TARGETED IMPROVEMENT OF THE WHEAT GENOME BY MARKER-ASSISTED SELECTION AND UNDERSTANDING ITS HOMOEOLOGOUS GENE EVOLUTION AND EXPRESSION BALANCE

By

JASDEEP SINGH MUTTI

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

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
Department of Crop and Soil Sciences

DECEMBER 2007

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of JASDEEP SINGH MUTTI find it satisfactory and recommend that it be accepted.

_____

Chair

_____

_____

_____

_____

ACKNOWLEDGMENT

I wish to thank Dr. Kulvinder Gill for providing me the opportunity to pursue my doctoral studies in Crop Science at Washington State University. His valuable guidance, whole hearted support and encouragement throughout my doctoral program were instrumental in shaping my academic career. It is not just the scientific skills he taught, but I picked up some finer points of life which reflected from his personality of vast professional experience. For my research progress, I owe a great to my committee members namely Drs. Camille Steber, Gary Thorgaard, Kim Campbell and Dorrie Main, who provided useful inputs from time to time to the best of my research interests. I would like to extend a word of special thanks to Dr. Steve Baenziger for his critical inputs and valuable guidance. It was indeed a pleasure for me to work in close association with him on field evaluation and release of the wheat germplasms. My special thanks are due to Dr. Ming-Shun Chen for screening of the germplasms with Hessian fly. I would also like to extend my gratitude to my erstwhile Lab colleagues Dr. Deepak Sidhu and Dr. Harpinder Randhawa. Deepak for her help in gel blot analysis of methylation pattern and Harpinder for marker analysis in introgression of *Yr15* gene into spring wheat cultivar, 'Zak'. My other Lab colleagues, especially Mustafa Shafqat, Harvinder Bennypaul, Sachin Rustgi and Neeraj Kumar, were real resource persons for creating a stimulating environment in the Lab for work. For folks at the front office - you are simply the best! Every time, they greeted me with a smile and helped me out a lot. I would like to thank the entire faculty and staff in the Departments of Crop Science for treating me as a part of the Crop Science family and making me feel very much at home. My wife, Meezy, deserves a special mention for her inspiration and providing me much needed moral support. Finally, I would be failing in my duty if I do not thank my brother Dr. Navdeep Mutti, sister-in-

law Dilpreet and my niece Navnoor for their love, constant support and encouragement.

TARGETED IMPROVEMENT OF THE WHEAT GENOME BY MARKER-ASSISTED SELECTION AND UNDERSTANDING ITS HOMOEOLOGOUS GENE EVOLUTION AND EXPRESSION BALANCE

Abstract

by Jasdeep Singh Mutti, PhD
Washington State University
December 2007

Chair: Kulvinder S Gill

Increasing global demand for wheat, due to burgeoning human population, increased use in feed, and ethanol production requires improved cultivars to substantially increase its productivity. The objectives of this study were to: i) understand mechanism(s) controlling the unique gene expression balance in the polyploid wheat that may have led to its adaptation and success as a major crop plant, and ii) to make cultivar improvement more targeted, precise, and fast through marker-assisted selection. *In silico* analysis, confirmed by single stranded conformation polymorphism (SSCP), revealed that more than 91% of the hexaploid wheat genes are expressed from two or more homoeologues. Expression of only 9% of the genes was from one of the three homoeologues. In general, the proportion of expressed copies decreased with the increase in structural copy number. Homoeologous gene expression was further influenced by the physical location of the genes on the chromosomes. Short arms of wheat chromosomes appear to be functionally different from long arms. Expresion of all three homoeologues was favored in the proximal regions of the short arms but in the long arms, the terminal regions were favored. The

most significant observation was that tissue specific expression patterns for 87% of the wheat genes were different for different homoeologues. More genes and more homoeologues/gene were expressed in the root tissue and the least number were expressed in meiosis and early stages of flowering. The maximum tissue-specific homoeologous gene expression was observed during flowering and seed development stages and the least was observed in roots. About 30% of the genes showed altered methylation or expression pattern in response to homoeologue copy number change, suggesting that expression of these genes is interdependent among homoeologues. Sequence changes among homoeologues showed a distinct pattern that was common to all genes studied. The middle of the coding region exhibited the least proportion of sequence changes. Accounting for codon usage, the ratio of synonymous to non-synonymous changes did not show any increase in synonymous changes which would be expected due to selection pressure. We therefore postulated that wheat and perhaps other higher eukaryotes possess a mechanism to protect translated parts of the genes from sequence change(s), irrespective of their function. The applied part of my research focused on targeted use of marker-assisted selection for rapid introgression of single genes into popular cultivars. Using marker assisted backcross breeding; we individually transferred seven pest resistance genes into seven winter wheat lines to develop 28 advanced breeding lines. We also optimized a marker-assisted background selection based method of gene introgressions. After testing various approaches by simulations, we incorporated the available information on the distribution of genes and recombination on wheat chromosomes to develop a marker-assisted background selection method to accomplish ≥96% recovery of recurrent parent genome (RPG) in two backcrosses.

TABLE OF CONTENTS

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

DEDICATION

Dedicated to my parents, Dr. Darshan Singh Mutti and Mrs. Narinder Kaur Mutti, for their constant encouragement and support.

CHAPTER 1

INTRODUCTION

Wheat (*Triticum aestivum* L. em. Thell) ranks second after rice, providing 23% of dietary energy and serving as a staple food for 40% of the total human population of the world (http://faostat.fao.org/). In accordance with its demand, it is the most widely grown crop throughout the world, accounting for 17% of the total cultivated area and production of 605 million tons (http://faostat.fao.org/). With the total annual production of 57 million tons in 2006, the United States accounts for 9.5% of the total wheat production in the world (http://faostat.fao.org/). To cope up with the increasing demand due to an ever increasing human population, it is projected that wheat grain production needs to increase at an annual rate of 2% on the same or reduced land area (Cassman 1999; Fufa et al. 2005). This increase in demand can be met either by increasing the yield of the wheat plant and/or by protecting the 25% of the crop that is lost each year to biotic (pests) and abiotic (drought, heat, cold and salinity) stress factors. To increase grain yield, it is imperative to understand the molecular mechanisms of the genes controlling agronomically important traits. Losses from biotic and abiotic stresses can also be best protected genetically by combining genes conferring resistance to various stresses into the elite varieties in the shortest possible time.

1.1 The Wheat Genome

Bread wheat is an allohexaploid with $2n = 6x = 42$ chromosomes, containing three homoeologous genomes (AABBDD). The hexaploid wheat, *T. aestivum* L. evolved from a hybridization event of a tetraploid species (*T. turgidum* L.; $2n = 28$,

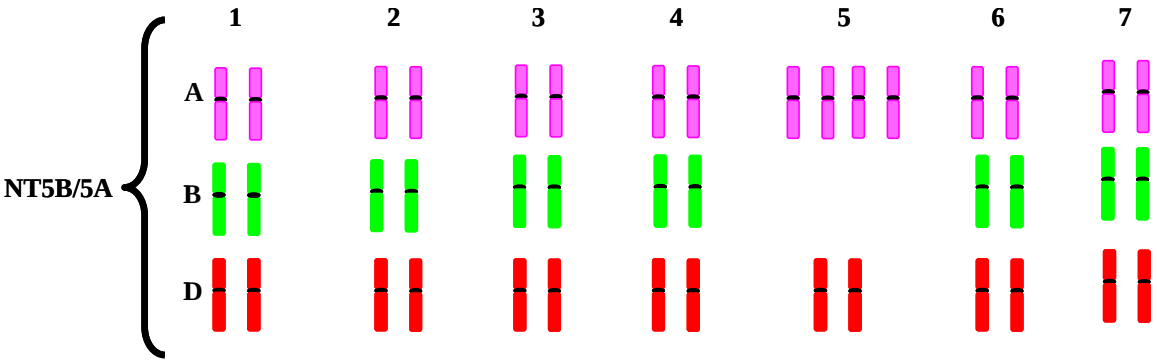
AABB) with a diploid species *Aegilops tauschii* (Coss.) Schmal (2*n* = 14, DD), about 8,000 years ago (Kihara 1944). *T. turgidum* resulted from a hybridization event between *T. urartu* Tumanian ex Gandilyan (2*n*=14, AA) and *A. speltoides* Tausch (2*n*=2*x*=14, SS»BB) or its closely related species about 500,000 years ago (Huang et al. 2002). Wheat is one of the most important crops belonging to the tribe Triticeae of the grass family Poaceae. Other major crop plants present in the family are rice (*Oryza sativa* L.), maize (*Zea mays* L.), barley (*Hordeum vulgare* L.), and oat (*Avena sativa* L.). The wheat and barley haploid genomes are about the same size except for bread wheat which is hexaploid and is thus about three times the size of the diploid cultivated barley. Despite intensive studies conducted over the past 70 to 80 years, little is known about the nature of the genomic changes that occurred following polyploidization, which facilitated harmonious coexistence of the different genomes in the same nucleus.

1.2 Wheat Aneuploid Stocks:

The polyploid nature of the wheat genome has enabled the development of aneuploid stocks.The most popular types include the nullisomic-tetrasomic (NT), ditelosomic (DT), and single-break deletion lines that are available in the background of a spring wheat cultivar, ‘Chinese spring’ (CS) (Sears 1954; Sears and Sears 1978; Endo 1988; Endo and Gill 1996). The NT lines lack a pair of homologous chromosomes, the deficiency of which is compensated for by a double dose of either of the two homoeologous chromosomes. For example, in nulli5B/tetra5A (NT5B/5A), the chromosome 5B pair is missing and replaced with four copies of 5A. Ditelosomics are the lines that lack a pair of chromosome arms (Figure 1.1). These types of aneuploid stocks (the NT and DT lines) are available for almost all of the wheat

chromosomes (www.k-state.edu/wgrc/). Another set of aneuploid stocks includes 436 chromosome deletion lines for all of the 21 wheat chromosomes (Endo 1990; Endo and Gill 1996). These aneuploids have been extensively used for various genetic studies including the precise physical mapping of DNA markers and genes. Recently, these aneuploid stocks were used to physically map ~9,400 wheat ESTs (Conley et al. 2004; Hossain et al. 2004; Linkiewicz et al. 2004; Munkvold et al. 2004; Miftahudin et al. 2004; Peng et al. 2004; Randhawa et al. 2004) (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi). Most of these aneuploid lines are stable and fertile. These aneupoloid stocks will be invaluable to understand the functional organization of the wheat genome and to analyze the effect of gene dosage and chromosomal breaks on wheat gene expression.

Nullisomic-tetrasomic lines:



Ditelosomic lines:

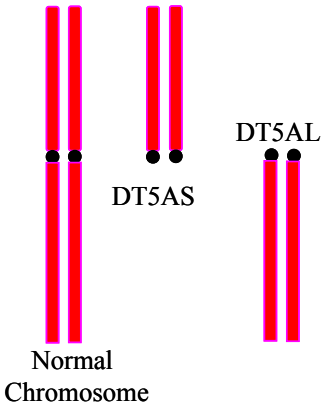


Figure 1.1. An example of nullitetra and ditelosomic lines of group 5 chromosomes of wheat.

1.3 Wheat Genome Organization, and Distribution of Genes and Recombination

Compared to model eukaryotic systems such as yeast, Arabidopsis and Drosophila, the genome size of most of the crop plants is large. In eukaryotes, the genome size ranges anywhere from 12 Mb for budding yeast (*Saccharomyces cerevisiae*) (Mewes et al. 1997) to 124,852 Mb in case of the plant *Fritillaria assyriaca* (Fritillary) (Jauhar et al. 1991). At 16,000 Mb, the hexaploid wheat genome is 41-fold larger than rice (*O. sativa*) (389 Mb) (Arumuganathan and Earle 1991). Discounting polyploidy, there is about 12- fold difference among these two, which is similar to that between barley (5000 Mb) and rice (Arumuganathan and Earle 1991).

Although the genome is large, the gene containing fraction appears to be only 1-5% (Sidhu and Gill 2004). Using deletion lines to reveal the physical location of gene markers combined with generation of consensus map for the three homoeologues chromosomes, revealed the distribution of genes in wheat which is best known among all poorly sequenced eukaryotes (Gill et al. 1993; Gill et al. 1996a; Gill et al. 1996c; Sandhu et al. 2001). These analyses localized more than 85% of the wheat genes to 48 gene rich regions (GRRs) that encompassed less than 10% of the chromosomal regions (Erayman et al. 2004). These GRRs are interspersed with gene-poor chromosomal regions predominantly consisting of retrotransposon-like repetitive DNA sequences and pseudogenes (SanMiguel et al. 1996; Barakat et al. 1997; Feuillet and Keller 1999; Sandhu and Gill 2002).

Along with gene distribution, recombination is also highly unevenly distributed along the chromosomes of eukaryotes (Dvorak and Chen 1984; Bollag et al. 1989; Curtis et al. 1991; Gill et al. 1993; Gill et al. 1996b; Lichten and Goldman 1995; Tranquilli et al. 1999; Sandhu and Gill 2002). The average frequency of

recombination in rice is about 0.003 cM kb^{-1} ranging from 0 to 0.06 cM kb^{-1}, with recombination hot spots interspersed with recombination cold spots. Accordingly, in the consensus of physical and genetic maps of wheat, a detailed and accurate estimate of recombination was provided (Erayman et al. 2004). At a resolution of ~7 Mb, the average recombination frequency was 0.0003 cM/kb with a range from 0 to 0.007 cM/kb. Recombination was further highly variable among the gene containing regions with gene poor regions accounting for only ~5% of recombination, and 95% of recombination observed in gene rich regions (Erayman et al. 2004). In general, a severe suppression of recombination was observed in the centromeric regions of the wheat chromosomes. Centromeres along with one-fourth of the wheat genome present around it, accounted for <1% of the total recombination. Recombination in the distal chromosomal regions was much higher. For example, 86% of of the recombination for the short arm of chromosome 1 occurred in the '1S0.8' gene rich region, which is only ~7 Mb in size.

1.4 Fate of Duplicated Genes in Polyploids

Polyploidy is a common phenomenon in the evolution of many eukaryotes. About 80% of plants are polyploids and polyploidy has played a major role in speciation, adaptation and genome expansion of these plants (Masterson 1994). In order to track the fate of duplicated genes following polyploidization, a younger polyploid like wheat is ideal particularly considering that its diploid progenitors are extant and phylogenetic relationship with its diploid progenitors is comparatively well understood (Sasanuma et al. 1996). Furthermore, the wealth of aneuploid stocks available for wheat provides a major resource for polyploidy research.

1.4.1 Structural changes

Very little is known about the structural changes that genes/genomes may undergo during polyploidization. Different studies done on interspecific and intergeneric hybrids of Brassica, Aegilops, and Triticum showed both loss of DNA in early generations and gain of some genomic fragments in later generations (Song et al. 1995; Han et al. 2003). Analysis of the F_5 generation of synthetic allopolyploids of *Brassica rapa* with *B. nigra*, and *B. oleracea* using 89 restriction fragment length polymorphism (RFLP) probes, showed a loss of paternal restriction fragments in the F_2 generation, some of which were later recovered in F_5 generation (Song et al. 1995; Axelsson et al. 2000). Similar observations of loss and/or gain of genomic segments were reported for synthetic allopolypoids of the Aegilops-Triticum group (Liu et al. 1998). Restriction fragment pattern analysis of six generations of amphiploids produced from a cross between *A. sharonensis* Eig and *T. monococcum* L. showed elimination of 14% of genomic loci of one of the parents in the F_1 generation (Liu et al. 1998; Shaked et al. 2001; Kashkush et al. 2002). On the contrary, nine sets of synthetic cotton allopolyploids showed no change in genome structure after analyzing 22,000 amplified fragment length polymorphism (AFLP) genomic loci (Liu et al. 2001).

1.4.2 Functional changes

Predicted fates of genes after duplication are: (i) nonfunctionalization- It refers to silencing of one of the duplicated copies due to accumulation of several mutations, (ii) neofunctionalization- One of the duplicated copies retains the original function and the second copy acquires a new function, and (iii) subfunctionalization- The combined expression of the two copies is reduced to the level of a single copy gene,

or tissue-specific activity is acquired (Haldane 1933; Watterson 1983; Walsh 1985; Holland et al. 1994; Hughes 1994; Sidow 1996; Force et al. 1999; Stoltzfus 1999; Lynch and Force 2000). Among these, nonfunctionalization is predicted to be most common because the majority of the mutations are deleterious, and thus are eliminated through the process of natural selection (Lynch and Force 2000). The following studies tested these models and the results are summarized below.

The first comprehensive study to look at gene expression immediately following polyploidization was reported by Comai et al. (2000). Synthetic allotetraploids from a cross between diploid or tetraploid *A. thaliana Landsberg erecta* (Ler) with *Cardaminopsis arenosa* (*C. arenosa* isolates are tetraploid) were developed and gene expression was analyzed using cDNA-AFLP. Confirmed by real time reverse transcription polymerase chain reaction (RT-PCR), the silencing of genes from either of the two parents was reported in 20 of the 700 (~3%) genes. Two transcripts that were not present in either of the parents were expressed in the allotetraploids. Similarly, expression analysis of 40 homoeologous gene pairs in natural and synthetic tetraploid cotton using cDNA-SSCP (single-stranded conformation polymorphism) documented silencing of 2% of the analyzed genes, while 18% exhibited differential expression patterns (Adams et al. 2003). In another study, gene expression in the synthetic polyploids in comparison to their diploid parents, *A. sharonensis* and *T. monococcum* was studied (Kashkush et al. 2002). Analysis of 3072 cDNA-AFLP fragments showed that 1.56% of the genes were silenced and the expression of 0.96% of the genes was activated by polyploidy.

The effect of polyploidization has been characterized for ribosomal RNA (rRNA) genes by a well-known phenomenon of nucleolar dominance. First observed in resynthesized allotetraploids of Brassica, nucleolar dominance refers to the

phenomenon in which rRNA genes from only one of the two parents are transcribed in their hybrid (Chen and Pikaard 1997; Pikaard 1999). Selection of the parent for rRNA gene expression was random and was independent of the direction of the cross. Except for the proportion of the silenced or reactivated genes, very little is known about the proportional contributions of two newly merged genomes to the transcriptome of the polyploid. Furthermore, most of these studies were conducted on synthetic polyploids therefore information on the fate of duplicated genes during evolution over an extended period of time, is still lacking for the most part.

In hexaploid wheat, differential contribution of the three homoeologues was studied for the genes controlling benzoxazinones (*Bxs*) biosynthesis and it was shown that the B genome had the maximum contribution to its synthesis (Nomura et al. 2005). Comparison with the diploid progenitor showed that the expression of *Bxs* gene was the highest in the B genome donor, indicating that genomic bias originated in the progenitors and it was retained even after the polyploidization.

Continuous influx of data on the fate of duplicated genes from different organisms has provided evidence for any of the three consequences suggested. Expression analysis of 40 homoeologous gene pairs in natural and synthetic tetraploid cotton using cDNA-SSCP (single-stranded conformation polymorphism) documented silencing of 2% of the analyzed genes, while 18% exhibited differential expression pattern (Adams et al 2003). In most cases, one or more homoeologues were silenced. Reciprocal organ-specific silencing and expression of the homoeologues in allotetraploid cotton was also reported. In a more comprehensive follow-up study, where ~2000 transcripts were analyzed for natural and synthetic allotetraploids of cotton, they showed that 7% of the duplicated genes showed silencing and about 8% of bands showed tissue or organ-specific expression or silencing in the allotetraploid

(Adams et al 2004). The phenomenon of sub-functionalization was first documented by Piatigorsky and Wistow (1991), who proposed a 'gene sharing model'. They suggested that two expression domains of Lens Crystallins could precede a gene duplication event, with each duplicate later losing one of the two expression domains. Evolution of a Hox gene cluster also followed the sub-functionalization model (McClintock et al 2002), with zebrafish *hoxb1b* sharing an early expression pattern in the hindbrain of gastrulating embryos, whereas *hoxb1a* shared later expression in a single segment of the neurulation-stage hindbrain of *hoxb1*. The neo-functionalization theory has been proposed in the evolution of *WUSCHEL*, the *NAM* and the *AP2* transcription factor families (Vandenbussche et al 2003). These transcription factor families all possess a highly conserved DNA binding domain and a divergent C-terminus. Its been hypothesized that after duplication of an ancestral gene, one of the copies may accumulate mutations in the C-terminus, while retaining features such as DNA binding, essential for its function as a transcription factor, in the upstream coding regions. Although some studies have incorporated these fates of duplicated genes to explain there results, comprehensive studies of duplicated genes on whole genome basis is still missing.

1.5 Tissue-Specific Expression of Homoeologues

Many functional and genomic reorganizational events following polyploidization have been documented in various plant systems (Wendel 2000; Rieseberg 2001; Liu and Wendel 2002). These include sequence elimination, methylation of DNA, and biased expression of homoeologues (Song et al. 1995; Liu et al. 1998; Ozkan et al. 2001; Shaked et al. 2001; Nomura et al. 2005). Most notable among those are reciprocal organ-specific silencing and expression of the

homoeologues in allotetraploid cotton. The expression of 18 homoeologues was examined in ten different organs of a cotton plant using cDNA-SSCP analysis. Ten of the 40 homoeologues showed tissue or organ-specific expression or silencing (Adams et al. 2003). In a more comprehensive follow-up study where ~2000 transcripts were analyzed for natural and synthetic allotetraploids of cotton, 7% of the duplicated genes showed silencing and about 8% of bands showed tissue or organ-specific expression or silencing in the case of allotetraploids (Adams et al. 2004).

Except for the *TaBx* gene, similar studies have not been performed in polyploid wheat mainly because it has been very difficult to distinguish transcripts of the three homoeologues because of their high sequence similarity. Therefore, one of the objectives of this proposed research was to develop a reliable method to resolve transcripts of the three homoeologues of wheat and to use that method for a comprehensive survey of homoeologue expression in different organs or tissues of polyploid wheat.

1.6 Evolution of Genes

Eukaryotic genomes are packed into two types of chromatin regions: heterochromatin and euchromatin. Euchromatin is generally associated with actively transcribing genes, whereas, heterochromatin is associated with inactive or pseudo-genes and repeated DNA. In general, the transcribed part of a genome is significantly more conserved than the non-transcribed part. However, intergenic regions, which comprise >98% of the vertebrate (Duret et al. 1993; Lipman 1997) and many plant genomes (Sidhu and Gill 2004), also contain numerous conserved stretches that are believed to contain functional signals (Duret and Bucher 1997; Kellis et al. 2003; Shabalina et al. 2004). Within the genic sequences, the untranslated parts are less

conserved compared to the protein-coding sequences (Pesole et al. 2001); Larizza et al. 2002). Nevertheless, highly conserved sequences of 30 to >50 bp are detected in the untranslated parts as well (Duret et al. 1993; Duret and Bucher 1997; Lipman 1997). In general, the functional parts of the genomes are less prone to changes compared to the non-functional DNA sequences. However, understanding the complete functional information encoded in a genome and the mechanisms controlling differential conservation of the sequences remain a central challenge in biological research.

Genes and genomes change over time and most of the changes are attributed to the process of natural selection as defined by Charles Darwin in 1859. According to the theory of natural selection, any change that increases the organism's survival and fitness will be selected, and thus will become widespread. On the other hand, any change that decreases survival and fitness gradually gets eliminated. Recently, Weinreich and coworkers (2006) demonstrated that very few mutations in the ß-lactamase gene leading to a fitter protein can be explained by Darwinian evolution. Authors were studying five point mutations in an allele of the ß-lactamase gene that jointly increases resistance against cefotaxime by a factor of 100,000. They showed that 102 of the 120 mutational trajectories leading to this allele were inaccessible to Darwinian selection. For the remaining 18 trajectories the probability of realization was negligible, as four of these five mutations in various combinations fail to increase drug resistance. Based on these results, they concluded that pervasive biophysical pleiotropy (one modification leading to a cascade of effects) is likely to be the cause for the evolution of this particular allele (Weinreich et al. 2006).

To date there are very few reports on genome evolution at the nucleotide level and the mechanisms driving DNA sequence changes. In yeast, orthologous regions of

four related species of the Saccharomyces *sensu stricto* group namely *S. paradoxus, S. mikatae* and *S. bayanus* were sequenced and aligned to compare the rate of sequence changes in genic vs. intergenic regions. About 14% of the insertions or deletions were found in intergenic regions, as compared to only 1.3% in genic regions. This difference between genic and intergenic regions was about 75-fold for insertion/deletion of more than 3 bp (Kellis et al. 2003). Although a comprehensive survey of changes among genic and intergenic regions has been documented, no report comparing untranslated regions (UTR) with the coding regions or different regions within the coding parts is available. Furthermore, no mechanism other than natural selection has yet been proposed for the differential susceptibilities of different parts of the genome to changes.

1.7 Chromosome Dosage and Gene Expression

Doubling of genetic material in a single nucleus comes as a genomic shock to the cell and it must respond to maintain the gene expression balance for its sustenance and propagation through the dynamic nature of the surrounding environment. With the changes in gene dosage, two types of effects on gene expression are reported. Some genes show either an increase or decrease in gene expression in response to copy number change (Grell 1962; Epstein 1986), whereas another sets of genes show dosage compensation, in which the expression stays the same regardless of the copy number of the respective genes (Schwartz 1971). Generally a linear correlation of gene expression with dosage is expected (Carlson 1972) such that with an increase in dosage, the expression also increases. However, linear gene dosage is not always the case as in *Drosophila* sex-linked genes (Muller 1950) and multigenic families of rDNA (Mohan and Ritossa 1970; Reeves 1977), where increase in dosage leads to

dosage compensation. Both types of effect are also found in plants. In a classical study done in wheat, 12 genes were analyzed on nullisomic-tetrasomic lines of seven chromosome groups of wheat (Brewer et al. 1969). Of the twelve, only the gene for alkaline phosphatase showed increased effect with extra dose, whereas in the case of the others, the effect of increased dose was compensated (Brewer et al. 1969). Differences in expression were seen, however, among their contemporary diploids for these 12 genes (Sing and Brewer 1969). In a related study, two genes had 30-80% more net protein output when the chromosome carrying its homoeologous counterpart was missing (Aragoncillo et al. 1978). Seed storage proteins, the high molecular weight (HWM) gliadins and glutenins, also have a positive correlation with gene dosage (Galili et al. 1986). In maize, gene dosage compensation is found for the alcohol dehydrogenase gene in one to four dosage series of the chromosome arm on which this gene is located (Birchler 1979), whereas in case of ploidy series involving monoploids, diploids, and tetraploids, the transcript level of these genes increases proportionally with increasing ploidy (Guo et al. 1996). Studied on some genes or artificially induced ploidy, the results do not provide a clear insight into how genes on the same or different chromosomes result in differential response to the same effect, nor the mechanism that regulates these changes.

1.8 Methylation and Gene Expression

DNA methylation is a chemical modification where the 5′ position of cytosine is methylated in a reaction catalyzed by DNA methyltransferases (DNMTs). In mammals, this modification occurs at CpG dinucleotides, whereas plants show extensive methylation of cytosine bases in sequences with CpG, CNG (where N denotes any nucleotide) and CHH (where H denotes A, C or T) bases (Yu et al. 2002;

Goll and Bestor 2005; Bernstein et al. 2006). DNA methylation along with histone modification, especially methylation of the lysine residues on H3, serves as an epigenetic mark for transcriptionally silent chromatin (Martienssen and Colot 2001). Histone modifications are involved in the silencing of genetic material but these modifications are reversible whereas DNA methylation is a more stable modification that is also inherited throughout the cellular divisions (McGarvey et al. 2007).

Epigenetic inheritance, especially via DNA methylation, has been extensively studied in plants. Confirming earlier reports, high-resolution whole genome microarray mapping of Arabidopsis localized cytosine methylation to the repeat sequences and centromeric regions of the genome (Zhang et al. 2006; Zilberman et al. 2007). About one third of the expressed genes were methylated in their open reading frames and less than 5% of the genes had their promoter sequence methylated (Zhang et al. 2006; Zilberman et al. 2007). The significance of methylation inside the coding region of the gene is not understood but it is correlated with highly transcribed and constitutively expressed genes (Zhang et al. 2006; Zilberman et al. 2007). Transcriptional silencing of rRNA genes in interspecific hybrids of Arabidopsis and Brassica was associated with increased DNA methylation and low histone acetylation (Chen and Pikaard 1997). Similarly, active nucleolus organizer regions (NORs) of rRNA genes were less methylated than the repressed NORs in wheat and maize (Barker et al. 1988; Houchins et al. 1997). In Xenopus species, active and inactive NORs did not show methylation differences (Macleod and Bird 1982). In most of the cases, however, DNA methylation is associated with inactivity of a gene.

1.9 Histone Modifications and Gene Expression

Chromatin of higher organisms is a complex of DNA with protein, both

present in about equal proportions. Distinguished cytologically by differential stain intensity, lightly stained chromatin was called euchromatin whereas darkly stained was termed heterochromatin (Heitz 1928). Later studies associated heterochromatin with a lack of gene expression (Schultz and Casperson 1939). Muller in 1930, recovered several chromosomal rearrangements of the *white+* (w+) gene in *Drosophila* that resulted in variegated eye color phenotypes. Each case demonstrated the repressive effect of heterochromatin on euchromatin genes, as chromosome rearrangements placed the genes adjacent to heterochromatin. The effect of variegated eye color could be alleviated by recombining the gene away from the heterochromatin-euchromatin junctions (Schultz 1936). This phenomenon was later termed as position effect variegation (Lewis 1950).

DNA in the nucleus of eukaryotes is complexed with histones to form chromatin. Chromatin is organized in arrays of nucleosomes with two copies of each histone protein, H2A, H2B, H3 and H4, assembled into an octamer that has 145–147 base pairs (bp) of DNA wrapped around it with a molecular mass of 206 kDa. The low-resolution (7-Å) structure of the nucleosome core particle revealed that the histone octamer is wrapped with 1.7 turns of a left-handed DNA superhelix (Richmond et al. 1984). A higher-resolution structure (2.8 Å) of the nucleosome core particle showed that each histone consists of a structured histone fold, as well as two unstructured tails (Luger et al. 1997). Heterodimers among H2A-H2B and H3-H4 histones are formed via a handshake-like arrangement. These oligomerize further to form the histone octamer. Of the 146 bp of DNA wrapped around, 121 bp is organized by the histone-fold domains. Each of the total four heterodimers binds about 30 bp of DNA. This structure also revealed the remarkably ordered nature of the histone amino-terminal tails (Luger et al. 1997). The histone tails have also been

shown to interact in nucleosome to nucleosome interactions. Tails of both H3 and H2B N-terminal protrude every 20 bp. The H3 tails extend considerably beyond the nucleosomal DNA surface and are stabilized by nucleosome to nucleosome contacts (Luger et al. 1997).

Any kind of stimuli or signal entering the nucleus encounters chromatin instead of 'naked' DNA. Changes in gene expression in response to stimuli are accomplished by alterations in chromatin structure but the exact mechanism that regulates this control is still unknown. Consistent with the expectation that DNA in chromatin needs to unwind before being accessible to RNA polymerase during gene expression, earlier observations showed that nucleosomes associated with active genes were more prone to DNA modifying enzymes compared to the heterochromatic chromatin (Weintraub and Groudine 1976). Three documented ways to alter chromatin structure are: nucleosome remodeling (Aalfs and Kingston 2000), covalent modification of histone tails (Zhang and Reinberg 2001), and replacement of core histones with their variants (Ahmad and Henikoff 2002; Redon et al. 2002; Smith 2002).

Some histone modifications are involved in disruption of chromatin structure to accommodate enzymatic complexes that are involved in mechanisms like transcription, DNA repair and recombination, whereas in others a chemical group is introduced which is recognized by additional regulatory proteins resulting in altered gene expression. These changes are heritable without altering the DNA sequence and are thus called epigenetic marks (Henderson and Jacobsen 2007). These modifications alter chromatin structure by influencing either histone-histone or histone-DNA interactions (Hansen et al. 1998; Wolffe 2001). Since the biological function of specific histone modifications is conserved in living organisms, the distinct histone

modifications are known as the 'Histone Code' (Strahl and Allis 2000; Groth et al. 2007; Kouzarides 2007; Li et al. 2007). Histone modification is a dynamic process with activated genes enriched in acetylated histones and the enzymes responsible for acetylation and deacetylation both recruited to the same site to activate or repress the expression, respectively (Felsenfeld and Groudine 2003).

Histone tails are subject to a vast array of posttranslational modifications including; methylation, acetylation, ubiquitination, sumoylation and ADP-ribosylation of lysines; methylation of arginine residues and phosphorylation of serines and threonines (Kouzarides 2007). In particular, amino-acid positions 9, 14, 18 and 23 of histone H3 lysine, and H4 lysines 5, 8, 12 and 16 are targeted for modifications (Roth et al. 2001). Most common euchromatin modifications that are associated with active chromatin, are acetylation of histone 3 and histone 4 or di- or trimethylation of H3K4 (Li et al. 2007). Heterochromatin modifications or modifications related to silent chromatin are H3 K9 methylation and H3 K27 methylation (Li et al. 2007).

The distribution of modifications on coding DNA is very distinct from some modifications present towards the promoter region. Even within the coding region, the 5' and the 3' ends of open reading frame (ORF) show distinct modifications as compared to the translated part of the gene (Li et al. 2007). Indeed, a recent finding has shown that the position of modification is tightly regulated and is crucial for function. For example, methylation of histone H3K36 within the ORF leads to activation whereas the presence of this modification in the promoter region represses transcription (Landry et al. 2003).

1.10 Molecular Marker Assisted Backcross Breeding

Development of new wheat cultivars involves a lengthy and expensive process of crossing followed by cycles of selection and evaluation. It usually takes 10-15 years to develop a new wheat cultivar. Due to population shifts and/or mutations, newly released cultivars can be rendered susceptible to one or more pests. The backcross breeding method has long been a valuable strategy in plant breeding for single gene introgression (Allard 1960). The objective of backcrossing is to recover most of the recurrent parent genome (RPG) along with the trait of interest from a donor parent. During backcrossing, the removal of donor parent segments linked or unlinked to the target gene occurs due to segregation and recombination (Young and Tanksley 1989).

Generally, more than six backcrosses are required to recover 99% of the RPG along with the introgressed trait (Allard 1960). Although unlinked donor segments are reduced to half in each generation, due to linkage drag the linked segments may take many generations to get eliminated. This occurs when the donor gene of interest is closely linked to other genes and significant segments of the donor genome can still be present after many backcross generations. Theoretical calculations estimated the linkage drag size in near-isogenic lines to be 32 cM after six backcrosses (Stam and Zeven 1981). In tomato, RFLP analysis found wide variation in the size of donor segments retained around the *Tm-2* locus (Young and Tanksley 1989). In the cultivar 'Vendor', the size of the donor segments was reduced to 4 cM after 10 backcrosses compared with the cultivar Craigella where a region of 51 cM was retained even after 11 cycles of backcrosses. Traditional backcrossing is less rewarding because of linkage drag of donor segments and time taken to introgress a single trait.

In backcross breeding, molecular marker-assisted backcrossing can be a powerful tool to accelerate recovery of the RPG either by foreground selection (Tanksley 1983) in which selection is for the presence of a target allele, or by background selection (Young and Tanksley 1989) in which selection is for the recurrent parent marker loci (Tanksley et al. 1989; Melchinger 1990; Hospital and Charcosset 1997; Frisch et al. 1999a; Frisch et al. 1999b). Multiple computer simulations are available; some compare different selection strategies used in breeding progams. Others study the effects of population size and number of marker data points over backcross generations with respect to the proportion of RPG retained (Hospital and Charcosset 1997; Frisch et al. 1999a; Frisch and Melchinger 2001; Frisch and Melchinger 2005). These simulations identify the number of total backcrossing generations required for successful completion of the program, size of the population to be screened in each backcrossing generation and number, location and nature of markers to be used in the program.

Despite the availability of these simulations, high throughput marker analysis systems and user-friendly molecular markers, use of marker-assisted backcrossing has been limited (Babu et al. 2004; Steele et al. 2006). Based on the information about the distribution of genes and recombination on the wheat chromosomes, our objective was to introgress various single gene traits into different locally adapted cultivars using marker-assisted foreground selection, and to the optimize marker-assisted background selection method in wheat.

