. L., and Sibbitt, L. D. 1962. Intergeneration Relations of Physical Properties of Dough and Carotenoid Pigment Content in Durum Wheat. Crop Science **2**, 277-281.

Breseghello, F. and Sorrells, M. E. 2006b. Association Analysis as a Strategy for Improvement of Quantitative Traits in Plants. Crop Science **46**, 1323-1330.

Breseghello, F. and Sorrells, M. E. 2006a. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. Genetics **172**, 1165-1177.

Buckler, E. S., IV and Thornsberry, J. M. 2002. Plant molecular diversity and applications to genomics. Current Opinion in Plant Biology **5**, 107-111.

Buckner, B., Miguel, P. S., Janick-Buckner, D., and Bennetzen, J. L. 1996. The *y1* Gene of Maize Codes for Phytoene Synthase. Genetics **143**, 479-488.

Buntjer, J. B., Sørensen, A. P., and Peleman, J. D. 2005. Haplotype diversity: the link between statistical and biological association. TRENDS in Plant Science **10**, 466-471.

Caldwell, K. S., Dvorak, J., Lagudah, E. S., Akhunov, E., Luo, M.-C., Wolters, P., and Powell, W. 2004. Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. Genetics **167**, 941-947.

Caldwell, K. S., Russell, J., Langridge, P., and Powell, W. 2006. Extreme Population-Dependent Linkage Disequilibrium Detected in an Inbreeding Plant Species, *Hordeum vulgare*. Genetics **172**, 557-567.

Camus-Kulandaivelu, L., Veyrieras, J.-B., Gouesnard, B., Charcosset, A., and Manicacci, D. 2007. Evaluating the Reliability of *Structure* Outputs in Case of Relatedness between Individuals. Crop Science **47**, 887-892.

Cardon, L. R. and Palmer, L. J. 2003. Population stratification and spurious allelic association. The Lancet **361**, 598-604.

Carrera, A., Echenique, V., Zhang, W., Helguera, M., Manthey, F., Schrager, A., Picca, A., Cervigni, G., and Dubcovsky, J. 2007. A deletion at the *Lpx-B1* locus is associated with low lipoxygenase activity and improved pasta color in durum wheat (*Triticum turgidum* ssp. *durum*). Journal of Cereal Science **45**, 67-77.

Cenci, A., Somma, S., Chantret, N., Dubcovsky, J., and Blanco, A. 2004. PCR identification of durum wheat BAC clones containing genes coding for carotenoid biosynthesis enzymes and their chromosome localization. Genome **47**, 911-917.

Cervigni, G., Zhang, W., Picca, A., Carrera, A., Helguera, M., Manthey, F., Miranda, R., Dubcovsky, J., and Echenique, V. QTL Mapping for LOX Activity and Quality Traits in Durum Wheat *Triticum turgidum* L. var. durum. 7th International Wheat Conference. 2005.

Chao, S., Zhang, W., Dubcovsky, J., and Sorrells, M. 2007. Evaluation of Genetic Diversity and Genome-wide Linkage Disequilibrium among U.S. Wheat (*Triticum aestivum* L.) Germplasm Representing Different Market Classes. Crop Science **47**, 1018-1030.

Chicaiza, O., Khan, I. A., Zhang, X., Brevis, J. C., Jackson, L., Chen, X., and Dubcovsky, J. 2006. Registration of five wheat isogenic lines for leaf rust and stripe rust resistance genes. Crop Science **46**, 485-487.

Ching, A., Caldwell, K. S., Jung, M., Dolan, M., Smith, O. S. H., Tingey, S., Morgante, M., and Rafalski, A. J. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics **3**.

Clarke, F. R., Clarke, J. M., McCaig, T. N., Knox, R. E., and DePauw, R. M. 2006. Inheritance of yellow pigment concentration in seven durum wheat crosses. Canadian Journal of Plant Science **86**, 133-141.

Clarke, F.R., Yates,S.M., Clarke,J.M., DePauw,R.M., McLaren,C.G. 2003. The international crop information system: a useful tool for wheat breeding programs. In: Proceedings of the Tenth International Wheat Genetics Symposium, 1-6 September 2003, Paestum, Italy. Vol. 2. (Eds: Pogna,N.E., Romano,M., Pogna,E.A., Galterio,G.) S.I.M.I., Rome, 684-686.

Clarke, J. M., Marchylo, B. A., Kovacs, M. I. P., Noll, J. S., McCaig, T. N., and Howes, N. K. 1998. Breeding durum wheat for pasta quality in Canada. Euphytica **100**, 163-170.

Clarke, J. M., McCaig, T. N., DePauw, R. M., Knox, R. E., Ames, N. P., Clarke, F. R., Fernandez, M. R., Marchylo, B. A., and Dexter, J. E. 2005a. Commander durum wheat. Canadian Journal of Plant Science **85**, 901-904.

Clarke, J. M., McCaig, T. N., DePauw, R. M., Knox, R. E., Clarke, F. R., Fernandez, M. R., and Ames, P. 2005b. Strongfield durum wheat. Canadian Journal of Plant Science **85**, 651-654.

Clarke, J. M., McLeod, J. G., DePauw, R. M., Marchylo, B. A., McCaig, T. N., Knox, R. E., Fernandez, M. R., and Ames, N. 2000. AC Navigator durum wheat. Canadian Journal of Plant Science **80**, 343-345.

Clarke, J. M., McLeod, J. G., DePauw, R. M., Marchylo, B. A., McCaig, T. N., Knox, R. E., Fernandez, M. R., and Ames, N. 2000. AC Pathfinder durum wheat. Canadian Journal of Plant Science **80**, 347-349.

Clarke, J. M., McLeod, J. G., McCaig, T. N., DePauw, R. M., Knox, R. E., and Fernandez, M. R. 1998. AC Avonlea durum wheat. Canadian Journal of Plant Science **78**, 621-623.

Cockram, J., White, J., Leigh, F. J., Lea, V. J., Chiapparino, E., Laurie, D. A., Mackay, I. J., Powell, W., and O'Sullivan, D. M. 2008. Association mapping of partitioning loci in barley. BMC Genetics **9**(16).

Commission Internationale de L'Eclairage. 1986. Colorimetry. (2nd Ed) Publication CIE No. 15.2 Central Bureau of the CIE, Vienna, Austria.

Crossa, J., Burgueño, 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., and Ortiz, R. 2007. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics **177**, 1889-1913.

Cunningham Jr., F. X. and Gantt, E. 1998. Genes and Enzymes of Carotenoid Biosynthesis in Plants. Annual Review of Plant Physiology and Plant Molecular Biology **49**, 557-583.

DeWan, A., Liu, M., Hartman, S., Zhang, S. S.-M., Liu, D. T. L., Zhao, C., Tam, P. O. S., Chan, W. M., Lam, D. S. C., Snyder, M., Barnstable, C., Pang, C. P., and Hoh, J. 2006. *HTRA1* promoter polymorphism in wet age-related macular degeneration. Science **314**, 989-992.

Dreisigacker, S., Melchinger, A. E., Zhang, P., Ammar, K., Flachenecker, C., Hoisington, D., and Warburton, M. L. 2005. Hybrid performance and heterosis in spring bread wheat, and their relations to SSR-based genetic distances and coefficients of parentage. Euphytica **144**, 51-59.

Dreisigacker, S., Zhang, P., Warburton, M. L., Van Ginkel, M., Hoisington, D., Bohn, M., and Melchinger, A. E. 2004. SSR and Pedigree Analyses of Genetic Diversity among CIMMYT Wheat Lines Targeted to Different Megaenvironments. Crop Science **44**, 381-388.

Elouafi, I., Nachit, M. M., and Martin, L. M. 2001. Identification of a microsatellite on chromosome 7B showing a strong linkage with yellow pigment in durum wheat (Triticum turgidum L. var. durum). Hereditas **135**, 255-261.

Eujayl, I., Sorrells, M., Baum, M., Wolters, P., and Powell, W. 2001. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRS. Euphytica **119**, 39-43.

Falush, D., Stephens, M., and Pritchard, J. K. 2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics **164**, 1567-1587.

Faris, J. D., Haen, K. M., and Gill, B. S. 2000. Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics **154**, 823-835.

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., and Georges, M. 2000. Extensive Genome-wide Linkage Disequilibrium in Cattle. Genome Research **10**, 220-227.

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S., IV. 2003. Structure of Linkage Disequilibrium in Plants. Annual Review of Plant Biology **54**, 357-374.

Fraser, P. D. and Bramley Peter M. 2004. The biosynthesis and nutritional uses of carotenoids. Progress in Lipid Research **43**, 228-265.

Fratianni, A., Irano, M., Panfili, G., and Acquistucci, R. 2005. Estimation of Color of Durum Wheat. Comparison of WSB, HPLC, and Reflectance Colorimeter Measurements. Journal of Agricultural and Food Chemistry **53**, 2373-2378.

Gallagher, C. E., Matthews, P. D., Li, F., and Wurtzel, E. T. 2004. Gene Duplication in the Carotenoid Biosynthetic Pathway Preceded Evolution of the Grasses. Plant Physiology **135**, 1776-1783.

Garris, A. J., McCouch, S. R., and Kresovich, S. 2003. Population Structure and Its Effect on Haplotype Diversity and Linkage Disequilibrium Surrounding the *xa5* Locus of Rice (*Oryza sativa* L.). Genetics **165**, 759-769.

Gaut, B. S. and Long, A. D. 2003. The Lowdown on Linkage Disequilibrium. The Plant Cell **15**, 1502-1506.

Gonzalez-Hernandez, J. L., Elias, E. M., and Kianin, S. F. 2004. Mapping genes for grain protein concentration and grain yield on chromosome 5B of *Triticum turgidum* L. var *dicoccoides*. Euphytica **139**, 217-225.

Groos, C., Robert, N., Bervas, E., and Charmet, G. 2003. Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theor Appl Genet **106**, 1032-1040.

Gupta, P. K., Rustgi, S., and Kulwal, P. L. 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Molecular Biology **57**, 461-485.

Hai, L., Wagner, C., and Friedt, W. 2007. Quantitative structure analysis of genetic diversity among spring bread wheats (*Triticum aestivum* L.) from different geographical regions. Genetica **130**, 213-225.

Hamblin, M. T., Mitchell, S. E., White, G. M., Gallego, J., Kukatla, R., Wing, R. A., Paterson, A. H., and Kresovich, S. 2004. Comparative Population Genetics of the Panicoid Grasses: Sequence Polymorphism, Linkage Disequilibrium and Selection in a Diverse Sample of *Sorghum bicolor*. Genetics **167**, 471-483.

He, X. Y., Zhang, Y. L., He, Z. H., Wu, Y. P., Xiao, Y. G., Ma, C. X., and Xia, X. C. 2008. Characterization of phytoene synthase 1 gene (*Psy1*) located on common wheat chromosome 7A and development of a functional marker. Theor Appl Genet **116**, 213-221.

Hentschel, V., Kranl, K., Hollmann, J., Lindhauer, M. G., Böhm, V., and Bitsch, R. 2002. Spectrophotometric Determination of Yellow Pigment Content and Evaluation of Carotenoids by High-Performance Liquid Chromatography in Durum Wheat Grain. Journal of Agricultural and Food Chemistry **50**, 6663-6668.

Hessler, T. G., Thomson, M. J., Benscher, D., Nachit, M. M., and Sorrells, M. E. 2002. Association of a Lipoxygenase Locus, *Lpx-B1*, with Variation in Lipoxygenase Activity in Durum Wheat Seeds. Crop Science **42**, 1695-1700.

Huang, X. Q., Cloutier, S., Lycar, L., Radovanovic, N., Humphreys, D. G., Noll, J. S., Somers, D. J., and Brown, P. D. 2006. Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). Theor Appl Genet **113**, 753-766.

Hyten, D. L., Choi, I.-Y., Song, Q., Shoemaker, R. C., Nelson, R. L., Costa, J. M., Specht, J. E., and Cregan, P. B. 2007. Highly Variable Patterns of Linkage Disequilibrium in Multiple Soybean Populations. Genetics **175**, 1937-1944.

