QTL AND TRANSCRIPTOMIC ANALYSIS BETWEEN RED WHEAT AND WHITE WHEAT DURING PRE-HARVEST SPROUTING INDUCTION STAGE

BY

Yuanjie Su

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics and Biotechnology-Crop and Soil Sciences - Doctor of Philosophy

2013

ABSTRACT

QTL AND TRANSCRIPTOMIC ANALYSIS BETWEEN RED WHEAT AND WHITE WHEAT DURING PRE-HARVEST SPROUTING INDUCTION STAGE

By

Yuanjie Su

Wheat pre-harvest sprouting (PHS) is a precocious germination of seed in the head when there are prolonged wet conditions occurs during the harvest period. Recent damage caused by PHS occurred in 2008, 2009 and 2011, resulting in severe losses to the Michigan wheat industry. Direct annual losses caused by PHS worldwide can reach up to US $1 billion. Breeding for PHS resistant wheat cultivars is critical for securing soft white wheat production and reducing the economic loss to Michigan farmers, food processors and millers. In general, white wheat is more susceptible to PHS in comparison to red wheat. However, the underlying mechanism connecting seed coat color and PHS resistance has not been clearly described. In this study, a recombinant inbred line population segregating for seed coat color alleles was evaluated for seed coat color and α-amylase activity in three years with two treatments. The genotyping results enabled us to group individuals by the specific red allele combinations and allowed us to examine the allelic contribution of each color loci to both seed coat color and α-amylase activity.

A high-density genetic map based upon Infinium 9K SNP array was generated to locate QTL in relatively narrow regions. A total of 38 Quantitative Trait Loci (QTL) for seed coat color and α-amylase activity were identified from this population and mapped on eleven chromosomes (1B, 2A, 2B, 3A, 3B, 3D, 4B, 5A, 5D, 6B and 7B) from three years and two post-harvest treatments. Most QTL explained 6-15% of the phenotypic variance while a major QTL on

chromosome 2B explained up to 37.6% of phenotypic variance of α-amylase activity in 2012 non-mist condition. Significant QTL × QTL interactions were also found between and within color and enzyme related traits.

Next generation sequencing (NGS) technology was used in current study to generate wheat transcriptome using Trinity with two methods: *de novo* assembly and Genome Guided assembly. Quality assessment of the two assemblies was conducted based on their concordance, completeness and contiguity. Three assembly scenarios were evaluated in order to find a balance between sample specificity and transcriptome completeness. Red wheat and white wheat lines from previous QTL population were collected under mist and non-mist conditions and their expression profiles were compared to identify differentially expressed (DE) genes. At non-mist condition, only around 1% of the genes were differentially expressed between physiologically matured red wheat and white wheat while the rate had a 10-fold increase after 48 hr misting treatment. Annotation of the DE genes showed signature genes involved in germination process, such as late embryogenesis abundant protein, peroxidase, hydrolase, and several transcription factors. They can be potential key players involved in the underlying genetic networks related to the PHS induction process. Gene Ontology (GO) terms enriched in DE genes were also summarized for each comparison and germination related molecular function and biological process were retrieved.

In conclusion, with the population segregating for seed coat color loci, the relationship between seed coat color and α-amylase activity were examined using biochemical methods, QTL analysis, and transcriptome profiling. The variation of seed coat color do closely linked with PHS resistance level at all three levels. DE genes and enriched GO terms identified were discussed for their potential role in bridging the gap between seed coat color and PHS resistance.

ACKNOWLEDGEMENTS

This is it! At the end of my Ph.D. study, I finally got a chance to document my sincere gratitude for all the help I received during the last four years at Michigan State University. Without their generous help, I can never go this far. First, I would like to thank my previous advisor, Dr. Janet Lewis, who first brought me to the field of wheat breeding. Her passion in plant breeding and agriculture truly inspired me to work in this exciting field. Then, I would like to express my greatest appreciation to my major advisor, Dr. David Douches, for his insightful suggestion and guidance on my dissertation study and career development, for his constant support and encouragement throughout my doctoral study and especially in the preparation of this dissertation. I would also like to thank my committee members, Dr. Mitch McGrath, Dr. Shinhan Shiu, and Dr. Dechun Wang, who spent significant amount of time discuss, critique, challenge, and encourage me through my dissertation work and the final dissertation editing. Discussions with them were always a pleasure and a great learning experience.

When it comes to research, I would like to first thank Dr. Robin Buell's group, especially John Johnston, for generously sharing with me their insights and experiences in bioinformatics. Then I would like to thank Dave Main, who helped me freeze dried my last two year's wheat samples in a timely manner. I would also like to thank MSU High Performance Computing Center, where all my bioinformatics analysis was done, and all the laboratories sharing equipment with me: USDA-ARS sugarbeet and bean research unit, soybean research lab, Dr. Ning Jiang's lab and Dr. Randy Beaudry's lab.

The last two years with potato breeding group was a great experience: Nathan Butler, Dr. Norma Manrique-Carpintero, Joe Coombs, Dr. Kim Felcher, Donna Kells, Dr. Alicia Massa, Dr. Dan Zarka, Kelly Zarka, … Working with them really made me feel like living with a big family, which is even more meaningful for an international student like me.

I am deeply indebted to Dr. Russ Freed for his leadership during the wheat lab transition period and his continuous support for the last years of my doctoral study. I am also grateful to my cohort in the wheat lab: graduate students, Esteban Falconi, Swasti Mishra, Chong Yu; technicians, Lee Siler, Sue Hammer, Randy Lorenz, and all the hardworking undergraduate crews, who offered a lot of help during my phenotype collection stage both in the field and in the lab. I would also like to thank Monsanto Fellows in Plant Breeding program for supporting my Ph.D. study.

Last but not the least, my mom, Xueqing Zhang, and my dad, Qinghe Su, whose unconditional love and support made me through the toughest time of the process. I also would like to thank my long-time friend, Ya Liu and all my friends here in the department: Valerio Hoyos, Guangxi Wu, Zixiang Wen, Jiazheng Yuan, Veronica Vallejo, Dongyan Zhao, Carmille Bales. Their companion, support, and friendship made my doctoral life exciting and memorable!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Literature review

1.1 White wheat as a commodity crop

1.1.1 Wheat market division and usage

Wheat (*Triticum aestivum* L.) is a staple crop after its domestication 9000 years ago (Peng et al., 2011). Wheat provides 20% of calories consumed by more than 40% of the world's population (Gill et al., 2004). Wheat is widely grown in all the continents and traded internationally as a commodity. In the US, wheat ranks third among field crops in both planted acreage and gross farm receipts (National Agricultural Statistics Service, 2012). Wheat can be divided into six categories based on vernalization requirement (winter, spring), seed coat color (red, white), and kernel texture (hard, soft) (McFall and Fowler, 2009). Winter wheat requires vernalization, a prolonged exposure to cold temperature, in order to flower and spring wheat can flower without vernalization. Red seed coat color is reddish brown and white seed coat color is yellowish and tan. For downstream usage, hard wheat is generally used for bread-making while soft wheat is used for pastry, snacks and breakfast cereals. The usage difference is due to the protein content. Major protein type in wheat is gluten, which holds carbon dioxide during fermentation and gives a soft texture. Thus, the higher protein content in hard wheat, when compared with soft wheat, makes it a preferred material for bread-making.

Recent years, whole-grain products are in high demand for its beneficial effects in prevention of diet-related disorders and cancers (Seal and Brownlee, 2010). White wheat is the preferred material for whole-grain products due to its higher flour yield and better tasting quality. However, the market demand for high quality white wheat is not fulfilled, which may due to the competition for acreage from genetically modified crops and farmers' concern about weather-induced quality issues, such as pre-harvest sprouting. White wheat production has declined, from

352 million bushels to 221 million bushels during last 15 years (Economic research services, United States department of agriculture, 2013).

1.1.2 White wheat in Michigan

Wheat is Michigan's third largest crop, grown on approximately half a million acres, with an economic impact of $2.7 billion (National Agricultural Statistics Service, 2012). Soft red (60% of acreage) and soft white (40% of acreage) winter wheats produced in Michigan are used as the primary ingredient of breakfast cereal and bakery goods (Peterson et al., 2006). During the past 15 years, white wheat production has declined rapidly in the eastern US, which has led to Michigan being the leading white wheat producer (78%) in this region of the US (Sutherland, 2011). This rapid decrease in white wheat acreage increased the production cost of major cereal and milling companies, such as Kellogg, Post, and Star of the West milling company, resulting in a premium pricing for high quality white wheat. Therefore, growing soft white wheat is a profitable and growing market for Michigan farmers.