1.10.1 Molecular markers

The past two decades marked an explosion in the availability of molecular markers of various types (Gupta et al. 1999). They can be broadly classified into

three groups (i) hybridization-based DNA markers, such as restriction fragment length polymorphisms (RFLPs); (ii) polymerase chain reaction (PCR) based DNA markers such as, random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs), sequence-tagged sites (STS) and cleaved amplified polymorphic sequences (CAPS) (iii) DNA Chip and sequence based markers like single nucleotide polymorphisms (SNPs).

Hybridization-based markers were among the first developed molecular markers and were initially used in human genome mapping (Botstein et al. 1980). These are highly reliable but are time consuming and labor intensive compared to PCR-based markers which are quick, cost-effective and amenable to high throughput analysis.

Among PCR based markers, SSRs have emerged as the most useful because of their reproducibility, multiallelic nature, codominant inheritance and relative abundance in the genome (Powell et al. 1996). In wheat, SSRs were first studied by Devos et al. (1995). They searched the sequence databases and converted two simple sequence repeats into PCR-based markers. These two markers were genome-specific and displayed a high level of polymorphism. Later, Röder et al. (1995), reported the relative abundance of SSRs in wheat. $(GA)_n/(GT)_n$ are found approximately every 270 kb of DNA. Similarly, $(AC)_n$ and $(AG)_n$ microsatellites were found every 292 and 212 kb, respectively (Röder et al. 1995). More recently a detailed genetic map of microsatellite loci was made available in wheat from different sources like GWM markers (Röder et al. 1998), CFA and CFD (Sourdille et al. 2001); (Guyomarch et al. 2002), WMC (Somers et al. 2004) and BARC markers (Song et al. 2005). Application

of these markers can serve as a valuable resource for a variety of applications in plant genetics and breeding programs.

1.11 OBJECTIVES

The specific objectives are:

1. To understand how homoeologous genes in polyploid wheat learn to co-exist.

2. Examine the dosage dependence of the hexaploid wheat gene expression.

3. To optimize single gene introgression using marker-assisted backcross procedures.

1.12 References

Aalfs, J.D. and R.E. Kingston. 2000. What does 'chromatin remodeling' mean? *Trends in Biochemical Sciences* **25:** 548-555.

Adams, K.L., R. Cronn, R. Percifield, and J.F. Wendel. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* **100:** 4649-4654.

Adams, K.L., R. Percifield, and J.F. Wendel. 2004. Organ-Specific Silencing of Duplicated Genes in a Newly Synthesized Cotton Allotetraploid. *Genetics* **168:** 2217-2226.

Ahmad, K. and S. Henikoff. 2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Molecular Cell* **9:** 1191-1200.

Allard, R.W. 1960. Principles of plant breeding. *John Wiley and Sons, Inc. New York, NY*.

Aragoncillo, C., M.A. Rodríguez-Loperena, G. Salcedo, P. Carbonero, and F. García-Olmedo. 1978. Influence of homoeologous chromosomes on gene-dosage effects in allohexaploid wheat (Triticum aestivum L.). *Proceedings of the Academy of Sciences* **75:** 1446-1450.

Arumuganathan, K. and E.D. Earle. 1991. Estimation of Nuclear DNA Content of Plants by Flow Cytometry. *Plant Molecular Biology Reporter* **9**.

Axelsson, T., C.M. Bowman, A.G. Sharpe, D.J. Lydiate, and U. Lagercrantz. 2000. Amphidiploid Brassica juncea contains conserved progenitor genomes. *Genome* **43:** 679-688.

Babu, R.E., V.P. Mani, and H.S. Gupta. 2004. Combining high protein quality and hard endosperm traits through phenotypic and marker assisted selection in maize. *4th International Crop Science Congress*.

Barakat, A., N. Carels, and G. Bernardi. 1997. The distribution of genes in the genomes of Graminae. *Proceedings of the National Academy of Sciences* **94:** 6857-6861.

Barker, R.F., N.P. Harberd, M.G. Jarvis, and R.B. Flavell. 1988. Structure and evolution of the intergenic region in a ribosomal DNA repeat unit of wheat. *Journal of Molecular Biology* **201:** 1-17.

Bernstein, B.E., A. Meissner, and E.S. Lander. 2006. The epigenome. *Cell* **128:** 669-681.

Birchler, J.A. 1979. A study of enzyme activities in a dosage series of the long arm of chromosome one in maize. *Genetics* **92:** 1211-1229.

Bollag, R.J., A.S. Waldman, and R.M. Liskay. 1989. Homologous recombination in mammalian cells. *Annual Reviews of Genetics* **23:** 199-225.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal of Human Genetics* **32:** 314-331.

Brewer, G.J., C.F. Sing, and E.R. Sears. 1969. Studies of isozyme patterns in nullisomic-tetrasomic combinations of hexaploid wheat. *Proceedings of the National Academy of Sciences* **64:** 1224-1229.

Carlson, P.S. 1972. Locating genetic loci with aneuploids. *Molecular General Genetics* **114:** 273.

Cassman, K.G. 1999. Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proceedings of the National Academy of Sciences* **96:** 5952-5959.

Chen, Z.J. and C.S. Pikaard. 1997. Transcriptional analysis of nucleolar dominance in polyploid plants: Biased expression/silencing of progenitor rRNA genes is developmentally regulated in Brassica. *Proceedings of the National Academy of Sciences* **94:** 3442-3447.

Conley, E.J., V. Nduati, J.L. Gonzalez-Hernandez, A. Mesfin, M. Trudeau-Spanjers, S. Chao, G.R. Lazo, D.D. Hummel, O.D. Anderson, L.L. Qi, B.S. Gill, B. Echalier, A.M. Linkiewicz, J. Dubcovsky, E.D. Akhunov, J. Dvorak, J.H. Peng, N.L.V. Lapitan, M.S. Pathan, H.T. Nguyen, X.-F. Ma, Miftahudin, J.P. Gustafson, R.A. Greene, M.E. Sorrells, K.G. Hossain, V. Kalavacharla, S.F. Kianian, D. Sidhu, M. Dilbirligi, K.S. Gill, D.W. Choi, R.D. Fenton, T.J. Close, P.E. McGuire, C.O. Qualset, and J.A. Anderson. 2004. A 2600-Locus Chromosome Bin Map of Wheat Homoeologous Group 2 Reveals Interstitial Gene-Rich Islands and Colinearity With Rice. *Genetics* **168:** 625-637.

Curtis, C.A., A.J. Lukaszewski, and M. Chrzastek. 1991. Metaphase I pairing of deficient chromosomes and genetic mapping of deficiency breakpoints in common wheat. *Genome* **34:** 553-560.

Devos, K.M., G.J. Bryan, A.J. Collins, P. Stephenson, and M.D. Gale. 1995. Application of two microsatellite sequences in wheat storage proteins as molecular markers. *Theoretical and Applied Genetics* **90:** 247-252.

Duret, L. and P. Bucher. 1997. Searching for regulatory elements in human noncoding sequences. *Current Opinion in Structural Biology* **7:** 399-406.

Duret, L., F. Dorkeld, and C. Gautier. 1993. Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression. *Nucleic Acid Research* **21:** 2315-2322.

Dvorak, J. and K.-C. Chen. 1984. Distribution of nonstructural variation between wheat cultivars along chromosome arm 6B: Evidence from the linkage map and physical map of the arm. *Genetics* **106:** 325-333.

Endo, T.R. 1988. Induction of chromosomal structural changes by a chromosome of *Aegilops cylindrica* L. in common wheat. *Journal of Heredity* **79:** 366-370.

Endo, T.R. 1990. Gametocidal chromosomes and their induction of chromosome mutations in wheat. *Japanese Journal of Genetics* **65:** 135-152.

Endo, T.R. and B.S. Gill. 1996. The deletion stocks of common wheat. *Journal of Heredity* **87:** 295-307.

Epstein, C.J. 1986. The Consequences of Chromosome Imbalance. *Cambridge University Press, New York.*

Erayman, M., D. Sandhu, D. Sidhu, M. Dilbirligi, P.S. Baenziger, and K.S. Gill. 2004. Demarcating the gene-rich regions of the wheat genome. *Nucleic Acids Research* **32:** 3546-3565.

Felsenfeld, G. and M. Groudine. 2003. Controlling the double helix. *Nature* **421:** 448-453.

Feuillet, C. and B. Keller. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. *Proceedings of the National Academy of Sciences* **96:** 8265-8270.

Force, A., M. Lynch, F.B. Pickett, A. Amores, Y.-l. Yan, and J. Postlethwait. 1999. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151:** 1531-1545.

Frisch, M., M. Bohn, and A. Melchinger. 1999a. Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. *Crop Science* **39:** 967-975.

Frisch, M., M. Bohn, and A.E. Melchinger. 1999b. Comparison of Selection Strategies for Marker-Assisted Backcrossing of a Gene. *Crop Science* **39:** 1295-1301.

Frisch, M. and A.E. Melchinger. 2001. The Length of the Intact Donor Chromosome Segment Around a Target Gene in Marker-Assisted Backcrossing. *Genetics* **157:** 1343-1356.

Frisch, M. and A.E. Melchinger. 2005. Selection Theory for Marker-Assisted Backcrossing. *Genetics* **170:** 909-917.

Fufa, H., S. Baenziger, P. , B.S. Beecher, R.A. Graybosch, K.M. Eskridge, and L.A. Nelson. 2005. Genetic Improvement Trends in Agronomic Performances and End-use Quality Characteristics Among Hard Red Winter Wheat Cultivars in Nebraska. *Euphytica* **144:** 187-198.

Galili, G., A.A. Levy, and M. Feldman. 1986. Gene-Dosage Compensation of Endosperm Proteins in Hexaploid Wheat Triticum aestivum. *Proceedings of the National Academy of Sciences* **83:** 6524-6528.

Gill, K.S., B.S. Gill, and T.R. Endo. 1993. A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. *Chromosoma* **102:** 374-381.

Gill, K.S., B.S. Gill, T.R. Endo, and E. Boyko. 1996a. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics* **143:** 1001-1012.

Gill, K.S., B.S. Gill, T.R. Endo, and E.V. Boyko. 1996b. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics* **143:** 1001-1012.

Gill, K.S., B.S. Gill, T.R. Endo, and T. Taylor. 1996c. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics* **144:** 1883-1891.

Goll, M.G. and T.H. Bestor. 2005. Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry* **74:** 481-514.

Grell, E.H. 1962. The dose effect of ma-l+ and ry+ on xanthine dehydrogenase activity in Drosophila melanogaster. *Z. Vererbungsl.* **93:** 371-377.

Groth, A., W. Rocha, A. Verreault, and G. Almouzni. 2007. Chromatin challenges during DNA replication and repair. *Cell* **128:** 721-733.

Guo, M., D. Davis, and J.A. Birchler. 1996. Dosage Effects on Gene Expression in a Maize Ploidy Series. *Genetics* **142:** 1349-1355.

Gupta, P.K., R.K. Varshney, S.P. C., and B. Ramesh. 1999. Molecular markers and their applications in wheat breeding. *Plant Breeding* **118:** 369-390.

Guyomarch, H., P. Sourdille, G. Charmet, K.J. Edwards, and M. Bernard. 2002. Characterization of polymorphic microsatellite markers from Aegilops tauschii and their transferability to the D-genome of bread wheat. *Theoretical and Applied Genetics* **104:** 1164-1172.

Haldane, J.B.S. 1933. The part played by recurrent mutation in evolution. *The American Naturalist* **67:** 5.

Han, F., G. Fedak, T. Ouellet, and B. Liu. 2003. Rapid genomic changes in interspecific and intergeneric hybrids and allopolyploids of Triticeae. *Genome* **46:** 716-723.

Hansen, J.C., C. Tse, and A.P. Wolffe. 1998. Structure and Function of the Core Histone N-termini: More than Meets the Eye. *Biochemistry* **37:** 17637-17641.

Heitz, E. 1928. Das Heterochromatin der Moose. *Jahrbuch Wiss Botanik* **69:** 762-818.

Henderson, I.R. and S.E. Jacobsen. 2007. Epigenetic inheritance in plants. *Nature* **447:** 418-424.

Holland, P.W., J. Garcia-Fernandez, N.A. Williams, and A. Sidow. 1994. Gene duplication and the origins of vertebrate development. *Development (supplement)***:** 125-133.

Hospital, F. and A. Charcosset. 1997. Marker-Assisted Introgression of Quantitative Trait Loci. *Genetics* **147:** 1469-1485.

Hossain, K.G., V. Kalavacharla, G.R. Lazo, J. Hegstad, M.J. Wentz, P.M.A. Kianian, K. Simons, S. Gehlhar, J.L. Rust, R.R. Syamala, K. Obeori, S. Bhamidimarri, P. Karunadharma, S. Chao, O.D. Anderson, L.L. Qi, B. Echalier, B.S. Gill, A.M. Linkiewicz, A. Ratnasiri, J. Dubcovsky, E.D. Akhunov, J. Dvorak, Miftahudin, K. Ross, J.P. Gustafson, H.S. Radhawa, M. Dilbirligi, K.S. Gill, J.H. Peng, N.L.V. Lapitan, R.A. Greene, C.E. Bermudez-Kandianis, M.E. Sorrells, O. Feril, M.S. Pathan, H.T. Nguyen, J.L. Gonzalez-Hernandez, E.J. Conley, J.A. Anderson, D.W. Choi, D. Fenton, T.J. Close, P.E. McGuire, C.O. Qualset, and S.F. Kianian. 2004. A Chromosome Bin Map of 2148 Expressed Sequence Tag Loci of Wheat Homoeologous Group 7. *Genetics* **168:** 687-699.

Houchins, K., MO'Dell, R.B. Flavell, and J.P. Gustafson. 1997. Cytosine methylation and nucleolar dominance in cereal hybrids. *Molecular and General Genetics* **255:** 294-301.

Huang, S., A. Sirikhachornkit, X. Su, J. Faris, B. Gill, R. Haselkorn, and P. Gornicki. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences* **99:** 8133-8138.

Hughes, A. 1994. The evolution of functionally novel proteins after gene duplication. *Proceedings: Biological Sciences* **256:** 119-124.

Jauhar, P.P., O. Riera-Lizarazu, W.G. Dewey, B.S. Gill, C.F. Crane, and J.H. Bennett. 1991. Chromosome pairing relationships among the A, B, and D genomes of bread wheat. *Theoretical and Applied Genetics* **82:** 441-449.

Kashkush, K., M. Feldman, and A.A. Levy. 2002. Gene Loss, Silencing and Activation in a Newly Synthesized Wheat Allotetraploid. *Genetics* **160:** 1651-1659.

Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E.S. Lander. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423:** 241-254.

Kouzarides, T. 2007. Chromatin Modifications and Their Function. *Cell* **128:** 693-705.

Landry, J., A. Sutton, T. Hesman, J. Min, R.M. Xu, M. Johnston, and R. Sternglanz. 2003. Set2-catalyzed methylation of histone H3 represses basal expression of GAL4 in Saccharomyces cerevisiae. *Molecular Cell Biology* **23:** 5972-5978.

Larizza, A., W. Makalowski, G. Pesole, and C. Saccone. 2002. Evolutionary dynamics of mammalian mRNA untranslated regions by comparative analysis of orthologous human, artiodactyl and rodent gene pairs. *Journal of Computational Chemistry* **26:** 479-490.

Lewis, E.B. 1950. The phenomenon of position effect. *Advances in Genetics* **3:** 73-115.

Li, B., M. Carey, and J. Workman. 2007. The Role of Chromatin during Transcription. *Cell* **128:** 707-719.

Lichten, M. and A.S. Goldman. 1995. Meiotic recombination hotspots. *Annual Reviews of Genetics* **29:** 423-444.

Linkiewicz, A.M., L.L. Qi, B.S. Gill, A. Ratnasiri, B. Echalier, S. Chao, G.R. Lazo, D.D. Hummel, O.D. Anderson, E.D. Akhunov, J. Dvorak, M.S. Pathan, H.T. Nguyen, J.H. Peng, N.L.V. Lapitan, Miftahudin, J.P. Gustafson, C.M. La Rota, M.E. Sorrells, K.G. Hossain, V. Kalavacharla, S.F. Kianian, D. Sandhu, S.N. Bondareva, K.S. Gill, E.J. Conley, J.A. Anderson, R.D. Fenton, T.J. Close, P.E. McGuire, C.O. Qualset, and J. Dubcovsky. 2004. A 2500-Locus Bin Map of Wheat Homoeologous Group 5 Provides Insights on Gene Distribution and Colinearity With Rice. *Genetics* **168:** 665-676.

Lipman, D. 1997. Making (anti)sense of non-coding sequence conservation. *Nucleic Acids Research* **25:** 3580-3583.

Liu, B., C.L. Brubaker, G. Mergeai, R.C. Cronn, and J.F. Wendel. 2001. Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* **44:** 321-330.

Liu, B., J.M. Vega, and M. Feldman. 1998. Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences. *Genome* **41:** 535-542.

Liu, B. and J.F. Wendel. 2002. Non-Mendelian phenomenon in allopolyploid genome evolution. *Current Genomics* **3:** 489-505.

Luger, K., A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389:** 231-233.

Lynch, M. and A. Force. 2000. The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics* **154:** 459-473.

Macleod, D. and A. Bird. 1982. DNAase I sensitivity and methylation of active versus inactive rRNA genes in Xenopus species hybrids. *Cell* **29:** 211-218.

Martienssen, R.A. and V. Colot. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293:** 1070-1074.

Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264:** 421-424.

McGarvey, K.M., E. Greene, J.A. Fahrner, T. Jenuwein, and B.S. B. 2007. DNA methylation and complete transcriptional silencing of cancer genes persist after depletion of EZH2. *Cancer Research* **67:** 5097-5102.

McClintock, J.M., M.A. Kheirbek, and V.E. Prince. 2002. Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development* **129:** 2339–2354.

Melchinger, A.E. 1990. Use of molecular markers in breeding for oligogenic disease resistance. *Plant Breeding* **104:** 1-19.

Mewes, H.W., K. Albermann, M. Bahr, D. Frishman, A. Gleissner, J. Hani, K. Heumann, K. Kleine, A. Maierl, S.G. Oliver, F. Pfeiffer, and A. Zollner. 1997. Overview of the yeast genome. *Nature* **387:** 7-65.

Miftahudin, K. Ross, X.-F. Ma, A.A. Mahmoud, J. Layton, M.A.R. Milla, T. Chikmawati, J. Ramalingam, O. Feril, M.S. Pathan, G.S. Momirovic, S. Kim, K. Chema, P. Fang, L. Haule, H. Struxness, J. Birkes, C. Yaghoubian, R. Skinner, J. McAllister, V. Nguyen, L.L. Qi, B. Echalier, B.S. Gill, A.M. Linkiewicz, J. Dubcovsky, E.D. Akhunov, J. Dvorak, M. Dilbirligi, K.S. Gill, J.H. Peng, N.L.V. Lapitan, C.E. Bermudez-Kandianis, M.E. Sorrells, K.G. Hossain, V. Kalavacharla, S.F. Kianian, G.R. Lazo, S. Chao, O.D. Anderson, J. Gonzalez-Hernandez, E.J. Conley, J.A. Anderson, D.-W. Choi, R.D. Fenton, T.J. Close, P.E. McGuire, C.O. Qualset, H.T. Nguyen, and J.P. Gustafson. 2004. Analysis of Expressed Sequence Tag Loci on Wheat Chromosome Group 4. *Genetics* **168:** 651-663.

Mohan, J. and F.M. Ritossa. 1970. Regulation of ribosomal RNA synthesis and its bearing on the bobbed phenotype in Drosophila melanogaster. *Developmental Biology* **22:** 495-512.

Muller, H.J. 1950. Evidence of the precision of genetic adaptation. *Harvey Lectures* **43:** 165-229.

Munkvold, J.D., R.A. Greene, C.E. Bermudez-Kandianis, C.M. La Rota, H. Edwards, S.F. Sorrells, T. Dake, D. Benscher, R. Kantety, A.M. Linkiewicz, J. Dubcovsky, E.D. Akhunov, J. Dvorak, Miftahudin, J.P. Gustafson, M.S. Pathan, H.T. Nguyen, D.E. Matthews, S. Chao, G.R. Lazo, D.D. Hummel, O.D. Anderson, J.A. Anderson, J.L. Gonzalez-Hernandez, J.H. Peng, N. Lapitan, L.L. Qi, B. Echalier, B.S. Gill, K.G. Hossain, V. Kalavacharla, S.F.

Kianian, D. Sandhu, M. Erayman, K.S. Gill, P.E. McGuire, C.O. Qualset, and M.E. Sorrells. 2004. Group 3 Chromosome Bin Maps of Wheat and Their Relationship to Rice Chromosome 1. *Genetics* **168:** 639-650.

Nomura, T., A. Ishihara, R.C. Yanagita, T.R. Endo, and H. Iwamura. 2005. Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proceedings of the National Academy of Sciences* **102:** 16490-16495.

Ozkan, H., A. Levy, and M. Feldman. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (Aegilops-Triticum) group. *Plant Cell* **13:** 1735-1747.

Peng, J.H., H. Zadeh, G.R. Lazo, J.P. Gustafson, S. Chao, O.D. Anderson, L.L. Qi, B. Echalier, B.S. Gill, M. Dilbirligi, D. Sandhu, K.S. Gill, R.A. Greene, M.E. Sorrells, E.D. Akhunov, J. Dvorak, A.M. Linkiewicz, J. Dubcovsky, K.G. Hossain, V. Kalavacharla, S.F. Kianian, A.A. Mahmoud, Miftahudin, E.J. Conley, J.A. Anderson, M.S. Pathan, H.T. Nguyen, P.E. McGuire, C.O. Qualset, and N.L.V. Lapitan. 2004. Chromosome Bin Map of Expressed Sequence Tags in Homoeologous Group 1 of Hexaploid Wheat and Homoeology With Rice and Arabidopsis. *Genetics* **168:** 609-623.

Pesole, G., F. Mignone, C. Gissi, G. Grillo, F. Licciulli, and S. Liuni. 2001. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* **276:** 73-81.

Piatigorsky, J. and G. Wistow. 1991. The recruitment of crystallins: new functions precede gene duplication. *Science* **252:** 1078–1079.

Pikaard, C.S. 1999. Nucleolar dominance and silencing of transcription. *Trends in Plant Science* **4:** 478-483.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2:** 225-238.

Randhawa, H.S., M. Dilbirligi, D. Sidhu, M. Erayman, D. Sandhu, S. Bondareva, S. Chao, G.R. Lazo, O.D. Anderson, Miftahudin, J.P. Gustafson, B. Echalier, L.L. Qi, B.S. Gill, E.D. Akhunov, J. Dvorak, A.M. Linkiewicz, A. Ratnasiri, J. Dubcovsky, C.E. Bermudez-Kandianis, R.A. Greene, M.E. Sorrells, E.J. Conley, J.A. Anderson, J.H. Peng, N.L.V. Lapitan, K.G. Hossain, V. Kalavacharla, S.F. Kianian, M.S. Pathan, H.T. Nguyen, T.R. Endo, T.J. Close, P.E. McGuire, C.O. Qualset, and K.S. Gill. 2004. Deletion Mapping of Homoeologous Group 6-Specific Wheat Expressed Sequence Tags. *Genetics* **168:** 677-686.

Redon, C., D. Pilch, E. Rogakou, O. Sedelnikova, K. Newrock, and W. Bonner. 2002. Histone H2A variants H2AX and H2AZ. *Current Opinion in Genetics and Development* **12:** 162-169.

Reeves, R. 1977. Structure of Xenopus ribosomal RNA gene chromatin during changes in genomic transcription rates. *Cold Spring Harbor Symposia on Quantative Biology* **42:** 709-791.

Richmond, T.J., T. Finch J, B. Rushton, D. Rhodes, and A. Klug. 1984. Structure of the nucleosome core particle at 7A resolution. *Nature* **311:** 532-537.

Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution* **16:** 351-358.

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

Roder, M.S., J. Plaschke, S.U. Konig, A. Borner, M.E. Sorrells, S.D. Tanksley, and M.W. Ganal. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. *Molecular General Genetics* **246:** 327-333.

Roth, S.Y., J.M. Denu, and C.D. Allis. 2001. Histone acetyltransferases. *Annual Review of Biochemistry* **70:** 81-120.

Sandhu, D., J.A. Champoux, S.N. Bondareva, and K.S. Gill. 2001. Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group *1S* chromosomes. *Genetics* **157:** 1735-1747.

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

SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova, and J.L. Bennetzen. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274:** 765-768.

Sasanuma, T., N.T. Miyashita, and K. Tsunewaki. 1996. Wheat phylogeny determined by RFLP analysis of nuclear DNA. 3. Intra- and interspecific variations of five Aegilops Sitopsis species. *Theoretical and Applied Genetics* **92:** 928-934.

Schultz, J. 1936. Variegation in Drosophila and the Inert Chromosome Regions. *Proceedings of the National Academy of Sciences* **22:** 27-33.

Schultz, J. and T. Casperson. 1939. Heterochromatic regions and the nucleic acid metabolism of the chromosomes. *Arch Exp Zellforsch* **22:** 650-654.

Schwartz, D. 1971. Genetic control of alcohol dehydrogenase--a competition model for regulation of gene action. *Genetics* **67:** 411-425.

Sears, E.R. 1954. The aneuploids of common wheat. *Research Bulletin, University of Missouri Agricultural Experiment Station***. 572:** 1-58.

Sears, E.R. and L.M.S. Sears. 1978. The telocentric chromosomes of common wheat. In *Proceedings of the 5th International Wheat Genetics Symposium* (ed. S. Ramanujam). 389-407.

Shabalina, S.A., A.Y. Ogurtsov, I.B. Rogozin, E.V. Koonin, and D.J. Lipman. 2004. Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. *Nucleic Acids Research* **32:** 1774-1782.

Shaked, H., K. Kashkush, H. Ozkan, M. Feldman, and A. Levy. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* **13:** 1749-1759.

Sidhu, D. and K.S. Gil. 2004. Distribution Of Genes And Recombination In Wheat And Other Eukaryotes. *Plant Cell Tissue and Organ culture* **79:** 257-270.

Sidow, A. 1996. Gen(om)e duplications in the evolution of early vertebrates. *Current Opinion in Genetics & Development* **6:** 715-722.

Sing, C.F. and G.J. Brewer. 1969. Isozymes of a polyploid series of wheat. *Genetics* **61:** 391-398.

Smith, M.M. 2002. Centromeres and variant histones: what, where, when and why? *Current Opinion in Cell Biology* **14:** 279-285.

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

Song, K., P. Lu, K. Tang, and T. Osborn. 1995. Rapid Genome Change in Synthetic Polyploids of Brassica and Its Implications for Polyploid Evolution. *Proceedings of the National Academy of Sciences* **92:** 7719-7723.

Song, Q.J., J.R. Shi, S. Singh, E.W. Fickus, J.M. Costa, J. Lewis, B.S. Gill, R. Ward, and P.B. Cregan. 2005. Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical and Applied Genetics* **110:** 550-560.

Sourdille, P., H. Guyomarch, C. Baron, B. Gandon, V. Chiquet, F. Artiguenave, K. Edwards, N. Foisset, P. Dufour, and M. Bernard. 2001. Wheat Genoplante SSR mapping data release: a new set of markers and comprehensive Genetic and physical mapping data.

Stam, P. and A.C. Zeven. 1981. The theoretical proportion of the donor genome in near-isogeneic lines of self-fertilizers bred by backcrossing. *Euphytica* **30:** 227-238.

Steele, K.A., A.H. Price, H.E. Shashidhar, and J.R. Witcombe. 2006. Marker-assisted selection to introgress of rice QTLs controlling root traits and aroma into an Indian upland rice variety. *Theoretical and Applied Genetics* **112:** 327-335.

Stoltzfus, A. 1999. On the possibility of constructive neutral evolution. *Journal of Molecular Evolution* **49:** 169-181.

Strahl, B.D. and C.D. Allis. 2000. The language of covalent histone modifications. *Nature* **403:** 41-45.

Tanksley, S.D. 1983. Molecular markers in plant breeding. *Plant Molecular Biology Reporter* **1:** 3-8.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. *Bio/Technology* **7:** 257-264.

Tranquilli, G., D. Lijavetzky, G. Muzzi, and J. Dubcovsky. 1999. Genetic and physical characterization of grain texture-related loci in diploid wheat. *Molecular General Genetics* **262:** 846-850.

Vandenbussche, M., G. Theissen, Y. Van de Peer, and T. Gerats. 2003. Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Research* **31:** 4401–4409.

Walsh, J.B. 1985. How many processed pseudogenes are accumulated in a gene family? *Genetics* **110:** 345-364.

Watterson, G.A. 1983. On the time for gene silencing at duplicate loci. *Genetics* **105:** 745-766.

Weinreich, D.M., N.F. Delaney, M.A. DePristo, and D.L. Hartl. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science* **312:** 111-114.

Weintraub, H. and M. Groudine. 1976. Chromosomal subunits in active genes have an altered conformation. *Science* **193:** 848-856.

Wendel, J.F. 2000. Genome evolution in polyploids. *Plant Molecular Biology* **42:** 225-249.

Wolffe, A.P. 2001. Transcriptional regulation in the context of chromatin structure. *Essays in Biochemistry* **37:** 45-57.

Young, N.D. and S.D. Tanksley. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theoretical and Applied Genetics* **77:** 95-101.

Yu, J., S. Hu, J. Wang, G.K. Wong, S. Li, B. Liu, Y. Deng, L. Dai, Y. Zhou, X. Zhang, M. Cao, J. Liu, J. Sun, J. Tang, Y. Chen, X. Huang, W. Lin, C. Ye, W.

Tong, L. Cong, J. Geng, Y. Han, L. Li, W. Li, G. Hu, J. Li, Z. Liu, Q. Qi, T. Li, X. Wang, H. Lu, T. Wu, M. Zhu, P. Ni, H. Han, W. Dong, X. Ren, X. Feng, P. Cui, X. Li, H. Wang, X. Xu, W. Zhai, Z. Xu, J. Zhang, S. He, J. Xu, K. Zhang, X. Zheng, J. Dong, W. Zeng, L. Tao, J. Ye, J. Tan, X. Chen, J. He, D. Liu, W. Tian, C. Tian, H. Xia, Q. Bao, G. Li, H. Gao, T. Cao, W. Zhao, P. Li, W. Chen, Y. Zhang, J. Hu, S. Liu, J. Yang, G. Zhang, Y. Xiong, Z. Li, L. Mao, C. Zhou, Z. Zhu, R. Chen, B. Hao, W. Zheng, S. Chen, W. Guo, M. Tao, L. Zhu, L. Yuan, and H. Yang. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science* **296:** 79-92.

Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S. Chan, H. Chen, I. Henderson, P. Shinn, M. Pellegrini, S. Jacobsen, and J.R. Ecker. 2006. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell* **126:** 1189-1201.

Zhang, Y. and D. Reinberg. 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & Development* **15:** 2343-2360.

Zilberman, D., M. Gehring, R.K. Tran, T. Ballinger, and S. Henikoff. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics* **39:** 61-69.

CHAPTER 2

Gene expression balance among homoeologues and its interdependence on gene dosage in polyploid wheat

Jasdeep S. Mutti, Deepak Sidhu[1], and Kulvinder S. Gill*

Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420.

[1] Current address: NCI-Frederick Building, 539/154 PO Box B, Frederick, MD 21702

**Corresponding author*

Kulvinder S. Gill

Department of Crop & Soil Sciences

277 Johnson Hall, P. O. Box 646420

Washington State University, Pullman, WA 99163

Email: ksgill@wsu.edu

Tel. (509) 335-4666, Fax: (509) 335-8674

2.1 Abstract

Gene duplication by polyploidy (homoeologues) or other means (paralogues) is a prominent feature of angiosperm evolution. We studied gene expression among three homoeologues of hexaploid wheat that evolved from a common progenitor about 3 million years ago (MYA) and came into a common nucleus at different times: ~0.5 and 0.01 MYA. Gene expression corresponding to each homoeologue was identified by sequence comparison of cultivar 'Chinese spring' (CS) ESTs and the results were confirmed by SSCP analysis of RNA using nulli-tetra lines. Of the 632 genes analyzed, 58% were expressed from all three homoeologues, 33% from two, and only 9% were expressed from one of the three homoeologues. The largest percentage of genes (14%) were expressed in anthers and the least (7%) were expressed in pistils. Whereas, the highest number of homoeologues/gene were expressed in roots (1.72 out of 3 homoeologues) and the lowest number were expressed from anthers (1.03 out of 3 homoeologues). In general, the proportion of expressed copies decreased with the increase in homoeologue copy number. The most significant observation was that homoeologues for 87% of the genes showed different expression patterns in different tissues and thus have likely evolved different gene expression controls. About 30% of the genes showed dosage dependence as the expression of homoeologues changed in response to changes in structural copy number.

2.2 Introduction

Polyploidy or the whole genome duplication has been proposed as the primary force shaping the evolution of flowering plants (Ohno 1970; Grant 1981). In addition to the 50 to 70% of the flowering plants that are believed to be polyploids, (Stebbins 1966; Grant 1981) DNA sequence analyses have suggested that even the diploid

organisms such as *Arabidopsis*, yeast and humans may in fact have undergone cycles of polyploidization (Henikoff 1997; The Arabidopsis Genome Initiative 2000; Venter et al. 2001).

Characterized as genome-wide duplication of either the same (autopolyploidy) or related genomes (allopolyploidy), polyploidization provides a massive reservoir for the evolution of novel gene functions (Grant 1981; Lynch and Force 2000). Doubling of genetic material in a nucleus comes as a genomic shock to the cell and it must respond to maintain the gene expression balance required for its survival and propagation in a dynamic environment. The fate of duplicated genes following polyploidization may include silencing of one of the duplicated copies (*nonfunctionalization)*, divergence of duplicated genes to result in new functions (*neofunctionalization),* reduction of the combined expression of the two copies to the level of a single copy gene, or acquisition of different tissue-specificities for gene expression (*subfunctionalization)* (Ohno 1970; Lynch and Force 2000).

Changes in gene expression leading to non-functionalization or sub-functionalization, have been reported to be a major consequence of polyploidization in Arabidopsis (Wang et al. 2004) and cotton allopolyploids (Adams et al. 2003b; Adams et al. 2004). Sequence elimination, appearance of novel genomic sequences and inactivation have been reported in polyploid wheat (Feldman et al. 1997; Shaked et al. 2001; Mochida et al. 2003) and *Tragopogon* (Tate et al. 2006), whereas chromosomal translocation and transposition are more common in *Brassica* allopolyploids (Song et al. 1995). These unexpected major differences in the fate of duplicated genes among different plants may be a result of there studies being based on artificially synthesized polyploids (Song et al. 1995; Feldman et al. 1997; Shaked et al. 2001) or are based on the study of very few genes (Sasakuma et al. 1995; Song

et al. 1995; Adams et al. 2004). The fate of duplicated genes in naturally occurring polyploids is largely unknown.

Tissue-specific differential expression of homoeologues has been documented. Of the 40 selected candidates, homoeologues for 10 allotetraploid cotton genes showed differential expression pattern in different tissues (Adams et al. 2003b). Homoeologues of the *TaBx* gene of wheat also showed a similar pattern of differential expression, as the expression of the B genome copy was higher in shoot as compared to the other two homoeologues (Nomura et al. 2005). The proportion of polyploid genes that follow this pattern of evolution is, however, not known.

Some reports suggest that at least for some genes the expression of homoeologues is regulated in a concerted manner. Deviating from the expected linear increase in expression, gene expression of some duplicated genes change in response to change in homoeologous copy number (Grell 1962; Epstein 1986), whereas for others the cumulative expression of the duplicated copies stays at the diploid level (Schwartz 1971; Carlson 1972). First documented for sex-linked genes in *Drosophila* (Muller 1950), dosage compensation has since been shown for many genes both in plants and animals, including multigenic families of rDNA (Mohan and Ritossa 1970; Reeves 1977), and for 11 of the 12 isozymes that were studied in wheat using nullisomic-tetrasomic lines (Brewer et al. 1969). In the latter study, only the alkaline phosphatase gene showed increased gene expression with an extra dose, whereas in other cases, the effect was compensated. Similarly, the alcohol dehydrogenase gene of maize showed dosage compensation in one to four dosage series of the chromosome arm on which this gene is located (Birchler 1979). However, in a whole genome ploidy series involving monoploids, diploids, and tetraploids, the transcript level of this gene increased proportionally with the increased ploidy level (Guo et al. 1996).

Non-linear changes in homoeologous gene expression have also been reported in wheat. Compared to the normal tetraploid, net protein output for a gene was 30-80% higher when one of the homoeologues was missing (Aragoncillo et al. 1978). On the other hand, the amount of the wheat seed storage proteins increased with the increase in gene dosage (Galili et al. 1986). The proportion of genes showing each of these expression patterns is not known in any of the natural polyploids. Also not known is the mechanism that dictates the expression pattern of different genes in response to polyploidization.