Hyten, D. L., Song, Q., Zhu, Y., Choi, I.-Y., Nelson, R. L., Costa, J. M., Specht, J. E., Shoemaker, R. C., and Cregan, P. B. 2006. Impacts of genetic bottlenecks on soybean genome diversity. PNAS **103**, 16666-16671.

Jannink, J.-L., Bink, M. C. A. M., and Jansen, R. C. 2001. Using complex plant pedigrees to map valuable genes. TRENDS in Plant Science **6**, 337-342.

Jing, H.-C., Kornyukhin, D., Kanyuka, K., Orford, S., Zlatska, A., Mitrofanova, O. P., Koebner, R., and Hammond-Kosack, K. 2007. Identification of variation in adaptively important traits and genome-wide analysis of trait-marker associations in *Triticum monococcum*. Journal of Experimental Botany **58**, 3749-3764.

Johnston, R. A., Quick, J. S., and Donnelly, B. J. 1980. Note on Comparison of Pigment Extraction and Reflectance Colorimeter Methods for Evaluating Semolina Color. Cereal Chemistry **57**, 447-448.

Johnston, R. A., Quick, J. S., and Hammond, J. J. 1983. Inheritance of Semolina Color in Six Durum Wheat Crosses. Crop Science **23**, 607-610.

Joppa, L. R., Du, C., Hart, G. E., and Hareland, G. A. 1997. Mapping gene(s) for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. Crop Science **37**, 1586-1589.

Joppa, L. R. and Williams, N. D. 1988. Langdon durum disomic substitution lines and aneuploid analysis in tetraploid wheat. Genome **30**, 222-228.

Jun, T.-H., Van, K., Kim, M. Y., Lee, S.-H., and Walker, D. R. 2008. Association analysis using SSR markers to find QTL for seed protein content in soybean. Euphytica **162**, 179-191.

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

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

Kim, H. S. and Ward, R. W. 1997. Genetic diversity in Eastern U.S. soft winter wheat (*Triticum aestivum* L. em. Thell.) based on RFLPs and coefficients of parentage. Theor Appl Genet **94**, 472-479.

Knowler, W. C., Williams, R. C., Pettitt, D. J., and Steinberg, A. G. 1988. $Gm^{3;5,13,14}$ and Type 2 Diabetes Mellitus: An Association in American Indians with Genetic Admixture. Am.J.Hum.Genet **43**, 520-526.

Kraakman, A. T. W., Martínez, F., Mussiraliev, B., van Eeuwijk, F. A., and Niks, R. E. 2006. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Molecular Breeding **17**, 41-58.

Kraakman, A. T. W., Niks, R. E., Van den Berg, P. M. M. M., Stam, P., and Van Eeuwijk, F. A. 2004. Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars. Genetics **168**, 435-446.

Kuchel, H., Langridge, P., Mosionek, L., Williams, K., and Jefferies, S. P. 2006. The genetic control of milling yield, dough rheology and baking quality of wheat. Theor Appl Genet **112**, 1487-1495.

Lewontin, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49**, 49-67.

Li, F., Vallabhaneni, R., and Wurtzel, E. T. 2008. *PSY3*, a new memeber of the phytoene synthase gene family conserved in the Poaceae and regulator of abiotic-stress-induced root carotenogenesis. Plant Physiology **146**, 1333-1345.

Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. 1996. SAS system for mixed models. SAS Institute Inc, Cary, p 633.

Liu, K., Goodman, M., Muse, S., Smith, J. S., Buckler, E., and Doebley, J. 2003. Genetic Structure and Diversity Among Maize Inbred Lines as Inferred from DNA Microsatellites. Genetics **165**, 2117-2128.

Liu, K. and Muse, S. V. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics **21**, 2128-2129.

Lu, H., Redus, M. A., Coburn, J. R., Rutger, J. N., McCouch, S. R., and Tai, T. H. 2005. Population Structure and Breeding Patterns of 145 U.S. Rice Cultivars Based on SSR Marker Analysis. Crop Science **45**, 66-76.

Maccaferri, M., Sanguineti, M. C., Donini, P., and Tuberosa, R. 2003. Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. Theor Appl Genet **107**, 783-797.

Maccaferri, M., Sanguineti, M. C., Noli, E., and Tuberosa, R. 2005. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Molecular Breeding **15**, 271-289.

Maccaferri, M., Sanguineti, M. C., Xie, C., Smith, J. S. C., and Tuberosa, R. 2007. Relationships among durum wheat accessions. II. A comparison of molecular and pedigree information. Genome **50**, 385-399.

Macdonald, S. J., Pastinen, T., Genissel, A., Cornforth, T. W., and Long, A. D. 2005. A low-cost open-source SNP genotyping platform for association mapping applications. Genome Biology **6**, R105.

Malosetti, M., van der Linden, C. G., Vosman, B., and van Eeuwijk, F. A. 2007. A Mixed-Model Approach to Association Mapping Using Pedigree Information With an Illustration of Resistance to *Phytophthora infestans* in Potato. Genetics **175**, 879-889.

Mares DJ and Campbell AW. 2001. Mapping components of flour and noodle colour in Australian wheat. Australian Journal of Agricultural Research **52**, 1297-1309.

Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., McCouch, S., and Purugganan, M. D. 2007. The Extent of Linkage Disequilibrium in Rice (*Oryza sativa* L.). Genetics **177**, 2223-2232.

Matsuo, R. R. and Dexter, J. E. 1980. Relationship Between Some Durum Wheat Physical Characteristics and Semolina Milling Properties. Canadian Journal of Plant Science **60**, 49-53.

McCaig, T. N., Gan, Y. T., Clarke, P., Clarke, J. M., and DePauw, R. M. 2006. Kernel colour changes associated with field weathering of spring wheat. Canadian Journal of Plant Science **86**, 371-377.

McCaig, T. N., McLeod, J. G., Clarke, J. M., and DePauw, R. M. 1992. Measurement of Durum Pigment with a Near-Infrared Instrument Operating in the Visible Range. Cereal Chemistry **69**, 671-672.

McCartney, C. A., Somers, D. J., Lukow, O., Ames, N., Noll, J., Cloutier, S., Humphreys, D. G., and McCallum, B. D. 2006. QTL analysis of quality traits in the spring wheat cross RL4452 x 'AC Domain'. Plant Breeding **125**, 565-575.

Morgante, M. and Salamini, F. 2003. From plant genomics to breeding practice. Current Opinion in Biotechnology **14**, 214-219.

Morrell, P. L., Toleno, D. M., Lundy, K. E., and Clegg, M. T. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. PNAS **102**, 2442-2447.

Nei, M. 1972. Genetic Distance between Populations. The American Naturalist **106**, 283-292.

Nei, M. 1978. Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. Genetics **89**, 583-590.

Neter, J., Kutner, M.H., Nachtshein, C.J., and Wasserman, W. 1996. Applied Linear Statistical Models. McGraw-Hill, New York.

Nordborg, M., Borevitz, J. O., Bergelson, J., Berry, C. C., Chory, J., Hagenblad, J., Kreitman, M., Maloof, J. N., Noyes, T., Oefner, P. J., Stahl, E. A., and Weigel, D. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics ***30***, 190-193.

Nordborg, M. and Tavaré, S. 2002. Linkage disequilibrium: what history has to tell us. TRENDS in Genetics **18**, 83-90.

Olmos, S., Distelfeld, A., Chicaiza, O., Schlatter, A. R., Fahima, T., Echenique, V., and Dubcovsky, J. 2003. Precise mapping of a locus affecting grain protein content in durum wheat. Theor Appl Genet **107**, 1243-1251.

Oliver, J. R., Blakeney, A. B., and Allen, H. M. 1992. Measurement of Flour Color in Color Space Parameters. Cereal Chemistry **69**, 546-551.

Page, R. 1996. Treeview: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences **12**, 357-358.

Palaisa, K. A., Morgante, M., Williams, M., and Rafalski, A. 2003. Contrasting Effects of Selection on Sequence Diversity and Linkage Disequilibrium at Two Phytoene Synthase Loci. The Plant Cell **15**, 1795-1806.

Panfili, G., Fratianni, A., and Irano, M. 2004. Improved Normal-Phase High-Performance Liquid Chromatography Procedure for the Determination of Carotenoids in Cereals. Journal of Agricultural and Food Chemistry **52**, 6373-6377.

Parker, G. D., Chalmers, K. J., Rathjen, A. J., and Langridge, P. 1998. Mapping loci associated with flour colour in wheat (*Triticum aestivum* L.). Theor Appl Genet **97**, 238-245.

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., and Tanksley, S. D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature **335**, 721-726.

Patil, R. M., Oak, M. D., Tamhankar, S. A., Sourdille, P., and Rao, V. S. 2008. Mapping and validation of a major QTL for yellow pigment content on 7AL in durum wheat (*Triticum turgidum* L. ssp. *durum*). Molecular Breeding **21**, 485-496.

Patterson, H. D. and Williams, E. R. 1976. A new class of resolvable incomplete block designs. Biometrika **63**, 83-92.

Peleg, Z., Fahima, T., Abbo, S., Krugman, T., and Saranga, Y. 2008. Genetic structure of wild emmer wheat populations as reflected by transcribed versus anonymous SSR markers. Genome **51**, 187-195.

Peng, J. H., Bai, Y., Haley, S. D., and Lapitan, N. L. V. 2008. Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. Genetica.

Plaschke, J., Ganal, M. W., and Röder, M. S. 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet **91**, 1001-1007.

Pozniak, C. J., Knox, R. E., Clarke, F. R., and Clarke, J. M. 2007. Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. Theor Appl Genet **114**, 525-537.

Prasad, M., Varshney, R. K., Kumar, A., Balyan, H. S., Sharma, P. C., Edwards, K. J., Singh, H., Dhaliwal, H. S., Roy, J. K., and Gupta, P. K. 1999. A microsatellite marker associated with a QTL for grain protein content on chromosome arm 2DL of bread wheat. Theor Appl Genet **99**, 341-345.

Price, A. H. 2006. Believe it or not, QTLs are accurate! TRENDS in Plant Science **11**, 213-216.

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

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. 2000b. Association Mapping in Structured Populations. Am.J.Hum.Genet **67**, 170-181.

Rafalski, J. A. 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science **162**, 329-333.

Ravel, C., Praud, S., Murigneux, A., Canaguier, A., Sapet, F., Samson, D., Balfourier, F., Dufour, P., Chalhoub, B., Brunel, D., Beckert, M., and Charmet, G. 2006a. Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome **49**, 1131-1139.

Ravel, C., Praud, S., Murigneux, A., Linossier, L., Dardevet, M., Balfourier, F., Dufour, P., Brunel, D., and Charmet, G. 2006b. Identification of *Glu-B1-1* as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (Triticum aestivum L.) by means of an association study. Theor Appl Genet **112**, 738-743.

Reif, J. C., Hamrit, S., Heckenberger, M., Schipprack, W., Maurer, H. P., Bohn, M., and Melchinger, A. E. 2005. Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. Theor Appl Genet **111**, 906-911.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., and Buckler, E. S., IV. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. PNAS **98**, 11479-11484.

Rhoné, B., Raquin, A.-L., and Goldringer, I. 2007. Strong linkage disequilibrium near the selected *Yr17* resistance gene in a wheat experimental population. Theor Appl Genet **114**, 787-802.

Röder, M. S., Plaschke, J., König, S. U., Börner, A., Sorrells, M. E., Tanksley, S. D., and Ganal, M. W. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet **246**, 327-333.

Rogers JS. 1972. Measures of genetic similarity and genetic distance. Studies in genetics. VII. Univ. Texas Publ. 7213:145-153.

Rostoks, N., Ramsay, L., MacKenzie, K., Cardle, L., Bhat, P. R., Roose, M. L., Svensson, J. T., Stein, N., Varshney, R. K., Marshall, D. F., Graner, A., Close, T. J., and Waugh, R. 2006. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. PNAS **103**, 18656-18661.