1.2 Pre-harvest sprouting damage in wheat production

Pre-harvest sprouting (PHS) is the precocious germination of seed induced by prolonged wet conditions prior to harvest. Sprout damage is mainly three-fold: loss in yield, reduction in test weight, and low-quality downstream products. The first two directly affect farmers profit and the last one is the biggest concern for milling and cereal companies. Sprouted wheat produces excess amounts of α-amylase. The final products are off-color with weak texture even after blending with batches containing low α-amylase activity. This causes substantial economic loss to food processors, therefore farmers get discounted price at grain receivers if grains have unacceptable sprout damage. Direct annual losses caused by PHS worldwide can reach up to US $1 billion (Black et al., 2006).

Based on the trend of climate change, Michigan is expected to experience more extreme weather and variable precipitation (Doll and Baranski, 2011), which poses a higher risk of PHS for wheat growers. However, few precautionary management practices has been established to avoid PHS except for a timely harvest, while an early harvest can increase drying cost and lower yield. On the other hand, swathing, a commonly used farm practice that cuts and windrows the wheat crop, can significantly increase PHS risk when rainfall is present (Derera, 1989). Moreover, a recent phenotypic screen at Michigan State University showed that few Michigan soft wheat varieties have PHS resistance (Yu, 2012). Therefore, enhancing PHS resistance in wheat varieties adapted to Michigan is in urgent need and is one of the most effective ways to reduce wheat growers' risk of PHS.

1.3 Seed dormancy

Pre-harvest sprouting is seed germination prior to seed harvest that is attributed to a lack of dormancy. Dormancy induction, maintenance and release in seed are closely related to PHS process. Dormancy can be defined as temporary growth arrest when environmental conditions are sub-optimal for germination. It is a strategy adopted by many species to survive adverse environmental conditions (Footitt and Cohn, 2001). It can be affected by internal balance of hormone levels and sensitivity, while the environmental signals can also modify the expression of related enzyme in a feedback-regulated fashion (Finkelstein et al., 2008; Footitt and Cohn, 2001). Seed dormancy has been previously reviewed from different aspects (Bentsink et al., 2007; Finkelstein et al., 2008; Foley, 2001; Graeber et al., 2012; Koornneef et al., 2002; Penfield and King, 2009).

1.3.1 Dormancy during seed development

Seed development initiates from double fertilization. A cereal seed includes embryo, endosperm and testa and dormancy occurs in both embryo and testa during seed development. The embryo development can be divided into two stages: 1. embryo morphogenesis, which is the formation of embryo and acquisition of polarity, and 2. embryo maturation, the stage that the embryo accumulates nutrients to prepare for the stress caused by desiccation (Ohto et al., 2007). The embryo gains the ability to germinate as soon as the embryo is fully developed. In order to continue the nutrient deposition and finish embryo maturation, abscisic acid (ABA) content increases to induce dormancy and stay high before grain desiccation to ensure the germination does not occur (Finkelstein et al., 2008). Evidences proved that ABA controls and coordinates various processes through multiple transcription factors and seed dormancy is closely related to the ABA content in the grain (Bentsink et al., 2007). Embryo sensitivity to ABA played an equally important role as ABA content in the induction of dormancy (Walker-Simmons, 1987). ABA insensitive mutants from Arabidopsis showed reduced seed dormancy (Koornneef et al., 1984), while ABA hyper-sensitive mutants showed enhanced seed dormancy (Cutler et al., 1996). In wheat, when ABA responsiveness of wheat embryo was restored by introducing a functional *VP1* ortholog from oat, the seed dormancy, or PHS resistance, increased (McKibbin et al., 2002).

Seed dormancy peaks when seed reach physiological maturity, which is characterized by the maximum dry weight. Then seed enters desiccation process, or dry-down, during which dormancy is slowly released. Several events happen during this period including a shift in the internal hormone balance from ABA to gibberellin acid (GA) which promotes seed germinability, a decrease in ABA content and sensitivity (Ali-Rachedi et al., 2004; Grappin et al., 2000), and an increased sensitivity to GA and light (Hilhorst, 2007). Alteration of membranes and protein

degradation also improves germination vigor (Angelovici et al. 2010). Hence, the longer the seed enters into dry-down process, the less dormant and more germinable the seed is, and the higher risk of PHS the seed has. When internal seed dormancy level is low, seed can quickly pass the dormancy threshold during the early stage of desiccation and PHS can occur when environmental conditions are favorable (Obroucheva and Antipova, 2000).

1.3.2 Genetic control of seed dormancy

Dormancy in plants is jointly regulated by embryo- and seed coat-imposed pathways, which are independently controlled by separate genetic systems. The embryo-imposed dormancy is controlled by both maternal and paternal parents while coat-imposed dormancy is adopted from female parents (Flintham, 2000; Himi et al. 2002).

Potential mechanisms of embryo-imposed dormancy include the balance of ABA and GA content, embryo sensitivity to these hormones, small molecules such as nitric oxide, and environmental cues, such as moisture, temperature and light (Walker-Simmons, 1987; Ohto et al., 2007; Bethke et al. 2007). After physiological maturity, balance between ABA and GA content shifts to GA after physiological maturity and germination occurs when outside conditions are optimal. Arabidopsis mutants identified in ABA and GA pathways, such as *ABA-insensitive 3* (*ABI3*) (Giraudat et al., 1992) and *GA-insensitive* (*GAI*) (Peng et al., 1997) are key players during the germination process (Koornneef et al., 2002). Their wheat orthologs, such as *viviparous 1* (*VP1)*, *ABI3* orthologs, (Flintham and Gale, 1982) and *reduced height 3* (*RHT3*), and other *GAI* orthologs (Peng et al., 1999) are also reported with similar functions. Recently cloned genes related to dormancy and germination are *delay of germination 1* (*DOG1*) in Arabidopsis, *seed dormancy 4* (*SDR4*) in rice and *TaPHS1*, a wheat homeolog of *mother of flowering time* (*MFT*) on short arm of Chromosome 3A (Bentsink et al., 2006; Liu et al., 2013;

Sugimoto et al., 2010). Functional studies of the causative mutations within these genes and allele mining from a diverse germplasm can be an effective approach to recover functional alleles and introduce into elite lines with genetic defects. One example lies in *VP1*. Common wheat, including ancestral varieties, are prone to PHS due to its mis-spliced *VP1* locus (McKibbin et al., 2002). After screening a diverse wheat germplasm, several functional alleles were found and proven to offer improved dormancy in wheat cultivars (Sun et al., 2012).

Seed-coat imposed dormancy can affect seed germination in three major ways: mechanical restriction to radical protrusion, testa permeability to water and gas exchange, and supply of germination inhibitor such as flavonoids (Debeaujon et al., 2007). Seed germination requires optimal environments. Water and oxygen are two major components. During germination, testa permeability interferes with seed imbibition and leaching of germination inhibitors, such as ABA (Bewley and Black, 1994). Several studies done in Arabidopsis showed that the more permeable the testa, the easier germination occurred (Debeaujon and Koornneef, 2000; Nesi et al., 2001). Testa permeability was also found to be inversely proportional to the phenolic content and their degree of oxidation (Debeaujon et al., 2007). Germination also involved active respiration process. Hence the limited oxygen supply to the embryo caused by the testa can slow down the germination process. This barrier to oxygen diffusion increases with increasing temperature and decreases during dry periods of after-ripening (Lenoir et al., 1986). In summary, seed coat imposed dormancy can come from the mechanical restriction provided by seed coat and by germination inhibitors, which are mainly phenolic compounds deposited in the seed coat.

1.4 Genetic control of seed coat color

Classical work on wheat kernel color by Nilsson-Ehle (1909) showed that grain color is controlled by three loci, termed *R* genes, with partial dominance. Each locus resides on the long arm of chromosome group 3 of hexaploid wheat (Sears, 1944; Metzger and Silbangh, 1970; Himi et al., 2011). However, the expression of color is more complex because of additional minor genes (Freed et al., 1976; Reitan, 1980), genotype x environment (G x E) interactions (Matus-Cadiz et al., 2003) which includes location effects (Wu et al., 1999), and soil nitrogen content (Kettlewell, 1999).

Wheat seed coat color is mainly composed of phelobaphene and proanthocyanidin (PAs). Phlobaphene is the flavonoid that provides reddish color in wheat seed coats, which is a derivative of catechin and catechin tannin and also is endogenous germination inhibitors (Miyamoto and Everson, 1958). These polyphenol compounds are synthesized through the flavonoid biosynthesis pathway (Debeaujon et al., 2007). A wheat *R* gene was cloned recently and found to be a Myb-type transcription factor (Himi and Noda, 2005), which is involved in activation of several flavonoid biosynthesis genes (*CHS, CHI, F3H, DFR*) and in turn controls phlobaphene biosynthesis (Himi et al., 2005).