Distribution of genes and recombination is highly uneven along the eukaryotic chromosomes (Sidhu and Gill 2004). In wheat and other grasses, genes are organized in clusters of varying sizes with the gene rich regions (GRRs) interspersed with gene-poor compartments predominantly consisting of retrotransposon-like repetitive DNA sequences (SanMiguel et al. 1996; Barakat et al. 1997; Feuillet and Keller 1999; Sidhu and Gill 2004). Uneven distribution of genes is a common feature of all eukaryotes although the unevenness is more pronounced in species with larger genomes. Distribution of recombination is also highly uneven in eukaryotes and the recombination is usually limited to genic regions (Dooner 1986; Lahaye et al. 1998; Wei et al. 1999; Stein et al. 2000; Tarchini et al. 2000; Brueggeman et al. 2002; Sidhu and Gill 2004). These observations suggest that different chromosomal regions are structurally and functionally distinct. The relationship between this uneven distribution of genes and recombination with the differential fate of duplication has not yet been established.

The effects of polyploidization on gene expression has been studied for synthetic polyploids, but the concerted evolution of homoeologues in natural polyploids, the interdependence of homoeologous gene expression and the

corresponding mechanisms controlling these processes, remain largely unexplored. The main questions about stabilization of a polyploid are: 1) is homoeologous gene expression interdependent or dependent? and 2) are the duplicated copies inactivated because they are redundant or do they develop enough functional differences to be retained due to natural selection? In allopolyploids, gene silencing has been reported to be a major consequence of polyploidization. 3) Does wheat also follows this pattern of homoeologous gene silencing or not?

In this study, we chose polyploid wheat to address these questions. Wheat, *Triticum aestivum* L. ($2n = 6x = 42$), is a natural allohexaploid with three relatively collinear genomes, designated A, B, and D. Two independent polyploidization events led to the evolution of hexaploid wheat. The first polyploidization step occurred about 500,000 years ago from a hybridization event between *T. urartu* Tumanian ex Gandilyan ($2n$=14, genome AA) and *Aegilops speltoides* Tausch ($2n$=$2x$=14, genome SS≈BB) or its closely related species, giving rise to allotetraploid *T. turgidum* L. (McFadden and Sears 1946; Huang et al. 2002b). The hexaploid wheat evolved from a hybridization event between *T. turgidum* L.; $2n = 28$, AABB) and a diploid species *A. tauschii* (Coss.) Schmal ($2n = 14$, DD), about 8,000 years ago (Kihara 1944). Furthermore, the diploid progenitors of hexaploid wheat are extant and the phylogenetic relationship with its diploid progenitors is comparatively well understood (Sasanuma et al. 1996, Huang et al. 2002a) Hexaploid wheat also has a wealth of aneuploid stocks available (Sears 1954; Sears and Sears 1978; Endo 1988; Endo and Gill 1996) that are ideal to understand the functional organization of the wheat genome (for example to study homoeologous gene expression balance) and to analyze the effect of gene dosage on gene expression.

2.3 Results

2.3.1 Expression pattern of the homoeologues

Gene expression corresponding to each homoeologue was studied by sequence comparison of cultivar 'Chinese spring' (CS) ESTs. Unique ESTs corresponding to 9,400 genes have been physically mapped to wheat chromosomes by gel-blot analysis of 164 wheat aneuploid lines including nullitetra, ditelosomic and deletion lines (http://wheat.pw.usda.gov/NSF/). Of these, gel-blot analysis images of 6000 that detected less than six bands were evaluated to select 854 ESTs that showed a clear hybridization pattern and where every restriction fragment band was physically mapped. The selected ESTs were compared with the full-length rice cDNAs and the rice homologues were identified by pairwise comparison using the Blast algorithm at a cut off e-value of e^{-70} (KOME, http://cdna01.dna.affrc.go.jp/cDNA/). The selected wheat ESTs and the rice full-length cDNAs were compared with a 'Chinese spring' (CS) wheat specific EST database using the megablast algorithm (http://www.ncbi.nlm.nih.gov/BLAST/) and were aligned with the 'ContigExpress' Module of the *Vector NTI Advance*™ 8.0 software. A cutoff value of ≥80% identity within minimum overlap of 30 bp was used for wheat-rice comparisons. These two comparisons yielded CS wheat specific EST contigs for each of the selected genes.

An example of the analysis and the approach to identify ESTs corresponding to each homoeologue is given in Figure 2.1a. For example, the EST *BE497160* (marked red) was physically mapped between FL 0.71 and 1.00 on the long arm of wheat homoeologous group *4* (http://wheat.pw.usda.gov/NSF/). The corresponding gene has three structural copies in the wheat genome as the deletion mapping detected three fragment bands, one on each of the group *4* homoeologue. Using megablast, 24 ESTs with more than 80% homology (green bars) were identified, upon comparison

with the CS EST database. A similar comparison using the full-length rice cDNA homologoue (Figure 2.1a, orange bar) identified additional 11 wheat ESTs (Figure 2.1a, blue bars) to extend the contig across the entire gene. Since this gene was represented by three distinct cDNA sequence patterns and all ESTs were from the same cultivar 'Chinese spring', we conclude that all three copies of this gene are expressed (Figure 2.1c). This approach was used to study the expression pattern of the 854 putative wheat genes. Full-length cDNA of rice homologues for 138 of these putative wheat genes were identified and were used for the analysis. Out of the 854 genes, 309 (36%) had less than five ESTs in their contigs, and were not included in the analysis. In order to account for sequencing errors, a sequence pattern was considered to be unique only if it was present in two or more ESTs and the differences/similarities among the ESTs within a pattern is consistent along the length of the assembled contig.

A summary of *in silico* expression analysis results for 854 wheat genes is given in Table 2.1. It was interesting to notice that genes with a higher copy number showed a higher proportion of genes with less than five ESTs. This proportion was highest (43/56, 76%) for genes with five structural copies followed by genes with four copies (46/76, 60%) (Figure 2.2). The least proportion showing less than five ESTs were the genes with three structural copies (194/632, 30%). The number of ESTs identified per structural copy also showed a range, with least number of EST/gene (4) for genes with five structural copies and maximum number in genes with one structural copy (30).

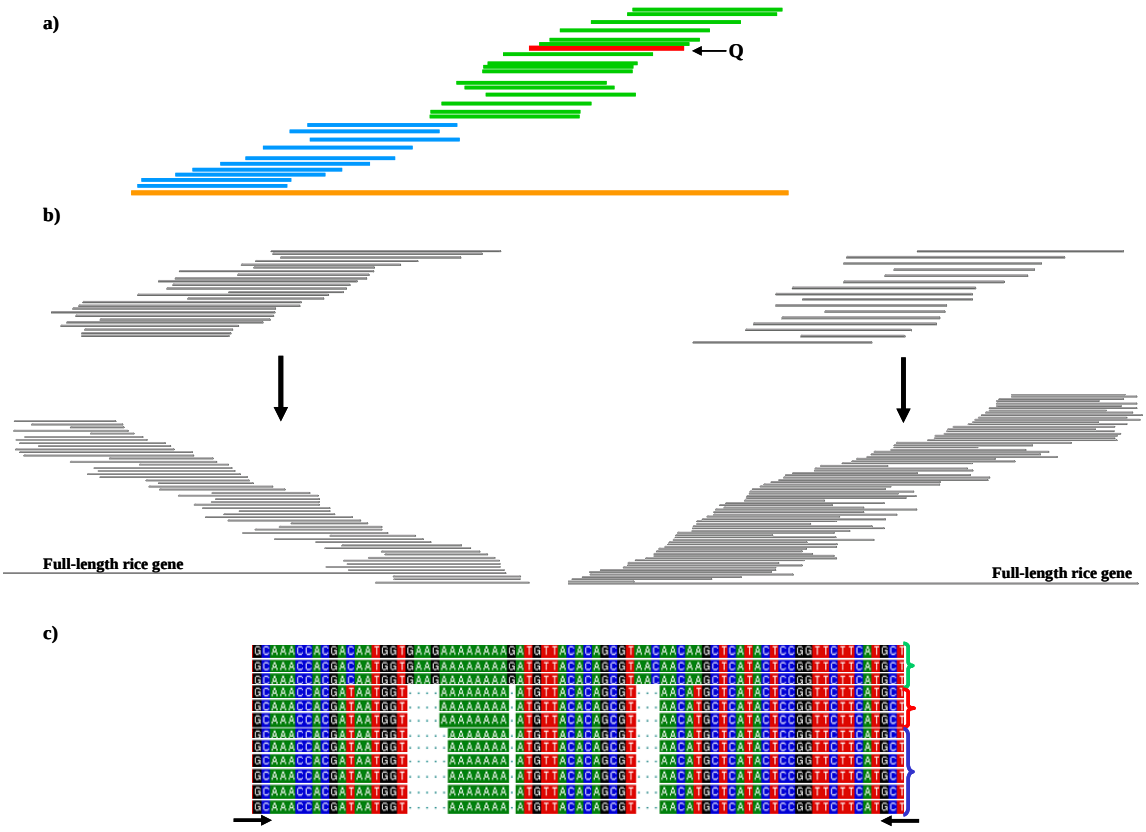


Figure 2.1. (a) Contig assembly of wheat homolog ESTs. Green bars denote the ESTs identified using wheat EST as a query (Q), whereas blue bars represent the additional wheat ESTs identified for the gene using rice full-length (orange bar) cDNA sequence. **(b)** Examples of genes, *BE352570* (Left)and *BE637763* (right), comparing the sequence alignment of ESTs, assembled using only wheat EST sequence as query, with alignment of wheat ESTs using full-length rice cDNA sequence as query. **(c)** Sequence alignment of the ESTs within the contig. ESTs corresponding to each of the three wheat homoeologues were distinguished based on different patterns as bracketed on right side by green, red and blue brackets. Brackets of different color represent ESTs of different patterns, whereas brackets of the same color represent ESTs with one pattern. Black arrows in the bottom represent the conserved region among the three patterns with no sequence differences, used to design primers for SSCP analysis.

Table 2.1. Expression pattern of homoeologues in hexaploid wheat**.**

Number of Structural copies	Number of genes analyzed	Number of copies expressed					
		5	**4**	**3**	**2**	**1**	**<5 ESTs**
5	56	-	7	3	2	1	43
4	76		11	3	9	7	46
3	632			254	146	38	194
2	50				26	8	16
1	40					30	10
Total	**854**		**18**	**260**	**183**	**84**	**309**

†More than one band mapping to one bin on a chromosome was scored as single loci.

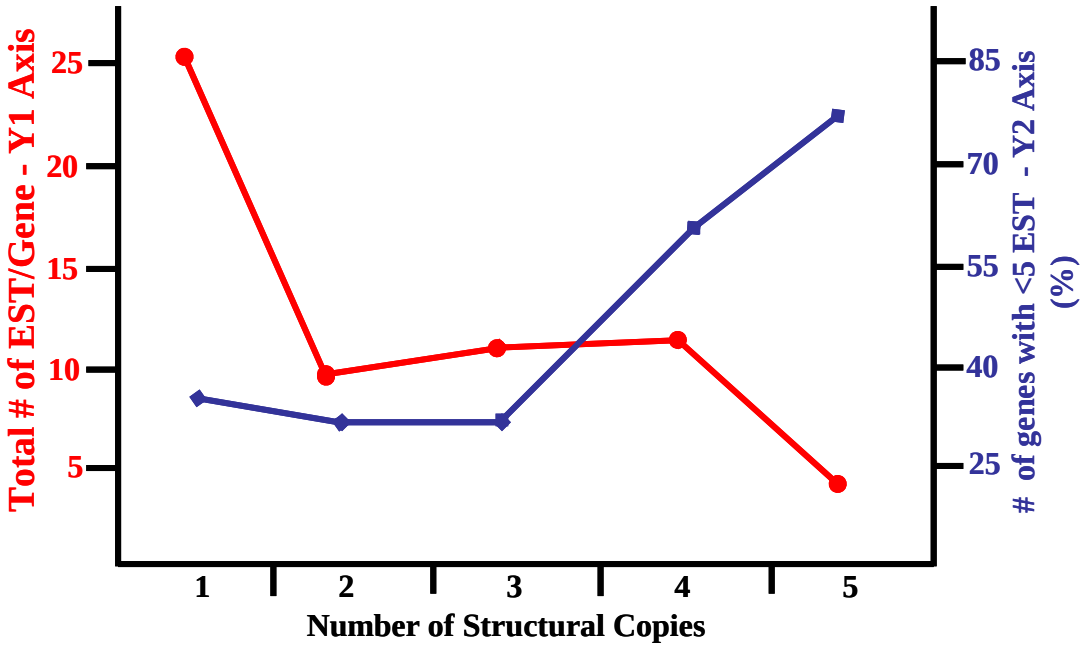


Figure 2.2. Total number of ESTs identified per gene (Y1-axis) and number of genes with <5 ESTs identified (Y2-axis). X-axis gives the structural copy number of the gene physically mapped to three homoeologous chromosomes of wheat. Red curve gives the distribution of total number of ESTs identified per gene per structural copy number, where as blue curve represents number of genes with <5 ESTs identified per structural copy.

The fraction of expressed copies was generally found to be inversely related to the overall number of structural copies in the genome as none of the genes with five structural copies expressed from all five (Table 2.1). Fifty-four percent of these expressed from four copies, 23% from three, 15% from two and only 8% from one of the five copies. In the case of genes with four structural copies, a larger proportion (23%) of the genes was expressed from one locus where as in case of genes with three copies only 8.6% expressed from one locus. The proportion of genes expressing from two copies was however very similar (30% and 33.3%) between the two groups. For genes with three copies, 58% expressed from all three loci. Proportion of genes expressing from both copies was 78% for the genes with two structural copies.

The *in silico* analysis results were evaluated by cDNA-SSCP and standard acrylamide/urea gel analysis for 31 randomly selected genes representing a range of structural copy number and expression patterns. This analysis was performed using RNA of various aneuploid stocks mixed in equal proportion from eight major developmental stages (see material and methods). Primers for the SSCP analysis were designed from the conserved regions flanking the sequences differentiating homoeologous (Figure 2.1c). Examples of the SSCP analysis are given in Figure 2.3 and are described below.

As an example we compare the *in silico* and SSCP analysis results of gene represented by EST *BE586090. BE586090* has three structural copies one each on the three homoeologous group *3* chromosomes (http://wheat.pw.usda.gov/NSF/). The *in silico* expression analysis showed three distinct sequence patterns suggesting that all three copies are expressing (data not shown). The SSCP analysis of CS and group *3* nulli-tetra (NT) lines showed three bands in CS with each mapping to a different group *3* homoeologue (Figure 2.3b). Similarly, the SSCP results for *BE490007,*

BE496983, BE442943, BE444392, BE498761, psr161, BE485127, BF485179, BE490149, BE443936, BE496983, BE442957 and *BF474347* matched with the *in silico* results Figure 2.3a-f.

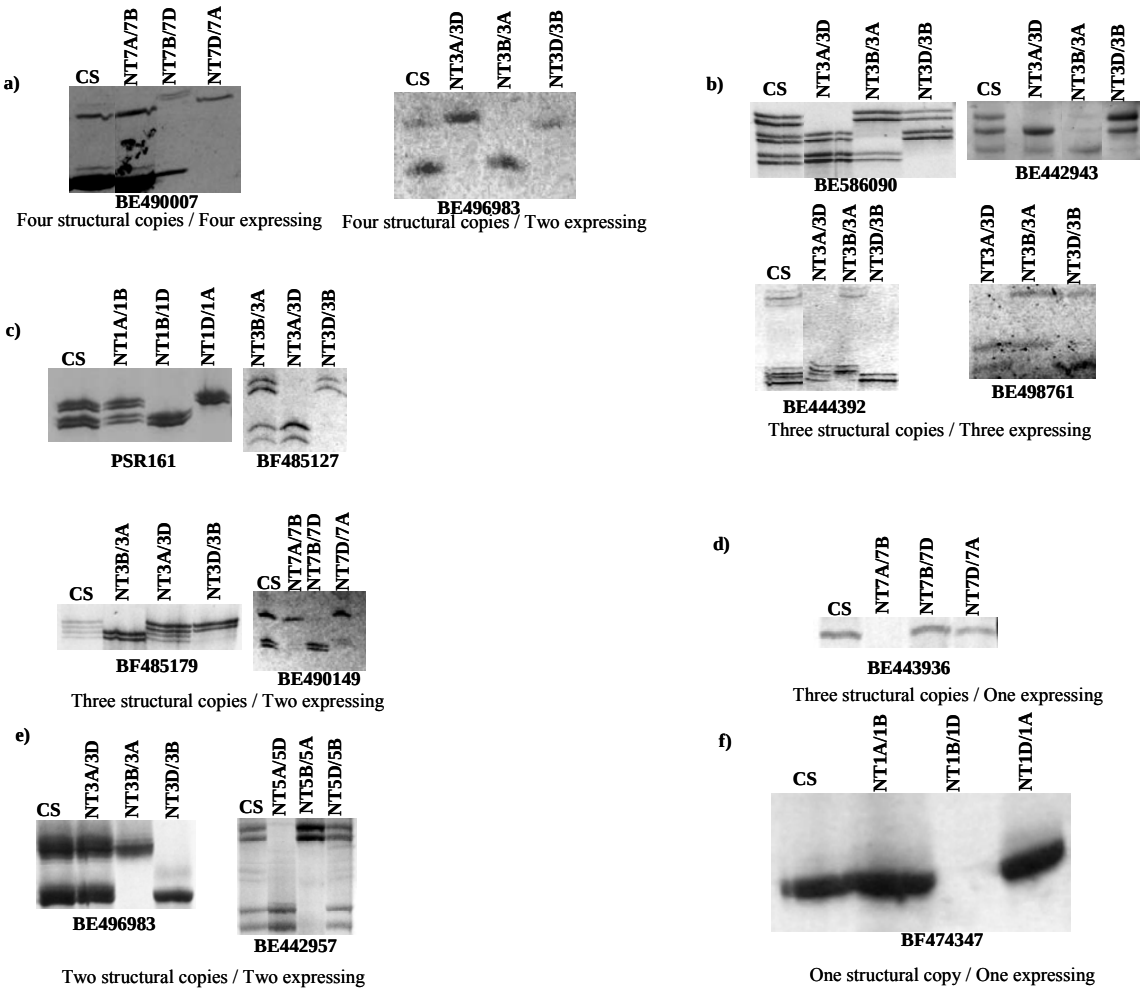


Figure 2.3. Confirmation of homoeologous gene expression patterns observed by EST sequence comparison (Figure 3.1a), by cDNA-SSCP analysis. CS denotes 'Chinese spring'; NTA, NTB, NTD indicates the nulli-tetra lines for A, B and D genomes. Gel electrophoresis pattern. **(a)** Indicates four structural copies and four expressing, alongwith four structural copies and two expressing. **(b)** Indicates three structural copies of the gene, three expresses. **(c)** Three structural copies of the gene, two expresses. **(d)** Three structural copies of the gene, one expressed. **(e)** Two structural copies, two expresses and **(f)** one structural copy of the gene and one expresses.

Both standard acrylamide/urea gels as well as SSCP gels were run for the 31 randomly selected genes. For 13 genes the SSCP gel worked better as it resolved all the bands where as for 18 genes the standard gel was able to better resolve the bands better. Between these two types of gels, all bands for 18 genes were resolved by the NT lines and the results of number of homoeologues expressed matched with the *in silico* analysis. Out of 31 genes that were tested, we were able to confirm the number of homoeologue expressed for 18 genes. For the remaining thirteen genes, primers for *BE606881, BE498241* and *BE426293* did not show amplification and for six genes not all bands were resolved by the NT lines *(BF292870, BG263064, BE493050*, *BE442957, BE403201, BF485127).* Three genes (*BE488246, BE488792* and *BE442943*) had more bands amplified than projected from the *in silico* analysis, ranging from seven to four bands. Out of these three, *BE488246* showed three structural copies and two patterns. Four bands were amplified, with two mapping on the A and D chromosomes of group *3. BE488792* had three structural copies and three patterns. It amplified five bands, with three mapping to A chromosome and two mapping to B chromosome of group *5.* In case of *BE442943* it had three structural copies and three patterns. It amplified seven bands, with three bands mapping to A, one to B and two to the D chromosome of group *3*. The autoradiogram of *BE442943,* showed in all five bands with only three of five mapping to three homoeologous chromosomes of wheat (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi).

2.3.2 Tissue Specific Expression of the Homoeologues

To study tissue specificity of the homoeologues, we randomly selected 220 genes out of the 632 with three structural copies. This analysis included 268,582 ESTs

from 41 cDNA libraries representing 14 different wheat tissues or developmental stages (Table 2.2).

Of the 268,582 ‘Chinese spring’ wheat ESTs, early flowering (developmental stage measured as Feekes scale: 6) (Large 1954) accounted for ~47,000 followed by anther at meiosis stage with 36,090 and root with 27,016; sheath and flag leaf had the lowest numbers with 2,552 and 1,314 respectively. Flag leaf and leaf sheath were not included in this analysis because the numbers of EST for these tissues were less than 5,000, an arbitirarily number we used as a cut off for this analysis. The ESTs corresponding to each homoeologue for the 220 genes were scored for expression in various tissues and the results are summarized in Figure 2.4. In total, 113 of the 220 genes expressed from all three homoeologues, 85 from two and 22 from only one of the three homoeologues.

Table 2.2. Tissue specific expression of the homoeologues.

Tissue	Number of homoeologous copies expressed			Total genes expressed	Total homoeologues expressed	Homoeologue/ gene	EST's/tissue (K)
	Three	Two	One				
Root	8	44	79	111	**191**	**1.7**	27
Shoot	3	11	36	50	67	1.3	15.5
Crown	2	12	49	63	89	1.4	15.7
Sheath	x	1	13	14	15	1	2.5
Leaf	3	10	41	54	70	1.2	23
Flag Leaf	x	1	20	21	22	1	1.3
Spike at FL	4	13	45	62	83	1.3	12.3
Anther	3	10	42	**121**	125	**1**	36.1
Spike at HD	2	15	51	68	87	1.3	11.3
Early FL	1	10	47	99	119	1.2	47.6
Pistil at HD	1	10	39	50	62	1.2	10.4
Pre-anthesis	4	9	57	70	87	1.2	14.3
DPA 5- 20	7	13	46	66	93	1.4	11.3
DPA 30	3	16	35	54	76	1.4	12.6
Total	**42**	**189**	**672**	**883**	**1186**	**1.3**	**269**

FL denotes flowering, HD heading date, DPA days post anthesis.

The maximum numbers of genes were expressed in anthers at meiosis stage (14%) and the least number were expressed in pistil at heading date (7%). The highest number of expressed homoeologues/gene (1.72 out of 3) was found in root tissue and the least number expressed in anthers (1.03 out of 3). It was also observed that the number of genes expressing in early stages of plant development, except in roots, is generally less than number of genes expressing after flower initiation stage.

For 87% of the genes analyzed, at least one of the homoeologues showed a difference in expression from the other two homoeologues in at least one tissue. In ~40% of analyzed genes, all three copies were expressed in some of the same tissues. Whereas in 57% of the genes analyzed, only a single copy was expressed in at least one tissue. Of the 220, homoeologues for 198 genes expressed from two or three copies. No difference was observed for the expression pattern of homoeologues for 25 of these genes (13%) in different tissues. Five of these expressed from all three copies and the remaining 20 from two. Most of the differential expression of the homoeologues was seen at flowering and seed development stages with least in roots.

Tissue specificity was more pronounced for the 22 genes that expressed from only one of the three homoeologues. On average, these 22 genes expressed from five different tissues per gene.

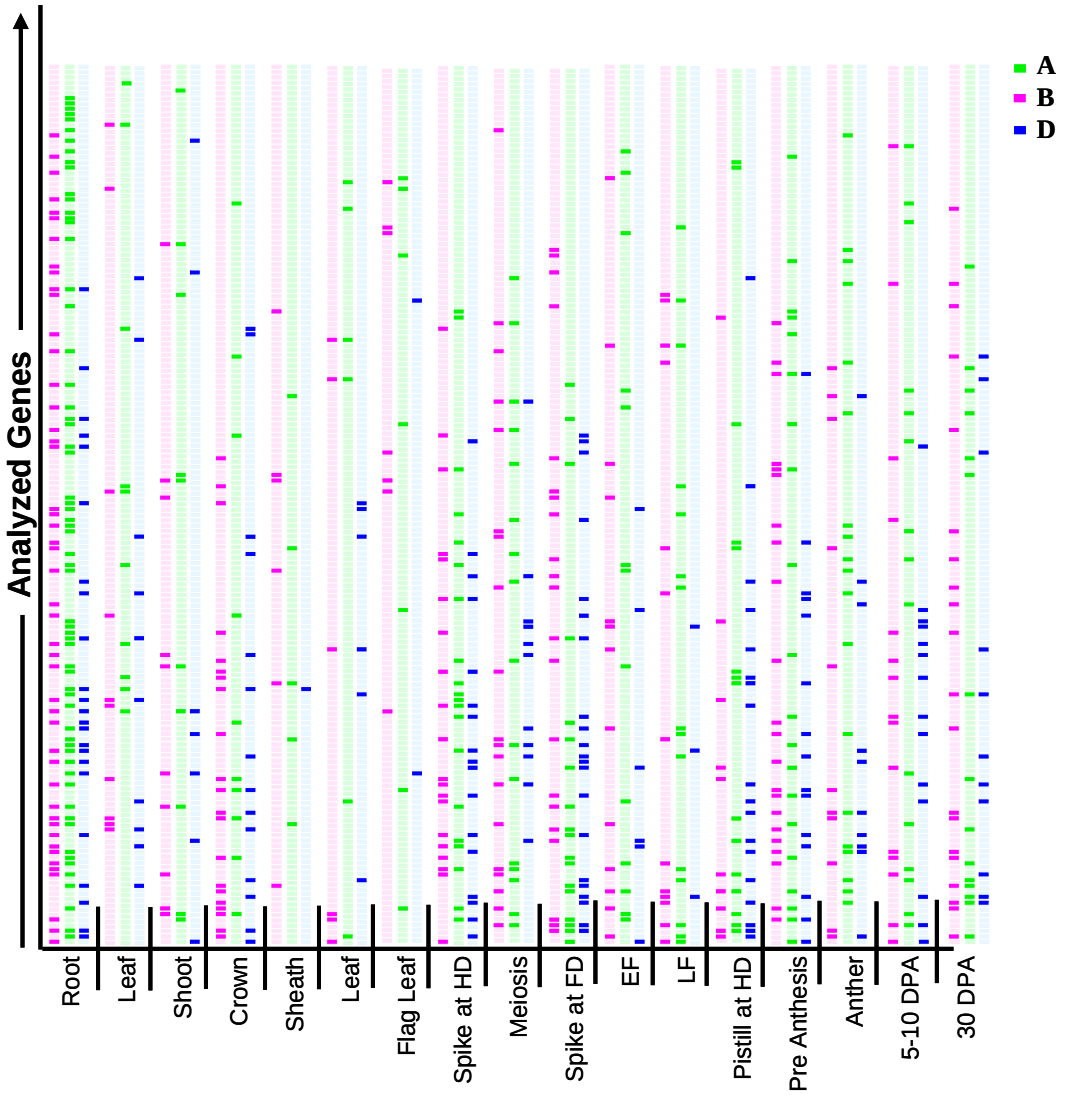


Figure 2.4 Virtual *northern* analysis of ESTs to show differential expression of homoeologues during different stages of plant development. Green, pink and blue bar corresponds to three homoeologues of wheat. The darker color bar indicates that the particular homoeologue expresses in that tissue. EF and LF denotes early flowering and late stages. Both stages were collected at Feekes scale: 6, so should be considered the same. FD denotes flowering date, HD heading date, where as DPA represent days after post anthesis.

Since the observed differential expression of homoeologues in different tissues might be attributed due to sampling error during EST sequencing. Tissue-specific expression of eight randomly selected genes was evaluated by SSCP analysis of cDNA from nine different tissues/developmental stages of CS. Out of eight genes tested, we were able to confirm the results of virtual *northern* analysis in most of the tissues analyzed, for five genes (*BE586090, BF485127, BE517931, BG263365* and *BE443936).* For example, the virtual *northern* analysis suggested that all three homoeologues for *BE586090* express in root, leaf, DPA 5-10 and DPA 30 whereas only two of the three expressed during anther at meiosis. The SSCP analysis confirmed these results as the B genome fragment band was missing during meiosis stage but all three bands were present during other stages (Figure 2.5a). For the remaining three, we were not able to resolve bands in the case of *BE17750.* For *BE444392*, virtual *northern* showed expression of all three homoeologues in root tissue only, with no expression in any other stage. Whereas, in the SSCP analysis all three homoeologues were expressed in pre-anthesis, post-anthesis and DPA5. One of the three homoeologue was expressed in root, leaf and meiosis stage. Flag leaf, early flowering and DPA30 showed no expression. In the case of *BE442943* virtual *northern* showed all three bands present in root tissue, with two bands expressing in DPA30 and post anthesis along with other stages showed no expression. SSCP analysis showed all three bands expressed in all of the tissues analyzed, although difference in expression level was observed among the homoeologues in different tissues.

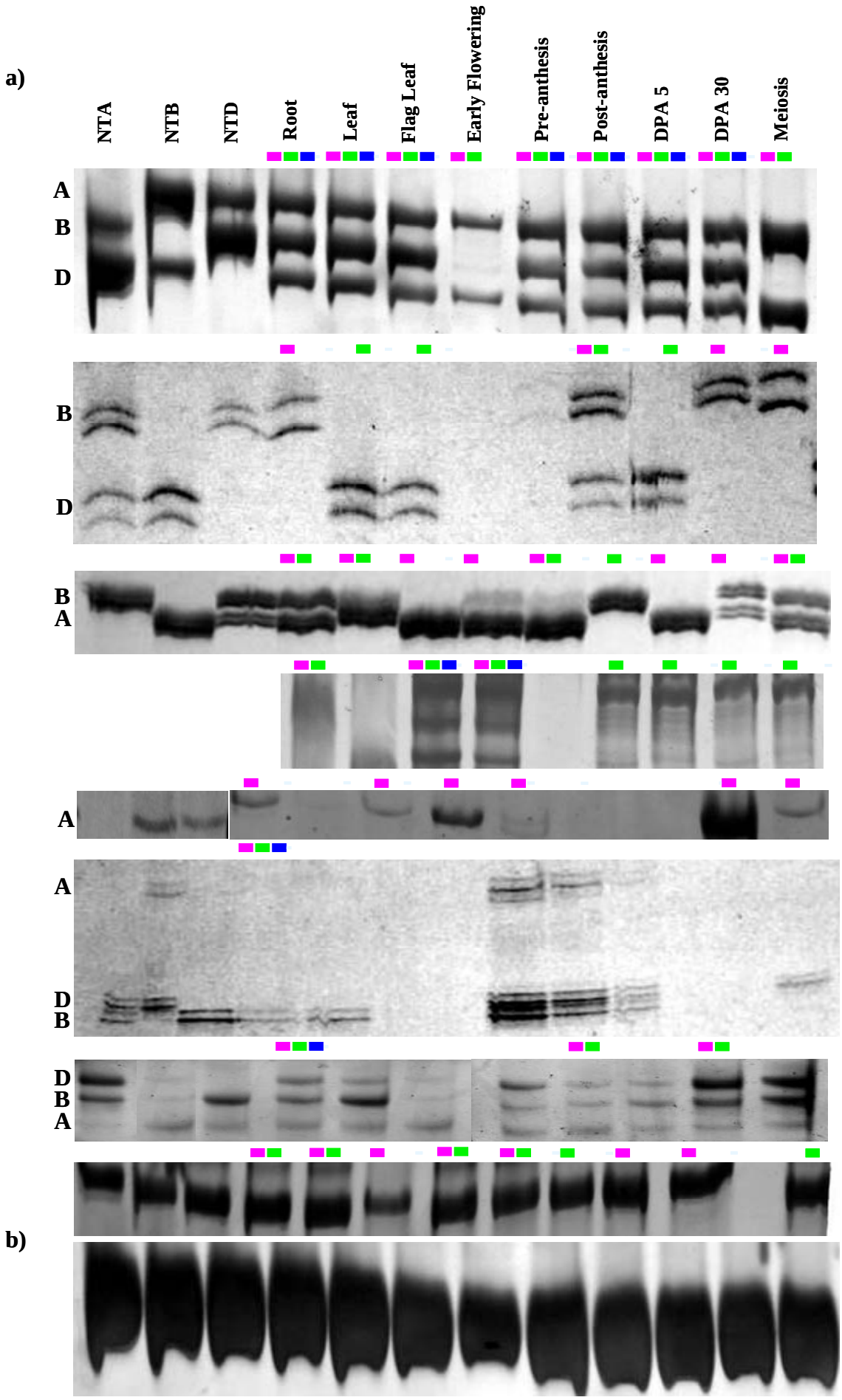

a)
NTA
NTB
NTD
Root
Leaf
Flag Leaf
Early Flowering
Pre-anthesis
Post-anthesis
DPA 5
DPA 30
Meiosis
A
B
D
B
D
B
A
A
A
D
B
D
B
A
b)

Figure 2.5. Tissue or stage specific expression of homoeologues on cDNA synthesized from RNA, extracted from NT lines (to localize the bands to respective chromosome location) and different developmental stages of CS plant, measured by SSCP analysis. The developmental stages included were root, leaf, flag leaf, early flowering, pre-anthesis, post-anthesis, DPA5, DPA30 and meiosis. **(a)** Tissue or stage specific expression of gene *BE586090*, *BF485127, BE517931, BG263365, BE443936, BE444392, BE442943* and *BE17750.* Above each gel picture; green, pink and blue bar corresponds to virtual *northern* results of expression of three homoeologues of wheat. Presence of one, two or three bars represent the number of homoeologues expressed. **(b)** The actin gene was used as a control.

2.3.3 Effect of chromosomal location on homoeologous gene expression

To study the possible effect of chromosomal physical location on homoeologous gene expression, 374 of 632 randomly selected genes were placed on the wheat consensus maps following the approach described by Randhawa et al. (2004). The proportion of various classes of gene expression pattern (number of homoeologues expressing, one, two or all three) was compared among various chromosomal regions using chi-square (χ^2) analysis. The number of ESTs averaged over the entire chromosome arm was used as the expected value for each of the expression pattern class. P <0.01 was considered significant for Chi-square test (see material and methods).

A summary of this analysis is shown in Figure 2.6. Chi-square goodness of fit revealed a non-random distribution of homoeologue gene expression patterns. Expression of all three homoeologues was significantly higher in bins 1S-0.50-0.59, 1L-0.69-0.61, 2S-0.47-0.53, 2L-0.50-0.49, 2L-0.89-0.85, 3S-0.23-0.24, 3L-0.78-0.81, 4S-0.37-0.43, 4L-0.86-1.00, 5L-0.75-0.76, 6S-0.35-0.45, 6L-0.36-0.40 and 6L-0.74-0.80. Expression from two of the three homoeologous was significantly higher in bins 1S-0.47-0.48, 2S-0.47-0.53, 2L-0.50-0.49, 3S-0.55-0.57, 3S-0.23-0.24, 3L-0.47-0.50, 4S-0.57-0.66, 4S-0.57-0.66, 4L-0.86-1.00, 5S-0.43-0.56, 5L-0.87-1.00, 7S-0.59-0.61

and 7S-0.36-0.45. The bin 2L-0.89-0.85 showed a significant preference for genes with only one expressed homoeologue. In general, genes present in the proximal regions on the short arm of the chromosomes expressed from all three copies whereas the genes present in the terminal parts of the chromosomes expressed from less than three copies.

Genes that were represented by less than five ESTs were also non-randomly distributed on the chromosome. Bins 1S-0.86-1.00, 1L-0.17-0.18, 4S-0.66-0.67, 6S-0.99-1.00 and 7S-0.59-0.61 showing a significant number of genes with <5 ESTs identified.