Roy, J. K., Bandopadhyay, R., Rustgi, S., Balyan, H. S., and Gupta, P. K. 2006. Association analysis of agronomically important traits using SSR, SAMPL, and AFLP markers in bread wheat. Current Science **90**, 683-689.

Sandhu, D. and Gill, K. S. 2002. Gene-containing regions of wheat and the other grass genomes. Plant Physiology **128**, 803-811.

Sanguineti, M. C., Li, S., Maccaferri, M., Corneti, S., Rotondo, F., Chiari, T., and Tuberosa, R. 2007. Genetic dissection of seminal root architecture in elite durum wheat germplasm. Annals of Applied Biology **151**, 291-305.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology **18**, 233-234.

Sharma, D. and Knott, D. R. 1964. The inheritance of seed weight in a wheat cross. Canadian Journal of Genetics and Cytology **6**, 419-425.

Skøt, L., Humphreys, J., Humphreys, M. O., Thorogood, D., Gallagher, J., Sanderson, R., Armstead, I. P., and Thomas, I. D. 2007. Association of Candidate Genes with Flowering Time and Water-Soluble Carbohydrate Content in *Lolium perenne* (L.). Genetics **177**, 535-547.

Somers DJ, Banks T, DePauw, R., Fox S, Clarke J, Pozniak C, and McCartney C. 2007. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome **50**, 557-567.

Somers, D. J., Isaac, P., and Edwards, K. 2004. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet **109**, 1105-1114.

Spielmeyer, W., Singh, R. P., McFadden, H., Wellings, C. R., Huerta-Espino, J., Kong, X., Appels, R., and Lagudah, E. S. 2008. Fine scale genetic and physical mapping using interstitial deletion mutants of *Lr34/Yr18*: a disease resistance locus effective against multiple pathogens in wheat. Theor Appl Genet **116**, 481-490.

Spielmeyer, W., Moullet, O., Laroche, A., and Lagudah, E. S. 2000. Highly Recombinogenic Regions at Seed Storage Protein Loci on Chromosome 1DS of *Aegilops tauschii*, the D-Genome Donor of Wheat. Genetics **155**, 361-367.

Stich, B., Maurer, H. P., Melchinger, A. E., Frisch, M., Heckenberger, M., van der Voort, J. R., Peleman, J., Sorenson, A. P., and Reif, J. C. 2006. Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. Molecular Breeding **17**, 217-226.

Stich, B., Melchinger, A. E., Frisch, M., Maurer, H. P., Heckenberger, M., and Reif, J. C. 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet **111**, 723-730.

Stich, B., Melchinger, A. E., Piepho, H.-P., Hamrit, S., Schipprack, W., Maurer, H. P., and Reif, J. C. 2007. Potential causes of linkage disequilibrium in a European maize breeding program investigated with computer simulations. Theor Appl Genet **115**, 529-536.

Tanksley, S. D. 1993. Mapping Polygenes. Annual Review of Genetics **27**, 205-233.

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

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

Tommasini, L., Schnurbusch, T., Fossati, D., Mascher, F., and Keller, B. 2007. Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties. Theor Appl Genet **115**, 697-708.

Townley-Smith, T. F., DePauw, R. M., Lendrum, C. W. B., McCrystal, G. E., and Patterson, L. A. 1987. Kyle Durum Wheat. Canadian Journal of Plant Science **67**, 225-227.

Troccoli, A., Borrelli, G. M., De Vita, P., Fares, C., and Di Fonzo, N. 2000. Durum Wheat Quality: A Multidisciplinary Concept. Journal of Cereal Science **32**, 99-113.

Turner, A. S., Bradburne, R. P., Fish, L., and Snape, J. W. 2004. New quantitative trait loci influencing grain texture and protein content in bread wheat. Journal of Cereal Science **40**, 51-60.

Van Becelaere, G., Lubbers, E. L., Paterson, A. H., and Chee, P. W. 2005. Pedigree- vs. DNA Marker-Based Genetic Similarity Estimates in Cotton. Crop Science **45**, 2281-2287.

van den Berg, H., Faulks R, Granado, H. F., Hirschberg J, Olmedilla B, Sandmann G, Southon S, and Stahl W. 2000. The potential for the improvement of carotenoid levels in foods and the likely systemic effects. Journal of the Science of Food and Agriculture **80**, 880-912.

Weber, A., Clark, R. M., Vaughn, L., Sánchez-Gonzalez, J. d. J., Yu, J., Yandell, B. S., Bradbury, P., and Doebley, J. 2007. Major Regulatory Genes in Maize Contribute to Standing Variation in Teosinte (*Zea mays* ssp. *parviglumis*). Genetics **177**, 2349-2359.

Weir, B. S., 1996: Genetic Data Analysis II, 2nd Ed. Sinauer Associates, Inc., Sunderland, MA.

Welsch, R., Wüst, F., Bär, C., Al-Babili, S., and Beyer, P. 2008. A Third Phytoene Synthase Is Devoted to Abiotic Stress-Induced Abscisic Acid Formation in Rice and Defines Functional Diversification of Phytoene Synthase Genes. Plant Physiology **147**, 367-380.

Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A., and Kilian, A. 2004. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. PNAS **101**, 9915-9920.

Wilson, L. M., Whitt, S. R., Ibáñez, A. M., Rocheford, T. R., Goodman, M. M., and Buckler, E. S., IV. 2004. Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Analysis. The Plant Cell **16**, 2719-2733.

Wong, J. C., Lambert, R. J., Wurtzel, E. T., and Rocheford, T. R. 2004. QTL and candidate genes phytoene synthase and zeta-carotene desaturase associated with the accumulation of carotenoids in maize. Theor Appl Genet **108**, 349-359.

Wurtzel, E. T., Luo, R., and Yatou, O. 2001. A simple approach to identify the first rice mutants blocked in carotenoid biosynthesis. Journal of Experimental Botany **52**, 161-166.

Youngs, V. L., 1988: Durum Lipids. In: G. Fabriani and C. Lintas (eds.), Durum Wheat: Chemistry and Technology, 139-148. American Association of Cereal Chemists, Inc., St. Paul, Minnesota.

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

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

Zhang, W. and Dubcovsky, J. 2008. Association between allelic variation at the Phytoene synthase 1 gene and yellow pigment content in the wheat grain. Theor Appl Genet **116**, 635-645.

Zhao, H., Nettleton, D., Soller, M., and Dekkers, J. C. M. 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. Genet Res Camb **86**, 77-87.

Zhao, J., Paulo, M.-J., Jamar, D., Lou, P., van Eeuwijk, F., Bonnema, G., Vreugdenhil, D., and Koornneef, M. 2007a. Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. Genome **50**, 963-973.

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., and Nordborg, M. 2007b. An *Arabidopsis* Example of Association Mapping in Structured Samples. PLOS Genetics **3**, 71-82.

Zhu, Q., Zheng, X., Luo, J., Gaut, B. S., and Ge, S. 2007. Multilocus Analysis of Nucleotide Variation of *Oryza sativa* and Its Wild Relatives: Severe Bottleneck during Domestication of Rice. Mol Biol Evol **24**, 875-888.

Zhu, Y. L., Song, Q. J., Hyten, D. L., Van Tassell, C. P., Matukumalli, L. K., Grimm, D. R., Hyatt, S. M., Fickus, E. W., Young, N. D., and Cregan, P. B. 2003. Single-Nucleotide Polymorphisms in Soybean. Genetics **163**, 1123-1134.

8.0 APPENDICES

Appendix 1. Association Mapping Population accession composition, origin and pedigree. Corresponding *Psy1-A1* and *Psy1-B1* alleles as determined by marker assignment are indicated for individual accessions.

Accession	Origin	Pedigree	*Psy1-A1*	*Psy1-B1*
Bonaerance Inta Cumenay	Argentina	Unknown	a*	a*
Bonaerance Quilaco	Argentina	MAGH72//GS/AA///RABI//D21563/AA	a*	a*
Bonaerance Valverde	Argentina	GIORGIO//CAPELLI/YUMA	a*	a*
Buck Ambar	Argentina	Unknown	c	a
Buck Topacio	Argentina	Unknown	a*	a*
920334	Australia	69850/ 86014	b	b
940030	Australia	Unknown	b	a*
940435	Australia	Unknown	a*	a*
950329	Australia	Unknown	a	a
950844	Australia	Unknown	b*	a
940955	Australia	Unknown	b*	b*
Tamaroi	Australia	RUFF/FLAMINGO-DW//MEXICALI-75///SHEARWATER/56113/TAM-1-B-17/KAMILAROI/56112/WELLS/56111//GUILLEMOT	a*	a*
Wollaroi	Australia	TAM-1-B-17/(SIB)KAMILAROI//ROKEL(S)/(SIB)KAMILAROI	b*	b*
AC Morse	Canada	RL 7196/DT 610	a	a
AC Napoleon	Canada	VIC/DT384//DT 471	a*	a*
9661-AF1D	Canada	W9262-260D3/ARUBA//DT 662	a*	a*
9661-CA5E	Canada	W9262-260D3/ARUBA//DT 662	a	a
AC Avonlea	Canada	8267-AD2A/DT 61	a*	a*
AC Melita	Canada	MEDORA/LLOYD	a*	a*
AC Navigator	Canada	KYLE/WESTBRED 881	a*	b*
AC Pathfinder	Canada	WESTBRED 881/DT 367	b*	a*
DT691	Canada	DT618/ 8667-D216C//DT 637	a	a*
DT695	Canada	DT 471/2*KYLE	a*	a*
DT696	Canada	DT618/DT 637//KYLE	a*	a*
Kyle	Canada	6962-92-8-5/ 6965-494-	a*	a

Accession	Origin	Pedigree	*Psy1-A1*	*Psy1-B1*
Commander	Canada	W9260-BK03/AC NAVIGATOR//AC PATHFINDER	b*	b*
DT704	Canada	AC AVONLEA/DT 665	a*	a*
DT705	Canada	AC AVONLEA/DT 665	a	a*
DT707	Canada	AC AVONLEA/DT 665	a*	a*
DT709	Canada	DT 674/DT 665	a*	b*
DT710	Canada	DT618/GREEN 27	a*	a*
DT711	Canada	WESTBRED 881/W9260-BK03	a*	b*
Strongfield	Canada	AC AVONLEA/DT 665	a*	a*
D24-1773	Canada	DT 520/D94078	a*	a*
DT513	Canada	DT 625/DT 612	b*	a*
DT532	Canada	D92269/D92413	a*	a*
DT536	Canada	D94350/D93108	a*	b
DT540	Canada	D95253/D95116	a*	a*
Carioca	France	CID 479402	a	a
RABD 93.40	France	Unknown	a*	a
Tetradur	France	EDMORE//CAPDUR/REGAL	a	a
Durabon	Germany	SIGNADUR/EDM//P 4312.86	a*	a*
Durafit	Germany	Unknown	a*	a*
44616	Iran	Unknown	a	a
44721	Iran	Unknown	c	a
CRDW17	Iran	Unknown	a	a
D-73-15	Iran	Unknown	c	a
Simeto	Italy	CAPEITI/VALNOVA	a*	b
Colosseo	Italy	CRESO/MEXA	a	a
Duilio	Italy	CAPPELLI//ANHINGA/FLAMINGO	a*	b*
Grazia	Italy	ISWRN-21/VALSELVA	a	a
Fortore	Italy	CAPEITI 8/VALFORTE	a	b
Lesina	Italy	Unknown	a	a
Varano	Italy	CAPEITI 8/CRESO//CRESO///VALFORTE/TRINAKRIA	a	b
Bronte	Italy	BERILLO/LATINO	c	a
Ciccio	Italy	APPULO/VALNOVA//VALFORTE/PATRIZIO	a	a
Demetra	Italy	MESSAPIA/GIOIA	a	a