Proanthocyanidins (PAs) are colorless phenolic oligomers or polymers synthesized during early stage of seed development and accumulate in the seed coat. During the dry-down period, PAs are oxidized to give brown derivatives that confer mature seed color (Koornneef et al., 2002). PAs can increase seed-coat dormancy by increasing the testa thickness and its mechanical strength (Meredith and Pomeranz 1985). During oxidation, PAs have a tendency to crosslink with proteins and carbohydrates in cell walls, thus reinforcing testa structure and also modifying its permeability properties (Marles et al., 2003; Marles and Gruber, 2004). Both

exogenous and endogenous PAs can inhibit seed germination by promoting *de novo* synthesis of ABA in Arabidopsis (Jia et al., 2012). PA-deficient Arabidopsis mutants were also found to have reduced dormancy (Winkel-Shirley, 2001), while in barley, *HvMYB10*, a key regulator for PAs accumulation, were found to be positively correlated with grain dormancy (Himi et al., 2012).

In summary, grain color is controlled by the presence and amount of phenolic compounds, and it is quantitatively expressed with a GxE interaction. Its relationship with seed dormancy is through phenolic compounds, which are known germination inhibitors.

1.5 Germination and α-amylase

Germination occurs when the environment, mainly moisture and temperature, is optimal and internal seed dormancy has been reduced to a low level during grain dry-down. Internal seed dormancy is generally controlled by the balance of ABA and GA. During later stages of dry-down, hormone balance shifts to GA. Excessive GA stimulates the degradation of DELLA proteins, a class of GA signaling repressor, via a ubiquitin-proteasome pathway (Silverstone, et al., 2001). This de-repression of GA responsiveness in seed further stimulates the downstream events of germination (Steber, 2007). Germination begins with a rapid increase in water uptake (imbibition), then follows with a lag phase that has reduced water uptake but more active metabolism and ends when the radicle protrudes from the pericarp (Davies et al., 2011). During germination, GA is released from embryo to aleurone layer, where hydrolytic enzymes are synthesized and secreted into the endosperm, causing reserve mobilization. These enzymes can be categorized as carbohydrate- and protein-degrading enzymes based on their targets (Kruger, 1989).

One of the major carbohydrate-degrading enzymes is α-amylase. It is synthesized in response to GA and further regulated by GA during germination process. The first element of α-

amylase promoter is part of a GA response element (Skriver et al., 1991). GA also activates a Myb transcription factor, *GAMyb*, which in turn activates α-amylase expression (Gubler et al., 1995). Considered as a signature of germination, α-amylase has been extensively studied over the past few decades (Sun and Gruber, 2004). In wheat, there are two major types of α-amylase, α-AMY-1 and α-AMY-2, each of which includes multiple forms (isozymes) (Kruger, 1989). The α-AMY-1 isozyme is the high iso-electric point (pI) group that is present mainly in the aleurone layer and/or scutellum of germinating grain. It is the primary α-amylase form in seed during germination. The other group, α-AMY-2, is the low pI group found in pericarp of immature grain. During seed development, it is the major form detected. It degrades continuously during maturation and its activity is low during germination (Kruger, 1989; Lunn et al., 2001).

Due to its deleterious effects in starch degradation, α-amylase content in sprouted grain is a concern for cereal companies. In a sprouted wheat grain, α-amylase cleaves α-(1→4) D-glucosidic linkages in starch components, which contributes to dextrin production during baking and forms a sticky crumb structure in the final product (Buchanan and Nicholas, 1980). In order to measure the sprouted damage caused by α-amylase, a number of quantitative methods have been developed, including viscometric, turbidometric, fluorometric, colorimetric, gel-diffusion, and reducing sugar assays (Kruger, 1989). In commercial practice, the falling number test has been widely accepted by elevators and mills (Hagberg, 1960). This method gives an indication of the amount of sprout damage that has occurred within a wheat sample by testing flour viscosity. Generally, a falling number value of 350 seconds or longer indicates a low α-amylase activity and sound grain quality. As the amount of α-amylase activity increases, the falling number decreases. Values below 200 seconds indicate an excessive α-amylase activity in the grain. Falling number test was found to be highly correlated (R^2 =0.975) with direct measurement of α-

amylase activity (Moot and Every, 1990; Verity et al., 1999; Perten 1964). The falling number test has three major disadvantages when adapted to a cultivar development program: 1. It has a relatively narrow detection range; 2. The measurement variation can be larger on the lower end of the detection range, which is due to the impact of base time, which includes the time for flour gelatinization and free-fall of the stirrer on the falling number value [The lower the falling number value, the larger the base time will impact the result]; and 3. The large sample size required by the test can be a limitation for early generation testing (Verity et al. 1999). In recent years, with the optimization of an enzyme-linked immunosorbent assay (ELISA) method, a direct measurement of flour α-amylase activity was adopted by regional wheat quality labs for PHS resistance screening (Dr. Edward Souza, Bayer CropSciences, *per comm.*).

In general, synthesis of α-amylase as a direct response to GA induction during germination is a phenotypic marker for grain germination. Hence, the measurement of α-amylase, either directly or indirectly, to evaluate PHS resistance in cereals is routine (Masojć et al., 2011; Ullrich et al., 2012; Yang et al., 2012; Zanetti, et al., 2000).

1.6 PHS in wheat

1.6.1 Evaluation of pre-harvest sprouting

In cereal crops, physiological maturity (PM) is defined as the time when seed reaches its maximum dry weight (Hanft and Wych, 1982). It is a transition point from seed maturation to seed dry-down. During dry-down, seed dormancy drops dramatically and seed is vulnerable to PHS. If seed moisture is low during this period, seed will enter dormancy as expected. However, if moisture remains high at this stage, PHS can happen, especially in varieties with low PHS resistance (King, 1976; Derera, 1989). Therefore, selection of wheat varieties with PHS resistance is usually conducted with materials at three to seven days after PM instead of at

harvest maturity. Complete loss of green pigments from glume and peduncle was found to be closely related with PM in hard red spring wheat (Hanft and Wych, 1982), and has been adopted as a consistent visual indicator of PM for wheat and barley, with little cultivar bias (Clarke, 1983; Copeland and Crookston, 1985). It is now widely used in wheat PHS research programs to reduce sample variation in plant maturity (Humphreys and Noll, 2002; Kulwal et al., 2012; Liu et al., 2008).

PHS resistance is quantitative and can be affected by multiple factors besides inherent dormancy level. Morphological traits that enable slower water-uptake have been studied for their potential impacts on PHS resistance such as awn or awnless, erectness of spike, openness of florets, tenacity of glumes, and germination inhibitors in the bracts (Gatford et al., 2002; King, 1984; Paterson et al., 1989). Conflicting results were shown between these studies.

Due to its complexity, PHS is phenotyped indirectly using different metrics. Four phenotypes, namely sprouting count (%), germination rate (%), α-amylase activity (units/gram flour), and falling number test (seconds), have been used in PHS related studies. Sprouting count is a direct measure of germination with the whole spike, which mimics field observations (Liu et al., 2008; Kulwal et al., 2012). Germination rate is a measure of seed dormancy *per se* (Chen et al., 2008). Both measurements are based on the visual evidence of sprouting, however information about the extent of internal enzyme damage to the starch integrity, which starts before visual sprouting, is missed. Thus, α-amylase activity (Mccleary and Seehan, 1987; Singh et al., 2008) and falling number (Perten, 1964; Rasul 2009) are often used to examine early stages of PHS.

QTL mapping and association mapping of PHS related traits has been done in multiple populations with different phenotypes (Imtiaz et al. 2008; Zanetti et al., 2000; Kulwal et al., 2005,

2012; Mares et al., 2005). QTL have been mapped on all 21 chromosomes of wheat (Kulwal et al., 2012), and can be divided into two categories: QTL do and do not collocate with color alleles on Chromosome Group 3. For QTL not linked to color, α-amylase is one of the traits of special interest. Wheat α-amylase has two major groups, α-AMY1 and α-AMY2. An earlier study mapped α-amylase to Chromosomes 6 and 7 in wheat (Gale et al., 1983). Recent QTL mapping and association methods have mapped α-amylase isozymes on the long arm of chromosome 6B in wheat (Mrva and Mares, 1999; Netsvetaev et al., 2012). Recently, studies in rye to evaluate the relationship between α-amylase and PHS (Masojć and Milczarski, 2005; Masojć et al., 2011; Masojć and Milczarski, 2009) showed QTL controlling α-amylase were found on all chromosomes. Some QTL can promote PHS while some can inhibit PHS (Masojć et al., 2011). These results indicate a complex genetic linkage between α-amylase and PHS. Based on synteny relationships between rye and wheat (Devos et al., 1993), a similar genetic structure linking PHS and α-amylase activity in wheat may be expected.