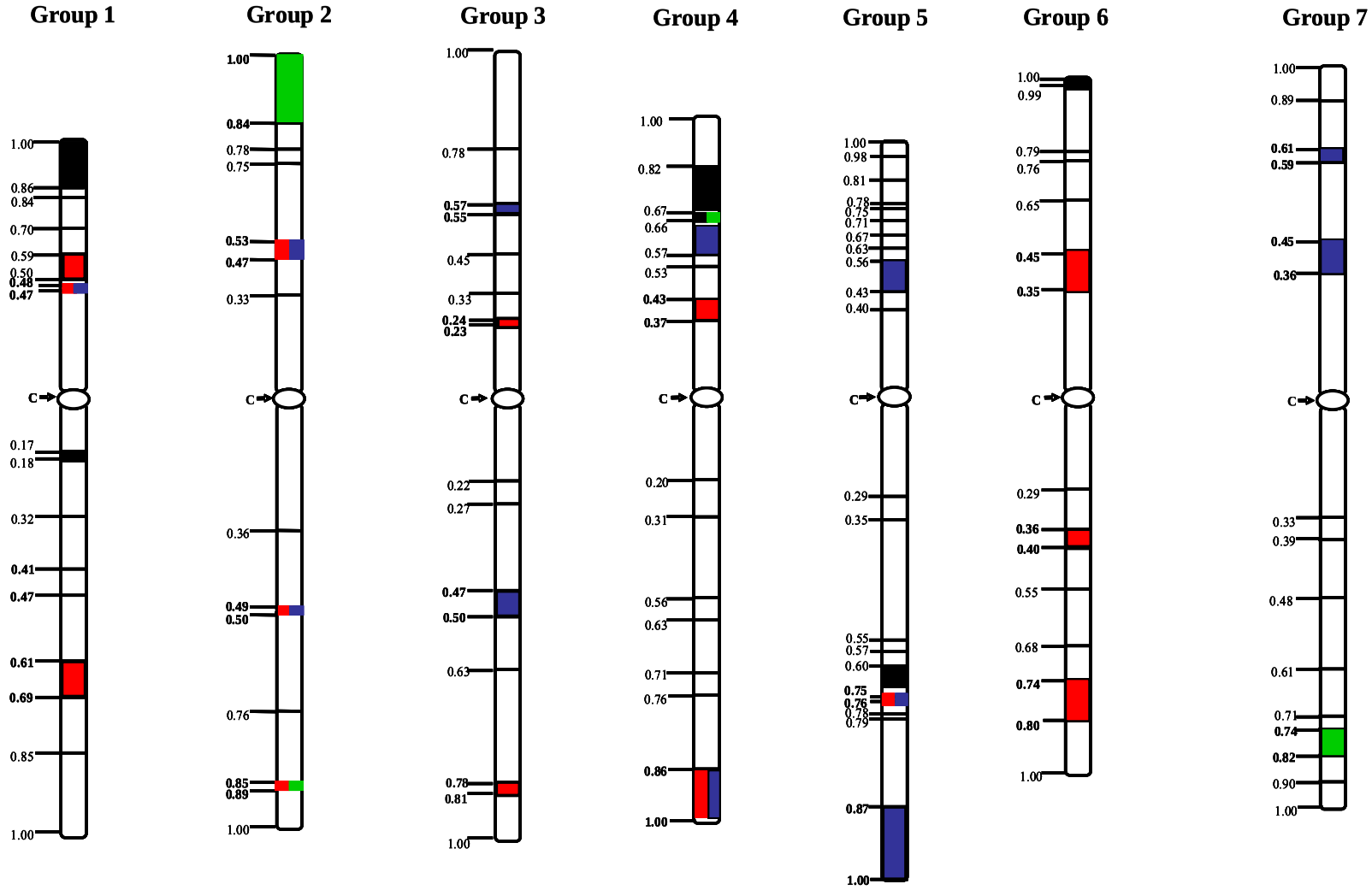

Group 1
1.00
0.86
0.84
0.70
0.59
0.50
0.48
0.47
C
0.17
0.18
0.32
0.41
0.47
0.61
0.69
0.85
1.00
Group 2
1.00
0.84
0.78
0.75
0.53
0.47
0.33
C
0.36
0.49
0.50
0.76
0.85
0.89
1.00
Group 3
1.00
0.78
0.57
0.55
0.45
0.33
0.24
0.23
C
0.22
0.27
0.47
0.50
0.63
0.78
0.81
1.00
Group 4
1.00
0.82
0.67
0.66
0.57
0.53
0.43
0.37
C
0.20
0.31
0.56
0.63
0.71
0.76
0.86
1.00
Group 5
1.00
0.98
0.81
0.78
0.75
0.71
0.67
0.63
0.56
0.43
0.40
C
0.29
0.35
0.55
0.57
0.60
0.75
0.76
0.78
0.79
0.87
1.00
Group 6
1.00
0.99
0.79
0.76
0.65
0.45
0.35
C
0.29
0.36
0.40
0.55
0.68
0.74
0.80
1.00
Group 7
1.00
0.89
0.61
0.59
0.45
0.36
C
0.33
0.39
0.48
0.61
0.71
0.74
0.82
0.90
1.00

Figure 2.6. A consensus expression map of 374 ESTs physically localized to bins on all 7 groups of wheat. Red color corresponds to the bins showing significant expression of genes with all three homoeologues expressed, blue color bins represent significant expression to genes with two homoeologues expressed, and green color represents to genes with only one homoeologue expressed. Bins with black color showed significant number of <5 ESTs. Fraction length for each bin is given on the left side of each chromosome. Number of ESTs/Bin was scored at a significance level of p <.01.

2.3.4 Dosage dependence of the homoeologues

The inter-dependence or balance of homoeologues for gene expression was first studied by gel blot analysis using isochizomer differing for CpG methylation sensitivity (*Hpa*II and *Msp*I), and expression differences by SSCP analysis in response to the homoeologue copy number change. The pair *Msp*I/*Hpa*II recognizes the CCGG sequence, with *Hpa*II sensitive to methylation of the second C in the site. Genes for wheat homoeologous group *7* were studied for this experiment and genes from other chromosomes were used as controls. A total of 66 gene probes specific to group *7* and 17 specific to groups *1* or *5* were used. The physical location of group *7* probes is shown on a chromosome map of deletion break points in Figure 2.7.

Of the 66 group *7* probes used for methylation analysis, 46 probes showed no methylation difference in any of the group *7* nullisomic-tetrasomic lines in comparison to CS, when digested with methylation sensitive enzyme *Hpa*II. The CS bands were observed in the NT lines and no additional band was observed. For the remaining probes, however, unique fragment bands were observed with *Hpa*II, in one or more NT or ditelosomic lines that were not present in CS and were not observed by the methylation insensitive enzyme *Msp*I. An example of a probe (*wsu102*) showing methylation difference in response to homoeologue copy number change is shown in Figure 2.8. The methylation pattern belonging to *7A* for *wsu102* gene fragment was

changed between CS and DT7DS. No variation in band pattern was observed in any of the lines and for any of the fragments when digested with *Msp*I. Of the 66 probes analyzed, 20 (30%) showed altered methylation pattern among the CS and NT or ditelosomic lines. Neither of the probes belonging to group *1* or *5* revealed any variation in methylation between CS, NT, or DT lines. We were unable to assign gene fragments corresponding to a particular homoeologue of group *7* chromosomes for *BE591273, BE474372*, and *RZ476*.

Of the 31 genes analyzed with SSCP on CS and NT lines, 9 (21%) showed differential expression with varying gene dosage. In probe psr161, the 3rd band is missing from both NT's of B and D chromosome. *BE404660* showed 4 bands in CS, 2nd and 3rd band mapping on D and B chromosome was missing from A, whereas the fourth band is missing from both of the B and D chromosome. In some of the cases, varying dose showed increased gene expression, as in the case of *BE403201*, the band mapping on B chromosome showed increased expression on A and D chromosome compared to its expression in 'Chinese spring'. In some of the cases, as in *BE488792,* an extra band was present on the NT's but absent in CS.

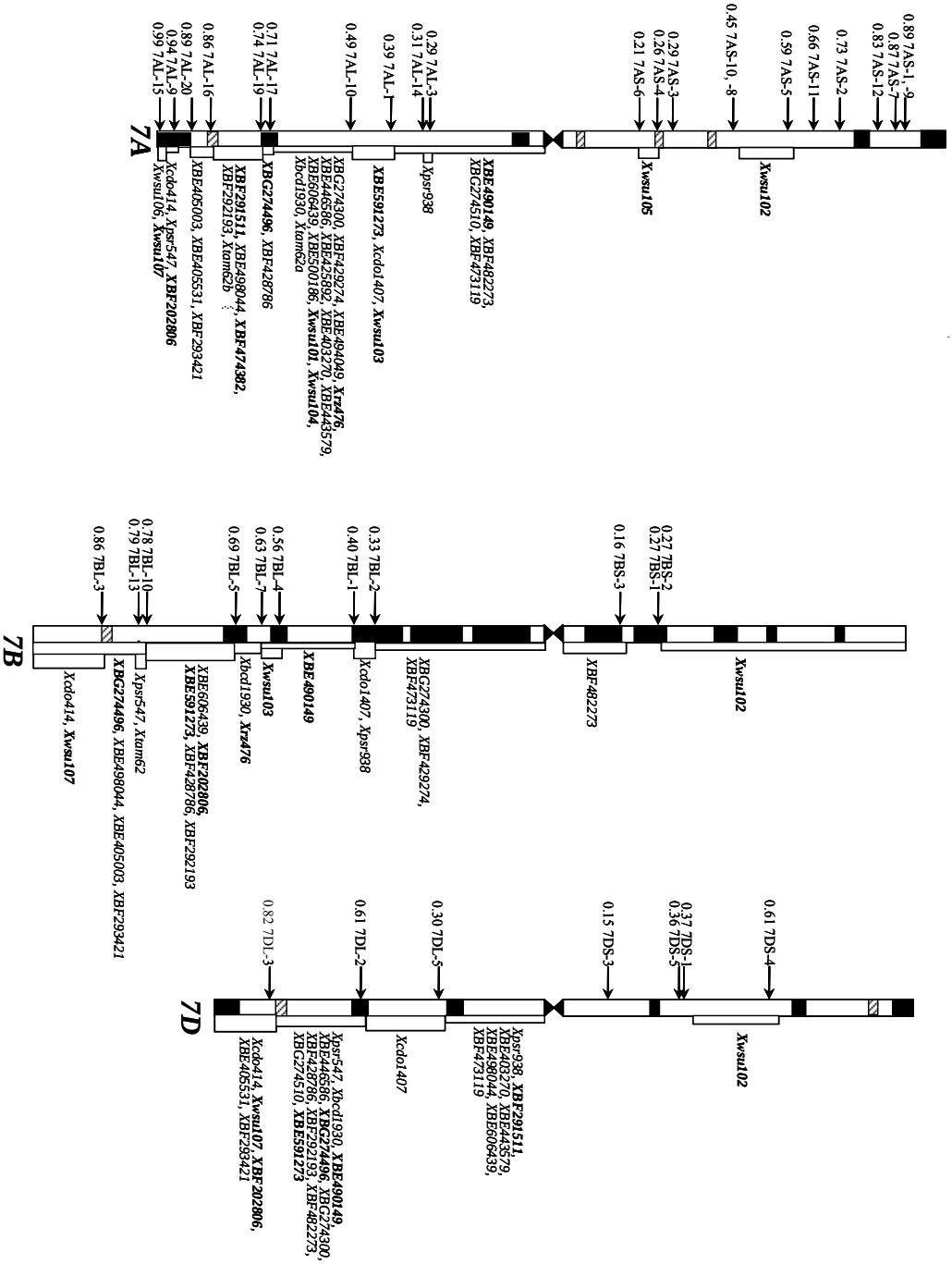


Figure 2.7. Physical map of group *7* probes. Relative chromosome length, arm ratios, and C-bands were drawn to scale. The chromosome deletion breakpoints with the fraction lengths of the retained arm are marked by arrows on the left of the chromosome diagrams. The probe loci and chromosome regions paneled by the deletion breakpoints are shown on the right of the drawings. Markers in '**bold'** identified variable methylation patterns among the NT or DT lines when compared to normal CS.

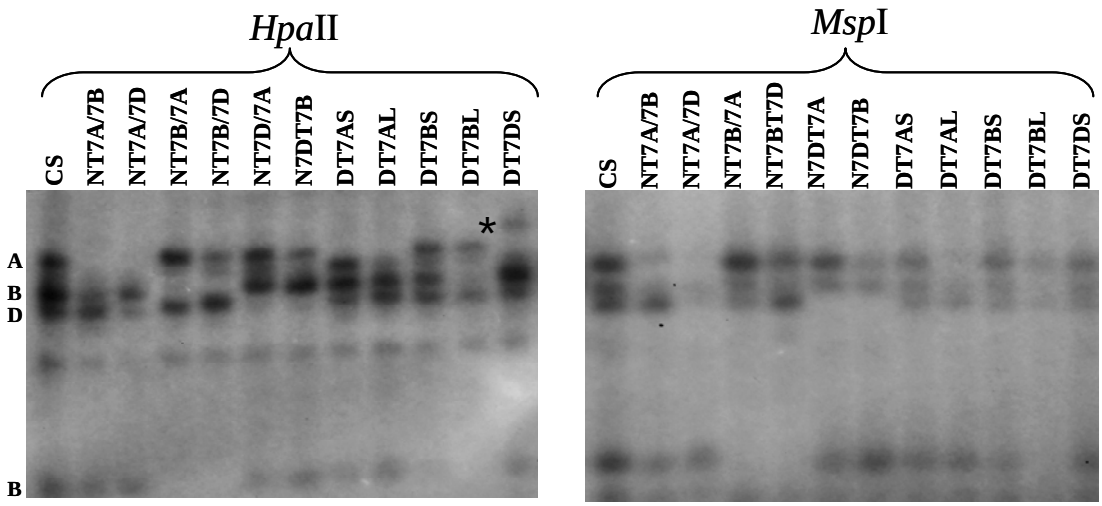


Figure 2.8. Methylation pattern between CS and aneuploid stocks of homeologous group *7* chromosomes. The autoradiographs of southern hybridization of genomic DNA of normal 'Chinese spring' (CS), nullisomic-tetrasomic (NT), and ditelosomic (DT) lines for group *7* chromosomes. The genomic DNA from CS, NT and DT lines was digested with a methylation-sensitive enzyme (*Hpa*II) and insensitive enzyme (*Msp*I) and hybridized with *wsu102*. The star indicates the fragments showing differences in methylation patterns. The genomic fragments identified by physical mapping are depicted as *A*, *B* and *D* on the left of the autoradiograph. The arrows indicate the fragments showing differences in methylation patterns

2.4 Discussion

Wheat is one of the most important cereal crops used for human consumption and thus any genetic insights into its genome can be translated into crop improvement. The three genomes in the hexaploid wheat evolved from a common progenitor about 3 MYA. Two of the genomes came into a common nucleus about ~0.5 MYA and the third one only ~10 thousand years ago (Kihara 1944). The three genomes are relatively intact and are well characterized. The diploid progenitors of wheat are also extant and the phylogenetic relationship with its diploid progenitors is comparatively well understood (McFadden and Sears 1946). Furthermore, wheat has a wealth of well-established, characterized aneuploid and deletion stocks, which are valuable resource to help, understand functional organization of the wheat genome and to analyze the effect of gene dosage on gene expression. These factors and the availability of an expansive EST dataset (~900,000), along with physical mapping of ~10,000 EST loci makes wheat a very good system to study the effect of polyploidy on evolution of gene expression.

Our *in silico* approach to identify ESTs corresponding to each of the three wheat homoeologues seems to be reliable as the standard acrylamide/urea gel or the SSCP analysis confirmed the *in silico* results. The gel analysis for 18 of the 31 genes resolved all of the fragments and the results matched with that observed with the *in silico* analysis. For the remaining 13 genes not all fragment bands were resolved by the nulli-tetrasomic lines with either of the two gel running approaches but the number of the resolved fragment bands were equal or less than the expected number. Only for three of the 31 genes, were the observed numbers of bands more than expected. Which could be due either to alternate splicing or the primer sequences for

these genes may be present in some other unrelated genes that were not detected by the gel blot analysis during physical mapping.

Resolving homoeologous copies in wheat is very difficult especially for important genes that are more conserved compared to the non-expressed sequences. So far, no good method is available to successfully differentiate the three homoeologues of wheat. Our acrylamide/urea gel based combination approach using both standard as well as SSCP runs, proved very effective as it resolved all mRNA's representing homoeolgues for 56% of the genes and 38% of the bands were resolved for the remaining genes. Some genes resolved better by the standard runs where for others the SSCP worked better. Therefore, both types of gels should be run to maximally resolve the homoeologues.

We are reporting that 55% of the wheat genes are expressing from all three homeologues, 36% from two, and 16% from only one. Based on the analysis of 116,232 ESTs, Mochida et al (2003), reported that 26% of the wheat genes express from all three loci whereas 46% express from two and 21% expressed from one of the three homoeologoues. Estimation of homoeologous gene expression based merely on ESTs can be misleading especially for genes exceeding 1 kbps in size, as two non-overlapping EST contigs corresponding to the same gene may be counted more than once. In our analysis, we used the full-length rice cDNA sequences to help assemble the full-length wheat contigs. No structural copy number estimation was included. Correct estimation of the number of homoeologues/gene expressed, cannot be done without structural copy number information. These flaws in analysis along with a limited number of ESTs used can provide us with estimates which can potentially be misleading for gene expression patterns of homoeologues in wheat. Therefore, precise

and full-scale analysis of fate of homoeologues is particularly important in wheat, as it serves as a model system for polyploids.

Expression of homoeologues in wheat is balanced with most of the homoeologues expressing at all times. More than one copy of a gene is expressed for the most part with silencing of two or more homoeologues kept at a minimum. Based on the sample of 652 genes with three structural copies of the gene mapped on three chromosomes of wheat, 91% of the time, two or three copies were expressing, with only 9% of genes showing silencing in two of three copies. Of the 91% of genes where more than one copy expresses, tissue specific expression and silencing seems to be common with stages like meiosis and seed development showing maximum specificity. A balance of homoeologue expression was also observed, as any deviation in copy number of the gene resulted in altered expression level. For example, on an average genes with five structural copies identified, four ESTs/gene compared to 30 ESTs/gene for genes with one structural copy. A balance of expression was seen in the case of genes with three structural copies, expressing 12 ESTs/gene. Any deviation in genomic copy number leads either to increase in expression of the gene (as in case of genes with one structural copy) or decrease in expression with increase in copy number (as in case of genes with five structural copies.

The most significant observation was that different homoeologues appear to have evolved different gene expression controls. In analysis of 198 genes where two or three homoeologues were expressed, 87% of genes showed differential control in different tissues. Differential expression of a homoeologue is recorded as the presence or absence of one homoeologue in different tissues. The question of what determines its differential expression was not explored. The sequence and functioning of the homoeologue is normal as it expresses normally in most of the tissues, thus what

determines it's silencing in one and subsequent expression in other tissues is unknown. 87% of the analyzed genes showing differential expression in different tissues suggest that it's not the sequence but a control mechanism which determines the fate of a particular homoeologue throughout the life cycle of an organism. Based on the stimuli or cell stage, this mechanism can use clues from epigenetic marks like various histone modifications and DNA methylation present on the sequence and thus specifically express or silence the particular copy of homoeologue.

As in recombination and gene distribution, the expression pattern of homoeologues also showed a correlation with chromosome location. For this analysis, 374 genes, with three structural copies and different expression patterns, were physically localized onto the three homoeologous genomes of wheat. After physical mapping, subsequent analysis using the Chi-square goodness of fit test, revealed a non-random distribution of homoeologue gene expression. While some regions on the chromosome favored expression of all three homoeologues other regions showed significant silencing of one or two of the three homoeologues present. In all, middle region on the short arm of the chromosome favored the expression of all three homoeologues whereas terminal region showed significant silencing of one or more homoeologues. Only the 2L-0.89-0.85 bin showed significant preference for genes with only one expressed homoeologue. This could be attributed to the fact that a very limited number of genes with one expressed homoeologue were detected. Genes that identified less than five ESTs were also non-randomly distributed on the chromosome. With bins 1S-0.86-1.00, 1L-0.17-0.18, 4S-0.66-0.67, 6S-0.99-1.00 and 7S-0.59-0.61 showing a significant number of genes with <5 ESTs identified.

Previously, the expression of genes that were physically present on both normal and deleted chromosomes was analyzed in response to terminal chromosomal

breaks in hexaploid wheat using RNAF/DD and cDNA-AFLP display performed on CS and 7AL-9 (FL0.94) lines (Sidhu and Gill 2004). The homozygous 7AL-9 is a terminal chromosomal deletion line generated in CS background (Endo and Gill 1996). These two lines differ only for the distal 6% of the chromosomal region of *7AL* which is missing in 7AL-9. A total of 83 fragments present in one and missing in the other were identified. Out of 83 differentially expressed fragments, 16 mapped to group *7*. These fragments along with 50 ESTs were subject to methylation pattern studies on NTs and DTs for group *7*. It was done to further elaborate that different chromosomal dosage is associated with altered methylation and differential gene expression. Further more this analysis was also looked in other groups of chromosomes by looking at the SSCP patterns of genes on NTs of various groups.

Altered methylation patterns was monitored as the variation in banding pattern between the NT, ditelosomic and CS when restriction digests with a methylation sensitive enzyme, such as, *Hpa*II. On the contrary, when restriction digests with its isoschizomer methylation insensitive enzyme *Msp*I, no differences in band pattern among CS, NT, and DT lines were observed. These probes were used to study the methylation pattern in CS, group *7* NT and DT lines. Changes in band pattern were observed between the corresponding NT lines and DT lines for the same chromosome. 30% of the analyzed genes showed altered methylation pattern and gene expression pattern when subjected to varying dosage.

There is evidence in support of the fact that transcription machineries are immobilized in the nucleus and the DNA slides past the polymerase while the transcript is being extruded. It would not be far-fetched if the disparity in gene expression between CS and the aneupliod lines correlates to improper accessibility of the appropriate transcription machinery by the deficient chromosome as compared to

the normal chromosome. The altered methylation pattern between normal and a chromosome deletion line indirectly suggests change in chromosome topology. Additional molecular analyses are required to test these hypotheses.

2.5 Material and Methods

2.5.1 Plant material

Wheat Nullisomic-tetrasomic (NT) lines and Ditelosomic (DT) lines produced in cultivar 'Chinese spring' (Sears, 1954), along with cultivar 'Chinese spring', grown at greenhouse conditions, were used for various experiments. For methylation studies and SSCP analysis, plants were grown under highly controlled conditions in a growth chamber [18/6 hr (day/night) light and 22°C/18°C (day/night) temperature]. Plants were started on the same day, spaced equally and were provided with equal amounts of water at regular intervals. The plant material was obtained from the Wheat Genetics Resource Center (Kansas State University, USA).

2.5.2. DNA and RNA procedures

The DNA extractions and gel blot analysis procedures were carried out as previously described (Sidhu, 2004; Sandhu et al. 2000). Total RNA was extracted using the guanidinium thiocyanate - cesium chloride density gradient method (Sandhu and Gill 2001; Sambrook et al. 1989). The poly(A)$^+$ RNA was selected from total RNA by affinity chromatography on oligo(dT)-cellulose using the standard protocol. For some of the stages used for looking at tissue specificity; the hot phenol RNA extraction method described in protocol optimized selection was used. For homoeologous gene expression analysis using SSCP analysis, RNA from eleven

different plant development stages namely five day old seedling, root from seedling and adult plants, 28 day old plant, flag leaf, early flowering (Feekes scale: 6) (Large 1954), meiosis, pre-anthesis, post-anthesis, Seed at 5 DPA and 30 DPA stage and adult plant (Feekes scale: 10.5) was extracted.

2.5.3 DNA methylation analysis

Gel blot analysis was used to evaluate changes in gene expression due to changes in homoeologue copy number. Genomic DNA from 28 day old leaf tissue of CS and wheat homoeologous group *7* NT and DT lines (N7AT7B, N7AT7D, N7BT7A, N7BT7D, N7DT7A, N7DT7B, DT7AL, DT7AS, DT7BL, DT7BS and DT7DS). The genomic DNA was cut with restriction enzymes sensitive to CpG and CpNpG methylation (*Hpa*II) and insensitive to methylation (*Msp*I). Gel blot analysis was performed using 15µg of genomic DNA and size separated on 0.8% agarose gels. All other steps were followed as previously described (Gill et al. 1993).

Methylation studies were conducted using 66 randomly selected wheat group *7* specific probes. An effort was made to select probes uniformly distributed over the group *7* chromosome. These probes included 50 ESTs present on various wheat genetic maps (http://wheat.pw.usda.gov/GG2/index.shtml). All of the probes used for the study were ESTs except for 16 group *7* probes that were previously identified by RNA fingerprinting/differential display or cDNA-AFLP display of CS and the 7AL-9 (FL0.94) deletion line (Sidhu and Gill 2004). As a control to see whether change in gene dosage of one chromosome has any effect on the genes on other chromosomes of wheat, randomly selected 12 group *5* probes (*abg314*, *abg387*, *wg530*, *mwg522*, *mwg52*, *mwg592*, *mwg2237*, *ksuG414*, *mwg768*, *mwg923*, *bcd508*, and *cdo1090)* and

5 group 1 probes (BE590674, BE495292, BE443071, BE443409 and *BE490291)* were also used.

2.5.4 SSCP analysis

Single stranded conformation polymorphism (SSCP) analysis was used to resolve fragment bands corresponding to the three homoeologues. We optimized the procedure of Cronn and Adams (2003) for use in wheat. Briefly, first stranded cDNA was synthesized from 2μg of DNaseI treated pooled poly(A)$^+$ RNA using Moloney Murine Leukemia Virus (MMLV) reverse transcriptase (Clontech Lab Inc., USA). The PCR reactions were performed with Advantage® PCR Kits & Polymerase Mixes (Clontech, Catalog #639101), in the presence of S^{35}dATP in a total volume of 20μl. The PCR product was mixed with an equal volume of loading buffer. About 5μl of this mixture was loaded onto 0.4mm thick denaturing 8% polyacrylamide/8M urea gels (Sambrook et al. 1989). Gels contained and were run in 0.5X TBE buffer at pH 8.3. Each sample was size separated both on gels run under standard conditions as well as on gels run for SSCP. For standard run, the gels were pre-run at a 33mA constant current for 30 mins and then at 70W constant power for 4 hours. For SSCP runs, the gels were pre-chilled at 4°C for at least 5-6 hr before running it at 10 W for 12-13 hrs at 4°C. An X-ray film was placed on the dried gels and was exposed for three to seven days.

2.5.5 Virtual northerns for tissue-specific expression analysis

Differential expression of homoeologues in different tissues was studies by manually performing virtual *northern* on 220 genes with three structural copies one each for the three genomes. All wheat ESTs (268,582) that were developed from 41

total cDNA libraries from 14 different tissues (http://www.ncbi.nlm.nih.gov/dbEST/), were used for this analysis. Different homoeologues were identified by sequence comparison of EST sequences for each gene and the source tissue for each of the EST representing different homoeologues was scored. The virtual *northern* results were confirmed by SSCP analysis of few randomly selected genes.

2.5.6 Expressed homoeologues distribution

Consensus physical maps for seven groups of wheat were downloaded (Conley et al. 2004; Hossain et al. 2004; Linkiewicz et al. 2004; Munkvold et al. 2004; Miftahudin et al. 2004; Peng et al. 2004; Randhawa et al. 2004). The physical size of the chromosome and chromosome arm were takem from Gill et al (1991). Physical size of chromosome intervals (bins) was calculated on the basis of their relative fraction length (FL). The distribution of expressed homoeologue was assumed to be uniform along the physical length of a chromosome and chromosome arm. The mean number of expressed EST loci per micrometer of chromosome arm was calculated by dividing the total EST loci mapped to the chromosome arm, by the total physical length of that arm (Gill et al. 1991). The numbers of expected expressed homoeologues (all three expressed, two of three expressed and only one of three expressed) per chromosome bin were calculated by multiplying the mean number with its physical length. The ratio of observed *vs.* expected EST loci was used to estimate the expressed homoeologues distribution. The chi-square goodness-of fit test was used to detect significant differences between observed and expected numbers of expressed EST loci. In chi-square distribution analysis $p < 0.05$ was considered significant. An example of the distribution of homoeologues and calculation of chi-square for group *1* chromosome is shown in Figure 2.9.

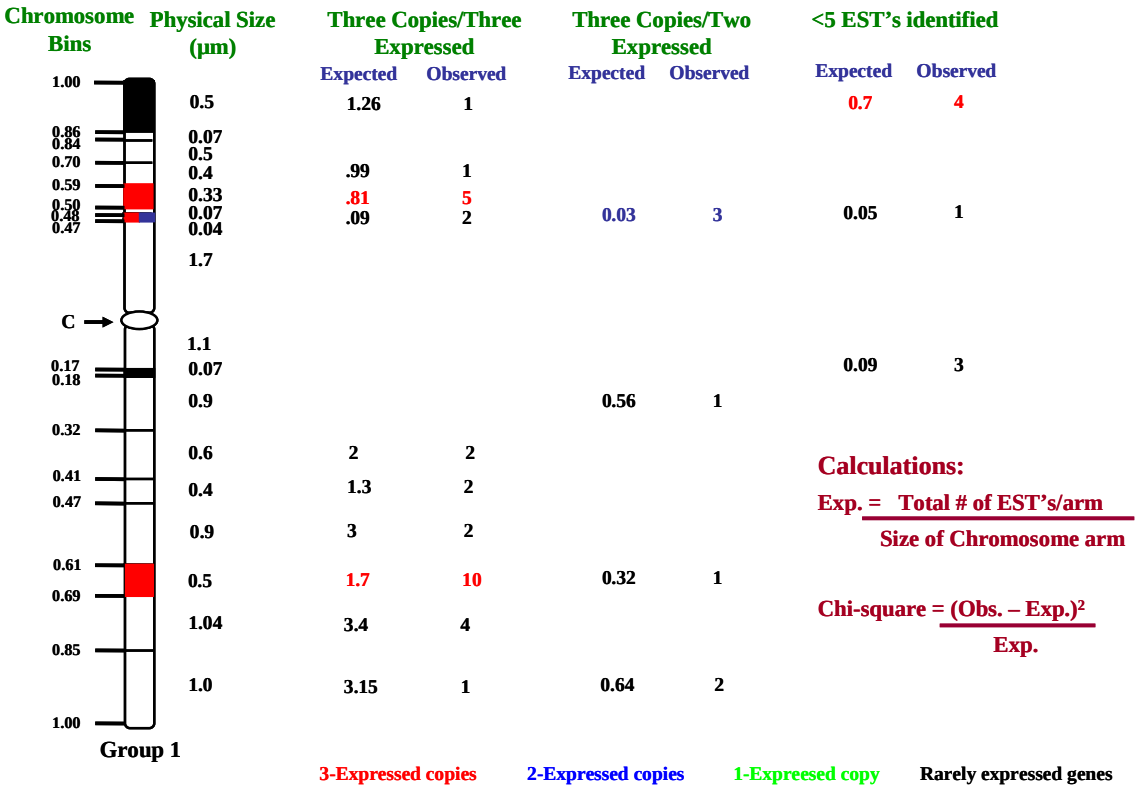


Figure 2.9. Demonstrate the distribution of expression pattern on consensus group *1* chromosome. Physical size of the chromosome bins were calculated from Gill et al (1991). Expected number of ESTs/bin was calculated and compared with the observed. Any deviation from the expected number was color coded. Red color for genes expressed from three copies, blue for genes expreesed from two copies, green for genes expressed from single copy and black color for rarely expressed genes (with less than 5 ESTs).

2.6 References

Adams, K.L., R. Cronn, R. Percifield, and J.F. Wendel. 2003a. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* **100:** 4649-4654.

Adams, K.L., R. Cronn, R. Percifield, and J.F. Wendel. 2003b. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* **100:** 4649-4654.

Adams, K.L., R. Percifield, and J.F. Wendel. 2004. Organ-Specific Silencing of Duplicated Genes in a Newly Synthesized Cotton Allotetraploid. *Genetics* **168:** 2217-2226.

Aragoncillo, C., M.A. Rodríguez-Loperena, G. Salcedo, P. Carbonero, and F. García-Olmedo. 1978. Influence of homoeologous chromosomes on gene-dosage effects in allohexaploid wheat (Triticum aestivum L.). *Proceedings of the Academy of Sciences* **75:** 1446-1450.

Barakat, A., N. Carels, and G. Bernardi. 1997. The distribution of genes in the genomes of Graminae. *Proceedings of the National Academy of Sciences* **94:** 6857-6861.

Birchler, J.A. 1979. A study of enzyme activities in a dosage series of the long arm of chromosome one in maize. *Genetics* **92:** 1211-1229.

Brewer, G.J., C.F. Sing, and E.R. Sears. 1969. Studies of isozyme patterns in nullisomic-tetrasomic combinations of hexaploid wheat. *Proceedings of the National Academy of Sciences* **64:** 1224-1229.

Brueggeman, R., N. Rostoks, D. Kudrna, A. Kilian, F. Han, J. Chen, A. Druka, B. Steffenson, and A. Kleinhofs. 2002. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Sciences* **99:** 9328-9333.

Carlson, P.S. 1972. Locating genetic loci with aneuploids. *Molecular General Genetics* **114:** 273.

Cook, P.R. 1999. The organization of replication and transcription. *Science* **284:** 1790-1795.

Dooner, H.K. 1986. Genetic fine structure of *bronze* locus in maize. *Genetics* **113:** 1021-1036.

Endo, T.R. 1988. Induction of chromosomal structural changes by a chromosome of *Aegilops cylindrica* L. in common wheat. *Journal of Heredity.* **79:** 366-370.

Endo, T.R. and B.S. Gill. 1996. The deletion stocks of common wheat. *Journal of Heredity* **87:** 295-307.

Epstein, C.J. 1986. The Consequences of Chromosome Imbalance. *Cambridge University Press, New York.*

Feldman, M., B. Liu, G. Segal, S. Abbo, A.A. Levy, and J.M. Vega. 1997. Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics* **147:** 1381-1387.

Feuillet, C. and B. Keller. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. *Proceedings of the National Academy of Sciences* **96:** 8265-8270.

Galili, G., A.A. Levy, and M. Feldman. 1986. Gene-Dosage Compensation of Endosperm Proteins in Hexaploid Wheat Triticum aestivum. *Proceedings of the National Academy of Sciences* **83:** 6524-6528.

Gill,K.S., B. S. Gill, and T. R. Endo. (1993) A chromosome region-specific mapping strategy reveals gene-rich telometric ends in wheat. *Chromosoma*, **102:** 374-381.

Grant, V. 1981. *Plant Speciation*. Columbia University Press, New York.

Grell, E.H. 1962. The dose effect of ma-l+ and ry+ on xanthine dehydrogenase activity in Drosophila melanogaster. *Z. Vererbungsl.* **93:** 371-377.

Guo, M., D. Davis, and J.A. Birchler. 1996. Dosage Effects on Gene Expression in a Maize Ploidy Series. *Genetics* **142:** 1349-1355.

Henikoff, S. 1997. Nuclear organization and gene expression: homologous pairing and long-range interactions. *Current Opinion in Cell Biology* **9:** 388-395.

Huang, S., A. Sirikhachornkit, X. Su, J. Faris, B. Gill, R. Haselkorn, and P. Gornicki. 2002a. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences* **99:** 8133-8138.

Huang, S., A. Sirikhachornkit, X. Su, J. Faris, B. Gill, R. Haselkorn, and P. Gornicki. 2002b. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. *PNAS* **99:** 8133-8138.

initiative, T.A.g. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408:** 796-815.

Kihara, H. 1944. Discovery of the DD-analyser, one of the ancestors of *vulgare* wheats. *Agriculture and Horticulture* **19:** 889-890.

Lahaye, T., K. Shirasu, and P. Schulze-Lefert. 1998. Chromosome landing at the barley *Rar1* locus. *Molecular General Genetics* **260:** 92-101.

Large, E.C. 1954. Growth stages in cereals, Illustration of the Feekes scale. *Plant Pathology* **3:** 128-129.

Lynch, M. and A. Force. 2000. The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics* **154:** 459-473.

McFadden, E.S. and E.R. Sears. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *Journal of Heredity* **37:** 81-89.

Mochida, K., Y. Yamazaki, and Y. Ogihara. 2003. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Molecular Genetic Genomics* **270:** 371-377.

Mohan, J. and F.M. Ritossa. 1970. Regulation of ribosomal RNA synthesis and its bearing on the bobbed phenotype in Drosophila melanogaster. *Developmental Biology* **22:** 495-512.

Muller, H.J. 1950. Evidence of the precision of genetic adaptation. *Harvey Lectures* **43:** 165-229.