Accession	Origin	Pedigree	*Psy1-A1*	*Psy1-B1*
Gianna	Italy	Unknown	c	a
Iride	Italy	ALTAR 84/ARES-SIB	a	a
Medda	Italy	TRINAKRIA/VALFORTE	a	a
Mongibello	Italy	TRINAKRIA/VALFORTE	a	a
Parsifal	Italy	INRA92-1/D81028	a	a
Svevo	Italy	SELEZIONE CIMMYT/ZENIT-SIB	a	a
Tresor	Italy	AMBER-DURUM/S-22-80	a	a
Green 27	Mexico	STERNA-DW 2/GRAVELOTE	a	a
Green 34	Mexico	STERNA-DW 2/GRAVELOTE	a	a
Nacori 97	Mexico	ALTAR 84/CMH82A.1062//CD58230-?	a	a
DHTON 1	Morrocco	Unknown	a	a
Gidara 17a	Morrocco	Unknown	a	a*
Marjak	Morrocco	Unknown	c*	a*
Arrivato	New Zealand	Unknown	a	a
CFR5001	New Zealand	Unknown	a*	a*
K-39099	Russia	LV-URAZOVSKII R-N,VORONEZHSKAYA OBL	a	a
Agridur	Spain	EDMORE//CIMMYT 303/CHANDUR	a	a
Altar-Aos	Spain	Unknown	a	a
Arcobelano	Spain	CHEN/ALTAR 84	a	a
Ariesol	Spain	Unknown	a	a
Borli	Spain	Unknown	a	a
Camacho	Spain	Unknown	a	a
Gallareta	Spain	RUFF/FLAMINGO-DW//MEXICALI-75/3/SHEARWATER/4/?	a	a
Mexa	Spain	GDOVZ469///JO 1//61.130/LDS	a	b
Vitron	Spain	TURCHIA-77///JORI-SIB/ANHINGA-SIB//FLAMINGO-SIB	c	a
Ocotillo	U.S.	Unknown	a	a
D940027	U.S.	D88104/D88207	a	a*
D940098	U.S.	D88450/D87436	a	a
D941038	U.S.	D86117/D88289	a	a
D95580	U.S.	BELZER/D88058//D88276	b	a
Plaza	U.S.	PLENTY/D8291	a	a
Durex	U.S.	AZ-MFSRS-86	a	b

Accession	Origin	Pedigree	*Psy1-A1*	*Psy1-B1*
Langdon	U.S.	LDN240/KHAPLI//LANGDON 308///MINDUM*3/VERNAL/4/VERNAL EMMER/3*MINDUM	a	a
Westbred 881	U.S.	WARD/WLS//CNDO/WCA///MEXI/WB1000	a*	b*
Kofa	U.S.	DICOCCUM ALPHA	a*	b*
Kronos	U.S.	APB MSFRS POP SEL (D03-21)	a	b

*Indicates accessions from which PCR product of approximately 1100bp was cloned and sequenced using primers Psy1F5 and Psy1R5.

Appendix 2. List of microsatellite markers used for genotyping of association mapping population. Markers used for determination of population structure are indicated by an (*).

Marker	Chromosome	Distance
gdm33*	1A	6
cfa2226	1A	24
cfd15	1A	25
wmc818	1A	29
wmc336	1A	35
barc83	1A	48
wmc24	1A	48
gwm357	1A	52
gwm164	1A	56
barc148	1A	57
cfd59	1A	57
gwm135	1A	61
wmc278	1A	62
wmc469	1A	64
wmc183	1A	65
wmc312	1A	69
cfa2129	1A	71
gwm497*	1A	86
wmc716	1A	91
wmc59	1A	98
barc158	1A	114
barc17	1A	115
cfa2219	1A	124
gwm264	1B	21
barc8*	1B	25
gwm413	1B	26
wmc128	1B	30
gwm498	1B	31
wmc419	1B	32
cfd65	1B	34
gwm11	1B	34
barc137	1B	34
wmc626	1B	35
wmc216	1B	35
wmc694	1B	37
barc181	1B	38
cfd48	1B	40
wmc416	1B	44
wmc134	1B	47
gwm274	1B	61
gwm124	1B	64
cfa2147.2	1B	75
wmc830	1B	90
wmc44*	1B	92

Marker	Chromosome	Distance
wmc367	1B	103
barc80	1B	106
barc124*	2A	8
wmc667	2A	10
wmc407	2A	15
wmc382	2A	16
wmc177	2A	28
wmc522	2A	45
wmc296	2A	49
gwm425	2A	52
gwm95	2A	53
gwm558	2A	54
gwm372	2A	60
barc5	2A	63
gwm312	2A	74
gwm294	2A	76
cfd168	2A	85
wmc181	2A	103
gwm356	2A	126
wmc658*	2A	140
wmc764*	2B	1
wmc382	2B	8
wmc25	2B	25
wmc154	2B	29
gwm257	2B	37
gwm429	2B	40
gwm148	2B	47
barc13	2B	50
barc18	2B	60
barc167	2B	61
wmc477	2B	63
gwm271	2B	65
gwm388	2B	72
barc101	2B	76
cfd73	2B	82
wmc332*	2B	93
wmc149	2B	100
barc159	2B	109
wmc11	3A	0
wmc532	3A	6
gwm369*	3A	14
barc45	3A	37
gwm5	3A	45
barc67	3A	47
wmc264	3A	61
cfa2262	3A	64
gwm497	3A	73
cfa2193	3A	74

Marker	Chromosome	Distance
wmc559	3A	83
wmc153	3A	87
cfa2076	3A	98
wmc594*	3A	105
barc75	3B	0
gwm389	3B	1
barc147*	3B	7
gwm493	3B	12
barc87	3B	14
wmc808	3B	39
wmc43	3B	49
wmc231	3B	56
wmc777	3B	58
wmc625	3B	59
wmc307	3B	65
wmc471	3B	69
barc164	3B	70
wmc418*	3B	72
gwm108	3B	94
barc84	3B	97
wmc687	3B	105
barc77	3B	111
gwm299	3B	123
wmc632	3B	143
gwm340	3B	148
wmc680	4A	8
wmc491*	4A	8
wmc48	4A	8
cfd71	4A	8
wmc96	4A	11
gwm610	4A	12
wmc650	4A	25
barc170	4A	27
wmc468	4A	38
wmc707	4A	40
gwm494	4A	49
wmc283	4A	59
wmc698	4A	62
wmc232	4A	68
gwm160*	4A	79
wmc313	4A	83
wmc219	4A	88
wmc125	4B	0
wmc47*	4B	10
wmc413	4B	17
wmc349	4B	19
gwm251	4B	25
gwm149	4B	28

Marker	Chromosome	Distance
gwm192	4B	30
Gwm495	4B	31
wmc238	4B	34
wmc89	4B	36
wmc48	4B	36
gwm368	4B	37
barc20	4B	38
wmc710*	4B	48
gwm443*	5A	23
wmc713	5A	28
cfa2104	5A	37
cfa2190	5A	45
gwm293	5A	52
cfa2250	5A	55
wmc805	5A	57
gwm304	5A	59
barc165	5A	63
gwm156	5A	73
wmc415	5A	80
barc151	5A	89
wmc445	5A	101
cfa2163	5A	104
cfa2141	5A	107
barc232	5A	111
cfa2185	5A	114
wmc110	5A	128
gwm126	5A	138
gwm595	5A	149
wmc524	5A	151
wmc727	5A	154
gwm291*	5A	163
cfd5	5B	0
cfd60	5B	14
cfd20	5B	38
wmc740*	5B	56
barc4	5B	58
wmc376	5B	60
wmc73	5B	63
gwm335	5B	68
gwm371	5B	73
wmc415	5B	80
wmc537	5B	84
gwm554	5B	89
cfd7	5B	94
wmc289	5B	112
gwm408	5B	117
gwm604	5B	124
wmc99	5B	128

Marker	Chromosome	Distance
wmc160*	5B	133
cfd86	5B	139
wmc508	5B	141
barc59	5B	154
gwm497	5B	164
gwm334*	6A	2
barc146	6A	37
wmc256	6A	40
barc3	6A	44
wmc201	6A	46
wmc553	6A	52
wmc417	6A	86
gwm617*	6A	95
wmc254	6A	148
wmc487*	6B	9
gwm705	6B	14
cfd13	6B	17
gwm518	6B	27
wmc494	6B	29
gwm193	6B	36
gwm311	6B	45
barc24*	6B	55
barc14	6B	
gwm79	6B	
gwm666	7A	0
gwm635	7A	8
gwm471*	7A	17
wmc593	7A	34
cfa2049	7A	37
wmc283	7A	40
barc127	7A	47
cfa2028	7A	48
wmc83	7A	55
barc174	7A	65
barc108	7A	71
wmc9	7A	72
wmc603	7A	73
barc121	7A	76
gwm276	7A	83
cfa2257	7A	92
cfd20	7A	92
wmc790	7A	102
cfa2019	7A	107
gwm554	7A	108
cfa2040*	7A	119
wmc809	7A	131
Psy1-A1	7A	
wmc606	7B	0

Marker	Chromosome	Distance
gwm537*	7B	41
gwm400	7B	47
wmc758	7B	57
wmc475	7B	60
wmc696	7B	60
wmc471	7B	61
gwm333	7B	65
wmc396	7B	68
wmc76	7B	78
wmc517	7B	93
wmc311*	7B	118
cfa2040	7B	143
Psy1-B1	7B	146
wmc10	7B	147
gwm146	7B	150

Appendix 3. Least square means of CIE b* analysis for endosperm colour of 93 accessions of the association mapping population grown at Swift Current and Saskatoon, 2005 and 2006.

	2005		2006		2005-2006
Accession	SC	SK	SC	SK	Combined
Bonaerance Inta Cumenay	17.0	15.9	15.9	15.6	16.1
Bonaerance Quilaco	17.5	16.3	16.0	15.7	16.4
Bonaerance Valverde	18.0	16.0	17.0	15.8	16.7
Buck Ambar	16.9	15.2	16.7	15.7	16.1
Buck Topacio	20.8	17.4	19.5	17.6	18.8
Ocotillo	18.4	16.6	16.8	14.9	16.6
920334	22.0	19.4	20.0	19.8	20.3
940030	21.5	18.5	20.4	19.7	20.0
940435	21.8	19.2	20.7	19.9	20.4
950329	20.8	17.9	18.7	18.1	18.9
950844	20.5	18.0	19.4	18.7	19.2
940955	22.2	18.7	20.5	20.0	20.3
Tamaroi	19.3	17.3	18.1	17.0	17.9
Wollaroi	20.4	18.6	19.3	19.4	19.4
AC Morse	19.7	17.3	18.6	17.6	18.3
AC Napoleon	20.0	17.3	19.5	18.0	18.7
9661-AF1D	19.4	16.7	18.5	17.2	17.9
9661-CA5E	19.5	16.9	18.4	17.5	18.1
AC Avonlea	20.0	17.1	19.4	17.3	18.4
AC Melita	19.8	17.2	18.2	16.9	18.0
AC Navigator	20.2	17.4	19.2	17.8	18.6
AC Pathfinder	19.6	17.4	18.8	17.7	18.4
DT691	19.9	17.4	18.8	17.0	18.3
DT695	18.6	16.5	17.9	17.0	17.5
DT696	20.4	17.4	18.2	17.7	18.4
Kyle	19.3	16.5	17.8	16.5	17.5
Commander	20.5	18.4	19.4	18.6	19.2
DT704	19.4	17.1	18.4	17.6	18.1
DT705	20.0	17.6	18.3	17.6	18.4
DT707	19.3	17.1	18.1	17.7	18.1
DT709	20.6	18.3	20.3	18.6	19.4
DT710	19.7	17.6	18.5	17.3	18.3
DT711	20.1	17.4	17.9	17.2	18.1
Strongfield	19.9	17.5	18.3	17.8	18.4
D24-1773	19.9	17.3	18.9	18.0	18.5
DT513	19.2	17.2	18.2	17.4	18.0
DT532	20.5	18.6	18.3	18.6	19.0
DT536	19.4	16.9	18.1	18.1	18.2
DT540	20.9	18.2	19.5	18.4	19.3
Carioca	17.6	16.0	16.2	16.0	16.4
RABD 93.40	19.7	17.6	18.6	17.3	18.3
Tetradur	19.8	17.3	17.7	17.1	18.0
Durabon	20.1	17.6	18.1	17.3	18.3
Durafit	18.3	16.4	16.9	16.7	17.1