1.6.2 Improving PHS resistance in wheat

Pre-harvest sprouting (PHS) in cereals has been recognized as an international problem since 1973. International meetings focusing on PHS have been hosted every three to four years around the world since then (Nyachiro, 2012). The direct economic loss worldwide due to PHS can reach up to US $1 billion annually (Black et al., 2006). Cultivar development for PHS resistance is challenging due to limited genetic resources, genotype x environment interactions, laborious sampling procedures, and inconsistent funding resources.

The intuitive way of breeding for PHS resistance is to improve seed dormancy. However, during crop domestication, dormancy is a trait that breeders strongly select against. In wheat, the mis-splicing of *viviparous 1* (*VP1*) in ancestral and modern varieties caused an even narrower

variation in seed dormancy (McKibbin et al., 2002). When compared with red wheat, the genetic variation of PHS resistance in white wheat is even narrower, which might be due to a lack of coat-imposed dormancy offered by red phenolic compounds in seed coat.

Phenotyping for PHS resistance requires careful sampling at physiological maturity and extra greenhouse space is required if artificial misting is conducted to provide high moisture conditions conducive for PHS. The timing, extra labor, and resources required by PHS breeding can significantly impact the harvest season of a breeding program and a lack of dedicated funding will limit the progress in this area. Moreover, PHS induction is highly dependent on surrounding environments. Known factors that can affect PHS include cold temperature during growth, moisture stress, and interaction between maturity and stress (Mares and Mrva, 2008; Biddulph et al., 2007). Therefore, the complex interactions between genotypes and surrounding environment further complicate the selection for PHS resistance (Joosen et al., 2013; Rasul et al., 2012).

With all these challenges, breeding for PHS is not an easy task. However, several powerful tools for genetic study have been developed during the last couple years, such as next-generation sequencing, high-throughput SNP genotyping, and genotyping-by-sequencing (Cavanagh et al., 2013; Elshire et al., 2011). In the meantime, the understanding of the wheat genome has been improved extensively while more and more genomic resources are available for the public, such as Cereals Data Base (http://www.cerealsdb.uk.net) and International Wheat Genome Consortium (http://www.wheatgenome.org). These tools will help identify candidate regions for complex trait discovery at a higher resolution than currently afforded while the annotation of the wheat genome of will facilitate identification of the underlying genes.

Some breeding schemes for selecting white seed color were suggested without a specific consideration for PHS resistance (Cooper and Sorrells, 1984; Knott et al., 2008). Recently, genomic selection (GS), a breeding scheme that uses all marker information to predict breeding value of each line, has been tested for efficiency in PHS breeding. Due to its low heritability, PHS using GS does not outperform traditional phenotypic selection significantly (Mark Sorrells, unpublished data). Except for conventional breeding, there are other ways to breed PHS resistance into wheat. Recently, a EMS mutated wheat cultivar increased seed dormancy by reducing ABA sensitivity (Schramm et al., 2013). Microarrays using after-ripened seed helped to identify genetic networks during the dormancy release period (Liu et al., 2013). A combination of candidate gene improvement with genome wide marker value prediction may help improve the efficiency of phenotypic selection, while multiple-environment screening is also critical to understand the genotype x environment interaction for PHS.

1.7 Project objectives

Breeding for PHS resistant wheat cultivars is critical for securing soft white wheat production and reducing the economic loss to Michigan wheat growers, food processors and millers. In Chapter 2, the allelic contributions of seed coat color and α-amylase activity were examined in a RIL population segregating for seed coat color ('Vida' × MTHW0471). The Infinium 9K SNP array was used for QTL mapping. QTL related to seed coat color and α-amylase activity were identified based on data collected over three years with two post-harvest treatments. Additive × additive gene interactions within and between traits were also identified in current study, which further demonstrated the genetic complexity of PHS resistance in wheat. In Chapter 3, wheat transcriptome data were generated for red and white wheat during misting process. The resulting transcripts assembly would be a valuable resource for future genetic

studies and genome annotation. Differential expression analysis conducted between red wheat and white wheat under mist and non-mist conditions for seeds at physiological maturity showed the similarity and differences of red and white wheat in response to misting treatment. GO enrichment test also showed multiple germination related GO terms. Both of them can be potential candidates for future analysis. In Chapter 4, the future directions related to current research were discussed.

REFERENCES

REFERENCES

Ali-Rachedi S., Bouinot D., Wagner M.H., Bonnet M., Sotta B., Grappin P., Jullien M. (2004) Changes in endogenous abscisic acid levels during dormancy release and maintenance of mature seeds: studies with the Cape Verde Islands ecotype, the dormant model of Arabidopsis thaliana. Planta 219:479-488.

Angelovici R., Galili G., Fernie A.R., Fait A. (2010) Seed desiccation: a bridge between maturation and germination, Trends in Plant Science 15:211-218.

Bentsink L., Jowett J., Hanhart C.J., Koornneef M. (2006) Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in Arabidopsis. Proceedings of the National Academy of Sciences 103:17042-17047.

Bentsink L., Soppe W., Koornneef M. (2007) Genetic aspects of seed dormancy Blackwell Publishing, Oxford, UK.

Bethke P.C., Libourel I.G.L., Aoyama N., Chung Y.Y., Still D.W., Jones R.L. (2007)The Arabidopsis aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. Plant Physiol. 143: 1173-1188.

Bewley J.D., Black M. (1994) Seeds: physiology of development and germination, 2nd Ed. Plenum Press, New York, London.

Biddulph T.B., Plummer, J.A., Setter, T.L., Mares, D.J. (2007) Influence of high temperature and terminal moisture stress on dormancy in wheat (Triticum aestivum L.). Field crops research 103(2): 139-153.

Black M., Bewley J.D., Halmer P. (2006) The encyclopedia of seeds science, technology and uses. CABI Publishing, Wallingford, Oxfordshire, p 528.

Buchanan A.M., Nicholas E.M. (1980) Sprouting, alpha-amylase and bread making quality. Cereal Res. Comm. 8:23-28.

Cavanagh C.R., Chao S., Wang S., Huang B.E., Stephen S., Kiani S., Forrest K., Saintenac C., Brown-Guedira G.L., Akhunova A., See D., Bai G., Pumphrey M., Tomar L., Wong D., Kong S., Reynolds M., da Silva M.L., Bockelman H., Talbert L., Anderson J.A., Dreisigacker S., Baenziger S., Carter A., Korzun V., Morrell P.L., Dubcovsky J., Morell M.K., Sorrells M.E., Hayden M.J., Akhunov E. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences 110:8057-8062.

Chen C.X., Cai S.B., Bai G.H. (2008) A major QTL controlling seed dormancy and pre-harvest sprouting resistance on Chromosome 4A in a Chinese wheat landrace. Molecular Breeding 21:351-358.

Clarke J.M. (1983) Time of physiological maturity and post-physiological maturity drying rates in wheat. Crop Science 23:1203-1205.

Cooper D.C., Sorrells M.E. (1984) Selection for white kernel color in the progeny of red/white wheat crosses. Euphytica 33: 227-232.

Copeland P.J., Crookston R.K. (1985) Visible indicators of phsiological maturity barley. Crop Science 25:843-847.

Cutler S., Ghassemian M., Bonetta D., Cooney S., McCourt P. (1996) A protein farnesyl transferase involved in abscisic acid signal transduction in Arabidopsis. Science 273:1239-1241.

Davies F.T., Geneve R.L., Kester D.E. (2011) Principles of propagation from seeds. In: Hartmann and Kester's Plant Propagation: Principles and Practices, 8th ed. Prentice Hall.

Debeaujon I., Lepiniec L., Pourcel L., Routaboul J.M. (2007) Seed coat development and dormancy. In: Bradford K.J., Nonogaki H. (Eds.) Annual Plant Reviews Vol. 27: Seed Development, Dormancy and Germination, Blackwell Publishing Ltd. p. 25-49.

Debeaujon I., Koornneef M. (2000) Gibberellin requirement for Arabidopsis seed germination is determined both by testa characteristics and embryonic abscisic acid. Plant Physiol. 122: 415-424.

Derera N.F. (Ed.) (1989) Preharvest field sprouting in cereals., CRC Press Inc., Boca Raton, Florida.

Devos K.M., Atkinson M.D., Chinoy C.N., Francis H.A., Harcourt R.L., Koebner R.M.D., Liu C.J., Masojc P., Xie D.X., Gale M.D. (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. Theoretical and Applied Genetics 85:673-680.