Nomura, T., A. Ishihara, R.C. Yanagita, T.R. Endo, and H. Iwamura. 2005. Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proceedings of the National Academy of Sciences* **102:** 16490-16495.

Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.

Qi, L.L., B. Echalier, S. Chao, G.R. Lazo, G.E. Butler, O.D. Anderson, E.D. Akhunov, J. Dvorak, A.M. Linkiewicz, A. Ratnasiri, J. Dubcovsky, C.E. Bermudez-Kandianis, R.A. Greene, R. Kantety, C.M. La Rota, J.D. Munkvold, S.F. Sorrells, M.E. Sorrells, M. Dilbirligi, D. Sidhu, M. Erayman, H.S. Randhawa, D. Sandhu, S.N. Bondareva, K.S. Gill, A.A. Mahmoud, X.-F. Ma, Miftahudin, J.P. Gustafson, E.J. Conley, V. Nduati, J.L. Gonzalez-Hernandez, J.A. Anderson, J.H. Peng, N.L.V. Lapitan, K.G. Hossain, V. Kalavacharla, S.F. Kianian, M.S. Pathan, D.S. Zhang, H.T. Nguyen, D.-W. Choi, R.D. Fenton, T.J. Close, P.E. McGuire, C.O. Qualset, and B.S. Gill. 2004. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* **168:** 701-712.

Reeves, R. 1977. Structure of Xenopus ribosomal RNA gene chromatin during changes in genomic transcription rates. *Cold Spring Harbor Symposia on Quantative Biology* **42:** 709-791.

Sambrook, J., E.F. Fristch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY.

Sandhu, D. and K.S. Gill. 2001. Candidate gene cloning and targeted marker enrichment of wheat chromosomal regions using RNA Fingerprinting/Differential Display. *Genome* **44**:1-7.

SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova, and J.L. Bennetzen. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274:** 765-768.

Sasakuma, T., Y. Ogihara, and H. Tsujimoto. 1995. Genome rearrangement of repetitive sequences in processes of hybridization and allopolyploidization in Triticinae. *In 8th International Wheat Genetics Symposium, Z. Li and Z. Xin, eds (Beijing: China Agriculture Scientech Press)***:** 563-566.

Sasanuma, T., N.T. Miyashita, and K. Tsunewaki. 1996. Wheat phylogeny determined by RFLP analysis of nuclear DNA. 3. Intra- and interspecific variations of five Aegilops Sitopsis species. *Theoretical and Applied Genetics* **92:** 928-934.

Schwartz, D. 1971. Genetic control of alcohol dehydrogenase--a competition model for regulation of gene action. *Genetics* **67:** 411-425.

Sears, E.R. 1954. The aneuploids of common wheat. *Research Bulletin, University of Missouri Agricultural Experiment Station***:** 1-58.

Sears, E.R. and L.M.S. Sears. 1978. The telocentric chromosomes of common wheat. In *Proceedings of the 5th International Wheat Genetics Symposium* (ed. S. Ramanujam). 389-407.

Shaked, H., K. Kashkush, H. Ozkan, M. Feldman, and A. Levy. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* **13:** 1749-1759.

Sidhu, D. 2004. Location of genes, and effect of gene dosage and chromosomal breaks on gene expression in wheat (*Triticum aestivum* L.).

Sidhu, D. and K.S. Gill. 2004. Distribution of genes and recombination in wheat and other eukaryotes. *Plant Cell Tissue and Organ culture* **79:** 257-270.

Song, K., P. Lu, K. Tang, and T. Osborn. 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences* **92:** 7719-7723.

Stebbins, G.L. 1966. Chromosomal variation and evolution. *Science* **152:** 1463-1469.

Stein, N., C. Feuillet, T. Wicker, E. Schlagenhauf, and B. Keller. 2000. Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.). *Proceedings of the National Academy of Sciences* **97:** 13436-13441.

Tarchini, R., P. Biddle, R. Wineland, S. Tingey, and A. Rafalski. 2000. The complete sequence of 340 kb of DNA around the rice Adh1-Adh2 region reveals interrupted colinearity with maize chromosome 4. *The Plant Cell* **12:** 381-391.

Tate, J.A., Z. Ni, A.C. Scheen, J. Koh, C.A. Gilbert, D. Lefkowitz, Z.J. Chen, P.S. Soltis, and D.E. Soltis. 2006. Evolution and expression of homeologous loci in Tragopogon miscellus (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* **173:** 1599-1611.

Venter, J.C. M.D. Adams E.W. Myers P.W. Li R.J. Mural G.G. Sutton H.O. Smith M. Yandell C.A. Evans R.A. Holt J.D. Gocayne P. Amanatides R.M. Ballew D.H. Huson J.R. Wortman Q. Zhang C.D. Kodira X.H. Zheng L. Chen M. Skupski G. Subramanian P.D. Thomas J. Zhang G.L. Gabor Miklos C. Nelson S. Broder A.G. Clark J. Nadeau V.A. McKusick N. Zinder A.J. Levine R.J. Roberts M. Simon C. Slayman M. Hunkapiller R. Bolanos A. Delcher I. Dew D. Fasulo M. Flanigan L. Florea A. Halpern S. Hannenhalli S. Kravitz S. Levy C. Mobarry K. Reinert K. Remington J. Abu-Threideh E. Beasley K. Biddick V. Bonazzi R. Brandon M. Cargill I. Chandramouliswaran R. Charlab K.

Chaturvedi Z. Deng V. Di Francesco P. Dunn K. Eilbeck C. Evangelista A.E. Gabrielian W. Gan W. Ge F. Gong Z. Gu P. Guan T.J. Heiman M.E. Higgins R.R. Ji Z. Ke K.A. Ketchum Z. Lai Y. Lei Z. Li J. Li Y. Liang X. Lin F. Lu G.V. Merkulov N. Milshina H.M. Moore A.K. Naik V.A. Narayan B. Neelam D. Nusskern D.B. Rusch S. Salzberg W. Shao B. Shue J. Sun Z. Wang A. Wang X. Wang J. Wang M. Wei R. Wides C. Xiao C. Yan A. Yao J. Ye M. Zhan W. Zhang H. Zhang Q. Zhao L. Zheng F. Zhong W. Zhong S. Zhu S. Zhao D. Gilbert S. Baumhueter G. Spier C. Carter A. Cravchik T. Woodage F. Ali H. An A. Awe D. Baldwin H. Baden M. Barnstead I. Barrow K. Beeson D. Busam A. Carver A. Center M.L. Cheng L. Curry S. Danaher L. Davenport R. Desilets S. Dietz K. Dodson L. Doup S. Ferriera N. Garg A. Gluecksmann B. Hart J. Haynes C. Haynes C. Heiner S. Hladun D. Hostin J. Houck T. Howland C. Ibegwam J. Johnson F. Kalush L. Kline S. Koduru A. Love F. Mann D. May S. McCawley T. McIntosh I. McMullen M. Moy L. Moy B. Murphy K. Nelson C. Pfannkoch E. Pratts V. Puri H. Qureshi M. Reardon R. Rodriguez Y.H. Rogers D. Romblad B. Ruhfel R. Scott C. Sitter M. Smallwood E. Stewart R. Strong E. Suh R. Thomas N.N. Tint S. Tse C. Vech G. Wang J. Wetter S. Williams M. Williams S. Windsor E. Winn-Deen K. Wolfe J. Zaveri K. Zaveri J.F. Abril R. Guigo M.J. Campbell K.V. Sjolander B. Karlak A. Kejariwal H. Mi B. Lazareva T. Hatton A. Narechania K. Diemer A. Muruganujan N. Guo S. Sato V. Bafna S. Istrail R. Lippert R. Schwartz B. Walenz S. Yooseph D. Allen A. Basu J. Baxendale L. Blick M. Caminha J. Carnes-Stine P. Caulk Y.H. Chiang M. Coyne C. Dahlke A. Mays M. Dombroski M. Donnelly D. Ely S. Esparham C. Fosler H. Gire S. Glanowski K. Glasser A. Glodek M. Gorokhov K. Graham B. Gropman M. Harris J. Heil

S. Henderson J. Hoover D. Jennings C. Jordan J. Jordan J. Kasha L. Kagan C. Kraft A. Levitsky M. Lewis X. Liu J. Lopez D. Ma W. Majoros J. McDaniel S. Murphy M. Newman T. Nguyen N. Nguyen M. Nodell S. Pan J. Peck M. Peterson W. Rowe R. Sanders J. Scott M. Simpson T. Smith A. Sprague T. Stockwell R. Turner E. Venter M. Wang M. Wen D. Wu M. Wu A. Xia A. Zandieh and X. Zhu. 2001. The sequence of the human genome. *Science* **291:** 1304-1351.

Wang, J., L. Tian, A. Madlung, H.S. Lee, M. Chen, J.J. Lee, B. Watson, T. Kagochi, L. Comai, and Z.J. Chen. 2004. Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. *Genetics* **167:** 1961-1973.

Wei, F., K. Gobelman-Werner, S.M. Morroll, J. Kurth, L. Mao, R.A. Wing, D. Leister, P. Schulze-Lefert, and R.P. Wise. 1999. The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* **153:** 1929-1948.

CHAPTER 3

Analysis of orthologous genes in polyploid wheat suggested that the translated part of genes is protected against changes by an unknown mechanism independent of selection pressure

(Prepared for Science)

Jasdeep S. Mutti and Kulvinder S. Gill*

Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420.

*Corresponding author:

Kulvinder S. Gill

Department of Crop & Soil Sciences

277 Johnson Hall, P. O. Box 646420

Washington State University, Pullman, WA 99163

Email: ksgill@wsu.edu

Tel. 509-335-4666

3.1 Abstract

A sequence comparison of orthologous genes was performed on the three homoeologous genomes of allohexaploid wheat that evolved from a common progenitor about 3 million years ago (MYA). Two genomes came into the same nucleus at different time: ~0.5 MYA and the third about 10,000 years ago. Sequence comparison of the three homoeologues for 43 genes showed a distinct pattern of changes that were common to all genes. As expected, the majority of these changes were localized to be present in the untranslated regions (UTRs) of the genes showing 3.5-fold higher single nucleotide polymorphisms (SNPs) and 85-fold more two base substitutions. The insertion and deletions were exclusively limited to the UTRs and were 1150-fold more abundant than those found in the translated region of the genes. The changes present within the translated part were however, unevenly distributed and the pattern observed was similar amongst all the analyzed genes. The mid part of the genes in the translated regions showed the least number of SNPs whereas the 3' end of translated genes showed no insertion/deletions in any of the genes. Accounting for codon usage, the ratio of synonymous to non-synonymous changes was as expected by random distribution and increased synonymous changes expected on account of selection pressure were not observed.

3.2 Introduction

Understanding the evolution of duplicated genes is of prime importance to the study of plant evolution since 50 to 70% of the angiosperms are believed to be polyploids (Stebbins 1966; Masterson 1994). Gene or genome duplication has been proposed as a primary source for evolution of new gene functions and expression patterns (Ohno 1970; Grant 1981). This can also lead to a situation of complete

redundancy, wherein one duplicated copy is likely to be expendable. Theoretical models have predicted three outcomes of gene duplication: i) silencing of one of the duplicated copies by degenerative mutation (*nonfunctionalization*); ii) divergence of duplicated genes resulting in new functions (*neofunctionalization*); or iii) changed gene expression in such a way that the combined expression of the two copies is reduced to the level of a single copy gene (*subfunctionalization*) (Ohno 1970; Lynch and Force 2000). Studies based on newly formed polyploids have reported a variety of structural and functional reorganization events for the duplicated genomes including chromosomal translocations, sequence elimination, and silencing or activation of duplicated genes relative to their diploid parents (Chen 2007, for review). Differential expression in different tissues seems to be a common fate of duplicated genes as it has been reported for a significant number of genes of both newly formed as well as natural polyploids in different organisms (Adams et al. 2003; see chapter 2, section 2.3.2). Homoeologues of 87% of the 220 wheat genes studied showed a differential expression pattern in different tissues (see chapter 2, section 2.3.2). However, it is still unclear how these duplicated copies are conserved during evolution, wherein one copy is likely to be expendable due to complete redundancy.

Only a small fraction of all higher eukaryotic genomes represent sequences that are involved in gene expression. In general, functionally relevant sequences in a genome are significantly more conserved as compared to the sequences without any known functions. Even though differentially conserved, many expressed genes don't seem to have any function. No function could be assigned to 21% of the yeast genes even after a systematic search (considering the sequence length of genes, sequence attributes like conserved syntenic positions in the sequence and there domain structure catalogued on Pfam) (Peña-Castillo and Hughes 2007). Similarly, many conserved

sequences are known to be non-functional as these results in truncated proteins that cannot be translated into a full-length protein. Nevertheless, highly conserved sequences of 30 to >50 bp are detected throughout the eukaryotic genome but it is not known if this conservation is driven by natural selection due to their functional significance (Duret et al.1993; Lipman 1997; Duret and Bucher 1997).

Gene and genome change over time and differential conservation of different sequences is attributed to the process of natural selection as defined by Charles Darwin (1859). According to the theory of natural selection, any change and/or mutation that increase the organism's fitness for survival will be selected and thus will become fixed. On the other hand, any change(s) that decrease fitness will gradually get eliminated or reduced in frequency. If natural selection is the only process regulating changes in the genome, a very strong correlation is expected between the function of a gene/sequences and its conservation through evolution. By contrast, authors studying five point mutations in an allele of the ß-lactamase gene, that jointly increase resistance against cefotaxime by a factor of 100,000 showed that 102 of the 120 mutational trajectories leading to this allele were inaccessible to natural selection. For the remaining 18 trajectories the probability of realization was negligible, as four of these five mutations in various combinations failed to increase drug resistance. Based on these results, they concluded that pervasive biophysical pleiotropy is likely to be the cause for the evolution of this particular allele (Weinreich et al. 2006).

Accounting for codon bias, about 30% of the base pairs in the coding region of genes are synonymous, and thus can undergo neutral changes. The rate of changes in these synonymous bases should be similar to those in non-coding/non-functional parts of the genome, and thus should occur significantly more frequently than non-

synonymous base changes. On the other hand, if the coding regions of genomes are protected against changes, then it should be equally protected against changes irrespective of their functional significance. Thus, the majority of mutations in the translated region are expected to be neutral or in synonymous bases such that only a minority of the mutations will change protein function. Secondly, with whole genome duplication or polyploidization, the redundant copies of the genes should be prone to inactivation by mutations unless subfunctionalization or neofunctionalization occurs rapidly. As a result, any comparison among the orthologous or paralogous genes should result in a plethora of changes randomly distributed along the length of the gene.

Accordingly, we studied the evolution of duplicated genes with respect to the distribution of changes within the genic region of the genomes along with codon usage of orthologous genes. If natural selection is the only driving force for the changes then the majority of these changes should result in synonymous codons as compared to the non-synonymous codons. For this purpose we studied hexaploid wheat (*Triticum aestivum* L., AABBDD, 2*n*=2*x*=42); which is a relatively young polyploid with its evolution from diploid to tetraploid at ~0.5 MYA and from tetraploid to hexaploid at only ~10,000 years ago. The three genomes evolved from a common progenitor ~3 MYA (Gill et al. 2004).

3.3 Sequence comparison of orthologues

Unique ESTs corresponding to 9,400 genes have been physically mapped to wheat chromosomes by gel-blot analysis of 164 wheat aneuploid lines including nullitetra, ditelosomic and deletion lines (http://wheat.pw.usda.gov/NSF/). Gel-blot analysis images of 6000 ESTs that detected less than six bands were evaluated to

select 854 ESTs that showed a clear hybridization pattern and where every restriction fragment band was physically mapped. The selected ESTs were compared with the full-length rice cDNAs and the rice homologues were identified using a cut off e-value of e^{-70} (see chapter 2, section 2.3.1). The selected wheat ESTs and the rice full-length cDNAs were compared with the 'Chinese spring' (CS) wheat-specific EST database via 'megablast' (http://www.ncbi.nlm.nih.gov/BLAST/) and were aligned using the 'ContigExpress' Module of *Vector NTI Advance*™ software. A cutoff value of ≥80% with minimum overlap of 30 bp was used for wheat-rice comparisons. These two comparisons yielded CS wheat specific EST contigs for each of the selected genes.

An example of the analysis and the approach to identify ESTs corresponding to each homoeologue is given in Figure 3.1a. For example, the EST *BE425898* was physically mapped between FL 0.47 and 0.86 on the short arm of wheat homoeologous group *1* (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi). The corresponding gene has three structural copies in the wheat genome as deletion mapping detected three fragment bands, one on each group *1* homoeologue. Megablast identified 11 ESTs with more than 80% homology, upon its comparison with the CS EST database. Similar comparison using its rice full-length cDNA homologoue identified an additional 12 wheat EST, to extend the contig across the entire gene. Represented by three distinct sequence patterns, all three copies of this gene were scored as being expressed (Figure 3.1b). This approach was used to study the expression pattern of the 854 putative wheat genes (see chapter 2, section 2.3).

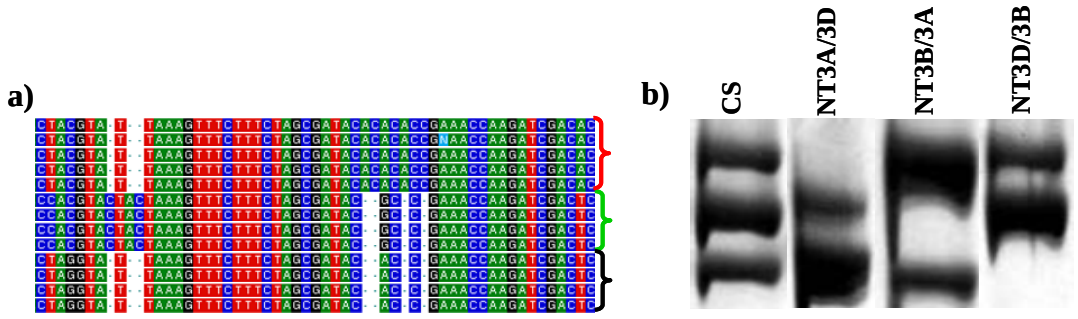


Figure 3.1. (a) Sequence alignment of the EST *BE425898* within the contig. ESTs corresponding to each of the three wheat homoeologues were distinguished based on different patterns as bracketed on right side by red, green and black brackets. **(b)** cDNA-SSCP analysis for gene *BE586090* to confirm the *in silico* results for homoeologous gene expression. Matching with the source of ESTs, RNA samples from 11 different developmental stages (5 days old seedling, root from seedling and adult plants, 28 day old plant, early flowering (Feekes scale: 6) (Large 1954), meiosis, pre-anthesis, post-anthesis. flag leaf, Seed at DPA 5 and 20 stage and adult plant) were used for the cDNA-SSCP analysis. CS denotes cultivar 'Chinese spring' that is the wild-type parent of the nullisomic tetrasomic lines; NT3A/3D, NT3B/3A, and NT3D/3B are the group *3* nullisomic-tetrasomic lines. The *in silico* and the cDNA-SSCP methods are described in detail in chapter 2 (section 2.3).

3.4 Distribution of changes within genes

Rice homologues were identified for 138 of the 854 putative wheat genes studied. The identified rice homologues were used to assemble full-length wheat contigs (Figure 3.2a). Fourty-three of the 138 assembled wheat contigs had full-length sequences generated for all three homoeologues and thus were used for this analysis. To score the sequence differences among the three homoeologues within the coding region, the full-length contig was further demarcated into the translated region, 5' and 3' UTR regions using the rice full-length cDNA sequence as a template (Figure 3.2a). Translation start and stop codons were identified using the rice full-length cDNA sequence and were observed to be conserved among the wheat sequences for all the 138 full-length contigs analyzed. To analyze any pattern of sequence conservation within the translated region, it was further divided into four quarters (covering 25% of sequence) as shown in Figure 3.2a. The number of insertion/deletions and single nucleotide polymorphisms (SNP) were scored separately for each demarcated region. An example of this analysis is shown in Figure 3.2a. Wheat EST (*BE446240*) was used to identify a full-length rice cDNA (shown as a green line). The identified full-length rice cDNA was further used to assemble a full-length wheat contig and to demarcate the UTR and translated region of the gene. Differentiated homoeologues were scored for sequence differences. These differences were scored separately among the demarcated regions on the full-length cDNA. Figure 3.2b summarizes the distribution of number for insertion/deletions and SNPs observed along the length of a gene.

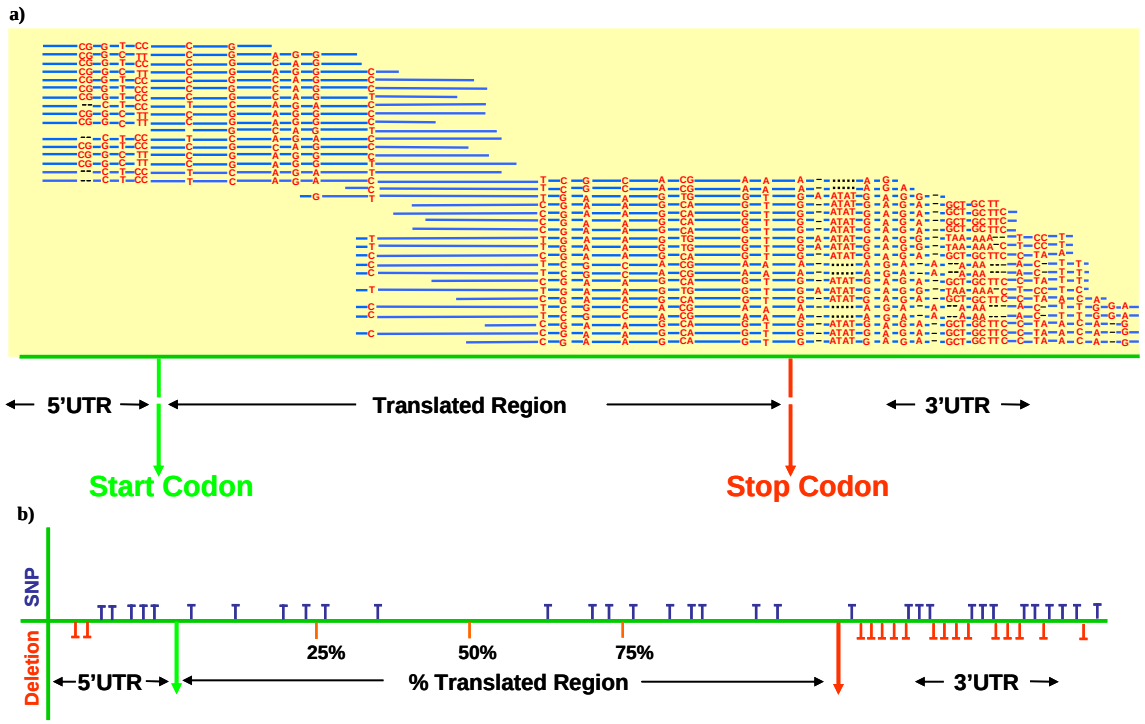


Figure 3.2. (a) Full-length EST contig of wheat gene (*BE446240*). Blue bars represent the sequence showing complete homology among the three homoeologues, whereas base pairs in red marks the differences among homoeologues. (**b)** Summary of differences found among UTRs and transcribed region. Marked above in blue are the number of single base pair changes found among homoeologues, whereas the bottom red bars represent the number of deletions found within the full-length gene divided into 5'UTR, transcribed region and 3'UTR.

Detailed sequence analysis of orthologues of the 43 genes analyzed showed that on an average there were 40 SNPs, 2.3 two-bp substitutions (BPS), and 1.7 one and more bp deletions (BPD) per kb of the translated region. The translated part of the gene contained only 20 SNP, 0.1 BPS, and 0.01 BPD per kb where as the 5' UTR region showed 70 SNP, 8 BPS, and 10 BPD per kb and the 3' UTR showed 60 SNP, nine BPS, and 13 BPD per kb. There were significant differences even within the translated part of the genes as no BPS or BPDs were observed in the 25% of the genes at the 3' end of open reading frame (ORF). Single base pair changes were prevalent throughout the translated region of the gene with their intensity maximum in the UTR regions, especially the 3' UTR indicated by red bar in Figure 3.3.

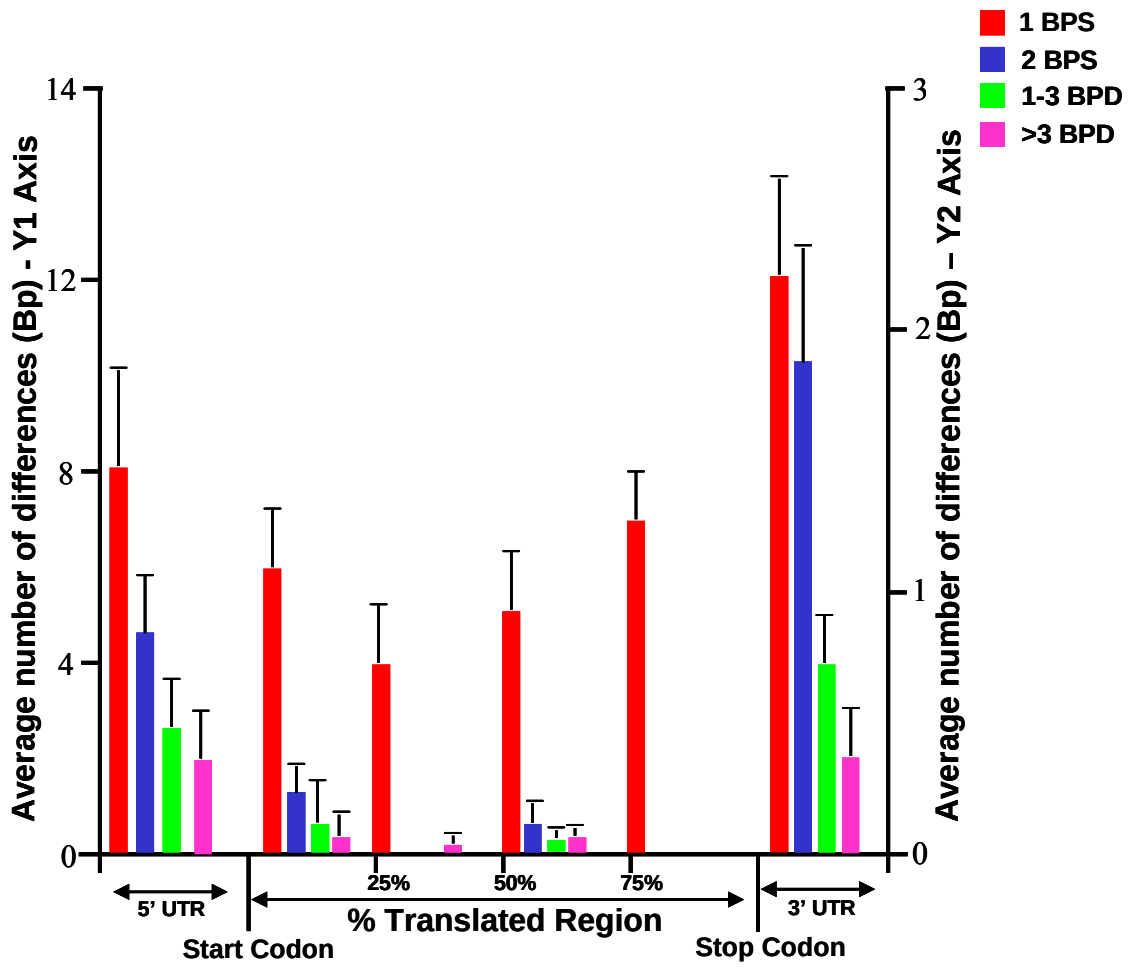


Figure 3.3. Differences in base pairs among the expressed part of homoeologues for 43 wheat genes. X-axis divides the gene into 5' UTR, % translated region (further divided into 0-25%, 25-50%, 50-75%, 75-100%) and 3' UTR region. The Y1-axis shows the average number of SNPs among the orthologs and Y2 axis presents the average number of 2 or more than 2 bp differences among the orthologous genes. The red and blue bars represent one and two base pair substitutions (BPS), whereas green and pink bars denote one to three and more than three base pair deletions (BPD). Error bars represent the standard deviation of the mean.

Two base pair substitutions and 1 to 3 or >3 base pair changes or deletions, indicated as blue, green and pink bars in Figure 3.3, were also more prevalent in the UTR regions of the gene. In the case of some of the genes, 2 or more than two BPD were found at the first 25% of the translated region, we suspect them as the genes that code for proproteins. The last 25% of the translated region seemed to be most protected as there were no ≥2 base pair changes found in any of the 43 gene pairs studied. Only one base-pair changes were prevalent in this region of the gene. Single-base pair changes as shown by the red bar in Figure 3.3 were prevalent throughout the translated region but their intensity as such was more in case of UTRs as compared to the ORF region. Within the translated region itself the frequency of single base pair differences was more on the first and last 25% region of the gene as compared to the middle part of the gene. These deletions and substitutions in the translated part of the gene seemed to be governed by some mechanism in the cell rather than as a result of chance mutation due to natural selection.

3.5 Synonymous vs non-synonymous changes

It was our expectation, that natural selection probably prevents mutations leading to loss of function in the coding region. If this were the case, then we would expect to find most of the mutations in the ORF on the synonymous codon. Thus, for each full-length contig, we examined the frequency of synonymous and non-synonomous base-pair changes. For each full-length wheat contig, an ORF starting with a start codon (ATG) and ending with a stop codon (TAA, TGA or TAG) was identified using *Vector NTI Advance*™ software. The ORF was then used to calculate two values: the number of synonymous substitutions and the number of non-synonymous substitutions. The fraction of synonymous and non-synonymous

differences was calculated for the translated part of each of the randomly selected homoeologues for 16 of 43 genes. After accounting for wheat codon usage 30.3% of the bases are synonymous (ratio of number of synonymous bases to total number of bases accounted in a genetic code), that is favorable for mutation. Surprisingly, our analysis showed that 69.9% of the differences in the 16 homoeologous genes were in non-synonymous base pairs as compared to 30.1% that were in synonymous base pairs. Comparison with the codon usage of wheat showed that the scarcity of changes in the translated part of the gene was not due to selection pressure as the number of changes in the wobble bases was not significantly higher than others.

3.6 Discussion

To investigate the pattern of sequence conservation among the UTR and translated part of the gene, we generated a profile wherein rice full-length cDNA sequences were used to identify and align three full-length homoeologues of wheat. Profiles of sequence conservation were drawn for the 5' and 3' UTR region as well as the translated part of the gene, demarcated using rice full-length cDNA. Genic boundaries, as demarcated by start and stop codon were seen to be highly conserved among the rice full-length cDNA and its counterpart, the wheat full-length gene sequence. The start codon represented a boundary where the changes were kept to minimum. Deletions were almost exclusively present outside in the UTR region. These patterns of changes were consistent among all the 43 analyzed genes thus illustrating the conservation of genic boundaries (the UTRs and translated region) among wheat and rice.

Sequence comparison of the three wheat homoeologues for 43 genes showed a distinct pattern of change, common to all the analyzed genes. The majority of the changes were localized in the UTRs of the genes, which showed 3.5-fold more SNPs and 85-fold more two base substitutions. The insertion/deletions were almost exclusively limited to the UTRs and were present 1150- fold more than that found in the translated part of the genes. The changes present within the translated region were unevenly distributed and the pattern was similar among all analyzed genes. Among the translated region, less number of SNPs was present in the middle of the gene. Whereas there frequency increased towards the peripheral region of the gene (both towards 5' and 3' end). Insertion/deletions also exhibited a unique pattern of distribution within the translated part. The 3' end of the translated part showed no insertion/deletions in any of the genes analyzed whereas the majority of the insertion/deletions observed were present towards the 5' end and few were in the middle of the gene.

Accounting for wheat codon usage wherein 30.3% of the bases are synonymous, that is favorable for mutation; our results showed that 69.9% of the changes were in non-wobble base pair while 30.1% were wobble base pairs. The number of changes on synonymous bases showed no significant increase and was as expected by random distribution. If selection pressure is the only mechanism governing these changes then the maximum number of the changes should be on the synonymous bases, with the minimum number of changes on non-synonymous bases. Our comparison showed that the majority (69.9%) of changes were on non-synonymous bases compared to 30.1% that were present on synonymous bases. This led us to conclude that the scarcity of changes in the translated part of the gene was not due to selection pressure as any increase in synonymous changes that was expected due to selection pressure,

was not observed. Based on these results and the pattern of changes of substitutions and insertion/deletions within the translated part of the gene, we postulate that wheat and perhaps other higher eukaryotes have a mechanism to differentiate and protect 'important' parts of the genes from change.

The cell has a precise mechanism to distinguish the genic and non-genic regions on the chromosome. These regions are decorated differently with specific histone tags and other epigenetic mechanisms like DNA methylation. Even the packaging of different regions on the chromosome is differential for heterochromatin and euchromatin regions. Histone modifications have distinctive patterns among major domains of the chromosome: the euchromatin and heterochromatin. Heterochromatin is distinguished by hypoacetylation and H3K9 methylation marks whereas euchromatin is characterized by histone H4 acetylation and methylation of histone H3 at lysine 4 (H3K4me) (Millar and Grumstein 2006). Even within euchromatin: the promoter, UTRs and the open reading frame or translated part of the gene has different patterns of histone modification. DNA methylation also seems to distinguish among the different regions of the chromosome and is preferentially present on the repeat sequences and centromeric regions of the genome (Zhang et al. 2006; Zilberman et al. 2007). About one third of the analyzed genes were also methylated in their open reading frames and less than 5% of the genes had their promoter sequence methylated (Zhang et al. 2006; Zilberman et al. 2007). These gradients of regulatory mechanisms with preferential presence or absence from particular regions or sequences suggest a central control which distinguishes and regulates the various regions of genomes differentially.

Numerous types of histone modifications are widely studied on a genomic scale (Liu et al. 2005). We analyzed the available data to determine if histone

modifications make any pattern within and outside the gene that correlates with the observed pattern of sequence conservation. These modifications were further divided into: transcription dependent and independent modifications. The summary of the analysis is shown in Figure 3.4a. These modifications showed a pattern which is consistent with the pattern of sequence conservation we established in comparison among wheat homoeologues. Transcription dependent modifications are shown as blue blocks in Figure 3.4a, H3K4 methylation showed gradient of modification from trimethylation at the 5' end to monomethyaltion at 3' end with dimethylation prevalent in the middle of the gene. Htz1K14 acetylation is enriched in promoter region and acetylation of H3K9 and H3K14 highest in 5'end of the gene (shown as orange blocks in Figure 3.4a). For transcription independent modifications the 5' coding region is primarily excluded from acetylation of H4K8, H4K16 and H2BK16 (represented by green lines in Figure 3.4a). Acetylation of H4K8 is absent from 5' UTR region, whereas it is present along the body of the gene thereafter. Acteylation of H2BK16 is highest in middle part of the gene and H4K16 actylation is absent in 5' UTR region but is present in middle and 3' region of coding sequence. The pattern of H2BK16 modification along the length of the translated part of the gene is consistent with the distribution of SNP's observed in our analysis. The 5' and 3' end of the translated region lacks acetylation of H2BK16 and has more of SNP's present. The middle part of the gene shows maximum acetylation and accordingly has less number of SNP's. This gradient of various histone modifications reflect the targeted recruitment of modifying and demodifying enzymes to different regions of the genome as well as within gene thus implying a mechanism which differentiates and modifies different regions differentially, depending upon its function.