	2005		2006		2005-2006
Accession	SC	SK	SC	SK	Combined
44616	17.4	17.1	16.0	16.1	16.6
44721	18.0	15.8	16.3	15.3	16.4
CRDW17	19.2	17.7	17.9	17.8	18.2
D-73-15	17.6	16.1	16.1	15.8	16.4
Simeto	18.2	16.4	17.8	16.9	17.3
Colosseo	16.9	16.2	16.3	15.2	16.2
Duilio	18.4	16.3	17.4	16.6	17.2
Grazia	18.4	16.2	17.0	16.2	17.0
Fortore	-	17.1	17.6	16.4	17.4
Lesina	17.7	15.9	16.9	16.3	16.7
Varano	17.1	16.0	16.3	16.4	16.4
Bronte	17.2	16.0	16.3	16.0	16.3
Ciccio	18.2	16.5	17.9	16.8	17.4
Demetra	15.9	14.6	15.3	14.6	15.1
Gianna	16.2	14.9	15.4	14.9	15.4
Iride	18.1	17.0	17.3	16.8	17.3
Medda	17.8	16.7	16.6	15.9	16.7
Mongibello	18.4	17.3	17.9	16.4	17.5
Parsifal	16.9	15.7	16.0	15.8	16.1
Svevo	20.1	17.7	18.0	18.4	18.5
Tresor	17.7	16.2	16.7	16.7	16.9
Green 27	18.8	16.8	16.5	15.5	16.9
Green 34	18.3	16.2	17.1	16.7	17.1
Nacori 97	18.7	16.1	17.5	16.2	17.1
DHTON 1	19.3	16.9	17.9	16.7	17.7
Gidara 17a	15.9	15.7	15.0	14.8	15.4
Marjak	15.5	14.9	14.0	13.8	14.6
D940027	18.8	16.3	17.6	16.6	17.4
D940098	18.6	16.8	18.2	16.7	17.5
D941038	20.4	17.6	19.5	18.5	19.0
D95580	18.8	17.1	18.1	17.5	17.9
Plaza	20.0	17.0	18.7	17.8	18.4
Arrivato	17.6	15.7	16.5	15.3	16.3
K-39099	17.5	15.2	16.4	15.7	16.2
Agridur	18.4	16.3	17.2	16.1	17.0
Altar-Aos	19.6	17.2	18.3	17.9	18.2
Arcobelano	18.9	17.0	18.0	16.4	17.5
Ariesol	18.5	16.5	17.4	16.7	17.3
Borli	18.4	16.7	16.9	16.6	17.2
Camacho	17.2	15.3	15.7	14.9	15.8
Gallareta	19.1	17.3	17.7	16.9	17.7
Mexa	18.5	16.2	17.5	16.7	17.2
Vitron	17.7	16.0	16.1	16.0	16.5
Durex	19.1	16.7	18.1	17.4	17.8
Langdon	17.3	15.5	16.4	15.5	16.2
Westbred881	19.0	17.1	18.5	17.1	17.9
Kofa	20.2	16.9	19.4	17.5	18.5
Kronos	19.8	17.4	18.6	18.5	18.6

	2005		2006		2005-2006
Accession	SC	SK	SC	SK	Combined
CFR5001	18.7	17.2	17.4	17.1	17.6
Ave	18.9	16.9	17.8	17.0	17.6
Min	15.5	14.6	14.0	13.8	14.6
Max	22.2	19.4	20.7	20.0	20.4
Ave $LSD_{0.05}$	0.51	0.47	0.80	0.61	0.71
Heritability	0.98	0.97	0.95	0.97	0.95

Appendix 4. Least square means for kernel weight (g) of association mapping population field trials conducted at Swift Current and Saskatoon, 2005 and 2006. Kernel weight was measured on a plot basis using a sample weight of 200 kernels multiplied by 5.

	Kernel Weight (g)				
	2005		2006		2005-2006
Line	SC	SK	SC	SK	Combined
Bonaerance Inta Cumenay	48.2	47.1	43.3	51.5	47.8
Bonaerance Quilaco	48.1	48.8	44.7	50.3	48.1
Bonaerance Valverde	43.2	42.6	43.1	55.0	45.9
Buck Ambar	44.5	48.2	37.2	49.4	44.9
Buck Topacio	41.8	42.7	37.1	46.5	42.1
Ocotillo	41.2	41.8	38.3	40.9	40.6
920334	39.6	42.0	35.6	46.9	41.0
940030	43.6	40.6	36.2	44.5	41.2
940435	38.2	39.3	37.4	43.8	39.7
950329	46.4	47.7	42.5	49.7	46.6
950844	44.8	51.2	40.3	52.3	47.3
940955	45.5	41.6	37.1	43.8	41.9
Tamaroi	42.4	38.2	39.1	47.4	41.8
Wollaroi	40.7	38.9	35.8	37.8	38.0
AC Morse	43.9	41.7	37.7	45.5	42.1
AC Napoleon	42.8	45.8	37.8	47.7	43.4
9661-AF1D	43.7	45.2	37.9	49.1	44.1
9661-CA5E	42.6	42.4	35.0	47.7	42.1
AC Avonlea	44.2	41.3	38.2	44.8	42.0
AC Melita	39.4	42.1	39.0	46.4	41.8
AC Navigator	45.0	44.5	37.2	48.0	43.9
AC Pathfinder	42.9	43.0	37.3	42.9	41.6
DT691	41.5	46.5	39.0	44.8	43.0
DT695	42.0	45.2	38.9	47.7	43.5
DT696	43.7	43.3	40.3	47.0	43.6
Kyle	36.9	41.4	35.8	49.0	40.8
Commander	45.5	43.6	41.1	47.4	44.3
DT704	40.2	42.6	33.8	45.2	40.5
DT705	41.9	37.9	35.4	46.8	40.6
DT707	40.7	40.0	33.4	42.3	39.0
DT709	43.9	42.1	35.0	47.2	42.0
DT710	40.6	42.1	35.7	47.8	41.6
DT711	45.3	46.7	41.2	49.1	45.8
Strongfield	41.6	42.3	35.8	44.4	40.9
D24-1773	42.0	44.3	35.5	47.0	41.7
DT513	41.2	40.8	34.9	42.4	39.9
DT532	42.7	38.8	37.1	45.5	41.1
DT536	40.5	43.0	37.4	45.2	41.1
DT540	41.2	40.6	37.6	44.9	41.0
Carioca	41.9	43.8	38.7	46.7	42.6
RABD 93.40	43.1	38.9	39.9	48.9	42.8
Tetradur	34.1	38.5	37.2	44.5	38.6

	Kernel Weight (g)				
	2005		2006		2005-2006
Line	SC	SK	SC	SK	Combined
Durabon	41.1	45.1	37.0	49.7	43.0
Durafit	40.7	40.6	36.8	44.9	40.9
44616	43.6	36.7	37.7	45.5	41.0
44721	41.0	42.6	37.9	46.8	41.8
CRDW17	38.9	43.6	36.8	48.5	42.0
D-73-15	44.2	43.2	39.5	46.6	42.9
Simeto	46.8	46.9	40.4	48.5	45.9
Colosseo	44.3	41.2	39.2	45.6	42.6
Duilio	44.4	48.5	44.1	52.5	47.2
Grazia	44.9	41.4	41.3	47.1	43.8
Fortore	-	41.9	37.8	50.7	43.5
Lesina	47.9	46.4	38.4	46.6	44.7
Varano	42.5	39.6	35.7	39.9	39.5
Bronte	47.6	46.4	39.6	47.0	45.2
Ciccio	45.0	41.0	41.0	47.8	43.8
Demetra	40.9	43.0	36.4	45.2	41.3
Gianna	40.5	45.3	37.7	46.3	42.5
Iride	39.4	37.3	34.0	43.3	38.5
Medda	43.8	36.8	40.4	47.6	42.2
Mongibello	45.0	41.6	41.8	50.5	44.8
Parsifal	40.0	39.0	36.7	42.9	39.6
Svevo	42.9	43.1	36.6	42.3	41.0
Tresor	39.6	37.6	36.4	40.6	38.4
Green 27	43.2	45.5	42.1	48.1	44.7
Green 34	35.7	44.3	37.8	46.5	41.2
Nacori 97	36.8	41.1	35.6	46.3	39.9
DHTON 1	41.8	43.6	39.5	44.9	42.5
Gidara 17a	43.7	37.0	39.4	43.9	41.0
Marjak	42.8	42.3	40.9	49.6	44.0
D940027	42.3	44.7	40.3	46.1	43.3
D940098	39.8	40.6	36.5	46.1	40.5
D941038	38.3	42.4	34.0	44.8	39.8
D95580	42.4	44.9	36.4	46.7	42.4
Plaza	35.2	41.6	33.1	44.7	38.8
Arrivato	38.0	39.3	35.2	49.0	40.4
K-39099	35.1	43.1	32.7	44.1	38.8
Agridur	46.6	44.2	42.8	48.9	45.8
Altar-Aos	36.4	33.2	35.2	39.6	36.2
Arcobelano	33.5	35.3	30.8	42.4	35.5
Ariesol	40.9	39.3	37.1	48.0	41.3
Borli	30.4	35.6	35.3	42.7	35.9
Camacho	43.0	46.6	40.7	51.9	45.7
Gallareta	31.6	32.6	30.9	39.1	33.7
Mexa	42.0	46.4	38.1	44.2	42.5
Vitron	40.9	39.9	36.7	44.1	40.2
Durex	41.0	39.7	36.0	41.8	39.5
Langdon	35.7	40.4	31.6	43.7	37.8

	Kernel Weight (g)				
	2005		2006		2005-2006
Line	SC	SK	SC	SK	Combined
Westbred881	41.7	41.1	39.5	43.1	41.4
Kofa	42.7	41.0	39.2	47.0	42.4
Kronos	40.1	43.7	38.0	43.1	41.3
CFR5001	30.4	33.3	32.2	39.0	33.7
Ave	41.5	42.0	37.7	46.0	41.8
Min	30.4	32.6	30.8	37.8	33.7
Max	48.2	51.2	44.7	55.0	48.1
Ave $LSD_{0.05}$	3.5	2.9	3.5	3.1	3.8
Heritability	0.90	0.91	0.83	0.90	0.76

Appendix 5. Q matrix as assigned by STRUCTURE for K=5 subpopulations based on an average of 3 independent runs of Structure using 28 unlinked microsatellite markers (one marker per chromosome arm; Appendix 2).