Doll J.E., Baranski M. (2011) Field crop agriculture and climate change. Climate Change and Agriculture Fact Sheet Series, E3149.

Economic research services, United States department of agriculture (2013) Table 1. Wheat: Planted acreage, harvested acreage, production, yield, and farm price. http://www.ers.usda.gov/data-products/wheat-data.aspx#.Uj5WWn-GtW8, accessed July 10th, 2013

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

Flintham J.E., Gale M.D. (1982) The tom thumb dwarfing gene, *Rht3* in wheat. 1. Reduced preharvest damage to breadmaking quality. Theoretical and Applied Genetics 62:121-126.

Flintham J.E. (2000) Different genetic components control coat-imposed and embryo-imposeddormancy in wheat. Seed Science Research 10:43-50.

Finkelstein R., Reeves W., Ariizumi T., Steber C. (2008) Molecular aspects of seed dormancy. Annual Review of Plant Biology 59:387-415.

Foley M.E. (2001) Seed dormancy: an update on terminology, physiological genetics, and quantitative trait loci regulating germinability. Weed Science 49:305-317.

Footitt S., Cohn M.A. (2001) Developmental arrest: from sea urchins to seeds. Seed Science Research 11:3-16.

Freed R.D., Everson E.H., Ringlund K., Gullord M. (1976) Seed coat in wheat and the relationship to seed dormancy at maturity. Cereal Res. Comm. 4:147-148.

Gale M.D., Law C.N., Chojecki A.J., Kempton R.A. (1983) Genetic control of α-amylase production in wheat. Theoretical and Applied Genetics 64:309-316.

Gatford K.T., Eastwood R.F., Halloran G.M. (2002) Germination inhibitors in bracts surrounding the grain of *Triticum tauschii*. Functional Plant Biology 29:881-890.

Giraudat J., Hauge B.M., Valon C., Smalle J., Parcy F., Goodman H.M. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. The Plant Cell 4:1251-1261.

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

Graeber K.A.I., Nakabayashi K., Miatton E., Leubner-Metzger G., Soppe W.J.J. (2012) Molecular mechanisms of seed dormancy. Plant, Cell & Environment 35:1769-1786.

Grappin P., Bouinot D., Sotta B., Miginiac E., Jullien M. (2000) Control of seed dormancy in *Nicotiana plumbaginifolia*: post-imbibition abscisic acid synthesis imposes dormancy maintenance. Planta 210:279-285.

Gubler F., Kalla R., Roberts J.K., Jacobsen J.V. (1995) Gibberellin-regulated expression of a *myb* gene in barley aleurone cells: evidence for Myb transactivation of a high-pI alpha-amylase gene promoter. The Plant Cell 7:1879-1891.

Hagberg S. (1960) A rapid method for determining alpha-amylase activity. Cereal Chemistry 37:218-222.

Hanft J.M., Wych R.D. (1982) Visual indicators of physiological maturity of hard red spring wheat. Crop Science 22:584-588.

Hilhorst H.W.M. (2007) Definitions and hypotheses of seed dormancy. Blackwell Publishing, Oxford, UK.

Himi E., Mares D.J., Yanagisawa A., Noda K. (2002) Effect of grain colour gene (R) on grain dormancy and sensitivity of the embryo to abscisic acid (ABA) in wheat. J. Exp. Bot. 53: 1569-1574.

Himi E., Maekawa M., Miura H., Noda K. (2011) Development of PCR markers for *Tamyb10* related to R-1, red grain color gene in wheat. Theoretical and Applied Genetics 122: 1561-1576.

Himi E., Nisar A., Noda K. (2005) Colour genes (*R* and *Rc*) for grain and coleoptile upregulate flavonoid biosynthesis genes in wheat. Genome 48:747-754.

Himi E. and Noda K. (2005) Red grain colour gene (*R*) of wheat is a Myb-type transcription factor. Euphytica 143:239-242.

Himi E., Yamashita Y., Haruyama N., Yanagisawa T., Maekawa M., Taketa S. (2012) *Ant28* gene for proanthocyanidin synthesis encoding the R2R3 MYB domain protein (Hvmyb10) highly affects grain dormancy in barley. Euphytica 188:141-151.

Humphreys, D.G. and J. Noll. 2002. Methods for characterization of preharvest sprouting resistance in a wheat breeding program. Euphytica. 126:61-65.

Imtiaz M., Ogbonnaya F.C., Oman J., van Ginkel M. (2008) Characterization of Quantitative Trait Loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. Genetics 178:1725-1736.

Jia, L.G., Wu, Q.Y., Ye, N.H., Liu, R., Shi, L., Xu, W.F., Zhi, H., Anmr B.R., Xia, Y.J., Zhang, J.H. (2012) Proanthocyanidins inhibit seed germination by maintaining a high level of abscisic acid in *Arabidopsis thaliana*. Journal of Integrative Plant Biology 54(9): 663-673.

Joosen, R.V.L., Arends, D., Li, Y., Willems, L. A.J., Keurentjes, J.J.B., Ligterink, W., Jansen, R.C., Hilhorst, H.W.M. (2013) Identifying genotype-by-environment interactions in the metabolism of germinating Arabidopsis seeds using generalized genetical genomics. Plant physiology 162(2): 553-566.

King R.W. (1976) Abscisic acid in developing wheat grains and its relationship to grain growth and maturation. Planta 132:43-51.

King R.W, Richards R.A. (1984) Water uptake in relation to pre-harvest sprouting damage in wheat: ear characteristics. Aust. J. Argic. Res. 35:327-336.

Kettlewell P.S. (1999) The response of alpha-amylase activity during wheat grain development to nitrogen fertiliser. Annals of Applied Biology 134:241-249.

Knott, C.A., Van Sanford, D.A., Souza, E.J. (2008) Comparison of selection methods for the development of white-seeded lines from red × white soft winter wheat crosses. Crop Sciences 48(5): 1807-1816.

Koornneef M., Bentsink L., Hilhorst H. (2002) Seed dormancy and germination. Current Opinion in Plant Biology 5:33-36.

Koornneef M., Reuling G., Karssen C.M. (1984) The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. Physiologia Plantarum 61:377-383.

Kruger J.E. (1989) Biochemistry of preharvest sprouting in cereals. In: N. F. Derera (Ed.) Preharvest field sprouting in cereals., CRC Press Inc., Boca Raton, Florida. p. 61-84.

Kulwal P., Ishikawa G., Benscher D., Feng Z.Y., Yu L.X., Jadhav A., Mehetre S., Sorrells M.E. (2012) Association mapping for pre-harvest sprouting resistance in white winter wheat. Theoretical and Applied Genetics 125:793-805.

Kulwal P., Kumar N., Gaur A., Khurana P., Khurana J.P., Tyagi A.K., Balyan H.S., Gupta P.K. (2005) Mapping of a major QTL for pre-harvest sprouting tolerance on Chromosome 3A in bread wheat. Theoretical and Applied Genetics 111:1052-1059.

Lenoir C., Corbineau F., Come D. (1986) Barley (*Hordeum vulgare*) seed dormancy as related to glumella characteristics. Physiologia Plantarum 68:301-307.

Liu S.B., Cai S.B., Graybosch R., Chen C.X., Bai G.H. (2008) Quantitative trait loci for resistance to pre-harvest sprouting in US hard white winter wheat Rio Blanco. Theoretical and Applied Genetics 117:691-699.

Liu S., Sehgal S.K., Li J., Lin M., Trick H.N., Yu J., Gill B.S., Bai G. (2013) Cloning and characterization of a critical regulator for pre-harvest sprouting in wheat. Genetics 195:263-273.

Lunn G.D., Major B.J., Kettlewell P.S., Scott R.K. (2001) Mechanisms leading to excess alpha-amylase activity in wheat (*Triticum aestivum* L) grain in the UK. J Cereal Sci. 33: 313-329.

Marles M.A.S., Ray H., Gruber M.Y. (2003) New perspectives on proanthocyanidin biochemistry and molecular regulation. Phytochemistry 64, 367-383

Marles, M. S., Gruber, M. Y. (2004) Histochemical characterisation of unextractable seed coat pigments and quantification of extractable lignin in the Brassicaceae. J. Sci. Food Agric., 84: 251-262.

Mares D., Mrva K., Cheong J., Williams K., Watson B., Storlie E, Sutherland M, Zou Y. (2005) A QTL located on Chromosome 4A associated with dormancy in white- and red-grained wheats of diverse origin. Theoretical and Applied Genetics 111: 1357–1364.