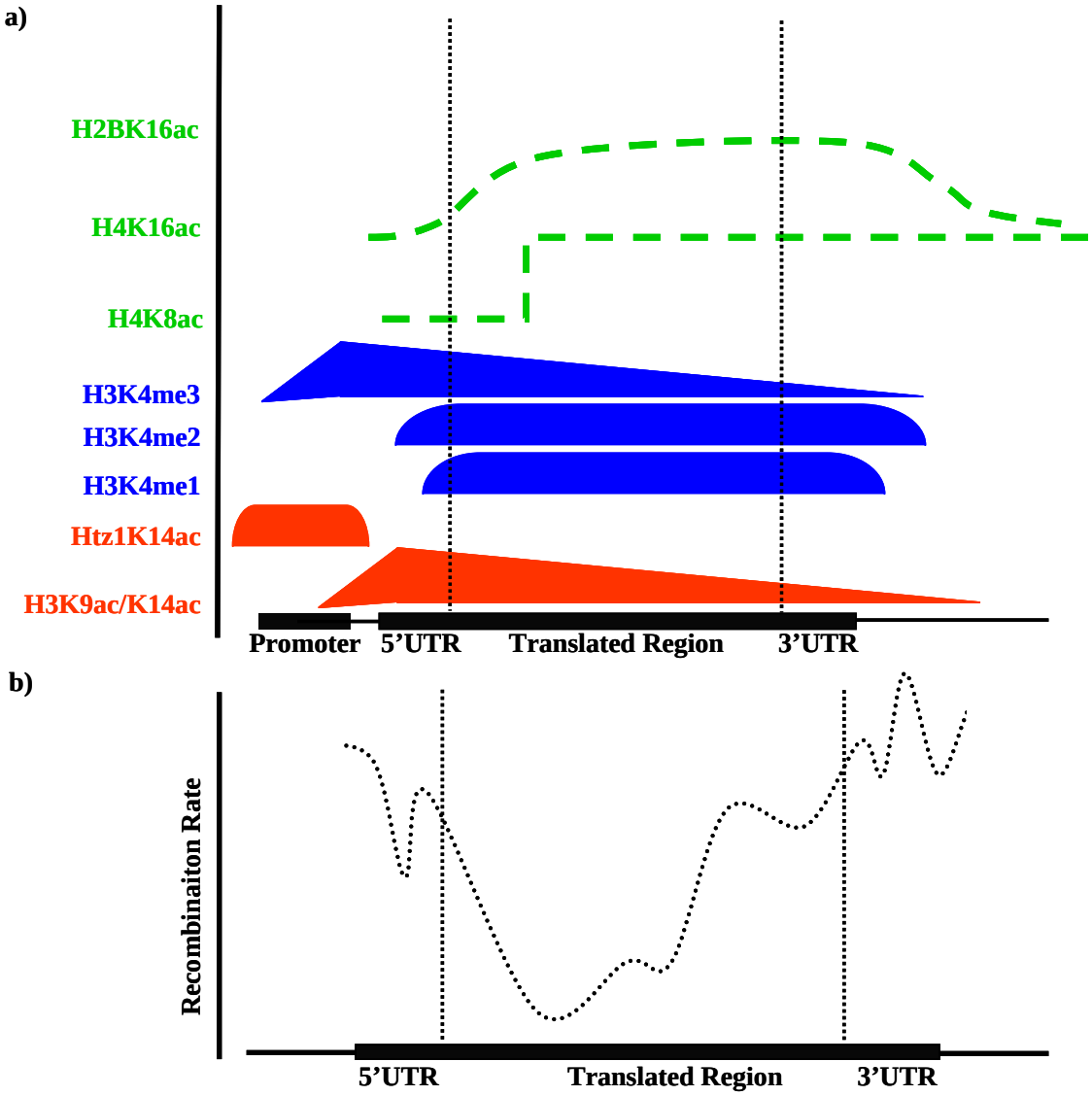


Figure 3.4. Patterns of recombination rate and various histone modifications within genes. **(a)** Pattern of different types of histone modifications found in yeast, averaged over 5324 genes. Green dotted lines represent transcription independent modifications, and blue and red solid lines represent transcription dependent histone methylation and acetylation modifications. Transcription dependent modifications were redrawn using data from Millar et al. (2006) and from Liu et al. (2005) for transcription independent modifications. **(b)** Distribution of recombination within the transcribed parts averaged over 25,000 human genes (Myers et al. 2005).

Recombination rates across the genome are also well studied in humans (Myers et al. 2005). Recombinantion rates are shown to be highly variable across the genome with 80% of the recombination occurring in 10 to 20% of the sequence. In all, 25,000 hot spots were identified which is comparable to the estimated number of genes on the human genome (Myers et al. 2005). Recombination rates tend to be higher in telomeric regions as compared to the centromeric regions. Recombination rates were lower within genes and increased with distance from the gene symmetrically in either side. Recombination hot spots in humans were preferentially located in regions within ~50 kb outside the transcribed domain of a gene. Even within the transcribed region, UTRs showed marked difference in recombination rates as compared to the middle of the gene (Figure 3.4b). Within the transcribed region, the first and the last exon of the gene showed, on average, greater recombination rates as compared to the average on internal exons (Figure 3.4b). This pattern of recombination distribution is exactly similar to what we observed for sequence conservation within the coding region of the wheat gene.

The selection pressure is main driving force to conserve gene sequences such that any change and/or mutation that increase the organism's survival and fitness will be selected. On the other hand, any change that decreases survival and fitness, will gradually be reduced in frequency. As such, all the functionally relevant gene/sequences will be expected to show strong sequence conservation throughout. Whereas, in our study we have demonstrated that the conservation pattern in the genic regions is unique. Maximum conservation of the sequence was seen on middle region of the gene as compared to the 5' and 3' end. The middle of the genes showed the least number of SNPs whereas the 3' end of the translated part showed no insertion/deletions in any of the genes. Accounting for codon usage, the ratio of

synonymous to non-synonymous changes did not show any increase in synonymous changes that was expected due to selection pressure. These results suggest that different parts of the genes are differentially susceptible to changes and these differences can not all be explained by natural selection as the number of non-synonymous changes was significantly more than the synonymous changes. These patterns of epigenetic mechanisms and molecular events like recombination and biased conservation of the sequences within genes, led us to believe in a mechanism which differentiate and protect certain functionally relevant regions of the genome from changes. This unknown mechanism may also utilize various histone modification patterns which are observed to be unique for different parts of the genes.

3.7 References

Adams, K.L., R. Cronn, R. Percifield, and J.F. Wendel. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* **100:** 4649-4654.

Chen, Z.J. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* **58:** 377-406.

Duret, L. and P. Bucher. 1997. Searching for regulatory elements in human noncoding sequences. *Current Opinion in Structural Biology* **7:** 399-406.

Duret, L., F. Dorkeld, and C. Gautier. 1993. Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression. *Nucleic Acid Research* **21:** 2315-2322.

Gill, B.S., R. Appels, A.-M. Botha-Oberholster, C.R. Buell, J.L. Bennetzen, B. Chalhoub, F. Chumley, J. Dvorak, M. Iwanaga, B. Keller, W. Li, R. McCombie, Y. Ogihara, F. Quetier, and T. Sasaki**.** 2004. A workshop report on wheat genome sequencing: international genome research on wheat consortium. *Genetics*, **168:** 1087–1096

Grant, V. 1981. *Plant Speciation*. Columbia University Press, New York.

Jensen-Seaman, M.I., T.S. Furey, B.A. Payseur, Y. Lu, K.M. Roskin, C.F. Chen, M.A. Thomas, D. Haussler, and H.J. Jacob. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research* **14:** 528-538.

Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E.S. Lander. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423:** 241-254.

Langley, C.H., E. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth. 1988. On the role of unequal exchange in the containment of transposable element copy number. *Genetical Research* **52:** 223-235.

Larizza, A., W. Makalowski, G. Pesole, and C. Saccone. 2002. Evolutionary dynamics of mammalian mRNA untranslated regions by comparative analysis of orthologous human, artiodactyl and rodent gene pairs. *Journal of Computational Chemistry* **26:** 479-490.

Li, B., M. Carey, and J. Workman. 2007. The Role of Chromatin during Transcription. *Cell* **128:** 707-719.

Lipman, D. 1997. Making (anti)sense of non-coding sequence conservation. *Nucleic Acids Research* **25:** 3580-3583.

Liu, X.M., A.K. Fritz, J.C. Reese, G.E. Wilde, B.S. Gill, and M.S. Chen. 2005. H9, H10, and H11 compose a cluster of Hessian fly-resistance genes in the distal

gene-rich region of wheat chromosome 1AS. *Theoretical and Applied Genetics* **110:** 1473-1480.

Lynch, M. and A. Force. 2000. The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics* **154:** 459-473.

Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264:** 421-424.

Millar, C.B. and M. Grumstein. 2006. Genome-wide patterns of histone modifications in yeast. *Nature Reviews Molecular Cell Biology* **7:** 657-666.

Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310:** 321-324.

Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.

Pesole, G., F. Mignone, C. Gissi, G. Grillo, F. Licciulli, and S. Liuni. 2001. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* **276:** 73-81.

Shabalina, S.A., A.Y. Ogurtsov, I.B. Rogozin, E.V. Koonin, and D.J. Lipman. 2004. Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. *Nucleic Acids Research.* **32:** 1774-1782.

Sidhu, D. and K.S. Gill. 2004. Distribution Of Genes And Recombination In Wheat And Other Eukaryotes. *Plant Cell Tissue and Organ culture* **79:** 257-270.

Stebbins, G.L. 1966. Chromosomal Variation and Evolution. *Science* **152:** 1463-1469.

Weinreich, D.M., N.F. Delaney, M.A. DePristo, and D.L. Hartl. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science* **312:** 111-114.

Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S. Chan, H. Chen, I. Henderson, P. Shinn, M. Pellegrini, S. Jacobsen, and J.R. Ecker. 2006. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell* **126:** 1189-1201.

Zilberman, D., M. Gehring, R.K. Tran, T. Ballinger, and S. Henikoff. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics* **39:** 61-69.

CHAPTER 4

Registration of 28 wheat germplasm lines developed in the background of seven winter wheat cultivars by transferring seven marker segments associated to genes conferring resistance to various viral, fungal and insect pests

Jasdeep S. Mutti[1], P. Stephen Baenziger[2], Ming-Shun Chen[3], and Kulvinder S. Gill[1*]

[1]Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420.

[2]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583-0940.

[3]Department of Entomology, Kansas State University, Manhattan, KS 66506.

[*]*Corresponding author*: Department of Crop & Soil Sciences, 277 Johnson Hall, PO Box 646420, Washington State University, Pullman, WA 99164-6420. E-mail: ksgill@wsu.edu

4.1 Abstract

Wheat (*Triticum aestivum* L.) germplasm lines carrying genes conferring resistance to: wheat streak mosaic virus - Wahoo/*wsm1,* Millennium/*wsm1*, Wesley/*wsm1,* Alliance/*wsm1,* Harry/*wsm1,* Goodstreak/*wsm1* and Arrowsmith/*wsm1*; Hessian fly - Wahoo/*H9,* Wesley/*H9,* Harry/*H9,* Wahoo/*H13,* Wesley/*H13,* Harry/*H13;* Russian wheat aphid - Wahoo/*Dn4,* Millennium/*Dn4*, Alliance/*Dn4,* Harry/*Dn4* and Arrowsmith/*Dn4*; wheat rusts - Wahoo/*Lr37,Yr17,Sr38,* Millennium/*Lr37,Yr17,Sr38,* Wesley/*Lr37,Yr17,Sr38,* Alliance/*Lr37,Yr17,Sr38,* Harry/*Lr37,Yr17,Sr38,* Arrowsmith/*Lr37,Yr17,Sr38,* Wesley/*Lr39,* Wahoo/*Lr39;* and fussarium head blight - Wesley/*Fhb1* and Harry/*Fhb1*; were co-developed by Washington State University, Pullman, WA and University of Nebraska, Lincoln, NE.

4.2 Introduction

Wheat (*Triticum aestivum* L.) germplasm lines carrying genes conferring resistance to: wheat streak mosaic virus - Wahoo/*wsm1,* Millennium/*wsm1*, Wesley/*wsm1,* Alliance/*wsm1,* Harry/*wsm1,* Goodstreak/*wsm1* and Arrowsmith/*wsm1*; Hessian fly - Wahoo/*H9,* Wesley/*H9,* Harry/*H9,* Wahoo/*H13,* Wesley/*H13,* Harry/*H13;* Russian wheat aphid - Wahoo/*Dn4,* Millennium/*Dn4*, Alliance/*Dn4,* Harry/*Dn4* and Arrowsmith/*Dn4*; wheat rusts - Wahoo/*Lr37,Yr17,Sr38,* Millennium/*Lr37,Yr17,Sr38,* Wesley/*Lr37,Yr17,Sr38,* Alliance/*Lr37,Yr17,Sr38,* Harry/*Lr37,Yr17,Sr38,* Arrowsmith/*Lr37,Yr17,Sr38,* Wesley/*Lr39,* Wahoo/*Lr39;* and head blight - Wesley/*Fhb1* and Harry/*Fhb1*; were co-developed by Washington State University, Pullman, WA and University of Nebraska, Lincoln, NE.

Recent developments in plant molecular biology and wheat genetics have a potential to instigate a second green revolution (http://maswheat.ucdavis.edu/). Molecular markers linked to genes controlling pest resistance, difficult to screen traits, and to traits of high agronomic values such as seed protein content are now available. Marker assisted selection (MAS) is a strategy to capitalize on available molecular markers and quickly move valuable traits into elite lines that are suitable for cultivar release (Paterson et al. 1991). These markers are particularly valuable for gene pyramiding that involves transfer of multiple genes conferring resistance to different biotypes of the same pest (Singh et al. 2001). Here, we demonstrate the power of this strategy by developing 28 germplasm lines using MAS.

Table 4.1 Band sizes for markers run on different parents.

Gene	Marker	Distance	Band Size (bp)		Gel Type
			Resistant Parent	Susceptible Parent	
Wsm1	STSJ15	Does not recombine	~400	-	0.8% agarose
H13	*Xgdm36*	2.7 cM	170	130	3% agarose
Dn4	*Xgwm106* and *Xgwm337*	7.4 and 1.3 cM	125 and 175	- and 150	1.2% and 3% agarose
Lr37,Yr17,Sr38	VENTRIUP-LN2	Does not recombine	260	-	1% agarose
Lr39	*Xgdm35*	0.9 cM	300	400	6.5% polyacrylamide
Qfhs.ndsu-3BS	*Xgwm493*	~11 cM	290	~200	3% agarose

The F_1's for all the crosses were made in January, 2002. Two generations per year of the backcross plants were screened. Rapid DNA extraction method described in Chapter 6 (section 6.5.1) was used for marker analysis. In each generation ten plants with marker linked to gene of interest were selected for further backcrossing. Only in case of Hessian fly resistance line, phenotypic selection was done and at least ten plants resistant to biotype L were used for further backcrossing.

4.3 Registration of seven wheat germplasm lines carrying *Wsm1* marker segment associated to the gene conferring resistance to wheat streak mosaic virus.

Seven germplasm lines carrying marker segment associated with *Wsm1* gene, conferring resistance to wheat streak mosaic virus were developed. The *Wsm1* gene donor was a hard red winter wheat line 'KS93WGRC27' (PI 583794) that carries the gene on an *Agropyron intermedium* (Host) P. Beauv. segment translocated on wheat chromosome 4DS (Friebe et al. 1991; Gill et al. 1995). This line was backcrossed five times to each of the hard red winter wheat cultivars: 'Wahoo' (Baenziger et al. 2002, PI 619098), 'Wesley' (Peterson et al. 2001, PI 605742,), 'Alliance' (Baenziger et al. 1995, PI 573096), 'Millennium' (Baenziger et al. 2001, PI 613099), 'Goodstreak' (Baenziger et al. 2004a, PI 632434) and 'Harry' (Baenziger et al. 2004b, PI 632435). A hard white winter wheat cultivar 'Arrowsmith' (Graybosch et al. 2005, PI 633911) was also used as one of the recurrent parents. The recurrent parents were always used as female parents for all crosses. After five backcrosses the plants were selfed for homozygosity. The pedigree of germplasms containing *Wsm1* gene is KS93WGRC27/*5 Wahoo, KS93WGRC27/*5 Millennium, KS93WGRC27/*5

Wesley, KS93WGRC27/*5 Alliance, KS93WGRC27/*5 Harry, KS93WGRC27/*5 Goodstreak and KS93WGRC27/*5 Arrowsmith.

Beginning with the BC_1 generation, the sequence-tagged-site (STS) marker STSJ15 that is specific for the translocated segment carrying *Wsm1* gene (Talbert et al. 1996), was used to select resistant plants. Since the gene is a translocated alien fragment, it no longer recombines with wheat chromatin. As such, the genetic distance of the marker to the gene is zero, even when the physical distance could be very large. The expected band sizes for the resistant and the susceptible plants are given in Table 4.1. In each backcross, plants heterozygous for STSJ15 primer set were selected. Ten plants from selfed progeny of each of the BC_4F_2 plants were screened with the marker in order to select BC_4F_2 homozygous for the gene. Ten plants homozygous for marker were used for seed increase of the germplasm line.

4.4 Registration of 11 wheat germplasm lines carrying genes conferring resistance against Hessian fly and marker segment associated to the gene conferring resistance to Russian wheat aphid.

Eleven germplasm lines carrying genes conferring resistance to Hessian fly (*Mayetiola destructor* Say) and marker segment associated to gene *Dn4* conferring resistance to Russian wheat aphid (*Diuraphis noxia* Mordvilko), were developed. Of these, three lines carry the *H9* and three carry *H13* genes for Hessian fly resistance. The remaining five lines carry the *Dn4* gene for Russian wheat aphid resistance. The gene *H9* was transferred from *Triticum turgidum* L. to hexaploid wheat (Patterson et al. 1994) and is located on chromosome arm 1AS (Liu et al. 2005). The *H13* gene located on chromosome arm 6DS was transferred from *Aegilops tauschii* (Coss.) Schmal. into hexaploid wheat via direct crosses (Hatchett and Gill 1981). Both the *H9*

and *H13* genes confer resistance to Hessian fly biotype L. 'Iris' (Patterson et al. 1994, PI 562615) was used as the donor for *H9,* whereas 'Molly' (Patterson et al. 1994, PI 562619) was the *H13* gene donor. Five backcrosses followed by selfing were done for each of the eleven germplasms. Cultivars 'Wahoo', 'Wesley' and 'Alliance' were recurrent parents for both *H9* and *H13* introgressions. The pedigree of germplasm lines containing Hessian fly resistance gene *H9* and *H13* is Iris/*5 Wahoo, Iris/*5 Wesley, Iris/*5, Molly/*5 Wahoo, Molly/*5 Wesley and Molly/*5 Harry. Using CO970547 (PI 372129) line as a donor parent, the Russian wheat aphid resistance gene *Dn4* was transferred to: hard red winter wheat 'Wahoo', 'Millennium', 'Harry', 'Alliance' and hard white winter wheat 'Arrowsmith'. The pedigree of the crosses involving the *Dn4* gene is CO970547/*5 Wahoo, CO970547/*5 Millennium, CO970547/*5 Alliance, CO970547/*5 Harry, and CO970547/*5 Arrowsmith.

Screening for both *H9* and *H13* was performed by phenotypic selection with biotype L except for *H13* where homozygous plants at the end of the backcrossing cycle were selected using a codominant DNA marker *Xgdm36* (Liu et al. 2005). Hessian fly resistance screening of the segregating populations along with the respective parents were carried out at Kansas State University in growth chambers set at 18 ± 1^0C with 14:10 h (Light:Dark) photoperiod. A previously described method (Hatchett et al. 1981) was used with modifications. Approximately 20 seeds of each backcross population were planted. Seedlings at the one-leaf stage were infested with ~200 newly mated Hessian fly females with a cheesecloth tent and the plants were scored for resistance after three weeks. Susceptible plants harbored live larvae and were stunted with dark green leaves. Resistant plants contained dead larvae and were normal with light green leaves. Selection for homozygous plants for *H13* was done on BC_4F_2 with a co-dominant marker *Xgdm36* that amplifies a 170 bp band from the

resistant parent and 130 bp band from the susceptible parent (Table 4.1). Screening of 20 BC_4F_2 identified four plants that only showed the donor parent band and thus were homozygous. Plants homozygous for *H9* were selected by phenotypic screening of 10 BC_4F_3 plants each from 20 BC_4F_2 plants. A BC_4F_2 plant was considered homozygous for resistance if all of the ten progeny plants were resistant to biotype L. Ten resistant plants were used for seed increase of the germplasm lines.

Screening for the Russian aphid resistance gene *Dn4* was performed using two flanking microsatellite markers *Xgwm106* and *Xgwm337* (Liu et al. 2002). In each backcross, plants were first screened with marker *Xgwm106*. A second round of screening was done with the *Xgwm337* marker on selected plants carrying *Xgwm106* marker. The expected fragment sizes for both markers are given in Table 4.1. In order to identify four homozygous resistant plants, 20 BC_4F_2 plants had to be screened for each cross. Ten homozygous resistant plants were further used for seed increase of the germplasm line.

4.5 Registration of ten wheat germplasm lines carrying marker segment associated to the genes conferring resistance to important pathogenic wheat fungi.

Ten germplasm lines resistant to wheat fungal pathogens; specifically, leaf (*Puccinia triticina* Eriks), stem (*P. graminis* Pers. f.sp. *tritic*i Eriks. & E. Henn.), stripe rust (*P. striiformis* West. f. sp. tritici) and fussarium head blight (*Fusarium graminearum* Schwabe), were developed. The *Lr37,Yr17,Sr38* gene cluster was transferred to 'Wahoo', 'Wesley', 'Alliance', 'Millennium', 'Harry' and 'Arrowsmith'; *Lr39* gene to 'Wahoo' and 'Wesley'; and a quantitative trait loci (QTL) resistant to Fusarium head blight to 'Wesley' and 'Harry'. The donor parent

for *Lr37,Yr17,Sr38* gene cluster was 'Madsen' (PI 511673, Allan et al. 1989), for *Lr39* gene was 'KS86WGRC02' (Gill et al. 1988, PI 504517) and for QTL resistant to Fusarium head blight was 'Sumai 3' (PI 481542). Five backcrosses followed by a selfing cycle were used for each cross. The pedigree of germplasm lines containing gene cluster *Lr37,Yr17,Sr38* is Madsen/*5 Wahoo, Madsen/*5 Millennium, Madsen/*5 Wesley, Madsen/*5 Alliance, Madsen/*5 Harry and Madsen/*5 Arrowsmith; for *Lr39;* KS86WGRC02/*5 Wahoo and KS86WGRC02/*5 Wesley; and for scab resistance; Sumai3/*5 Wesley and Sumai3/*5 Harry, respectively.

A segment of chromosome 2AS of cultivar Madsen contains a cluster of genes conferring resistance to leaf rust (*Lr37*), stripe rust (*Yr17*) and stem rust (*Sr38*) (Bariana and McIntosh 1993). The gene cluster was first introgressed from *Triticum ventricosum* Tausch (Bariana and McIntosh 1993) into cultivar 'VPM1' (Maia 1967) and later into 'Madsen' (Allan et al. 1989). The *Lr39* was transferred from *A. tauschii* accession TA 1675 into wheat germplasm 'KS86WGRC02' (Gill et al. 1988). The gene is located on the short arm of chromosome 2DS (Raupp et al. 2001). The QTL for scab resistance *Qfhs.ndsu-3BS* (synonymous with *Fhb1*) was derived from 'Sumai 3', located on chromosome arm 3BS (Waldron et al. 1999).

Molecular markers linked to the genes were used to introgress these genes into different recurrent parents. For gene cluster *Lr37,Yr17,Sr38,* VENTRIUP-LN2 dominant marker linked to 2NS segment was used (Helguera et al. 2003). The expected fragment size is given in Table 4.1. Expected ratio of 50% resistance to 50% susceptible plants in each backcross was not seen, instead marker gave 25% resistance to 75% susceptible ratio. For *Lr39* the microsatellite marker *gdm35* (Pestsova et al. 2000) was used and for scab resistance microsatellite markers *Xgwm493* was used (Anderson et al. 2001). The expected fragment size is given in Table 4.1. Selection for

homozygous plant for germplasm lines with *Lr37,Yr17,Sr38* resistance was done in BC_4F_3 generation. For each cross, ten progeny from ten BC_4F_2 plants were screened for homozygosity. The presence of the diagnostic marker(s) in all progeny plants indicated the parent plant was homozygous for resistance gene. The parent plant was considered hetereozygous if there were plants with and without the marker(s) in the progeny. The parent plant was considered as homozygous for the recurrent parent alleles (absence of resistance) if the diagnostic marker was absent in all progeny plants. Ten plants homozygous for the diagnostic marker were used for seed increase of the germplasm line.

In case of *Lr39*, the *Xgdm35* marker amplified ~300 bp band in donor parent as compared to ~400 bp band in susceptible parent. Whereas in case of *Qfhs.ndsu-3BS* resistance, marker *Xgwm493* amplified 290 bp band in resistant parent compared to ~200 bp band in susceptible parent (Table 4.1). Marker *Xgwm533* did not showed any polymorphism among the recurrent parents used in the study and thus was not included for screening. For homozygosity 20 BC_4F_2 plants for each germplasm line containing *Lr39* and *Qfhs.ndsu-3BS (Fhb1)* resistance were screened. Out of 20 plants screened, we found at least 5 plants amplifying only donor parent bands which were used for seed increase of the germplasm line.

These 28 germplasm lines are grown in the greenhouse and in the field for seed increase. Twenty grams of seed of each germplasm line has been deposited in National Small Grains Collection (NSGC) and BC_4F_4's are grown in Nebraska for field evaluations and seed increase. Small quantities of germplasm seeds can be obtained on request from NSGC and the corresponding author.

4.6 References

Allan, R.E., C.J. Peterson, Jr., G.L. Rubenthaler, R.F. Line, and D.E. Roberts. 1989. Registration of 'Madsen' wheat. *Crop Science* **29:**1575.

Anderson, J.A., R.W. Stack, S. Liu, B.L. Waldron, A.D. Fjeld, C. Coyne, B. Moreno-Sevilla, F.J. M., Q.J. Song, P.B. Cregan, and R.C. Frohberg. 2001. DNA markers for Fusarium head blight resistance QTLs its two wheat populations. *Theoretical and Applied Genetics* **102:** 1164-1168.

Baenziger, P.S., B. Beecher, R.A. Graybosch, D.D. Baltensperger, L.A. Nelson, J.M. Krall, D.V. McVey, J.E. Watkins, J.H. Hatchett, and Ming-Shun Chen. 2004. Registration of 'Goodstreak' wheat. *Crop Science* **44:** 1473-1474

Baenziger, P. S., B. Moreno-Sevilla, C. J. Peterson, J. W. Schmidt, D. R. Shelton, D. D. Baltensperger, L. A. Nelson, D. V. McVey, J. E. Watkins, and J. H. Hatchett. 1995. Registration of 'Alliance' wheat. *Crop Science* **35:** 938-939.

Baenziger, P. S., B. Moreno-Sevilla, R. A. Graybosch, J. M. Krall, M. J. Shipman, R. W. Elmore, R. N. Klein, D. D. Baltensperger, L. A. Nelson, D. V. McVey, J. E. Watkins, and J. H. Hatchett. 2002. Registration of 'Wahoo' Wheat. *Crop Science* **48:** 1752-1753.

Bariana, H.S. and R.A. McIntosh. 1993. Cytogenetic studies in wheat XIV. Location of rust resistance genes in VPM1 and their genetic linkage with other disease resistance genes in chromosome 2A. *Genome* **36:** 476-482.

Friebe, B., Y. Mukai, H.S. Dhaliwal, T.J. Martin, and B.S. Gill. 1991. Identification of alien chromatin specifying resistance to wheat streak mosaic and greenbug in wheat germ plasm by C-banding and in situ hybridization. *Theoretical and Applied Genetics* **81:** 381-389.

Gill, B.S., B. Friebe, D.L. Wilson, T.J. Martin, T.S. Cox. 1995. Registration of KS93WGRC27 wheat streak mosaic virus resistant T4DL-4Ai#2S wheat germplasm. *Crop Science* **35:**1236-1237.

Gill, B.S., W.J. Raupp, L.E. Browder, and T.S. Cox. 1988. Registration of KS86WGRC02 Leaf Rust Resistant Hard Red Winter Wheat Germplasm. *Crop Science* **28:** 207.

Graybosch, R.A., C.J. Peterson, P.S. Baenziger, L.A. Nelson, B.B. Beecher, D.B. Baltensperger, and J.M. Krall. 2005. Registration of 'Arrowsmith' Hard White Winter Wheat. *Crop Science* **45:** 1662-1663

Hatchett, J.H. and B.S. Gill. 1981. D-genome sources of resistance in *Triticum tauschii* to Hessian fly. *Journal of Heredity* **72:** 126-127.

Helguera, M., I.A. Khan, J. Kolmer, D. Lijavetzky, L. Zhong-qi, and J. Dubcovsky. 2003. PCR assays for the Lr37-Yr17-Sr38 cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. *Crop Science* **43:** 1839-1847.

Liu, X.M., B. S. Gill, and M. S. Chen. 2005. Hessian fly resistance gene *H13* is mapped to a distal cluster of resistance genes in chromosome 6DS of wheat**.** *Theoretical and Applied Genetics* **111:** 243-249.

Liu, X.M., A.K. Fritz, J.C. Reese, G.E. Wilde, B.S. Gill, and M.S. Chen. 2005. H9, H10, and H11 compose a cluster of Hessian fly-resistance genes in the distal gene-rich region of wheat chromosome 1AS. *Theoretical and Applied Genetics* **110:** 1473-1480.

Liu, X. M., C. M. Smith, and B. S. Gill. 2002. Identification of microsatellite markers linked to Russian wheat aphid resistance genes *Dn4* and *Dn6*. *Theoretical and Applied Genetics* **104:** 1042-1048.

Paterson, A.H., S.D. Tanksley, and M.E. Sorrells. 1991. DNA markers in plant improvement. *Advances in Agronomy* **46:** 39-90.

Patterson, F.L., F.B. Mass III, J.E. Foster, R.H. Ratcliffe, S. Cambron, G. Safanski, P.L. Taylor, and H.W. Ohm. 1994. Registration of eight Hessian fly resistant common winter wheat germplasm lines (Carol, Erin, Flynn, Iris, Joy, Karen, Lola, and Molly). *Crop Science* **34:** 314-316.

Patterson, F.L., F.B. Maas III, J.E. Foster, R.H. Ratcliffe, S. Cambron, G. Safranski, P.L. Taylor, H.W. Ohm. 1994. Registration of eight Hessian fly resistant common winter wheat germplasm lines (Carol, Erin, Flynn, Iris, Joy, Karen, Lola, and Molly. *Crop Science* **34:**315-316.

Pestsova, E., M.W. Ganal, and M.S. Röder. 2000. Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. *Genome* **43:** 689-697.

Peterson, C.J., D. R. Shelton, P.S. Baenziger, D. D. Baltensperger, R. A. Graybosch, W. D. Worrall, L.A. Nelson, D. V. McVey, J. E. Watkins, and J. Krall. 2001. Registration of 'Wesley' Wheat. *Crop Science* **41:** 260-261.

Raupp, W.J., Sukhwinder-Singh, G.L. Brown-Guedira, and B. S. Gill. 2001. Cytogenetic and molecular mapping of the leaf rust resistance gene *Lr39* in wheat**.** *Theoretical and Applied Genetics* **102:**347-352.

Singh, S., J.S. Sidhu, N. Huang, Y. Vikal, Z. Li, D.S. Brar, H.S. Dhaliwal, and G.S. Khush. 2001. Pyramiding three bacterial blight resistance genes (Xa5, xa13 and Xa21) using marker-assisted selection into indica rice cultivar PR106. *Theoretical and Applied Genetics* **102:** 1011-1015.

Talbert, L.E., P.L. Bruckner, L.Y. Smith, R. Sears, and T.J. Martin. 1996. Development of PCR markers linked to resistance to wheat streak mosaic virus in whea. *Theoretical and Applied Genetics* **93:** 463-467.

Waldron, B.L., B. Moreno-Sevilla, J. A. Anderson, R. W. Stack, and R. C. Frohberg. 1999. **R**FLP mapping of QTL for fusarium head blight resistance in wheat**.** *Crop Science* **39:** 805-811.

Chapter 5

Rapid introgression of single genes into popular cultivars using marker-assisted background selection

Harpinder S. Randhawa*, Jasdeep S. Mutti*, Kim Kidwell and Kulvinder S. Gill[1]

Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420

*** Authors with equal contribution thus should be considered joint first authors**

[1]*Corresponding author*:

Department of Crop & Soil Sciences,

277 Johnson Hall, PO Box 646420,

Washington State University, Pullman, WA 99164.

E-mail: ksgill@wsu.edu

5.1 Abstract

Backcross (BC) breeding is a powerful tool to introgress genes controlling simply inherited traits in order to extend life and to increase profitability of popular cultivars. This method will become even more important as additional value-added traits become available that will need to be quickly transferred into popular cultivars. However, a major hurdle for the successful use of this approach in crop breeding has been the time required. Four to six backcrosses are usually required to recover a high proportion of the recurrent parent genome (RPG), making it difficult for the improved breeding lines to compete with the newly released cultivars. We, therefore, developed a marker-assisted background selection based method wherein we recovered more than 96% of the fragments associated with the RPG in just two backcrosses. After testing various approaches by simulation, the available information on the distribution of genes and recombination on wheat chromosomes was incorporated, to develop a marker-assisted backcross method. The aim of the first BC was to select plants with proximal recombinants between the target gene and the closest marker and the target for BC_2 was to recover the recombinant in the distal end. For background selection, at least two markers flanking each of the 48 wheat gene-rich regions (GRRs) were selected except for the GRRs with more than 50 cM genetic length, for which an additional marker present in the middle of the GRR was also selected. This method was tested to introgress the *Yr15* gene resistance to yellow rust (*Puccinia striiformis* Westend f. sp. *tritici*) into the spring wheat cultivar 'Zak'. After screening 1,000 plants during each of the two BC cycles with 251 selected polymorphic SSR markers, four plants carrying the maximum number of homozygous recurrent parent alleles were selected. Thirty-seven homozygous resistant $BC_2F_{2:3}$ families were field evaluated and two plants of one family carrying about 99% of RPG were selected.

5.2 Introduction

Backcross breeding has been a valuable strategy to introgress genes controlling simply inherited traits in order to extend life and to increase profitability of popular cultivars (Allard 1999). This method will become even more important as additional value-added traits become available that will have to be quickly transferred to popular cultivars. A good example is the *GPC-B1* gene for high grain protein content (Uauy et al. 2006). The backcross breeding method has long been a valuable strategy in plant breeding for introgression of favorable alleles from the donor genotype into a recipient elite genotype possessing many desirable but one or few undesirable trait(s) (Allard 1999). The objective of backcrossing is to recover most of the recurrent parent genome (RPG) along with a trait of interest from the donor parent. The backcross cycles are repeated until a near-isogenic to the recurrent parent background is obtained. Generally, more than six backcrosses are required to recover 99% of the RPG along with the introgressed trait (Allard 1999). A major hurdle in a successful application of this approach has been that it takes a long time to complete the four to six backcrosses required to recover a high proportion of the recurrent parent genome (RPG). In the 6+ years (in case of winter wheat) that the transfer would take, the new introgression line may not be able to compete with the newly released varieties with enhanced insect pest resistance and other agronomically important traits.

Ever-evolving insect populations and changes in disease population structures are the most limiting factors shortening the life of a cultivar. Newly released cultivars are rendered susceptible with emergence of new races/biotypes of one or more pests caused by a sudden population shift and/or mutations. Along with these, other traits related to agronomic performance (e.g. herbicide tolerance), consumer preference

(white wheats, yellow pasta durum), and superior end-use quality (protein content, value-added traits) could also affect the performance and success of a cultivar.

Molecular marker-assisted backcrossing can prove to be a powerful tool in accelerating the recovery of the RPG (Tanksley et al. 1989). In marker-assisted backcross breeding, molecular markers serve as a tremendous resource, first by targeting the gene of interest (foreground selection), and then by accelerating the recovery of recurrent parent genome and reducing the linkage drag (background selection) (Young and Tanksley 1989). However, developing an efficient strategy for a crop system is not a straightforward task considering that there are significant differences among various crop systems, for the genome structure, available genomic information, type and number of available markers and other unique features of each crop system.

Recently, a plethora of computer simulations has been made available to design an efficient and successful marker-assisted backcross breeding program. The efficiency of marker-assisted selection (MAS) is evaluated by considering the effects of (i) population size in different backcross generations, (ii) the number of marker data points required with respect to the proportion of RPG attained (Hospital and Charcosset 1997; Frisch et al. 1999a; Frisch and Melchinger 2001; Frisch and Melchinger 2005), and (iii) by monitoring the distance of markers flanking the target loci and on non-carrier chromosomes (Hospital 2001). Among these, the required population size during each generation has been reported as the most important factor affecting the efficiency of marker-assisted backcross program (Zhang and Smith 1992; Gimelfarb and Lande 1994; Whittaker et al. 1995; Hospital and Charcosset 1997; Frisch et al. 1999b). Theoretically, any proportion of RPG recovery is possible in the BC_1 generation if a population of infinite size is considered. Frisch et al. (1999)

estimated the minimum population size required for each generation, calculated for a known distance of the flanking markers from the target gene. These simulations were run to find with probability (0.90, 0.95 and 0.99) of single or double recombinants between the target gene and the flanking markers. The minimum marker distance required to obtain at least one backcrossed individual with a single recombinant in a population of 50 individuals is 22 cM. For a population of 500 individuals, however, the minimum marker distance drops to 2 cM. Similar numbers for a double recombinant are 92 cM for 50 individuals and 16 cM for 500.