		Population					Population
Accession	Origin	1	2	3	4	5	Assignment
AC Napoleon	Canada	0.723	0.212	0.029	0.012	0.024	1
9661-AF1D	Canada	0.803	0.059	0.027	0.018	0.093	1
9661-CA5E	Canada	0.909	0.023	0.025	0.008	0.034	1
AC Avonlea	Canada	0.965	0.019	0.009	0.004	0.003	1
AC Navigator	Canada	0.968	0.009	0.008	0.005	0.009	1
AC Pathfinder	Canada	0.925	0.019	0.007	0.006	0.043	1
Kyle	Canada	0.968	0.012	0.006	0.008	0.007	1
Commander	Canada	0.932	0.010	0.008	0.005	0.044	1
DT704	Canada	0.948	0.009	0.008	0.031	0.004	1
DT705	Canada	0.977	0.006	0.007	0.007	0.003	1
DT707	Canada	0.982	0.006	0.005	0.004	0.003	1
DT709	Canada	0.966	0.008	0.007	0.012	0.008	1
DT711	Canada	0.857	0.089	0.036	0.012	0.006	1
Strongfield	Canada	0.974	0.009	0.006	0.007	0.005	1
DT513	Canada	0.663	0.312	0.011	0.006	0.007	1
DT532	Canada	0.955	0.021	0.011	0.007	0.005	1
DT540	Canada	0.700	0.245	0.035	0.005	0.016	1
Durex	U.S.	0.785	0.081	0.032	0.053	0.048	1
Ocotillo	U.S.	0.794	0.056	0.027	0.008	0.116	1
Westbred881	U.S.	0.973	0.008	0.008	0.005	0.006	1
Kofa	U.S.	0.953	0.021	0.013	0.006	0.008	1
AC Morse	Canada	0.142	0.831	0.013	0.006	0.006	2
AC Melita	Canada	0.401	0.563	0.015	0.011	0.010	2
DT695	Canada	0.150	0.795	0.037	0.005	0.013	2
Buck Topacio	Argentina	0.012	0.710	0.027	0.124	0.127	2
D24-1773	Canada	0.283	0.677	0.015	0.015	0.010	2
DT536	Canada	0.277	0.616	0.028	0.019	0.060	2
Tetradur	France	0.032	0.758	0.108	0.021	0.082	2
Wollaroi	Australia	0.067	0.666	0.232	0.009	0.025	2
Durabon	Germany	0.023	0.913	0.038	0.006	0.019	2
Durafit	Germany	0.028	0.929	0.020	0.004	0.019	2
D940027	U.S.	0.011	0.945	0.015	0.007	0.022	2
D940098	U.S.	0.010	0.850	0.040	0.075	0.025	2
D941038	U.S.	0.017	0.896	0.052	0.015	0.020	2
D95580	U.S.	0.023	0.896	0.041	0.011	0.029	2
Plaza	U.S.	0.015	0.959	0.009	0.012	0.006	2
Langdon	U.S.	0.036	0.938	0.013	0.004	0.008	2
K-39099	Russia	0.016	0.441	0.331	0.013	0.200	2
Agridur	Spain	0.019	0.763	0.085	0.038	0.094	2
Ariesol	Spain	0.025	0.915	0.028	0.009	0.023	2
DT710	Canada	0.014	0.022	0.943	0.011	0.009	3
Bonaerance Quilaco	Argentina	0.123	0.021	0.829	0.014	0.013	3
920334	Australia	0.012	0.094	0.843	0.034	0.017	3
940030	Australia	0.011	0.016	0.963	0.005	0.004	3

		Population					Population
Accession	Origin	1	2	3	4	5	Assignment
940435	Australia	0.027	0.022	0.942	0.005	0.005	3
950329	Australia	0.016	0.024	0.907	0.045	0.008	3
950844	Australia	0.006	0.012	0.954	0.018	0.009	3
940955	Australia	0.009	0.014	0.967	0.007	0.004	3
Green 27	Mexico	0.007	0.011	0.969	0.007	0.006	3
Green 34	Mexico	0.010	0.011	0.944	0.029	0.005	3
Nacori 97	Mexico	0.008	0.025	0.702	0.209	0.055	3
DHTON 1	Morrocco	0.009	0.043	0.491	0.383	0.075	3
Gidara 17a	Morrocco	0.005	0.022	0.804	0.100	0.069	3
Marjak	Morrocco	0.007	0.026	0.894	0.006	0.066	3
D-73-15	Iran	0.018	0.115	0.682	0.060	0.125	3
Demetra	Italy	0.005	0.017	0.724	0.005	0.249	3
Parsifal	Italy	0.016	0.141	0.696	0.092	0.056	3
Svevo	Italy	0.025	0.087	0.755	0.116	0.016	3
Tresor	Italy	0.096	0.076	0.786	0.012	0.030	3
Buck Ambar	Argentina	0.014	0.107	0.208	0.506	0.164	4
Simeto	Italy	0.004	0.007	0.011	0.971	0.007	4
Duilio	Italy	0.004	0.004	0.006	0.982	0.004	4
Gianna	Italy	0.004	0.005	0.025	0.961	0.005	4
Iride	Italy	0.004	0.004	0.004	0.984	0.005	4
Altar-Aos	Spain	0.004	0.004	0.004	0.984	0.004	4
Arcobelano	Spain	0.005	0.006	0.006	0.977	0.006	4
Borli	Spain	0.005	0.006	0.009	0.971	0.009	4
Gallareta	Spain	0.009	0.010	0.005	0.966	0.010	4
Vitron	Spain	0.003	0.004	0.005	0.984	0.003	4
CFR5001	New Zealand	0.008	0.014	0.007	0.960	0.009	4
Bonae Inta Cumenay	Argentina	0.003	0.008	0.014	0.028	0.947	5
Bonaerance Valverde	Argentina	0.014	0.166	0.010	0.016	0.794	5
RABD 93.40	France	0.015	0.384	0.045	0.011	0.545	5
Colosseo	Italy	0.005	0.217	0.162	0.020	0.597	5
Grazia	Italy	0.030	0.188	0.036	0.011	0.735	5
Fortore	Italy	0.006	0.008	0.008	0.010	0.969	5
Lesina	Italy	0.007	0.009	0.007	0.004	0.973	5
Varano	Italy	0.007	0.007	0.006	0.004	0.976	5
Kronos	U.S.	0.164	0.092	0.247	0.070	0.426	5
Arrivato	New Zealand	0.011	0.021	0.080	0.167	0.721	5
44616	Iran	0.027	0.027	0.011	0.031	0.904	5
44721	Iran	0.020	0.293	0.232	0.054	0.402	5
Ciccio	Italy	0.004	0.009	0.004	0.004	0.979	5
CRDW17	Iran	0.010	0.029	0.411	0.017	0.533	5
Medda	Italy	0.007	0.021	0.013	0.007	0.952	5
Mongibello	Italy	0.007	0.006	0.005	0.006	0.976	5
Camacho	Spain	0.007	0.017	0.009	0.005	0.961	5
Mexa	Spain	0.029	0.228	0.143	0.099	0.501	5
DT696	Canada	0.488	0.457	0.030	0.007	0.017	1/2
DT691	Canada	0.473	0.503	0.014	0.004	0.006	2/1
Tamaroi	Australia	0.030	0.061	0.446	0.448	0.015	4/3
Carioca	France	0.018	0.035	0.447	0.457	0.043	4/3

		Population					Population
Accession	Origin	1	2	3	4	5	Assignment
Bronte	Italy	0.007	0.010	0.104	0.434	0.446	5/4

Appendix 6. Significant marker-YP trait associations identified using TASSEL. Markers significantly ($P \leq 0.05$) associated with YP are presented including chromosome location, degrees of freedom (df), significance level (p-permutations marker) and r^2 of the marker.

Environment	Locus	Chromosome	Distance (cM)	df Marker	p-permutations Marker	r² Marker
2006SC_WSB	wmc818	1A	28.9	5	2.00E-04	0.1923
20052006Combined_WSB	wmc818	1A	28.9	5	6.00E-04	0.1792
2006SK_WSB	wmc818	1A	28.9	5	0.0012	0.1954
2005SC_WSB	wmc818	1A	28.9	5	0.0021	0.1453
2005SK_WSB	wmc818	1A	28.9	5	0.006	0.1426
20052006Combined_WSB	barc83	1A	47.6	3	1.00E-04	0.16
2006SC_WSB	barc83	1A	47.6	3	2.00E-04	0.1524
2006SK_WSB	barc83	1A	47.6	3	3.00E-04	0.1755
2005SK_WSB	barc83	1A	47.6	3	4.00E-04	0.1553
2005SC_WSB	barc83	1A	47.6	3	9.00E-04	0.1366
2005SK_WSB	gwm135	1A	60.7	2	2.00E-04	0.1607
2006SC_WSB	gwm135	1A	60.7	2	3.00E-04	0.1339
2006SK_WSB	gwm135	1A	60.7	2	3.00E-04	0.1459
20052006Combined_WSB	gwm135	1A	60.7	2	4.00E-04	0.1449
2005SC_WSB	gwm135	1A	60.7	2	8.00E-04	0.1109
2005SK_WSB	wmc312	1A	68.5	1	9.00E-04	0.1024
20052006Combined_WSB	wmc312	1A	68.5	1	0.0034	0.0729
2006SK_WSB	wmc312	1A	68.5	1	0.006	0.0728
2006SC_WSB	wmc312	1A	68.5	1	0.0062	0.0601
2005SC_WSB	wmc312	1A	68.5	1	0.0086	0.0537
2006SC_WSB	cfa2129	1A	70.8	4	0.0035	0.1181
20052006Combined_WSB	cfa2129	1A	70.8	4	0.0213	0.096
2005SK_WSB	cfa2129	1A	70.8	4	0.0265	0.0977
2005SC_WSB	cfa2129	1A	70.8	4	0.035	0.0818
2006SC_WSB	wmc716	1A	91	3	0.0381	0.0741
2006SK_WSB	wmc716	1A	91	3	0.0487	0.0789
2006SK_WSB	barc158	1A	113.8	1	0.0068	0.0679
2006SC_WSB	barc158	1A	113.8	1	0.0075	0.0602
20052006Combined_WSB	barc158	1A	113.8	1	0.0083	0.0608
2005SC_WSB	barc158	1A	113.8	1	0.0118	0.0509
2005SK_WSB	barc158	1A	113.8	1	0.0184	0.0537
2005SC_WSB	barc17	1A	114.8	1	0.0044	0.0625
2006SC_WSB	barc17	1A	114.8	1	0.0048	0.0681
20052006Combined_WSB	barc17	1A	114.8	1	0.0049	0.0713
2006SK_WSB	barc17	1A	114.8	1	0.0054	0.079
2005SK_WSB	barc17	1A	114.8	1	0.0105	0.0636
2005SK_WSB	barc8	1B	24.9	3	0.0067	0.1047
2006SC_WSB	barc8	1B	24.9	3	0.0182	0.0779
20052006Combined_WSB	barc8	1B	24.9	3	0.0196	0.0824
2006SK_WSB	barc8	1B	24.9	3	0.0239	0.0876
2005SC_WSB	wmc419	1B	31.8	2	0.0368	0.0551
20052006Combined_WSB	wmc419	1B	31.8	2	0.04	0.0555
2006SC_WSB	wmc419	1B	31.8	2	0.0415	0.0528

Environment	Locus	Chromosome	Distance (cM)	df Marker	p-permutations Marker	r^2 Marker
2005SC_WSB	barc137	1B	34.3	5	1.00E-04	0.3014
2005SK_WSB	barc137	1B	34.3	5	1.00E-04	0.3129
2006SC_WSB	barc137	1B	34.3	5	1.00E-04	0.3066
2006SK_WSB	barc137	1B	34.3	5	1.00E-04	0.3648
20052006Combined_WSB	barc137	1B	34.3	5	1.00E-04	0.3332
2005SC_WSB	wmc626	1B	34.8	4	1.00E-04	0.3018
2005SK_WSB	wmc626	1B	34.8	4	1.00E-04	0.3103
2006SC_WSB	wmc626	1B	34.8	4	1.00E-04	0.2723
2006SK_WSB	wmc626	1B	34.8	4	1.00E-04	0.3342
20052006Combined_WSB	wmc626	1B	34.8	4	1.00E-04	0.3151
2005SC_WSB	cfd48	1B	39.8	3	0.0051	0.0992
2006SK_WSB	cfd48	1B	39.8	3	0.008	0.1099
20052006Combined_WSB	cfd48	1B	39.8	3	0.0113	0.0943
2006SC_WSB	cfd48	1B	39.8	3	0.025	0.0765
2005SK_WSB	cfd48	1B	39.8	3	0.039	0.0795
2005SC_WSB	wmc44	1B	91.5	3	1.00E-03	0.1172
20052006Combined_WSB	wmc44	1B	91.5	3	0.0029	0.1134
2006SK_WSB	wmc44	1B	91.5	3	0.0042	0.1258
2005SK_WSB	wmc44	1B	91.5	3	0.0049	0.1152
2006SC_WSB	wmc44	1B	91.5	3	0.0122	0.0855
2005SK_WSB	wmc407	2A	14.5	6	0.0447	0.1158
2006SC_WSB	wmc177	2A	28.3	3	0.0063	0.0991
20052006Combined_WSB	wmc177	2A	28.3	3	0.0143	0.0904
2006SK_WSB	wmc177	2A	28.3	3	0.0168	0.0963
2005SK_WSB	wmc177	2A	28.3	3	0.0233	0.0871
2005SC_WSB	wmc177	2A	28.3	3	0.0311	0.0699
2005SC_WSB	wmc522	2A	45	5	1.00E-04	0.253
2005SK_WSB	wmc522	2A	45	5	1.00E-04	0.2777
2006SC_WSB	wmc522	2A	45	5	1.00E-04	0.2827
2006SK_WSB	wmc522	2A	45	5	1.00E-04	0.3482
20052006Combined_WSB	wmc522	2A	45	5	1.00E-04	0.3055
20052006Combined_WSB	wmc296	2A	48.8	2	1.00E-04	0.1493
2005SC_WSB	wmc296	2A	48.8	2	2.00E-04	0.155
2006SC_WSB	wmc296	2A	48.8	2	4.00E-04	0.1308
2006SK_WSB	wmc296	2A	48.8	2	7.00E-04	0.1434
2005SK_WSB	wmc296	2A	48.8	2	8.00E-04	0.1339
2005SC_WSB	gwm95	2A	52.5	2	1.00E-04	0.1442
2005SK_WSB	gwm95	2A	52.5	2	1.00E-04	0.1782
20052006Combined_WSB	gwm95	2A	52.5	2	1.00E-04	0.1607
2006SC_WSB	gwm95	2A	52.5	2	2.00E-04	0.1317
2006SK_WSB	gwm95	2A	52.5	2	3.00E-04	0.1676
2006SK_WSB	gwm312	2A	73.7	2	0.0019	0.115
2005SC_WSB	gwm312	2A	73.7	2	0.0034	0.0866
20052006Combined_WSB	gwm312	2A	73.7	2	0.0049	0.0945
2006SC_WSB	gwm312	2A	73.7	2	0.0065	0.0785
2005SK_WSB	gwm312	2A	73.7	2	0.0156	0.0738
2006SK_WSB	gwm294	2A	76.3	2	4.00E-04	0.1414
20052006Combined_WSB	gwm294	2A	76.3	2	8.00E-04	0.1186
2005SC_WSB	gwm294	2A	76.3	2	0.0025	0.0977