Masojć P., Milczarski P. (2005) Mapping QTLs for alpha-amylase activity in rye grain. Journal of Applied Genetics 46:115-123.

Masojć P., Wisniewska M., Lan A., Milczarski P., Berdzik M., Pedziwiatr D., Pol-Szyszko M., Galeza M., Owsianicki R. (2011) Genomic architecture of alpha-amylase activity in mature rye grain relative to that of preharvest sprouting. Journal of Applied Genetics 52:153-160.

Masojć P., Milczarski P. (2009) Relationship between QTLs for preharvest sprouting and alpha-amylase activity in rye grain. Molecular Breeding 23:75-84.

Matus-Cadiz M.A., Hucl P., Perron C.E., Tyler R.T. (2003) Genotype X environment interaction for grain color in hard white spring wheat. Crop Science 43:219-226.

McCleary, B.V., Sheehan H. (1987) Measurement of cereal alpha-amylase: a new assay procedure. Journal of Cereal Science 6:237-251.

McFall, K. L., Fowler, M. E. (2009) Overview of wheat classification and trade, in Wheat Science and Trade (Carver B.F. ed), Wiley-Blackwell, Oxford, UK.

McKibbin R.S., Wilkinson M.D., Bailey P.C., Flintham J.E., Andrew L.M., Lazzeri P.A., Gale M.D., Lenton J.R., Holdsworth M.J. (2002) Transcripts of Vp-1 homeologues are misspliced in modern wheat and ancestral species. Proceedings of the National Academy of Sciences 99:10203-10208.

Meredith P., Pomeranz Y. 1985.Sprouted grain. In: Advances in cereal Science and technology, Pomeranz Y. (Ed.) Vol VII: 239-299.

Metzger, R. J., Silbaugh, B. A. (1970) Location of genes for seed coat color in hexaploid wheat, *Triticum aestivum* L. Crop Sci. 10: 495-496.

Miyamoto T., Everson E.H. (1958) Biochemical and physiological studies of wheat seed pigmentation. Agron Jour 50:733-734.

Moot D.J., Every D. (1990) A comparison of bread baking, falling number, α-amylase assay and visual method for the assessment of pre-harvest sprouting in wheat. Journal of Cereal Science 11:225-234.

Mrva K., Mares D.J. (1999) Regulation of high pI alpha-amylase synthesis in wheat aleurone by a gene(s) located on Chromosome 6B. Euphytica 109:17-23.

National Agricultural Statistics Service. (2012) http://www.nass.usda.gov/Statistics_by_Subject/index.php?sector=CROPS, Accessed on July 10, 2013

Nesi N., Jond C., Debeaujon I., Caboche M., Lepiniec L. (2001) The Arabidopsis *TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell 13: 2099-2114.

Netsvetaev V.P., Akinshina O.V., Bondarenko L.S. (2012) Genetic control of several α-amylase isozymes in winter hexaploid wheat. Russian Journal of Genetics 48:347-349.

Nilsson-Ehle, H. 1909. Einige Ergebnisse von Kreuzungen bei Hafer und Weizen. Botaniska Notiser 1:257-294.

Nyachiro J.M. (2012) Pre-harvest sprouting in cereals. Euphytica 188: 1-5.

Obroucheva NV, Antipova OV. (2000) The distinct controlling of dormancy release and germination commencement in seeds. In: Viemont J.D., Crabbe J. (Eds.) Dormancy in plants: from whole plant behaviour to cellular control. Willingford: CABI Publishing. p. 35-46.

Ohto M.A., Stone S.L., Harada J.J. (2007) Genetic control of seed development and seed mass. In: Bradford K. J., Nonogaki H. (Eds.) Annual Plant Reviews Vol. 27: Seed Development, Dormancy and Germination. Blackwell Publishing Ltd. p. 1-24.

Paterson A.H., Sorrells M.E., Obendorf R.L.(1989) Methods of evaluation for pre-harvest sprouting resistance in wheat breeding programs. Can. J. Plant Sci. 69: 681-689.

Penfield S., King J. (2009) Towards a systems biology approach to understanding seed dormancy and germination. Proceedings of the Royal Society B-Biological Sciences 276:3561-3569.

Peng J., Carol P., Richards D.E., King K.E., Cowling R.J., Murphy G.P., Harberd N.P. (1997) The Arabidopsis *GAI* gene defines a signaling pathway that negatively regulates gibberellin responses Genes & Development 11:3194-3205.

Peng J., Sun D., Nevo E. (2011) Domestication evolution, genetics and genomics in wheat. Molecular Breeding 28:281-301.

Peterson, H., Knudson W.A., Abate G. 2006. The economic impact and potential of Michigan's Agri-food system. The Strategic Marketing Institute Working Paper. Department of Ag. Economics, Michigan State University.

Perten H. (1964) Application of the falling number method for evaluating alpha-amylase activity. Cereal Chemistry 41:127-139.

Rasul G., Humphreys D.G., Brule-Babel A., McCartney C.A., Knox R.E., DePauw R.M., Somers D.J. (2009) Mapping QTLs for pre-harvest sprouting traits in the spring wheat cross 'RL4452/AC Domain'. Euphytica 168:363-378.

Rasul G., Humphreys D.G., Wu J.X., Brule-Babel A., Fofana B., Glover K.D. (2012) Evaluation of preharvest sprouting traits in a collection of spring wheat germplasm using genotype and genotype environment interaction model. Plant Breeding 131:244-251.

Reitan, L. 1980. Genetical aspects of seed dormancy in wheat related to seed coat color in wheat and the relationship to seed dormancy at maturity. Cereal Res. Comm. 8:275-276.

Schramm, E.C., Nelson, S.K., Kidwell, K.K., Steber, C.M. (2013) Increased ABA sensitivity results in higher seed dormancy in soft white spring wheat cultivar 'Zak'. Theoretical and Applied Genetics 126(3): 791-803.

Seal C.J., Brownlee I.A. (2010) Whole grains and health, evidence from observational and intervention studies. Cereal Chemistry Journal 87: 167-174.

Sears, E.R. 1944. Cytogenetic studies with polyploid species of wheat: II. Additional chromosome aberrations in *Triticum vulgare*. Genetics 29:232-246.

Skriver K., Olsen F.L., Rogers J.C., Mundy J. (1991) *cis*-acting DNA elements responsive to gibberellin and its antagonist abscisic acid. Proceedings of the National Academy of Sciences 88:7266-7270.

Silverstone AL, Jung H-S, Dill A, Kawaide H, Kamiya Y, Sun TP (2001) Repressing a repressor: gibberellin-induced rapid reduction of the RGA protein in Arabidopsis. Plant Cell 13**:** 1555-1566.

Singh, R., Matus-Cadiz M., Baga M. Hucl P., Chibbar R.N. 2008. Comparison of different methods for phenotyping preharvest sprouting in white-grained wheat. Cereal Chemistry. 85:238-242.

Steber, C. M. (2007). De-repression of seed germination by GA signaling. In: Bradford K.J., Nonogaki H. (Eds.) Annual Plant Reviews Vol. 27: Seed Development, Dormancy and Germination. Oxford, UK: Blackwell Publishing Ltd. p. 248-263.

Sugimoto K., Takeuchi Y., Ebana K., Miyao A., Hirochika H., Hara N., Ishiyama K., Kobayashi M., Ban Y., Hattori T., Yano M. (2010) Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice. Proceedings of the National Academy of Sciences 107:5792-5797.

Sun Y.W., Jones H.D., Yang Y., Dreisigacker S., Li S.M., Chen X.M., Shewry P.R., Xia L.Q. (2012) Haplotype analysis of Viviparous-1 gene in CIMMYT elite bread wheat germplasm. Euphytica 186:25-43.

Sutherland B. (2011) Wheat Market Outlook: trends & direction for Michigan. 2011 MABA Winter Outlook Conference. http://www.miagbiz.org/images/e0186601/sutherland.pdf , Accessed on July 10th, 2013.

Sun TP, Gubler F (2004) Molecular mechanism of gibberellin signaling in plants. Annu. Rev. Plant Biol. 55: 197-223.

Ullrich S.E., Clancy J.A., del Blanco I.A., Lee H., Jitkov V.A., Han F., Kleinhofs A., Matsui K. (2012) Genetic analysis of preharvest sprouting in a six-row barley cross. Molecular Breeding 21: 249-259.

Verity J.C.K., Hac L., Skerritt J.H. (1999) Development of a field enzyme-linked immunosorbent assay (ELISA) for detection of α-amylase in preharvest-sprouted wheat. Cereal Chemistry Journal 76: 673-681.

Walker-Simmons M. (1987) ABA levels and sensitivity in developing wheat embryos of sprouting resistant and susceptible cultivars. Plant Physiol. 84: 61-66.