For a fixed distance spanned by the flanking markers, the minimum population size can easily be simulated. The minimum population size required to obtain one backcross individual with a single recombinant ranged from 192 for the flanking markers spanning 5 cM to 27 individuals for 50 cM. The same estimates for a double recombinant were 4066 individuals for 5 cM and 90 for 50 cM (Frisch et al. 1999a). The minimum number of individuals required to obtain a double recombinant in the BC_1 generation for the markers spanning 2 cM were simulated to be 24,000 (Decoux and Hospital 2002). If only 150 individuals were used in a BC_1 generation, a 0.99 % probability can only be insured if the flanking markers spanned at least 34 cM (Frisch 1999a).

Recent advances in polymerase chain reaction based molecular markers, for example, simple sequence repeats (SSRs) (Röder et al. 1998) and single nucleotide polymorphisms (SNPs) (Gilles et al. 1999), have imparted substantial improvement in the efficient screening of large populations. Screening of thousands of genotypes no longer poses an intractable technical problem, and can be considered in optimizing MAS strategies (Ribaut and Ragot 2007).

The available simulations estimate a very large number of marker data points that would be needed to recover a high proportion of RPG in three or less backcrossing cycles. The required number of marker data points were estimated to be 20,000 to obtain >96% RPG content in three backcross generation (Frisch et al. 1999a). It is intuitive that, for close flanking markers, double recombinant genotypes are highly unlikely in a single backcross generation (BC_1) such that more than two backcross generations are required for a successful backcrossing program (Young and Tanksley 1989). The experimental plans generated using these computer simulations estimated population size and marker data point numbers that are not feasible in a breeding program. The main reason for the unrealistic numbers generated by these simulations is probably the assumptions made while designing the mathematical models. Some of the common assumptions such as uniform distribution of genes and uniform rates of recombination on chromosomes may be grossly incorrect (Sidhu and Gill 2004).

Development of high-density genetic and physical maps in many crops in the past two decades, has significantly improved our knowledge on the distribution of crossover or recombination events on the chromosomes. The distribution of recombination is highly uneven on chromosomes of higher eukaryotes, negating the very assumption made in all these simulation models (Dvorak and Chen 1984; Bollag et al. 1989; Curtis and Doyle 1991; Gill et al. 1993; Lichten and Goldman 1995; Gill et al. 1996a; Gill et al. 1996b; Tranquilli et al. 1999; Sandhu and Gill 2002; Erayman et al. 2004). For example the average recombination rate in yeast is ~262 cM/chromosome, in rice ~127 cM/chromosome, humans ~154 cM/chromosome, whereas in the case of wheat it is ~206 cM/chromosome (Sidhu and Gill 2004). Erayman et al. (2004) provided a detailed estimation of the recombination at the

chromosomal level in wheat by comparing the consensus physical maps of wheat with the genetic linkage maps of 428 common markers. In general, recombination was severely suppressed in the centromeric regions, accounting for <1% of the total recombination events. Recombination was restricted to the 48 gene-rich regions (GRRs) that account for more than 85% of the genes but span less than 29% of the genome. However, different GRRs showed significantly different recombination rates.

The information on the distribution of recombination rates and genes has not been incorporated into any simulations or in the design of any breeding program. It is particularly important for GRRs with low recombination rates. The linkage drag for these regions would be very high if the recombination rate information was not considered during simulations or backcross program design.

The objective of this study was to develop a highly efficient, marker-assisted background selection method based on a backcross scheme for wheat by incorporating the available information on the genome structure and the distribution of genes and recombination rate. First, we optimized a backcrossing model using computer simulations and then we tested this model to transfer the *Yr15* gene into the soft white spring wheat cultivar 'Zak' that has an excellent grain yield and superior end-use quality. 'Zak' was released in year 2002 and was grown on about 94,000 acres. A severe epidemic of stripe rust caused by change in virulent *P. striiformis* Westend f. sp. *tritici* race of wheat occurred in the Pacific Northwest (PNW) during 2002 causing severe yield losses. By the year 2005, the area under 'Zak' cultivation was reduced to 8300 acres due to stripe rust sensitivity (http://www.usda.gov/nass/pubs/agstats.htm). Stripe rust (yellow rust), caused by the fungus *P. striiformis*, is the most common disease of wheat (*Triticum aestivum* L.), in

the cooler and humid regions of production (Jones and Clifford 1983). The stripe rust resistant gene *Yr15* was introgressed from *T. dicoccoides* Koern., to cultivated hexaploid wheat (Grama and Gerechter-Amitai 1974) and was found to be highly resistant to broad spectrum races of the *P. striiformis* races prevalent in PNW (Chen 2005).

5.3 Results

5.3.1 Comparison of Selection Strategies by Computer Simulations

Simulation results are summarized in Figure 5.1. The control simulation runs involving selection for only the target gene showed an RPG recovery at a Q10 value of 76.2% (means that with probability 0.90 an RPG proportion greater than 76.2% is attained), that is very close to the 75% ratio expected if no background selection is applied. The same value after six backcross generations reached 97.3%, which is also very similar to the expected value of 96%. In case of the two-stage selection a Q10 value of 96% was reached in the BC_3 generation provided the background selection was carried out using 11228 marker data points (MDPs). Three-stage selection required 5225 MDPs, whereas four-stage selection required only 1534 MDPs to reach a set goal of 96% RPG in BC_3 (Figure 5.1a). These simulations clearly showed that the four-stage selection required the least number of MDPs to achieve the goal of 96% RPG with a minimum number of MDPs. Simulation to check the effect of unequal population sizes on the number of MDP required to achieve 96% RPG in three backcrosses using four stage selection, are given in Figure 5.1c. A population size of 100, 150 and 200 plants per subsequent backcrosses required 7500 MDPs compared to 1534 if a constant population of 100 plants per backcross was used.

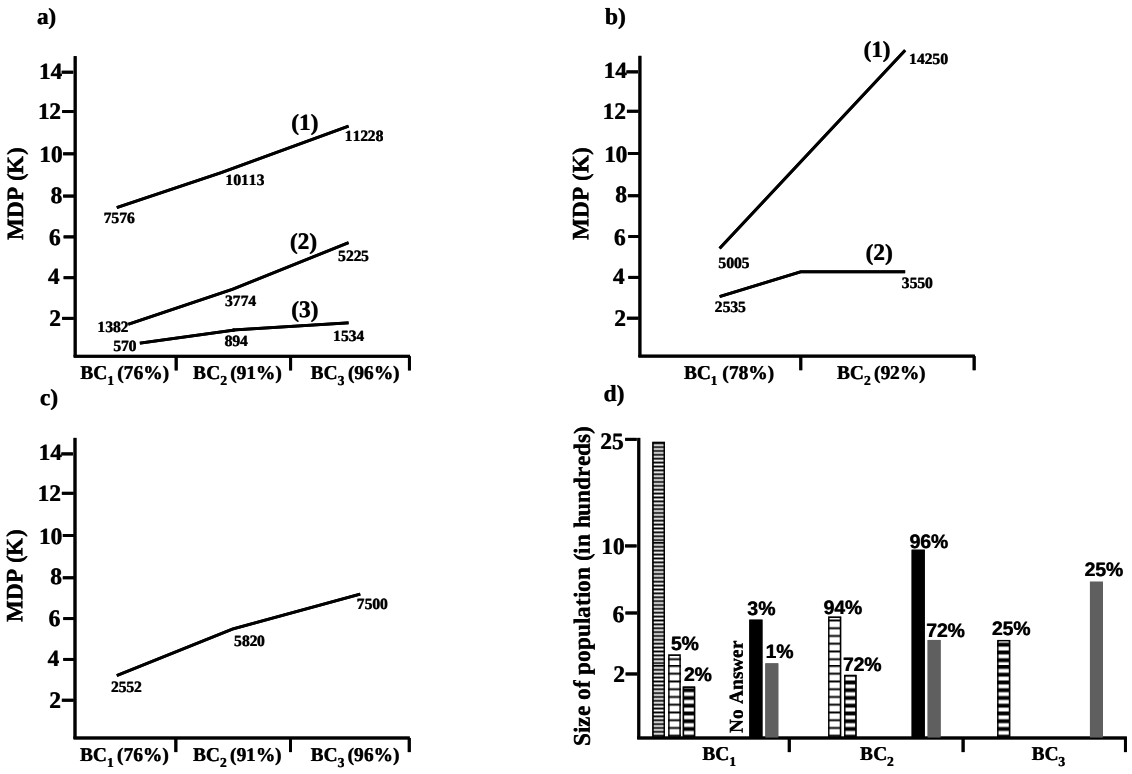


Figure 5.1. Computer simulations to estimate the number of marker data point (MDP) and size of population required for marker assisted background selection program in wheat. In (a-c) the X-axis represents the number of backcross generations along with the percent RPG recovered and the y-axis represents the number of MDP required in each generation. **(a)** Plabsim results for number of MDP required using two-stage (1), three-stage (2) and four-stage (3) selection strategy in three backcross generations. **(b)** Number of MDP required in three-stage (1) and four-stage (2) selection strategies for two backcross generations. **(c)** Four-stage selection for three backcross generations with population size of 100, 150 and 200, respectively during the three generations. **(d)** The x-axis represents the number backcross generations and y-axis represents the size of population required. Probability of finding a double homozygote for markers flanking the gene of interest is given in percentage on top of each bar. Dashed bars represent probability and size of population required for markers at 2 cM intervals, whereas solid bars give probability for markers at 1 cM interval.

After establishing the four-stage selections to be the most efficient, we ran simulations to obtain the number of plants and MDPs that would be needed to recover the same proportion of RPG in two backcrosses. Simulations run with a population size of 200 plants per backcross gave no results (data not shown). Simulation results with 1000 plants per backcross are summarized in Figure 5.1b. With the three-stage selection 14250 MDPs were needed to recover a Q10 value of 92% for RPG recovery whereas only 3550 MDPs were required for the four stage selection (Figure 5.1b).

In order to answer the question if a double recombinant should be recovered in BC_1 or in two generations, the probability and number of plants required to obtain a double recombinant for markers flanking the target gene, was calculated using Popmin software (Figure 5.1d). A BC_1 population of 24,000 is required to obtain a double recombinant for flanking markers spaced at a 2 cM interval. The simulations failed to give an answer if the markers were assumed to be 1 cM apart. With the target to obtain a double crossover in two backcrosses for flanking markers at a 2 cM interval, the probability of finding a double recombinant was 5% in BC_1 and 94% in BC_2 with population sizes of 290 and 500, respectively. The probability of finding a double recombinant for markers at a 1 cM interval is 3% in BC_1 and 96% in BC_2 with population sizes of 579 and 996 (Figure 5.1d). Similar numbers for three backcross generations for markers at 2 cM intervals are 2% in BC_1, 72% in BC_2 and 25% in BC_3 with population sizes of 117, 171 and 370, respectively. At 1 cM interval the similar probability numbers are 1%, 72% and 25% with population sizes of 233, 344 and 721.

5.3.2 Marker Assisted Background Selection Scheme

Based on the computer simulation results and by taking into account the structural and functional organization of the wheat genome, we developed a

background selection scheme based on the four-stage selection strategy (Figure 5.2). Our strategy to select markers for the background selection is described in the 'methods' section and examples are given in Figure 5.4. There are 48 gene rich regions in wheat that account for only about 29% of the genome but contain more than 85% of the genes (Erayman et al. 2004). Two markers flanking each of the GRR were selected except for the GRRs like 1S0.8 (46.1 cM) and 1L0.9 (50.0 cM) that has a very high rate of recombination. In those cases, another marker mapping in the middle of the GRR was also selected.

In order to increase the probability to obtain a double recombinant in two BC generations and a high proportion of RPG, we proposed to use 1000-3000 plants for each of the BC generations. Since the simulations showed a low probability to find a double recombinant in a single generation, we propose to identify recombinants between the target gene and one of the flanking markers in BC_1 and for the second flanking marker in BC_2 (Figure 5.1d). The selected BC_1 plants should then be screened using the available markers for the carrier chromosome (chromosome carrying the target gene) to select plants with all the recurrent parent alleles. If the number of selected plants was more than four, markers for the other chromosomes could be used to select two to four plants with the highest RPG that could be used to develop the BC_2 population. After selecting recombinants for the second flanking marker, the selected BC_2 plants should be screened with the selected markers for the remaining chromosomes, except for the markers that were homozygous in BC_1. Selection for agronomically important traits such as yield, plant height, kernel size, kernel weight and protein content was incorporated, wherever possible. Two to four BC_2 plants were selected and then made homozygous either using double haploid

methods or by selfing. The progeny of the selected plants was subjected to field evaluation in order to select the most desirable genotype.

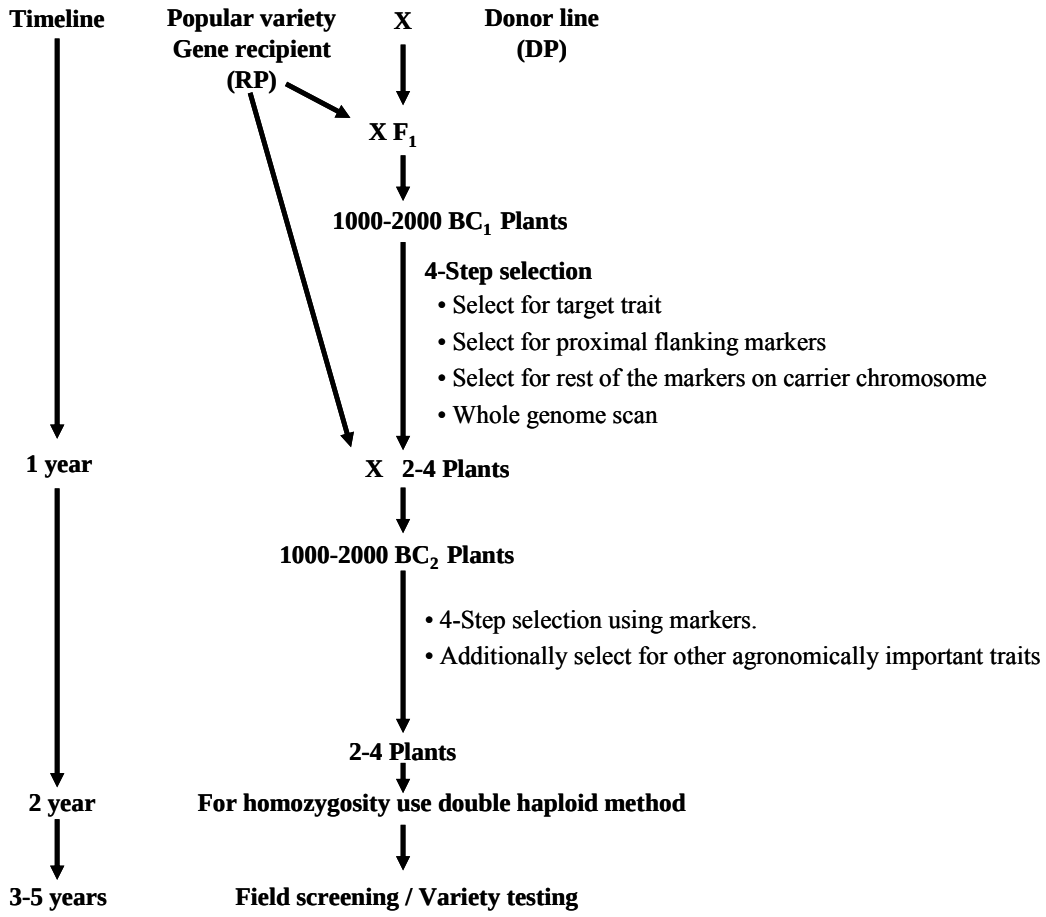


Figure 5.2. Strategy for marker assisted background selection for wheat.

5.3.3 Model Testing: Introgression of Yr15 gene into cultivar 'Zak'

The actual step-wise methodology that we used to introgress the *Yr15* gene into 'Zak', along with the number of plants selected during each step, is given in Figure 5.3. First the polymorphism survey of the two parents ('Zak' and the donor parent Avocet S*6/*Yr15*) was performed with 639 SSR markers covering the entire wheat genome. About 50% (314/639) of these markers detected polymorphism that included 35 markers for chromosome 1B that carries *Yr15* gene (Table 5.1). Using the marker selection strategy outlined earlier, 251 of these markers were selected for the background selection as well as for constructing the genetic linkage map used for the simulation studies. Out of 314 polymorphic markers effort was made to select for the markers flanking the gene rich regions (GRRs) as described in the materials and methods section.

Table 5.1. Polymorphism survey results for the SSR markers on 'Zak' and the *Yr15* donor line.

Genome	**A, B, D**	**Total**
Group	Poly/Analyzed	Poly/Analyzed
1	14/35, 35/55, 18/42	67/132
2	24/52, 8/24, 15/33	47/109
3	29/49, 30/47, 20/49	79/145
4	10/28, 6/16, 12/22	28/66
5	24/47, 10/23, 13/28	47/98
6	6/16, 9/18, 6/12	21/46
7	7/10, 6/12, 12/21	25/43
Total	**114/237, 104/195, 96/207**	**314/639**

5.3.4 Yr15 segregation

Segregation of *Yr15* in all the BC generations was distorted. Instead of the expected 1:1 ratio, only 156 plants showed complete resistance out of 1131 BC_1F_1 plant. The same number for BC_2F_1, was 204 plants out of 1056 plant that were screened with race *PST-78*. Reciprocal BC_1F_1 with 'Zak' as a pollen parent showed only 24 resistant plants out of 90. However, an $F_{2:3}$ population derived from the same cross with 'Zak' as a female parent showed the expected single gene ratio of 1:2:1 (data not shown).

Selection strategies used to introgress *Yr15* into cultivar 'Zak' by marker assisted background selection are shown in Figure 5.3. The 156 completely resistant BC_1F_1 plants were first genotyped with the *Yr15* gene flanking markers *gwm11* and *gwm33* in order to identify plants with recombination between *Yr15* and one or both of these markers. Twelve plants that were homozygous for recurrent parent type allele for the distal flanking marker *gwm33* were identified. No double crossover plant was recovered as all 12 plants were heterozygous for the proximal flanking marker *gwm11*. These 12 plants were genotyped with the remaining 33 1B specific markers to select four plants with the maximum number of homozygous loci. Two of these plants (1-3, 1-38) were heterozygous only for the flanking marker *gwm11* but the other two were heterozygous for *gwm11* and *gwm413*. These four plants were used as male parents to generate 1056 BC_2F_1 plants, of which only 204 showed immune response when screened with the race *PST-78*. Genotyping of the resistant plants with the flanking marker *gwm11* identified five plants that were homozygous for the marker. Genotyping of these plants with the remaining 205 selected markers (described earlier) identified a plant that was homozygous for 188 markers and was heterozygous for 17 (Table 5.2). Its selfed progeny of 300 BC_2F_2 plants were

screened for resistance to stripe rust to identify 150 resistant plants. Screening of the $F_{2:3}$ progeny of the resistant plants identified 37 homozygous resistant families. Multiple plants for each of these resistant families were screened with the 17 markers that were heterozygous in the selected BC_2F_1 plant. Two plants with 96.8% RPG were selected and were used for field evaluation.

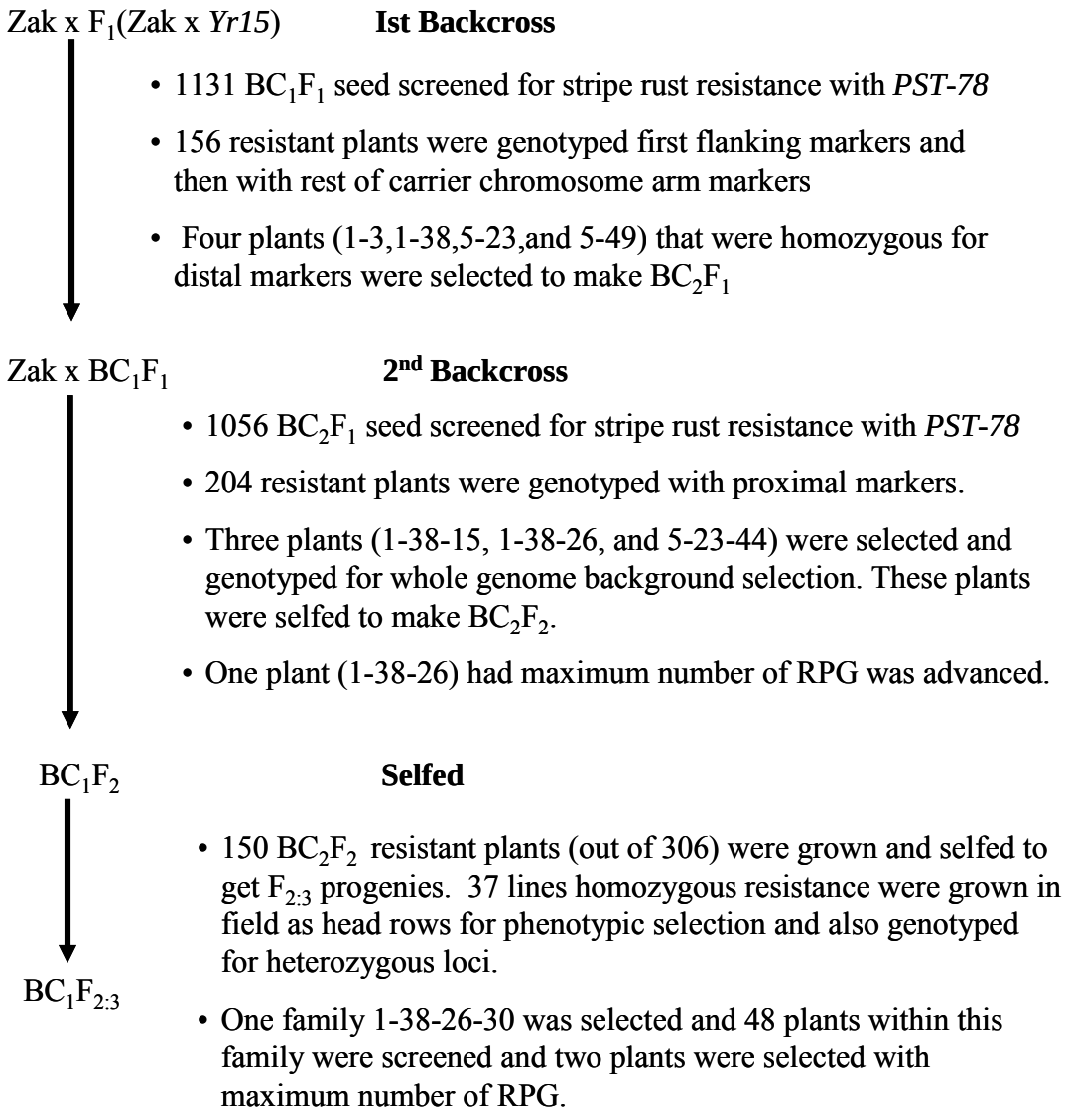


Figure 5.3. Schematic representation of introgression of stripe rust resistance gene *Yr15* into 'Zak' cultivar using marker assisted backcrossing.

Table 5.2. Proportion of recurrent and donor parents genome in different generations of backcrosses

Generation	BC_1F_1	BC_2F_1	$BC_2F_{2:3}$
Number of loci analyzed	115	205	251
Homozygous loci for RP	84	188	241
Homozygous loci for DP	-	-	2
Heterozygous loci	31	17	8
% of RPG	73	92	96
% of DPG	-	-	0.8
%Heterozygous loci	27	8	3.2

5.4 Discussion

While transferring a single gene by repeated backcrossing, the physical size of the retained donor segment surrounding the target gene depends upon the recombination rate for the region and the genetic length of the region depends upon the number of plants screened during each backcross cycle. Breaking a tight linkage would require a larger number of plants during each backcrossing cycle compared to a situation where genetic distance is higher between the marker and the target gene. Even a single recombinant between the target gene and a tightly linked marker is a very rare event and thus is less likely to be selected if the backcrossing is carried out without any background selection. As shown by simulations, the probability of recovering a double recombinant for the markers flanking the target gene is very low even after six backcrosses if only a few plants are selected during each backcrossing cycle.

Optimum population size required during each backcross cycle has been analyzed using different computer simulations. Using Popmin software package, Hospital (2001) determined that 62 BC_1 and 100 BC_2 plants are required to recover a single recombinant in at least one plant when the markers flanking the target gene span 10 cM. The number of plants increased to 118 and 200 for BC_1 and BC_2, respectively when the spanning distance was reduced to 5 cM. The number of plants increased to 4000 if a double recombinant is desired (Frisch et al. 1999). Although the required number of plants decreased significantly for a sequential strategy to obtain a double recombinant using two BC cycles, it was still a rare event warranting a need for marker assisted background selection.

Our simulations clearly showed that a four-step selection strategy involving a sequential selection of the two recombinants between the target gene and its flanking

markers is the most efficient approach as it provided a very high level of RPG recovery with relatively fewer marker data points. Based on the simulations, we propose a marker assisted background selection strategy for wheat that insures more than 96% of the RPG recovery in two BC cycles. Testing of this strategy to transfer *Yr15* gene into cultivar 'Zak' showed the efficiency of the approach. We were able to recover more than 96% of the RPG (based on the markers analyzed) in two backcrossing cycles by screening 156 resistant (total 1131) BC_1 and 204 resistant (total 1056) BC_2 plants with 251 total polymorphic markers.

The severe segregation distortion that we observed for the *Yr15* gene in the two backcross generations was unexpected. Only about 14% plants were resistant in BC_1 and 19% in BC_2, which is significantly less than the expected 50%. Selection of only those plants that showed an immune response could explain getting less than expected number of resistant plants but the segregation was the normal 1:2:1 in the F_2 of the same cross using the same disease screening rating. We do not have any clear explanation of these results as the difference is only in BC but not in the F_2 generation of the same cross. However, presence of a suppressor gene that functions only in a heterozygous condition or an enhancer that functions only in a homozygous condition can be postulated.

The background selection was particularly efficient for the non-carrier chromosomes. Ours and previously reported simulations suggested that at least two markers per non-carrier chromosome are mandatory for an RPG recovery above 97% (Servin, 2005; our data not shown). However, increasing the per chromosome marker number beyond two did not significantly reduce RPG (Hospital and Charcosset 1997; Frisch et al. 1999a). An increase from one to eight markers reduced the expected donor genome content from 0.13 to 0.07 in BC_2 (Stam 2003). Even during the

introgression of *Yr15* into cultivar 'Zak', only 12 plants during BC_1 and 5 plants during BC_2 were used for background selection using markers for the non-carrier chromosomes yet we were able to recover more than 96% of the RPG in two backcrossing cycles based on the markers screened.

5.4.1 Conclusions and implications for plant breeding

Avoiding linkage drag on the carrier chromosome is the most difficult part of the backcross breeding. Because of a very low frequency of occurence, double crossovers for the flanking markers and the gene of interest are less likely to be selected even after many backcross cycles involving random selection of plants. As suggested by simulations and shown by our results, it is possible to obtain double recombinants and to recover a high percentage of RPG just in two BC cycles if a carefully chosen marker assisted background selection scheme is followed. This approach can be a boon to the introgressions of desirable traits from unadapted germplasm into popular varieties without a yield penalty usually associated with linkage drag. Although the desirable plants can be selected in two BC cycles, additional selfing generations required to achieve a desired level of homozygosity is still a 'bottleneck' to further reduce the time taken by the marker assisted background selection approach. Although not utilized in this study, we propose to alleviate that constraint by the use of doubled haploid methods to instantly fix homozygosity.

5.5 Materials and Methods

5.5.1 Plant materials

The soft white spring wheat cultivar 'Zak' (Kidwell et al. 2002) that is susceptible to stripe rust but otherwise has an excellent grain yield and superior end-

use quality was used as a recurrent parent. The near isogenic lines of spring wheat cultivar 'Avocet S*6/*Yr15*' (PI 640428) was used as a *Yr15* donor parent. The cultivar 'Zak' was used as female parent for all crosses. All the crossing and stripe rust screening were carried out in the Washington State University (WSU) wheat plant growth facilities.

5.5.2 Disease screening

Plants from various generations were screened for stripe rust resistance in the growth chamber using *P. striformis* race *PST-78* that is a highly virulent and prevalent race in the PNW (Chen 2005). Single seeds were planted in 96 cell plastic tray (20" X 10") (McConkey Co. USA) and were inoculated at the two leaf stage (about 10 days). Spores were mixed with talcum powder and applied on the plant leaf, as described by Chen and Line (1992). A fine mist of water was sprayed on the plants and a high level of relative humidity was ensured by placing plant trays in larger water containing tray at 10ºC and covering with plexiglass covers so as to accomplish water seal. After 16 hrs of incubation, plants were moved to temperature regime of 8ºC and 16ºC cycled every 6 hours starting at midnight with 8ºC. Disease scoring was performed after 21 days using the scale and method described by Chen and Line (1992). Only the plants showing immune response were selected.

5.5.3 DNA extraction

A quick method using 96-2ml deep well plates (Axygen Scientific, Inc. USA) was first optimized for DNA extractions for this project. Briefly, 3-4 fresh leaves of about one inch length were placed into plate wells and lyophilized for 3-4 days. A metallic bead (3mm, V & P Scientific, USA) was then added to each of the wells to

grind the tissue using 'Qiagen Mixer Mill' (Model MM 301) (Qiagen, USA) for 5-10 min. Using multi channel pipette, 750 µl of CTAB buffer solution (100mM Tris pH 7.5, 1.4M NaCl, 20mM EDTA pH 8.0, 2% CTAB, 0.2% Mercaptoethanol) was added and the plates were incubated at 60°C for 60-90 mins. Equal volume of Chloroform: Octanol (24:1) was added and mixed thoroughly by pipeting. Plates were centrifuged at room temperature at 4000 rpm for 15 minutes and DNA was precipitated by adding 2/3 volume of isopropanol to the supernatant in a new plate. The DNA pellet was washed with 70% ethanol; air dried, was suspended in 400 µl of TE and stored at 4°C.

5.5.4 Marker selection

Based on the computer simulations and incorporating the gene rich regions and recombination hot/cold spots in wheat, we outlined our strategy as shown in Figure 5.2, with particular emphasis on selection of markers using the strategy shown in Figure 5.4. For the selection of markers particular emphasis was given on gene rich regions and recombination hot/cold spots, demarcated by Erayman et al. (2004). As an example, shown in Figure 5.4 for group *1* chromosome, two markers flanking each gene rich region was selected for MAS. In some cases, where exceptionally high recombination rates were reported, three markers per region were selected. For instance, for GRR 1S0.8 (46.1 cM) and 1L0.9 (50.0 cM) three markers each were selected, with two markers flanking each GRR and one to its interior. Accordingly, markers were selected to cover the whole wheat genome.

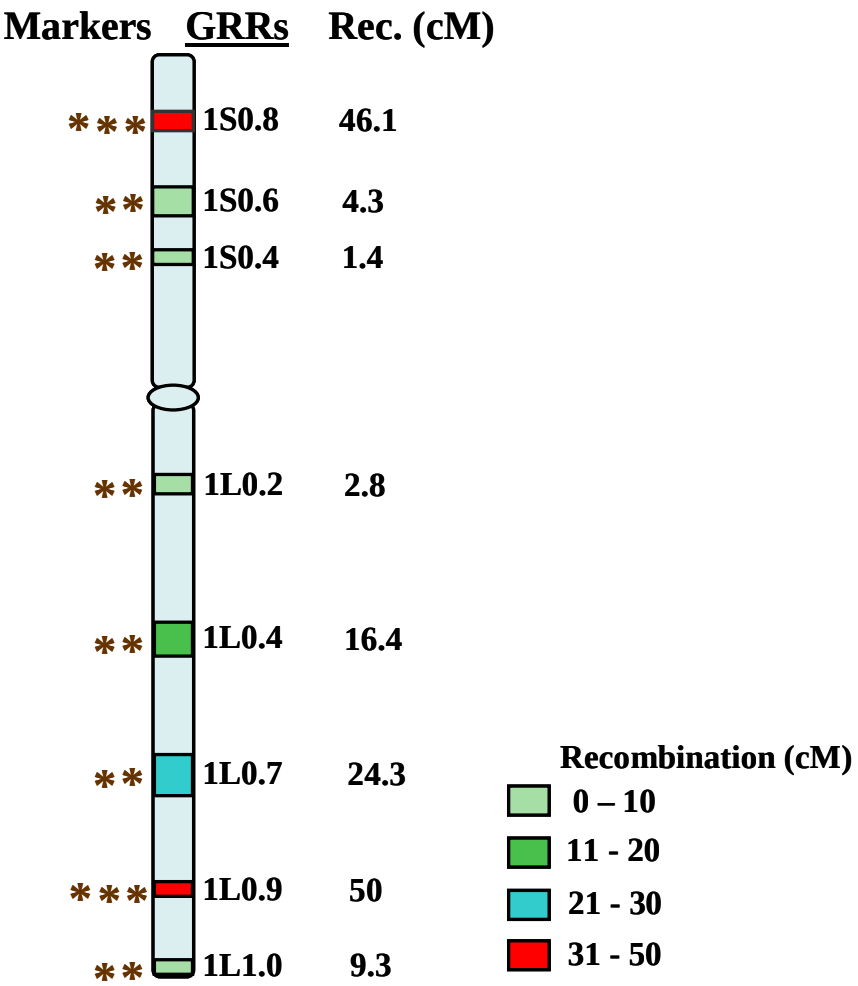


Figure 5.4. Marker selection strategy for the marker assisted background selection scheme. The consensus physical of group *1* is given as an example and was adapted from that given in Erayman et al. (2004). The GRRs are drawn to scale and their names are given on the right side of the consensus map. In the nomenclature of GRRs (e.g. '1S0.8'), the first digit represents wheat homoeologous group followed by the arm location either as short arm (S) or long arm (L). The last two numeral numbers represent GRR location as fraction length (FL) of the chromosome (e.g. 0.8 for '1S0.8'). Recombination for each of the GRRs was as calculated by Eryman et al. (2004). On the left side of the concensus chromosome, * denotes the number of markers selected for a particular GRR region.

5.5.5 Marker analysis

Wheat SSR markers acquired from different sources GWM (Röder et al., 1998a, 1998b), CFA and CFD (Sourdille et al. 2001; Guyomarch et al. 2002), WMC (Somers et al. 2004) and BARCs (Song et al. 2005) were used for the project. Primer sequences and all the other information about the SSR markers are available at the Graingenes website (http://wheat.pw.usda.gov/GG2/quickquery.shtml). The forward primers were synthesized with the M13 tail (CACGACGTTGTAAAACGAC) at the 5′ end in order to detect the PCR products on the IR2 4200 DNA Analyzer (Li-Cor Biosciences, USA). PCR reaction volume for SSR marker analysis was 10 µL that contained ~25 ng of genomic DNA, 1x PCR buffer, 2.5mM dNTP, 50mM $MgCl_2$, 1 pmol of IRDye M-13 forward tail primer (Li-cor), 1 pmol each of Forward and Reverse primers and 0.1 U *Taq* DNA polymerase. PCR amplification was performed in 96-well plates on Px2 Thermal Cycler (Thermo Electron, USA) with an initial cycle of denaturation at 94°C for 3 min, annealing at 60°C for 1 minute and extension at 72°C for 1 minute followed by 30 cycles at 94°C for 30 sec; 60°C for 30 sec; and 72°C for 30 sec. The PCR products were held at 4°C after a final extension step at 72°C for 5 minutes. The PCR products were diluted with nine volumes of loading dye (Formamide, 10ml + Bromophenol Blue, 5mg + 0.5M EDTA, 200µl) and denatured just before loading at 94º C for 2 min and chilled on ice. Electrophoresis was carried out using an IR2 4200 DNA Analyzer (Li-Cor) on 6% polyacrylamide gel.

5.5.6 Simulation to optimize a selection strategy

To get an estimate on the size of population required along with number of marker data points for marker assisted background selection in wheat, we performed

computer simulations using Plabsim (Druka et al. 2000) and Popmin (Decoux and Hospital 2002) softwares. The Plabsim software was used to get the estimate on size of population and number of MDP required for completion of backcrossing in two or three backcross generations. Whereas, Popmin software was run to estimate the probability and number of population required to select for the plant with single or double recombinant event for markers flanking the gene of interest in first or second backcross generation.