Environment	Locus	Chromosome	Distance (cM)	df Marker	p-permutations Marker	r^2 Marker
2005SK_WSB	gwm294	2A	76.3	2	0.0039	0.1042
2006SC_WSB	gwm294	2A	76.3	2	0.004	0.0975
2006SK_WSB	cfd168	2A	85.2	2	0.0091	0.0918
20052006Combined_WSB	cfd168	2A	85.2	2	0.02	0.0666
2006SC_WSB	cfd168	2A	85.2	2	0.0255	0.0598
2005SK_WSB	cfd168	2A	85.2	2	0.0399	0.0605
2005SC_WSB	wmc25	2B	25	2	0.0016	0.0978
2005SK_WSB	wmc25	2B	25	2	0.0031	0.1011
20052006Combined_WSB	wmc25	2B	25	2	0.0052	0.0917
2006SK_WSB	wmc25	2B	25	2	0.0097	0.0871
2006SC_WSB	wmc25	2B	25	2	0.0122	0.0725
2005SC_WSB	wmc154	2B	28.5	4	1.00E-04	0.1946
2005SK_WSB	wmc154	2B	28.5	4	1.00E-04	0.2563
2006SC_WSB	wmc154	2B	28.5	4	1.00E-04	0.191
20052006Combined_WSB	wmc154	2B	28.5	4	1.00E-04	0.2124
2006SK_WSB	wmc154	2B	28.5	4	2.00E-04	0.1846
2006SC_WSB	gwm429	2B	40.4	3	0.0099	0.0916
20052006Combined_WSB	gwm429	2B	40.4	3	0.0466	0.0664
2006SC_WSB	gwm148	2B	46.5	3	0.0505	0.0646
2005SK_WSB	cfd73	2B	82.4	3	0.0107	0.0966
2005SC_WSB	cfd73	2B	82.4	3	0.0199	0.0763
20052006Combined_WSB	cfd73	2B	82.4	3	0.0208	0.0828
2006SC_WSB	cfd73	2B	82.4	3	0.021	0.0778
2005SC_WSB	gwm5	3A	44.9	4	0.0154	0.0994
2006SC_WSB	cfa2193	3A	73.8	2	1.00E-04	0.1524
2006SK_WSB	cfa2193	3A	73.8	2	1.00E-04	0.1851
20052006Combined_WSB	cfa2193	3A	73.8	2	1.00E-04	0.1592
2005SC_WSB	cfa2193	3A	73.8	2	2.00E-04	0.1455
2005SK_WSB	cfa2193	3A	73.8	2	0.0012	0.123
2006SK_WSB	wmc559	3A	83.3	1	9.00E-04	0.1022
2006SC_WSB	wmc559	3A	83.3	1	0.0014	0.0851
20052006Combined_WSB	wmc559	3A	83.3	1	0.0019	0.0803
2005SC_WSB	wmc559	3A	83.3	1	0.0068	0.0574
2005SK_WSB	wmc559	3A	83.3	1	0.0096	0.0588
2006SK_WSB	cfa2076	3A	97.7	1	1.00E-03	0.0965
20052006Combined_WSB	cfa2076	3A	97.7	1	0.0065	0.0629
2005SC_WSB	cfa2076	3A	97.7	1	0.0072	0.0595
2006SC_WSB	cfa2076	3A	97.7	1	0.0197	0.0446
2005SK_WSB	cfa2076	3A	97.7	1	0.0447	0.0379
2005SC_WSB	wmc594	3A	105	4	0.0245	0.0888
2006SC_WSB	wmc594	3A	105	4	0.0346	0.0821
20052006Combined_WSB	wmc594	3A	105	4	0.0438	0.082
2006SK_WSB	wmc594	3A	105	4	0.0475	0.0899
2005SK_WSB	gwm389	3B	0.7	3	1.00E-04	0.1857
2006SC_WSB	gwm389	3B	0.7	3	1.00E-04	0.2036
2006SK_WSB	gwm389	3B	0.7	3	1.00E-04	0.2206
20052006Combined_WSB	gwm389	3B	0.7	3	1.00E-04	0.204
2005SC_WSB	gwm389	3B	0.7	3	2.00E-04	0.163
2005SC_WSB	wmc418	3B	72.4	3	0.049	0.0671

Environment	Locus	Chromosome	Distance (cM)	df Marker	p-permutations Marker	r^2 Marker
2006SK_WSB	wmc418	3B	72.4	3	0.0515	0.0751
2005SK_WSB	wmc418	3B	72.4	3	0.0533	0.0721
2006SC_WSB	barc84	3B	97.1	1	0.0505	0.0315
2005SC_WSB	barc77	3B	111.2	1	0.0058	0.064
2005SK_WSB	barc77	3B	111.2	1	0.0087	0.0653
20052006Combined_WSB	barc77	3B	111.2	1	0.0088	0.0591
2006SC_WSB	barc77	3B	111.2	1	0.0153	0.0519
2006SK_WSB	barc77	3B	111.2	1	0.0359	0.0437
2006SC_WSB	gwm299	3B	122.5	3	2.00E-04	0.1486
2006SK_WSB	gwm299	3B	122.5	3	2.00E-04	0.1976
20052006Combined_WSB	gwm299	3B	122.5	3	2.00E-04	0.1696
2005SK_WSB	gwm299	3B	122.5	3	4.00E-04	0.1736
2005SC_WSB	gwm299	3B	122.5	3	0.0011	0.1293
20052006Combined_WSB	wmc632	3B	142.9	5	0.0037	0.1379
2005SK_WSB	wmc632	3B	142.9	5	0.0047	0.1441
2006SK_WSB	wmc632	3B	142.9	5	0.0054	0.1479
2005SC_WSB	wmc632	3B	142.9	5	0.0073	0.1221
2006SC_WSB	wmc632	3B	142.9	5	0.0091	0.1208
2005SK_WSB	gwm340	3B	147.6	4	1.00E-04	0.2378
2006SC_WSB	gwm340	3B	147.6	4	1.00E-04	0.1836
2006SK_WSB	gwm340	3B	147.6	4	1.00E-04	0.2359
20052006Combined_WSB	gwm340	3B	147.6	4	1.00E-04	0.207
2005SC_WSB	gwm340	3B	147.6	4	2.00E-04	0.1599
2005SC_WSB	cfd71	4A	8.4	1	0.0529	0.0304
2005SC_WSB	wmc96	4A	10.5	2	0.0027	0.0893
20052006Combined_WSB	wmc96	4A	10.5	2	0.005	0.0855
2006SK_WSB	wmc96	4A	10.5	2	0.0053	0.0999
2006SC_WSB	wmc96	4A	10.5	2	0.0089	0.074
2005SK_WSB	wmc96	4A	10.5	2	0.0446	0.0578
2005SC_WSB	barc170	4A	26.6	4	1.00E-04	0.2037
2005SK_WSB	barc170	4A	26.6	4	1.00E-04	0.2426
2006SC_WSB	barc170	4A	26.6	4	1.00E-04	0.2337
2006SK_WSB	barc170	4A	26.6	4	1.00E-04	0.2813
20052006Combined_WSB	barc170	4A	26.6	4	1.00E-04	0.252
2005SC_WSB	gwm251	4B	24.5	2	0.0438	0.0507
2005SK_WSB	wmc238	4B	33.9	1	0.009	0.0592
2006SK_WSB	wmc238	4B	33.9	1	0.0226	0.0469
20052006Combined_WSB	wmc238	4B	33.9	1	0.0367	0.038
2005SC_WSB	wmc238	4B	33.9	1	0.0502	0.0306
2005SC_WSB	gwm368	4B	36.9	3	1.00E-04	0.1901
2005SK_WSB	gwm368	4B	36.9	3	1.00E-04	0.2461
2006SC_WSB	gwm368	4B	36.9	3	1.00E-04	0.1912
2006SK_WSB	gwm368	4B	36.9	3	1.00E-04	0.2277
20052006Combined_WSB	gwm368	4B	36.9	3	1.00E-04	0.2247
2005SK_WSB	barc20	4B	37.7	1	2.00E-04	0.1061
20052006Combined_WSB	barc20	4B	37.7	1	1.00E-03	0.0853
2005SC_WSB	barc20	4B	37.7	1	0.0012	0.0748
2006SC_WSB	barc20	4B	37.7	1	0.0027	0.0631
2006SK_WSB	barc20	4B	37.7	1	0.0035	0.0791