Winkel-Shirley B. (2001) Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiology: 126(2): 485-493.

Wu J.M., Carver B.F., Goad C.L. (1999) Kernel color variability of hard white and hard red winter wheat. Crop Science 39:634-638.

Yang R.C., Ham B.J. (2012) Stability of genome-wide QTL effects on malt α-amylase activity in a barley doubled-haploid population. Euphytica 188:131-139.

Yu C. (2012) Evaluation of alpha-amylase activity and falling number for soft white and soft red wheat varieties adapted to Michigan. MS Thesis, Michigan State University (Publication No. UMI 1510084)

Zanetti S., Winzeler M., Keller M., Keller B., Messmer M. (2000) Genetic analysis of pre-harvest sprouting resistance in a wheat x spelt cross. Crop Science 40:1406-1417.

CHAPTER 2 QTL analysis for seed color components and α-amylase activity in a spring wheat recombinant inbred line population segregating for seed color

Abstract

Pre-harvest sprouting (PHS) in wheat (*Triticum aestivum* L.) is precocious grain germination induced by prolonged wet conditions during the harvest period. Sprouted kernels significantly downgrade flour quality and cause significant financial losses to farmers and downstream processors. It is commonly known that red wheat is more resistant to PHS than white wheat. The objective of this study is to 1) Assess the allelic contribution of seed coat color loci on Chromosome Group 3 to seed coat color and α-amylase activity, and 2) Identify QTL for these phenotypes using a high-throughput genotyping platform. An F_5-derived recombinant inbred line population of 165 individuals from spring wheat cross 'Vida' × MTHW0471 was used as the mapping population. A 9K single nucleotide polymorphism (SNP) array was used to construct a high-density genetic map with 1,692 SNPs and three simple sequence repeats (SSR). Strong correlations were found between color components and α-amylase activity within and across environments while a strong year effect was found in α-amylase activity from 2010 to 2012. The different combinations of color alleles and the dosage of dominant alleles at color loci significantly affected both the seed color components (b, L^*) and α-amylase activity. The dominant color allele on Chromosomes 3A and 3B showed a significant effect in reducing α-amylase activity. A total of 38 quantitative trait loci (QTL) were identified on eleven chromosomes (1B, 2A, 2B, 3A, 3B, 3D, 4B, 5A, 5D, 6B and 7B) from three years with two post-harvest treatments in this population. Most QTL explained 6-15% of the phenotypic variation while a major QTL on Chromosome 2B explained up to 37.6% of phenotypic variation of α-amylase activity in 2012 non-misted conditions. A total of 21 significant QTL × QTL interactions were also found in both the color and PHS resistance traits while Chromosome Group 3 were identified as a hotspot for QTL × QTL interactions. In conclusion, the population

segregating for color allele is critical for the evaluation of allelic contribution to phenotype. QTL identified in current study showed a pleiotropic effect between seed color loci and α-amylase activity.

Introduction

Wheat (*Triticum aestivum* L.) is a staple crop, first domesticated 9,000 years ago (Peng et al., 2011). Wheat provides 20% of calories consumed for more than 40% of the world's population (Gill et al. 2004). In the US, wheat ranks third among field crops in both planted acreage and gross farm receipts (National Agricultural Statistics Service, 2012). Wheat is categorized as red wheat and white wheat based on seed coat color. In recent years, industry demand for white wheat has increased rapidly due to the popularity of whole grain products and white wheat is the preferred material due to its higher flour yield and better end-use quality. However, due to the susceptibility to pre-harvest sprouting, white wheat is not favored for commercial production. Sprouted wheat has an extensive amount of α-amylase, which degrades the endosperm reserve and severely limits grain end-use. Growers receive discounted price for sprouted wheat while food processors also need to make adjustments when processing with batches contain sprouted wheat. The direct annual losses caused by PHS can reach up to US $1 billion worldwide (Black et al., 2006).

Pre-harvest sprouting (PHS) is the precocious seed germination induced by prolonged wet conditions during harvest season. It has been documented as a worldwide issue since the 1970s (Derera, 1989). Red wheat is known to be categorically more resistant to PHS than white wheat (Gale and Lenton, 1987). However, the debate of either the relationship is a causal pleotropic or a close linkage between red allele and seed dormancy still await further study (Gale, 1989). Recent evidence showed that the red gene was a *Myb* type transcription factor in

flavonoid biosynthesis pathway (Himi and Noda, 2005), but the mapping resolution is still not enough to differentiate the R genes from surrounding genes.

Both phenotypic evaluation and QTL analysis has been done on PHS related traits. Phenotypic evaluations can be divided into two categories: measurement of single seed dormancy after regular harvest, and measurement of in-head sprouting resistance at physiological maturity (Bewley, 1997; Mares et al., 2005). The latter category is done by the measurement of sprout damage mainly caused by α-amylase. Therefore, a direct measure of α-amylase activity or an indirect measure of flour viscosity through falling number test are commonly used (Rasul et al., 2009). QTL studies have been done with germination index, sprout index, α-amylase activity, and falling number test to represent PHS resistance level in bi-parental populations and QTL were identified from different genetic background on all 21 chromosomes (Imtiaz et al. 2008; Rasul et al. 2009; Kulwal et al. 2012). Recently, association mapping has been adapted to identify genetic factors providing PHS resistance specifically within white wheat germplasm (Kulwal et al., 2012). All these studies identified PHS related QTL in various regions of wheat genome, which might offer us a way to pyramid different PHS resistance resources in the near future.

Seed coat color is known to be closely related to seed dormancy and red wheat is generally more resistant in PHS than white wheat. Proanthocynidin (PA) is the red pigment in wheat seed coat (Himi et al., 2005). Studies in Arabidopsis found that exogenous PA enhances abscisic acid (ABA) biosynthesis, which inhibits germination, and PA-deficient mutants had reduced dormancy (Himi et al., 2005; Winkel-Shirley, 2001). Grain color is mainly controlled by three loci on Chromosome Group 3. Red is dominant to white and the degree of redness can be affected by environment and management practices. However, the dosage effect of the three

alleles has not been examined before, which may be due to the availability of allelic specific markers.

In this study, a recombinant inbred line population derived from a red by white spring wheat cross was used to examine the dosage effect of specific color allele and their combinations to seed coat color and α-amylase activity. QTL identification was conducted with a high-density linkage map and QTL by QTL interaction were explored. This study provided us more details for dissecting the relationship between seed coat color and PHS resistance.

Materials and Methods

Plant Materials

A recombinant inbred line (RIL) population consisting of 165 lines, kindly shared by Dr. Jamie Sherman at Montana State University, was developed from a cross between two elite spring wheat lines that segregated for seed coat color homeologs on Chromosome Group 3. The female parent, 'Vida', is a hard red spring wheat cultivar with all three color homeologs dominant. The male parent, MTHW0471, is an elite hard white spring wheat breeding line with all three color homeologs recessive. F_2 population was genotyped with simple sequence repeats (SSR) markers closely linked to the color loci. F_2 individuals that were homozygous (dominant or recessive) at all three seed color loci were selected and inbred until F_5 using single-seed-descent (Sherman et al., 2008). Seed increase was done in the greenhouse in 2009. Field evaluations were conducted on $F_{5:6}$, $F_{5:7}$ and $F_{5:8}$ seeds in 2010, 2011, and 2012, respectively.

Field Design and Greenhouse Misting

Field trials were conducted in three consecutive years (2010, 2011, and 2012) at Michigan State University Mason Research Farm, Mason, MI. Soil type is sandy loam. Within each experimental field, the RIL population was planted in 1.5-meter twin rows (2010 was single

row) with 1.2-meter alley in a randomized complete block design with three replications, each contains 165 RIL and two parents. Growth-related dates and precipitation information are summarized in Table A.1. Physiological maturity was visually determined by the date when 80% heads within the plot were loss-green at peduncle. Due to differences in maturity, lines were hand-harvested individually at 3 days after physiological maturity. The wheat heads with its 20 cm stem were transported to the greenhouse for artificial misting treatments.

Harvested wheat heads were divided into two subsamples, one mist group, the other non-mist group. Two groups were kept in the same greenhouse with a temperature ranging from 25-30°C, and processed in an identical manner except for the misting treatment received. The plants were arranged in plastic tubes (5.3 cm in diameter and 25 cm in depth), which were placed upright in racks to simulate PHS field conditions. The misted group was placed on the greenhouse bench under a misting system made up of brass seedling nozzles (1.4 meter interval) attached to the pipelines 1.5 meter above the bench. The misting system was controlled by a timer (Model 15079, General Electric), which set to mist for 45 min every 6 h. Both groups were collected after 48 hr misting and stored at -20 °C freezer prior to freeze-drying. The freeze-dried spikes were hand-threshed and stored at -20 °C to keep enzyme activity.