Simulations were performed for various selection strategies that differed for the selection stages and the order of selection applied to the carrier (chromosome carrying the gene under selection) vs non-carrier (remaining) chromosomes. The two-stage approach involved selection for the target gene followed by selection of a plant with the maximum number of markers homozygous for the recurrent parent alleles across the whole genome. The three-stage approach involved selection for the target gene, followed by selection for markers flanking the gene of interest. The third stage involved selection of a plant with a maximum number of markers homozygous for the recurrent parent alleles across the rest of the genome. The four-stage approach was similar to the three stage selection approach except for an extra step of selection for markers present on the carrier chromosome, after the second step (Frisch 1999).

Each simulation started with a cross between two homozygous parents that were assumed to be polymorphic at the target locus as well as for the loci used for the background selection. All F_1's of a cross were assumed to be the same. The best BC_1 individual was selected for crossing to produce BC_2. These simulations were performed for both two as well as three backcross generations. The percentage of the RPG was determined by dividing the number of loci homozygous for the recurrent parent by the total number of loci analyzed. In the BC_1, the entire set of markers was

analyzed, whereas in the following generations, only the markers not fixed for the recurrent parent in the previous cycle, were considered. The simulation for each run was repeated 5,000 times.

The simulations were performed using the wheat genetic linkage map of SSR markers (Somers et al. 2004). The map has 1,235 markers with a total length of 2,569 cM. A survey of polymorphism between 'Zak' and the *Yr15* donor line with 639 SSR markers identified 251 polymorphic markers, with an average distance of 10 cM between the markers. We constructed a genetic linkage map of the 251 markers using the cM data from the wheat SSR genetic linkage map (Somers et al. 2004). Previous simulations have shown that average marker density of 20 cM is sufficient for a good marker assisted selection (MAS) program (Frisch 1999). The target gene *Yr15* is located ~30 cM from the telomeric end of chromosome 1BS.

5.6 References

Allard, R.W. 1999. Principles of plant breeding. *2nd Edition, John Wiley and Sons, Inc. New York, NY*.

Bollag, R.J., A.S. Waldman, and R.M. Liskay. 1989. Homologous recombination in mammalian cells. *Annual Reviews of Genetics* **23:** 199-225.

Chen, X. M., and R. F. Line. 1992. Identification of stripe rust resistance genes in wheat used to differentiate North American races of *Puccinia striiformis*. *Phytopathology* **82:**1428-1434.

Chen, X.M. 2005. Epidemiology and control of stripe rust on wheat. *Canadian Journal of Plant Pathology* **27:**314-337.

Curtis, C.A. and G.G. Doyle. 1991. Double meiotic mutants of maize: implications for the genetics regulation of meiosis. *Journal of Heredity* **82:** 156-163.

Decoux, G. and F. Hospital. 2002. Popmin: a program for the numerical optimization of population sizes in marker-assisted backcross programs. *Journal of Heredity* **93:** 383-384.

Druka, A., D. Kudrna, F. Han, A. Kilian, B. Steffenson, D. Frisch, J. Tomkins, R. Wing, and A. Kleinhofs. 2000. Physical mapping of the barley stem rust resistance gene rpg4. *Molecular General Genetics* **264:** 283-290.

Dvorak, J. and K.-C. Chen. 1984. Distribution of nonstructural variation between wheat cultivars along chromosome arm 6B: Evidence from the linkage map and physical map of the arm. *Genetics* **106:** 325-333.

Erayman, M., D. Sandhu, D. Sidhu, M. Dilbirligi, P.S. Baenziger, and K.S. Gill. 2004. Demarcating the gene-rich regions of the wheat genome. *Nucleic Acids Research* **32:** 3546-3565.

Frisch, M., M. Bohn, and A. Melchinger. 1999a. Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. *Crop Science* **39:** 967-975.

Frisch, M., M. Bohn, and A.E. Melchinger. 1999b. Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Science* **39:** 1295-1301.

Frisch, M. and A.E. Melchinger. 2001. The length of the intact donor chromosome segment around a target gene in marker-assisted backcrossing. *Genetics* **157:** 1343-1356.

Frisch, M. and A.E. Melchinger. 2005. Selection theory for marker-assisted backcrossing. *Genetics* **170:** 909-917.

Gill, K.S., B.S. Gill, and T.R. Endo. 1993. A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. *Chromosoma* **102:** 374-381.

Gill, K.S., B.S. Gill, T.R. Endo, and E.V. Boyko. 1996a. Identification and high-density mapping of gene-rich regions in chromosome group *5* of wheat. *Genetics* **143:** 1001-1012.

Gill, K.S., B.S. Gill, T.R. Endo, and T. Taylor. 1996b. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics* **144:** 1883-1891.

Gilles, P.N., D.J. Wu, C.B. Foster, P.J. Dillon, and S.J. Chanock. 1999. Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nature Biotechnology* **17:** 365-370.

Gimelfarb, A. and R. Lande. 1994. Simulation of marker-assisted selection in hybrid populations. *Genetical Research* **63:** 39-47.

Grama, A. and Z.K. Gerechter-Amitai. 1974. Inheritance of resistance to stripe rust (Puccinia striiformmis) in crosses between wild emmer (Triticum dicoccoides) and cultivated tetraploid and hexaploid wheat. *Euphytica* **23:** 393-398.

Guyomarch, H., P. Sourdille, G. Charmet, K.J. Edwards, and M. Bernard. 2002. Characterization of polymorphic microsatellite markers from Aegilops tauschii and their transferability to the D-genome of bread wheat. *Theoretical and Applied Genetics* **104:** 1164-1172.

Hospital, F. 2001. Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics* **158:** 1363-1379.

Hospital, F. and A. Charcosset. 1997. Marker-Assisted Introgression of Quantitative Trait Loci. *Genetics* **147:** 1469-1485.

Jones, D.G. and B.C. Clifford. 1983. *Cereal Diseases. Their Pathology and Control.* John Wiley and Sons.

Kidwell, K.K., G.B. Shelton, V.L. Demacon, C.F. Morris, D.A. Engle, J.W. Burns, R.F. Line, C.F. Konzak, J.H. Hatchett. 2002. Registration of 'Zak' wheat. Crop Science **42:** 661-662.

Lichten, M. and A.S. Goldman. 1995. Meiotic recombination hotspots. *Annual Reviews of Genetics* **29:** 423-444.

Ribaut, J.M. and M. Ragot. 2007. Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *Journal of Experimental Botany* **58:** 351-360.

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

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

Servin B. Using markers to reduce the variation in the genomic composition in marker assisted backcrossing. *Genetical Research* **85:** 151-157.

Sidhu, D. and K.S. Gill. 2004. Distribution Of Genes And Recombination In Wheat And Other Eukaryotes. *Plant Cell Tissue and Organ culture* **79:** 257-270.

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

Song, Q.J., J.R. Shi, S. Singh, E.W. Fickus, J.M. Costa, J. Lewis, B.S. Gill, R. Ward, and P.B. Cregan. 2005. Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical and Applied Genetics* **110:** 550-560.

Sourdille, P., H. Guyomarch, C. Baron, B. Gandon, V. Chiquet, F. Artiguenave, K. Edwards, N. Foisset, P. Dufour, and M. Bernard. 2001. Wheat Genoplante

SSR mapping data release: a new set of markers and comprehensive Genetic and physical mapping data.

Stam, P. 2003. Marker-assisted introgression: Speed at any cost? *Eucarpia Leafy Vegetables***:** 117-124.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. *Bio/Technology* **7:** 257-264.

Tranquilli, G., D. Lijavetzky, G. Muzzi, and J. Dubcovsky. 1999. Genetic and physical characterization of grain texture-related loci in diploid wheat. *Molecular General Genetics* **262:** 846-850.

Uauy, C., A. Distelfeld, T. Fahima, A. Blechl, and J. Dubcovsky. 2006. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* **314:** 1298-1301.

Whittaker, J.C., R.N. Curnow, C.S. Haley, and R. Thompson. 1995. Using marker-maps in marker-assisted selection. *Genetical Research* **66:** 255-265.

Young, N.D. and S.D. Tanksley. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theoretical and Applied Genetics* **77:** 95-101.

Zhang, W. and C. Smith. 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theoretical and Applied Genetics* **83:** 813-820.

CHAPTER 6

Protocols Optimized

6.1 Hot Phenol RNA Extraction

RNA extraction of all the developmental stages collected to look at the tissue specificity of homoeologues was done by hot phenol RNA extraction. Before homogenizing the samples, set the water bath to 80°C. Make 10 ml of extraction buffer by mixing 1 ml of 1M Tris-Hcl (pH = 8.0), 0.125 ml of 8M Licl, 0.2 ml of 0.5M EDTA (pH = 8.0), 0.1 gm of SDS and 8.64 ml of water. To the extraction buffer add 10 ml phenol (pH = 4.7) such that the final concentration is 1:1. Heat the above mixture to 80°C. Grind the tissue in liquid nitrogen and add 500 µl of 1Extraction buffer : 1Phenol (for 0.3 gm of tissue). Homogenize by vigorous vortexing for 15 to 20 mins. Add 1/2 volume of chloroform to the samples and vigorously vortex again. Centrifuge for 10 min @14000 rpm at room temperature (RT) and transfer the supernatant to a new tube. Add 1/3 volume of 8M lithium chloride (LiCl) to the supernatant and mix well. Precipitate on ice for at least 2 hrs. Balance the tubes using 2M LiCl and centrifuge at 14,000 rpm for 15 mins at 4°C. Discard the supernatant and wash the RNA pellet with 200 µl of 2M LiCl (RT) and centrifuge at 14,000 RPM for 5 min (RT). Discard the supernatant and wash the RNA pellet with 200 µl of 70% DEPC treated ethanol (EtOH). Centrifuge again at 14,000 RPM for 5 min (RT). Discard the EtOH and dissolve the dried pellet in 20 µl of sterile DEPC treated water. To get rid of the debris after dissolving the pellet, centrifuge at 14,000 RPM for 5 min at 4°C and transfer the supernatant to a new tube. For precipitation add, $1/10^{th}$ and 2.5 the volume of 3M sodium acetate (3M, pH-5.2) and absolute EtOH (RT) respectively. Precipitate the RNA at –80°C for 20 min and centrifuge at 14,000 RPM for 10 min at 4°C. Discard the supernatant and wash the pellet with 70% EtOH and centrifuge again

at 14,000 rpm for 2 mins at 4°C. Air dry and dissolve the RNA in 20 µl of sterile DEPC treated water.

6.2 Poly(A)+ RNA Selection

The poly(A)+ RNA selection from total RNA was done following a standard protocol (Sambrook et al. 1989). 0.05 g (little) of oligo (dT)-cellulose was suspended in 0.1N NaOH. Pour the oligo (dT)-cellulose into a dispocolumn. Generally 0.1 ml oligo column is enough for 1 mg of total RNA. Wash the column with sterile water 2 to 3 times followed by 2 or 3 washes with 1X column loading buffer (20mM Tris.Cl pH 7.6, 0.5M NaCl, 1mM EDTA pH 8.0, 0.1% sodium lauryl sarcosinate) until the pH of the effluent was less than 8.0. Dissolve the RNA in sterile water and heat the sample to 65°C for 5 minutes. Mix total RNA sample with the same amount of 2X column loading buffer. Make the ratio 1:1. Apply the solution to the column and collect the eluate in a tube followed by applying about 1ml of 1X column loading buffer to the column and collect the elute in the same tube. Heat the collected elute again to 65°C for 5 minutes and apply to the column. Wash the column with 5-10 column volume of 1X Column Loading buffer. Elute the poly A RNA with 2-3 column volumes of freshly prepared elution buffer (10mM Tris.Cl pH 7.6, 1mM EDTA pH 8.0, 0.05% SDS) and collect the final solution in an ependorf tube. Add 1/10 amount of Sodium acetate (3 M and pH 5.2) into the RNA solution. Mix well. Add 2.5 volumes of ice cold ethanol, mix and store for atleast 30 mins on ice. Recover poly(A)$^+$ RNA by centrifugation at 10,000 rpm for 15 mins at 4°C. Carefully discard the supernatant, and wash the pellet with 70% ethanol. Recentrifuge briefly and allow the pellet to dry in air and redissolved in sterile water.

6.3 First Strand cDNA Synthesis:

First strand cDNA synthesis was started immediately after poly(A)+ RNA selection. For the synthesis combine 2 ug of RNA sample with 1 ul of oligo (dT) cDNA synthesis primer (1 uM), in a 0.5 ul of tube. Add sterile H_2O to make the volume 5 ul. Incubate the mixture at 70°C for 3 min. Cool the tube on ice for 2 min and briefly spin the tube to collect the content. In the meantime prepare a cDNA master mix by combining 2 µl of 5X first strand buffer, 2 µl of dNTP mix (5 mM) and 1 µl of MMLV reverse transcriptase (200 units/µl). Add 5 µl of master mix to RNA sample, to make the final volume of 10 ul. Mix the tube by pipetting gently and spin briefly. Incubate the mixture at 42^{o}C for 1 hour (either in waterbath or PCR machine). Terminate the reaction by incubating at 75°C for 10 min. Place the tube on ice, centrifuge briefly and collect the content. Transfer 1 ul reaction into a new tube. Dilute 1 ul of cDNA in tube containing 99 ul of sterile H_2O (1/100 dilution). Store in –20°C freezer until use.

6.4 Chromatin Immunoprecipitation Analysis (CHIP):

CHIP analysis was done on chromatin extracted from 28 day old leaf tissue of CS, N7AT7B, N7AT7D, N7BT7A, N7BT7D, N7DT7A, N7DT7B, Dt7AS, Dt7AL, Dt7BS, Dt7BL, Dt7DS, Dt7DL, 7AL-15, 7AL-9, 7AL-10, 7BL-1, and 5BL-11 lines. Chromatin extraction and subsequent immunoprecipitation using different antibodies was adopted with modifications from Arabidopsis (Bowler et al. 2004). CHIP technique is used to map the sites of protein-DNA interaction *in vivo* by combining immunoprecipitation of chromatin fragments with polymerase chain reaction (PCR). CHIP relies on cross-linking of histone/DNA complex by formaldehyde, follwed by chromatin isolation, subsequent shearing and immunoprecipitation with the antibodies specific for a particular modification of the histone tail. The relative enrichment or

depletion of DNA fragment in co-immunoprecipitated DNA is analysed by PCR and as such, reflects its association with particular histone modification. The detailed protocol is described below:

6.4.1 Cross linking, isolation and sonication of chromatin

For CHIP analysis, wheat plants were grown under controlled conditions of 16/8 hr light/dark and 22°C/18°C day / night temperature in the growth chambers. About 5 grams of wheat leaves of 'Chinese spring' (CS) and 17 different aneuploid lines of group 7 were harvested. Transfer the leaves to 50 ml conical centrifuge tubes and rinse with 40 ml deionized (Milli-Q) water; repeat two times. Holes were poked on the cap of the conical tube for vacuum infiltration. Add 37 ml of 1.0% formaldehyde and stuff the top of each 50 ml conical tube (containing the formaldehyde-soaked leaves) with nylon mesh to keep the leaves immersed in the liquid during vacuum. Leaves were vacuum infiltrated for 15 minutes at 37°C in a freeze dryer. The formaldehyde solution should boil. 2.5 ml of 2M glycine (Final Concentration = 0.125M) was added to the formaldehyde solution to quench the cross-linking with an additional 5 minutes vacuum infiltration. Leaves were again rinsed 2 times with Milli-Q (deionized) water. Grind the tissue to a fine powder in liquid nitrogen and add 30 ml of Extraction Buffer 1 in a 50 ml conical tube, vortex and place on ice.

Extraction Buffer 1	**250 ml**	**Final Concentration**
2M Sucrose	50ml	0.4M
1M Tris-HCl pH 8	2.5ml	10mM
14.3M BME (2 (beta)-mercaptoethanol)	87.5 µl	5mM
0.1M PMSF	2.5ml	1mM
+ Cocktail Protease Inhibitors	2.5ml	1:100 dilution

Filter the slurry through 3 layers of Miracloth, into a 50mL conical tube and centrifuge solution 20 minutes @ 4000rpm (1940 x g) at 4°C in a Beckman JA-20 rotor. Remove supernatant and resuspend pellet in 1 ml extraction buffer 2 at 4°C (transfer to 1.5ml eppendorf tube).

Extraction Buffer 2	**10 ml**	**Final Concentration**
2M Sucrose	1.25 ml	0.25M
1M Tris-HCl pH 8	100 µl	10mM
1M MgCl2	100 µl	10mM
100%Triton X-100	100 ul	1%
14.3M BME	3.5 µl	5mM
.1M PMSF	100 ul	1mM
+ Cocktail Protease Inhibitors	100 ul	1:100 dilution

Dissolved pellet was centrifuged at 14,000 RPM for 10 minutes at 4°C. Remove supernatant and resuspend the pellet in 300 ul of extraction buffer 3 at 4°C.

Extraction Buffer 3	**10 ml**	**Final Concentration**
2M Sucrose	8.5 ml	1.7M
1M Tris-HCl pH 8	100 µl	10mM
1M MgCl2	100 µl	10mM
100%Triton X-100	15 ul	0.15%
1M MgCl2	20 µl	2mM
14.3M BME	3.5 µl	5mM
.1M PMSF	100 ul	1mM
+ Cocktail Protease Inhibitors	100 ul	1:100 dilution

In seperate eppendorf tube, add 300ul extraction buffer 3 and overlay the resuspended pellet on top of this 300 ul cushion. Spin solution 1 hour @ top speed in microfuge at 4°C (this step is done to remove the chlorophyll or if the samples are still green). Remove supernatant and resuspend chromatin pellet in 500 ul of nuclear lysis buffer (keep on ice). Take an aliquot for later examination.

Nuclear Lysis Buffer	**5 ml**	**Final Concentration**
1M Tris-HCl pH 8	0.25 ml	50mM
0.5M EDTA	100 µl	10mM
20% SDS	1 ml	1%
.1M PMSF	50 µl	1mM
+ Cocktail Protease Inhibitors	50 µl	1:100 dilution

Once resuspended sonicate chromatin solution on ice 5 times, 10 seconds each @ power setting 8 on a Fisher Scientific sonic dismembrator (Model 100) — pause 1 minute between each 10 sec pulse. After sonication samples were microfuge two times to get rid of the debris, for 10 minutes at 14,000 RPM at 4°C. Take another aliquot and check for chromatin quality and sonication by reverse cross linking following the procedure described below. Shear the chromatin approximately to 0.5 to 3 Kb DNA fragments. Chromatin can be frozen at –80°C at this point for latter use.

6.4.2 Chromatin immunoprecipitation:

Transfer the sonicated chromatin to two 100 ul aliquots in separate siliconized microfuge tubes (using siliconized tubes significantly reduces carry over of non-antibody bound chromatin). Add 900 ul of ChIP dilution buffer to each tube.

ChIP Dilution Buffer	**10 ml**	**Final Concentration**
100%Triton X-100	110 µl	1.1%
0.5 M EDTA	24 µl	1.2mM
1M Tris-HCl pH 8	167 µl	16.7mM
5M NaCl	334 µl	167mM
.1M PMSF	100 µl	1mM
+ Cocktail Protease Inhibitors	100 µl	1:100 dilution

Add 40 µl of equilibrated Protein A agarose beads blocked with salmon sperm DNA (Upstate Biotechnology to the chromatin sample and mix on a rotating mixer (Mini Labroller, Labnet International Inc.) at 4°C for 1 hour. Pellet the beads by

spinning the samples of chromatin and beads at 0.8 RPM, 2 min at 4°C. Combine both 1 ml tubes and split into 4 siliconized microcentrifuge tubes (e.g. for mock, H3K4me, and H3K9me IPs) of 600ul and one of 60ul (10 % input) using. Add desired antibodies to each tube except in the mock control. Antibodies and amounts are:

anti- H3 K9 dimethyl (Abcam) = 10 µl
anti- H3 K4 dimethyl (Abcam) = 10 µl
anti- H3 K27 dimethyl (Upstate) = 5 µl

Incubate chromatin plus antibodies on a rotating mixer wheel, 4 hours to overnight at 4°C. Capture immune complexes by adding 40 µl equilibrated Protein A agarose beads and rotating at 4°C for 1 hour. Again pellet beads by centrifuging 0.8 RPM, 2 min at 4°C. Pelleted beads was washed with 1 ml of each of the following wash buffers 5 min at 4°C each followed by centrifuging at 0.8 RPM, 2 min at 4°C in each wash.

Low Salt Wash Buffer	**50 ml**	**Final Concentration**
5M NaCl	1.5 ml	150mM
20% SDS	0.5 ml	0.2%
100% Triton X-100	0.25 ml	0.5%
0.5M EDTA	200 µl	2mM
1M Tris-HCl pH 8	1 ml	20mM

High Salt Wash Buffer	**50 ml**	**Final Concentration**
5M NaCl	5ml	500mM
20%SDS	0.5 ml	0.2%
100%Triton X-100	0.25 ml	0.5%
0.5M EDTA	200 µl	2mM
1M Tris-HCl pH 8	1 ml	20mM

LiCl Wash Buffer	**50 ml**	**Final Concentration**
4M LiCl	3.125 ml	0.25M
100% NP-40	0.25 ml	0.5%
Sodium deoxycholate salt	0.25 g	0.5%
0.5M EDTA	100 µl	1mM
1M Tris-HCl pH 8	0.5 ml	10mM

TE Buffer	**50 ml**	**Final Concentration**
1M Tris-HCl pH 8	0.5 ml	10mM
0.5M EDTA	100 µl	1mM

After the final wash remove TE completely.

6.4.3 Elution and reverse cross-linking of chromatin

Bead bound complexes were released by adding 250 µl of elution buffer to the pelleted beads.

Elution Buffer	**20 ml**	**Final Concentration**
20% SDS	1 ml	1%
Sodium bicarbonate ($NaHCO_3$)	0.168 gm	0.1M

Vortex and incubate the mixture at 65°C for 15 min with gentle agitation. Centrifuge and transfer the supernantant to new tube repeat two times and combine two eluates. For reverse cross linking incubate the the eluate at 65°C overnight with 20 µl of 5M NaCl. Treat the mixture with 1.5 µl of proteinase K (14 mg/ml), 10 µl of 0.5M EDTA and 20 µl of 1M Tris-HCl, pH 6.5 and incubate for 1 hr at 45°C. DNA was recovered by adding equal volume of phenol:chloroform:isoamylalcohol (25:24:1) and 2 volumes of ethanol, precipitated in -80°C for 2 hrs. Spin the DNA samples at 14,000 RPM for 25 mins, air dry and dissolve the pellets in 25 µl of TE. 1 µl was used for PCR analysis.

6.5 High Throughput Marker Analysis

6.5.1 Quick method of DNA extraction

A quick method using 96 well 2ml deep plates was used for DNA extraction. Briefly, 3-4 fresh leave of about one inch length were placed into plate well and

lyophilized for 3-4 day. After drying, metallic beads were used to grind leaves tissue using Qiagen Mixer Mill for 5-10min. Using multi channel pipette, 750 µl of CTAB buffer solution (100mM Tris pH 7.5, 1.4M NaCl, 20mM EDTA pH 8.0, 2%CTAB, 0.2% Mercaptoethanol) was added and plates were place into levitation machine at 60°C for 1-11/2 hrs. Equal volume of Chloroform:Octanol (24:1) was added and mixed thoroughly using levitation machine for about 30 min. Plates were centrifuged at room temperature at 4000 rpm for 15 minutes and the supernatant was transferred to new plates and DNA was precipitated by adding 2/3 volume of isopropanol. After washing with 70% ethanol, the DNA pellet was air dried and suspended in about 300 µl of TE and stored at 4°C overnight.

Since DNA extraction methods available at the time were expensive, time and labour consuming, and were not compatible with a high throughput operation, we first developed a DNA extraction method that overcame these problems. Starting with 1-2 leaves, our current protocol yields 20 to 100µg of DNA per sample which is sufficient for hundreds of PCR reactions. One person can easily isolate DNA from 192 samples per day (Figure 6.1). Since we use 'home made' chemical reagents and solutions rather than company kits, our cost of DNA extraction per sample is significantly lower than the previously available protocols.

Figure 6.1: High throughput DNA extraction. DNA pellets from 96 plants shown at the bottom of a 96 well plate.

Availability of a large number of user-friendly markers and a high throughput marker analysis system were the other two requirements for the success of our marker assisted background selection method. There are now more than 15,000 DNA markers available for wheat of which about 1,500 are very user-friendly as can be assayed using polymerase chain reaction (PCR).

6.5.2 Marker analysis

Wheat SSR markers acquired from different sources GWM (Röder et al. 1998); CFA and CFD (Sourdille et al. 2001); (Guyomarch et al. 2002), WMC (Somers et al. 2004) and BARCs (Song et al. 2005) were used for background selection. Primer information for all these SSR is available at the Graingenes web site (http://wheat.pw.usda.gov/GG2/quickquery.shtml). All these primers were synthesized with M13 tail (CACGACGTTGTAAAACGAC) at the 5′ end on the forward primer to detect on the Li-cor genetic analyzer. PCR reaction mixture contained (total volume 10 µL) of 25 ng of genomic DNA ,1x PCR Buffer, 2.5mM dNTP, 50mM $MgCl_2$, 1 pmol of IRDye M-13 forward tail primer (Li-cor), 1 pmol each of Forward and Reverse primer and 0.1 unit of Taq DNA polymerase. PCR amplification was performed in 96 well plates on Px2 Thermal Cycler (Thermo Electron). PCR cycling consist of initial cycle of denaturation at 94°C for 3 min, annealing at 60°C for 1 minute and extension at 72°C for 1 minute followed by 30 cycles (94°C for 30sec; 60°C for 30 sec; 72°C for 30 sec). The final extension cycle was at 72°C for 5 minutes and then the PCR product was held at 4°C. PCR products were diluted to 1:10 with 9µl of loading dye (Formamide 10ml+ Bromophenol Blue 5mg+ 0.5M EDTA 200µl). Samples were denatured just before loading at 94º C for 2

min and chilled on ice. Electrophoresis was carried out using an IR2 4200 DNA Analyzer (Li-Cor) on 6% polyacrylamide gel.

6.6 Fate of Homoeologous Genes

About 9,400 ESTs have been physically mapped to wheat chromosomes by deletion mapping (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi). The physical mapping was performed by DNA gel blot analysis with restriction digested DNA from various aneuploid and deletion lines. The genetic stock used included 21 nullisomic-tetrasomic (NT) lines (Sears 1954; Sears 1966), 24 ditelosomic (DT) lines (Sears and Sears 1978) and 119 deletion lines (Endo and Gill 1996). The NT lines were used to map ESTs to individual chromosomes, DT lines for arm and centromere mapping whereas the deletion lines were used for subarm mapping. The number of loci detected by the gel blot analysis of the ESTs ranged from 1 to 24 (Qi et al. 2004). On an average each EST detected 4.8 restriction fragments representing 2.8 loci. About 46% of the ESTs detected 3 or 4 fragments, 12% detected 1 or 2 fragments and remaining 42% detected 5 or more restriction fragments (Qi et al. 2004) http://wheat.pw.usda.gov/NSF/. To study the homoeologous gene expression, gel-blot analysis images of about 6000 ESTs that detected less than five well-characterized restriction fragments all of which were resolved by aneuploids, was downloaded from a local mirror site of the wEST database (http://wheat.pw.usda.gov/wEST/). Out of these, ~1000 ESTs showing very clear hybridization pattern where every restriction fragment band was physically mapped, were selected. Sequence comparisons at a cut off e-value of e^{-70} will be used to identify full-length rice cDNA homologs for the selected ESTs (http://cdna01.dna.affrc.go.jp/cDNA/). The identified rice full-length cDNA sequences was used to identify additional ‘Chinese spring’ wheat ESTs via

'megablast' (blastn tool in NCBI for nucleotide to nucleotide search of highly similar sequences) with a cutoff value of ≥80% sequence similarity. The identified wheat EST sequences were assembled into a contig using ContigExpress Module of *Vector NTI Advance*™ software. EST sequences in a full-length contig were compared for differences in order to identify unique sequence patterns. For example, any full-length contig assembled using an EST physically mapped to all three homoeologues, showing three distinct sequence patterns, would suggest that all three structural copies of the gene express. Similarly, two patterns would suggest that two out of the three copies express. Using this approach, expression of about 1000 genes was studied. These gene expression results for about 30 genes was confirmed using single stranded conformation polymorphism (SSCP) analysis (Adams et al. 2003). For SSCP analysis the primers were made from conserved (that is regions showing no differences among the three homoeologues scored) regions flanking the variable regions. These primers were used to amplify cDNA from CS and the NT lines of the respective chromosome groups, where the EST is physically mapped.

6.7 Modified Single Stranded Conformation Polymorphism (SSCP) Analysis

Presence of three distinct yet similar genomes in a common nucleus, creates a situation whereby to resolve highly identical gene products or homoeologues could be very difficult. Several techniques have been used to study the polyploid gene expression such as RNA gel blot analysis, EST based microarrays, and cDNA-AFLP (Bachem et al. 1996; Bachem et al. 1998). These techniques are successful to study accumulative expression of all copies of a gene but are not suited to discriminate the expression of highly similar homoeologous genes. Recently, Cronn and Adams used SSCP analysis to resolve tetraploid cotton homoeologues with a sequence similarity

of >98% (Cronn et al. 1999, Adams et al. 2003). This approach is sensitive and was able to resolve gene products which differ only by single base pair insertion or deletion (Adams et al. 2003).

RNA from different developmental stages of the wheat plants was extracted using the guanidinium thiocyanate - cesium chloride density gradient method (Sambrook et al. 1989; Sandhu and Gill 2001). After extraction, the RNA was treated with DNaseI enzyme, and then pooled in equal amounts. The first stranded cDNA was synthesized from 2 μg of pooled poly(A)$^+$ RNA using Moloney Murine Leukemia Virus (MMLV) reverse transcriptase. About 20 ng of first stranded cDNA was PCR amplified using oligo 'dT' primer (Sandhu and Gill 2001). The PCR reactions were performed in a total volume of 20μl. The PCR product was mixed with 20 μl of loading buffer (95% formamide, 20mM EDTA, 10mM NaOH, 0.05% bromophenol blue and 0.05% xylene cyanol) (Adams et al. 2003), heated to 94°C for 3 to 5 min. About 5 μl of this mixture was loaded onto 0.4mm thick denaturing 8% polyacrylamide/8M urea gels (Sambrook et al. 1989). Gels contain and were run in 0.5x buffer containing Tris and EDTA with pH 8.3. In standard run, the gels were pre-run at 33mA constant current for 30min and then at 70W constant power for 4 hours. For SSCP analysis, the procedure was modified from Cronn and Adams (2003). In their procedure, the PCR product was run on 1x mutation detection enhancement gel (*MDE*; Cambrex BioScience) containing 2 or 10% urea on model S2 sequencer. After loading the samples, gel was run at 5 W for 16-18hrs at 4°C (Adams et al. 2003). In our case, the PCR product was size separated on a 8% polyacrylamide/8M urea gel. Before loading the samples, the gel was pre-chilled to 4°C for at least 5-6 hrs. After that the gel was run at 10 W for 12-13 hrs. An X-ray film was placed on the dried gel and exposed for three to seven days.

6.8 Data Mining Procedures

To study evolution of genes, the above-mentioned wheat full-length contigs will be further analyzed. The Open reading frames (ORF), and 3' and 5' untranslated regions (UTR) were demarcated using the rice full-length sequence. The ORF regions of the genes was further divided into four equal quarters, to see if there is difference in deletion or substitution pattern within the transcribed region. The differences in sequences among the three homoeologues or patterns aligned were scored manually for each contig and for each of the demarcated regions. One or more base pair substitutions and deletions were scored separately.

In addition to wheat, the differences within the transcribed region were also studied in sequenced and well characterized model genomes. Included in this study was the orthologs of four yeast species available at ftp://genome-ftp.stanford.edu/pub/yeast/sequence/fungal_genomes/. As no software is available for the type of analysis and comparisons proposed in this project, we developed specific software. To study sequence differences among the ortholog species, software first differentiate among the transcribed and intergenic regions and further divide the transcribed portion of the gene into four equal parts. Secondly, the software provided an output in the form of number of insertion/deletion ranging anywhere from 1bp to ≥10bp among the aligned sequences. Another software was developed which counted synonymous and non-synonymous changes within each of the transcribed sub-regions. This software scanned three nucleotides at a time within the transcribed region of the aligned orthologs.

6.9 References

Adams, K.L., R. Cronn, R. Percifield, and J.F. Wendel. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* **100:** 4649-4654.

Bachem, C.W.B., R.J.F.J. Oomen, and R.G.F. Visser. 1998. Transcript imaging with cDNA-AFLP: A step-by-step protocol. *Plant Molecular Biology Reporter* **16:** 157-173.

Bachem, C.W.B., R.S. Van der Hoeven, S.M. De Bruijn, D. Vreugdenhil, M. Zabeau, and R.G.F. Visser. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. *The Plant Journal* **9:** 745-753.

Bowler, C., G. Benvenuto, P. Laflamme, D. Molino, A.V. Probst, M. Tariq, and J. Paszkowski. 2004. Chromatin techniques for plant cells. *The Plant Journal* **39:** 776-789.

Cronn, R.C., R.L. Small, and J.F. Wendel. 1999. Duplicated genes evolve independently after polyploid formation in cotton. *Proceedings of the National Academy of Sciences* **96:** 14406-14411.

Endo, T.R. and B.S. Gill. 1996. The deletion stocks of common wheat. *Journal of Heredity* **87:** 295-307.

Guyomarch, H., P. Sourdille, G. Charmet, K.J. Edwards, and M. Bernard. 2002. Characterization of polymorphic microsatellite markers from Aegilops tauschii and their transferability to the D-genome of bread wheat. *Theoretical and Applied Genetics* **104:** 1164-1172.

Qi, L.L., B. Echalier, S. Chao, G.R. Lazo, G.E. Butler, O.D. Anderson, E.D. Akhunov, J. Dvorak, A.M. Linkiewicz, A. Ratnasiri, J. Dubcovsky, C.E.

Bermudez-Kandianis, R.A. Greene, R. Kantety, C.M. La Rota, J.D. Munkvold, S.F. Sorrells, M.E. Sorrells, M. Dilbirligi, D. Sidhu, M. Erayman, H.S. Randhawa, D. Sandhu, S.N. Bondareva, K.S. Gill, A.A. Mahmoud, X.-F. Ma, Miftahudin, J.P. Gustafson, E.J. Conley, V. Nduati, J.L. Gonzalez-Hernandez, J.A. Anderson, J.H. Peng, N.L.V. Lapitan, K.G. Hossain, V. Kalavacharla, S.F. Kianian, M.S. Pathan, D.S. Zhang, H.T. Nguyen, D.-W. Choi, R.D. Fenton, T.J. Close, P.E. McGuire, C.O. Qualset, and B.S. Gill. 2004. A Chromosome Bin Map of 16,000 Expressed Sequence Tag Loci and Distribution of Genes Among the Three Genomes of Polyploid Wheat. *Genetics* **168:** 701-712.

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

Sambrook, J., E.F. Fristch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY.

Sandhu, D. and K.S. Gill. 2001. Candidate gene cloning and targeted marker enrichment of wheat chromosomal regions using RNA Fingerprinting/Differential Display. *Genome* **44**: 1-7.

Sears, E.R. 1954. The aneuploids of common wheat. *Research Bulletin 572 University of Missouri Agricultural Experiment Station***:** 1-58.

Sears, E.R. 1966. Nullisomic-tetrasomic combinations in hexaploid wheat. *Chromosome Manipulations and Plant Genetics***:** 29-45.

Sears, E.R. and L.M.S. Sears. 1978. The telocentric chromosomes of common wheat. In *Proceedings of the 5th International Wheat Genetics Symposium* (ed. S. Ramanujam), pp. 389-407.

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

Song, Q.J., J.R. Shi, S. Singh, E.W. Fickus, J.M. Costa, J. Lewis, B.S. Gill, R. Ward, and P.B. Cregan. 2005. Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical and Applied Genetics* **110:** 550-560.

Sourdille, P., H. Guyomarch, C. Baron, B. Gandon, V. Chiquet, F. Artiguenave, K. Edwards, N. Foisset, P. Dufour, and M. Bernard. 2001. Wheat Genoplante SSR mapping data release: a new set of markers and comprehensive Genetic and physical mapping data.