Environment	Locus	Chromosome	Distance (cM)	df Marker	p-permutations Marker	r^2 Marker
2005SC_WSB	wmc713	5A	27.8	1	1.00E-04	0.1831
2005SK_WSB	wmc713	5A	27.8	1	1.00E-04	0.2314
2006SC_WSB	wmc713	5A	27.8	1	1.00E-04	0.2269
2006SK_WSB	wmc713	5A	27.8	1	1.00E-04	0.2042
20052006Combined_WSB	wmc713	5A	27.8	1	1.00E-04	0.2205
2005SC_WSB	cfa2104	5A	36.7	1	1.00E-04	0.195
2005SK_WSB	cfa2104	5A	36.7	1	1.00E-04	0.2452
2006SC_WSB	cfa2104	5A	36.7	1	1.00E-04	0.2396
2006SK_WSB	cfa2104	5A	36.7	1	1.00E-04	0.2157
20052006Combined_WSB	cfa2104	5A	36.7	1	1.00E-04	0.2338
2005SC_WSB	cfa2190	5A	44.9	2	1.00E-04	0.1969
2005SK_WSB	cfa2190	5A	44.9	2	1.00E-04	0.2462
2006SC_WSB	cfa2190	5A	44.9	2	1.00E-04	0.239
2006SK_WSB	cfa2190	5A	44.9	2	1.00E-04	0.2301
20052006Combined_WSB	cfa2190	5A	44.9	2	1.00E-04	0.2365
2006SC_WSB	gwm293	5A	52	3	0.0079	0.0972
20052006Combined_WSB	gwm293	5A	52	3	0.0089	0.099
2006SK_WSB	gwm293	5A	52	3	0.0103	0.1092
2005SC_WSB	gwm293	5A	52	3	0.0112	0.0914
2005SK_WSB	gwm293	5A	52	3	0.0449	0.0736
2006SC_WSB	gwm304	5A	59	3	0.0068	0.095
2006SK_WSB	gwm304	5A	59	3	0.0073	0.1088
20052006Combined_WSB	gwm304	5A	59	3	0.0096	0.0961
2005SC_WSB	gwm304	5A	59	3	0.0117	0.0875
2005SK_WSB	gwm304	5A	59	3	0.0534	0.0699
2005SK_WSB	gwm595	5A	149.4	4	0.038	0.0908
2005SC_WSB	wmc524	5A	150.6	1	0.0193	0.0431
2006SK_WSB	wmc524	5A	150.6	1	0.0534	0.0352
2005SK_WSB	gwm291	5A	162.9	4	6.00E-04	0.156
2005SC_WSB	gwm291	5A	162.9	4	0.0014	0.1312
20052006Combined_WSB	gwm291	5A	162.9	4	0.0026	0.1283
2006SC_WSB	gwm291	5A	162.9	4	0.0105	0.1059
2006SK_WSB	gwm291	5A	162.9	4	0.0171	0.11
2005SK_WSB	wmc740	5B	55.6	4	0.0073	0.1223
2005SC_WSB	wmc740	5B	55.6	4	0.0074	0.1089
20052006Combined_WSB	wmc740	5B	55.6	4	0.0095	0.1129
2006SC_WSB	wmc740	5B	55.6	4	0.0126	0.1031
2006SK_WSB	wmc740	5B	55.6	4	0.0279	0.1026
2006SC_WSB	gwm335	5B	67.6	3	0.0067	0.0936
20052006Combined_WSB	gwm335	5B	67.6	3	0.0131	0.0887
2005SC_WSB	gwm335	5B	67.6	3	0.0187	0.079
2006SK_WSB	gwm335	5B	67.6	3	0.0189	0.0932
2005SK_WSB	gwm335	5B	67.6	3	0.0424	0.074
2006SK_WSB	wmc415	5B	80.3	2	0.0319	0.0654
2005SC_WSB	wmc415	5B	80.3	2	0.032	0.0549
20052006Combined_WSB	wmc415	5B	80.3	2	0.0346	0.0575
2006SC_WSB	wmc415	5B	80.3	2	0.05	0.0495
2006SK_WSB	cfd86	5B	139.2	1	0.0258	0.046
2006SC_WSB	cfd86	5B	139.2	1	0.031	0.037

Environment	Locus	Chromosome	Distance (cM)	df Marker	p-permutations Marker	r^2 Marker
20052006Combined_WSB	cfd86	5B	139.2	1	0.0391	0.0369
2005SC_WSB	cfd86	5B	139.2	1	0.0458	0.0323
2005SK_WSB	wmc508	5B	140.9	1	1.00E-04	0.1584
2006SK_WSB	wmc508	5B	140.9	1	1.00E-04	0.1354
20052006Combined_WSB	wmc508	5B	140.9	1	1.00E-04	0.1294
2005SC_WSB	wmc508	5B	140.9	1	2.00E-04	0.1244
2006SC_WSB	wmc508	5B	140.9	1	0.0014	0.0882
2006SC_WSB	gwm334	6A	1.8	3	8.00E-04	0.1218
20052006Combined_WSB	gwm334	6A	1.8	3	0.0049	0.111
2006SK_WSB	gwm334	6A	1.8	3	0.0058	0.1175
2005SK_WSB	gwm334	6A	1.8	3	0.0113	0.1013
2005SC_WSB	gwm334	6A	1.8	3	0.0147	0.0876
2005SC_WSB	barc146	6A	36.66	3	1.00E-04	0.175
2006SC_WSB	barc146	6A	36.66	3	1.00E-04	0.2092
20052006Combined_WSB	barc146	6A	36.66	3	2.00E-04	0.1751
2006SK_WSB	barc146	6A	36.66	3	4.00E-04	0.1702
2005SK_WSB	barc146	6A	36.66	3	0.0034	0.1169
2006SC_WSB	wmc201	6A	46.361	3	0.0227	0.078
20052006Combined_WSB	wmc201	6A	46.361	3	0.0378	0.071
2005SK_WSB	wmc201	6A	46.361	3	0.0478	0.0708
2005SC_WSB	wmc201	6A	46.361	3	0.0539	0.0612
2005SC_WSB	wmc487	6B	9.2	4	0.0492	0.0792
2006SC_WSB	wmc487	6B	9.2	4	0.0492	0.0794
2005SK_WSB	gwm705	6B	13.8	4	0.0286	0.0993
20052006Combined_WSB	gwm705	6B	13.8	4	0.0498	0.0846
2006SC_WSB	gwm705	6B	13.8	4	0.0529	0.0783
2005SC_WSB	gwm193	6B	36.3	2	0.0012	0.1151
2006SC_WSB	gwm193	6B	36.3	2	0.0013	0.1098
20052006Combined_WSB	gwm193	6B	36.3	2	0.0019	0.1087
2006SK_WSB	gwm193	6B	36.3	2	0.0038	0.1045
2005SK_WSB	gwm193	6B	36.3	2	0.0132	0.0819
2005SK_WSB	gwm471	7A	17.4	3	0.0034	0.1169
20052006Combined_WSB	gwm471	7A	17.4	3	0.0046	0.111
2006SC_WSB	gwm471	7A	17.4	3	0.0058	0.0987
2006SK_WSB	gwm471	7A	17.4	3	0.0066	0.1156
2005SC_WSB	gwm471	7A	17.4	3	0.0083	0.0952
2005SC_WSB	wmc283	7A	39.6	3	1.00E-04	0.1789
20052006Combined_WSB	wmc283	7A	39.6	3	1.00E-04	0.1695
2005SK_WSB	wmc283	7A	39.6	3	2.00E-04	0.1952
2006SK_WSB	wmc283	7A	39.6	3	6.00E-04	0.1662
2006SC_WSB	wmc283	7A	39.6	3	0.0012	0.1246
2005SC_WSB	barc127	7A	47.1	2	1.00E-04	0.1781
20052006Combined_WSB	barc127	7A	47.1	2	3.00E-04	0.1363
2005SK_WSB	barc127	7A	47.1	2	5.00E-04	0.1265
2006SC_WSB	barc127	7A	47.1	2	5.00E-04	0.1102
2006SK_WSB	barc127	7A	47.1	2	0.0022	0.1163
2005SK_WSB	cfa2028	7A	47.7	2	0.0393	0.0593
2006SC_WSB	cfa2028	7A	47.7	2	0.0531	0.0479
2005SK_WSB	barc108	7A	70.8	1	1.00E-04	0.1336

Environment	Locus	Chromosome	Distance (cM)	df Marker	p-permutations Marker	r^2 Marker
2006SC_WSB	barc108	7A	70.8	1	2.00E-04	0.1152
20052006Combined_WSB	barc108	7A	70.8	1	2.00E-04	0.1189
2006SK_WSB	barc108	7A	70.8	1	3.00E-04	0.1205
2005SC_WSB	barc108	7A	70.8	1	9.00E-04	0.0883
20052006Combined_WSB	wmc603	7A	72.7	3	4.00E-04	0.1337
2006SC_WSB	wmc603	7A	72.7	3	5.00E-04	0.1397
2006SK_WSB	wmc603	7A	72.7	3	0.0018	0.1319
2005SC_WSB	wmc603	7A	72.7	3	0.0021	0.1106
2005SK_WSB	wmc603	7A	72.7	3	0.0025	0.1277
2005SC_WSB	wmc790	7A	102	4	1.00E-04	0.2529
2005SK_WSB	wmc790	7A	102	4	1.00E-04	0.2816
2006SC_WSB	wmc790	7A	102	4	1.00E-04	0.2752
2006SK_WSB	wmc790	7A	102	4	1.00E-04	0.2716
20052006Combined_WSB	wmc790	7A	102	4	1.00E-04	0.2813
2005SK_WSB	cfa2019	7A	106.5	3	1.00E-04	0.1793
2006SK_WSB	cfa2019	7A	106.5	3	0.0013	0.135
20052006Combined_WSB	cfa2019	7A	106.5	3	0.0013	0.1227
2006SC_WSB	cfa2019	7A	106.5	3	0.0035	0.1052
2005SC_WSB	cfa2019	7A	106.5	3	0.0181	0.0818
2006SK_WSB	*Psy1-A1*	7A		1	6.00E-04	0.1083
2006SC_WSB	*Psy1-A1*	7A		1	0.0018	0.0752
20052006Combined_WSB	*Psy1-A1*	7A		1	0.0018	0.0816
2005SK_WSB	*Psy1-A1*	7A		1	0.0045	0.0717
2005SC_WSB	*Psy1-A1*	7A		1	0.0074	0.0581
2005SK_WSB	wmc606	7B	0	7	1.00E-04	0.2462
2006SC_WSB	wmc606	7B	0	7	1.00E-04	0.2267
20052006Combined_WSB	wmc606	7B	0	7	1.00E-04	0.2415
2005SC_WSB	wmc606	7B	0	7	2.00E-04	0.2091
2006SK_WSB	wmc606	7B	0	7	4.00E-04	0.2414
2005SK_WSB	gwm537	7B	41.2	5	0.0217	0.118
2006SC_WSB	gwm537	7B	41.2	5	0.0401	0.0956
20052006Combined_WSB	gwm537	7B	41.2	5	0.0488	0.096
2005SK_WSB	wmc758	7B	56.6	1	0.004	0.0693
2006SK_WSB	wmc758	7B	56.6	1	0.0059	0.0698
2005SC_WSB	wmc758	7B	56.6	1	0.006	0.0585
20052006Combined_WSB	wmc758	7B	56.6	1	0.0069	0.0626
2006SC_WSB	wmc758	7B	56.6	1	0.0191	0.043
2006SK_WSB	wmc475	7B	59.7	2	1.00E-03	0.1289
20052006Combined_WSB	wmc475	7B	59.7	2	0.0022	0.1055
2005SK_WSB	wmc475	7B	59.7	2	0.0027	0.1086
2006SC_WSB	wmc475	7B	59.7	2	0.005	0.0846
2005SC_WSB	wmc475	7B	59.7	2	0.0055	0.0846
2006SC_WSB	gwm333	7B	65	6	0.0017	0.1636
2005SK_WSB	gwm333	7B	65	6	0.0025	0.1775
2006SK_WSB	gwm333	7B	65	6	0.0026	0.1802
20052006Combined_WSB	gwm333	7B	65	6	0.0026	0.1619
2005SC_WSB	gwm333	7B	65	6	0.011	0.1284
2006SK_WSB	wmc396	7B	68.2	4	4.00E-04	0.1874
20052006Combined_WSB	wmc396	7B	68.2	4	4.00E-04	0.1538

Environment	Locus	Chromosome	Distance (cM)	df Marker	p-permutations Marker	r^2 Marker
2006SC_WSB	wmc396	7B	68.2	4	7.00E-04	0.1523
2005SC_WSB	wmc396	7B	68.2	4	0.0033	0.1213
2005SK_WSB	wmc396	7B	68.2	4	0.0039	0.1331
2006SC_WSB	wmc10	7B	147.2	1	0.0207	0.0415
2006SK_WSB	wmc10	7B	147.2	1	0.0215	0.0519
2005SC_WSB	wmc10	7B	147.2	1	0.0224	0.0424
20052006Combined_WSB	wmc10	7B	147.2	1	0.0252	0.0427
2006SK_WSB	gwm146	7B	149.9	3	0.0091	0.1069
2006SK_WSB	*Psy1-B1*	7B		1	1.00E-04	0.1556
20052006Combined_WSB	*Psy1-B1*	7B		1	0.0011	0.0811
2006SC_WSB	*Psy1-B1*	7B		1	0.0038	0.0639
2005SC_WSB	*Psy1-B1*	7B		1	0.0151	0.0478
2005SK_WSB	*Psy1-B1*	7B		1	0.0264	0.0452