Seed coat color measurement

Grain color of each recombinant inbred line and two parents, from the mist and non-mist groups, was measured using a chroma meter (Knoica Minolta, CR-400) with a 20 g grain sample. This equipment decomposes color into a 3-dimension color space, L^*a^*b (Smith and Guild, 1931). 'L^*' measures black (0) to white (100), 'a^*' measures green when negative and red when positive, and 'b' measures blue when negative and yellow when positive. The measurement was done by pressing the chroma meter onto the cover of a 60 × 15 mm petri plate (Falcon 351007)

filled with grains from one genotype. Then three readings were taken from different regions of the plate and averaged to represent each sample. Three biological replicates were measured and averaged to represent that genotype.

Determination of α-amylase activity

Whole grain flour was milled from each RIL and the two parents using a UDY Laboratory Mill with a 0.5 mm sieve. Flours were analyzed for α-amylase activity based on American Association of Cereal Chemists International (AACCI) approved method 22-02.01, Ceralpha method (AACC International, 2002). The enzyme extraction and assay was performed using the Megazyme kit (K-CERA 08/05, Megazyme International Ltd., Ireland) with a modified protocol established by the USDA Soft White Wheat Quality Lab, Wooster, Ohio and was described in details below (Dr. Edward Souza, *pers. comm.*).

Three grams of whole grain flour and 20 mL of extraction buffer (1 M sodium malate, 1 M sodium chloride, 40 mM calcium chloride, 0.1% sodium azide , provided by Megazyme kit) were added into a 50 mL centrifuge tube followed by vigorous shaking and incubated at 42°C in a water bath (Fisher Scientific, Pittsburgh, PA) for 20 minutes. Samples were taken out of the water bath and were shaken vigorously every 5 min. The samples were then centrifuged at 3100 rpm for 10 minutes (SORVALL RT7, Kendro Laboratory Products, Newton, CT). At the same time, twenty microliter (μL) aliquots of Ceralpha substrate (non-reducing-end blocked pnitrophenyl maltoheptaoside, BPNPG7, provided by Megazyme kit) solution were dispensed into each well of a 96-well plate based on sample layout and pre-incubated at 42°C for 5 minutes. After centrifuge, twenty microliter α-amylase extract from the supernatant of each sample was added to the bottom of each well and mixed with the pre-heated Ceralpha substrate. The enzyme reaction took place for precisely 20 minutes at 42°C. For processing efficiency, three replicates

per sample were assigned to the same column next to each other on a 96-well plate and a 30-second interval was kept between each triplicate aliquots pipetted. A 300 µL Stopping Reagent (1% (w/v) Trizma base, provided by Megazyme kit), was added to the well at the end of the 20-minute period for each triplicate. Sample controls were prepared by adding the substrate after the stopping reagent while plate controls were 340 µL of distilled water. The absorbance of each well was read using a spectrophotometer (Synergy HT Multi-Mode Microplate Reader, BioTek, Winooski, VT) at 400 nm. If the absorbance value was greater than 1.2, the enzyme extract was diluted to a proper range due to the possible saturation of limited amount substrate and a second run for the sample was done. The standard curve was generated based on the dilution series of pnitrophenol standard solution (Sigma, 104-1) in 1% tri-sodium phosphate. The extinction coefficient (E_{mM} = 14.1) was obtained following instructions in Megazyme manual. The α-amylase activity (Unit/g) was then calculated based on Formula 2.1.

$$\frac{\text{Absorbance (reaction-blank)}}{\text{Incubation Time}} \times \frac{\text{Total Volume in Cell}}{\text{Aliquot Assayed}} \times \frac{1}{E_{mM}} \times \frac{\text{Extraction Volume}}{\text{Sample Weight}} \quad (2.1)$$

Statistical analyses

Data was analyzed using SAS statistical software 9.2 (SAS Institute, Cary, NC). Due to the large variation within the population, α-amylase activity was transformed using Box-Cox transformation with SAS procedure, PROC TRANSREG, recorded as EnzBC, to restore residual normality (Box and Cox, 1964). Pearson's correlations were calculated for phenotypes within each environment and between environments. Significance level was chosen (0.05 or smaller) and p-value was adjusted by Tukey's procedure.

For seed color components (L^*, a^*, b) and transformed α-amylase activity (EnzBC), analysis of variance (ANOVA) was carried out using PROC MIXED to check the effect of fixed

factors (genotype group, dosage group, allele group, treatment) and to calculate least square means between different genotype groups, dosage groups or allele groups. The genotype groups of the RIL population were provided by Dr. Sherman (Table 2.1) and dosage groups were generated by combining genotype groups with same number of dominant red alleles: 0R (group 1), 2R (group 2, 3, 4), 4R (group 5, 6, 7), 6R (group 8). For each allele (R-3A, R-3B, R-3D), allele groups divided the population into two groups, one contains the specific allele as dominant, the other contains the same allele as recessive.

Within each environment (2010 mist; 2011 mist, non-mist; 2012 mist, non-mist), contrasts were established between either dosage groups or allele groups to measure the dosage effects or effects of specific allele on color components and α-amylase activity by comparing. Pairwise comparisons were adjusted with Tukey's procedure. Dunnett's tests were also used to compare the significance of difference between selected groups and 0R (white) or 6R (red) groups. Mist treatment effects on α-amylase activity were measured by comparing groups having the same R allele combinations between mist and non-mist conditions. Results across environments were recorded and compared over years to see if there were Genotype × Year interactions.

DNA isolation and SNP genotyping

DNA from the parents (two replicates each) and 165 RILs were extracted from young leaf tissues of greenhouse plants using Wizard Genomic DNA Purification Kit (Promega, A1120), quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, San Diego, CA) and adjusted to a concentration of 50 ng/ul before SNP genotyping. SNP genotyping was conducted on an Illumina iScan Reader utilizing the Infinium HD Assay Ultra (Illumina, Inc., San Diego, CA) and the Infinium 9k Wheat array (Cavanagh et al., 2013).

Table 2.1 Mapping population used in this study with red homeolog combinations determined by SSR marker linked to color loci

Group	Number of individuals*	Grain color	Red dosage	R-3A#	R-3B	R-3D
Parents						
'Vida'	1	Red	6	b	b	b
MTHW0471	1	White	0	a	a	a
RILs						
1	16	White	0	a	a	a
2	24	Red	2	b	a	a
3	18	Red	2	a	b	a
4	16	Red	2	a	a	b
5	20	Red	4	b	b	a
6	22	Red	4	b	a	b
7	25	Red	4	a	b	b
8	24	Red	6	b	b	b

* These are the individuals having full set of genotyping and phenotyping data and used for QTL analysis;

Upper case letter represents the chromosome where the SSR marker is located and lower case letter were named following the recommended rules for gene symbolization in wheat to represent dominant (b) and recessive (a) color alleles according to McIntosh et al. (2008).

SNP calling, filtering, map construction

Raw SNP genotypes obtained from Illumina GenomeStudio software (Illumina, San Diego, CA) were first categorized based on the custom cluster file to filter out low quality SNP calls (Cavanagh et al., 2013). Prior to mapping, SNPs were further filtered to remove non-informative SNPs or SNPs with a call rate less than 90% (greater than 15 progeny with missing genotypes) were removed (Table A.2). The three SSR markers linked to color loci (Sherman et al. 2008) were added to the 1692 SNP markers for map construction. The linkage analysis was performed using Joinmap 4.1 (Van Ooijen 2006). The genotype data was coded as RIL population type. Markers were first grouped into linkage groups based on a minimum logarithm of odds (LOD) score of three. The grouping result was validated by custom defined chromosome assignment for each SNP (Cavanagh et al., 2013). Then each linkage group was ordered by regression method with Kosambi function.

QTL mapping for color measurements and α-amylase activity

QTL mapping was conducted by MapQTL6 (Van Ooijen 2004) using Multiple QTL Mapping (MQM) method and validated by WinQTL Cartographer,V2.5 (Wang et al. 2008) with composite interval mapping (CIM) using standard model Zampqtl6. A permutation test of 1000 runs was conducted to identify genome-wide threshold of LOD at 5% significance level for declaring significant QTLs (Doerge and Chuchill 1996). QTL by QTL interaction (additive effects only) were explored with Multiple Interval Mapping (MIM) method using WinQTL Cartographer V2.5. Broad sense heritability was obtained from MIM results in WinQTL Cartographer V2.5.

Results

Map construction with Infinium 9k wheat SNP array

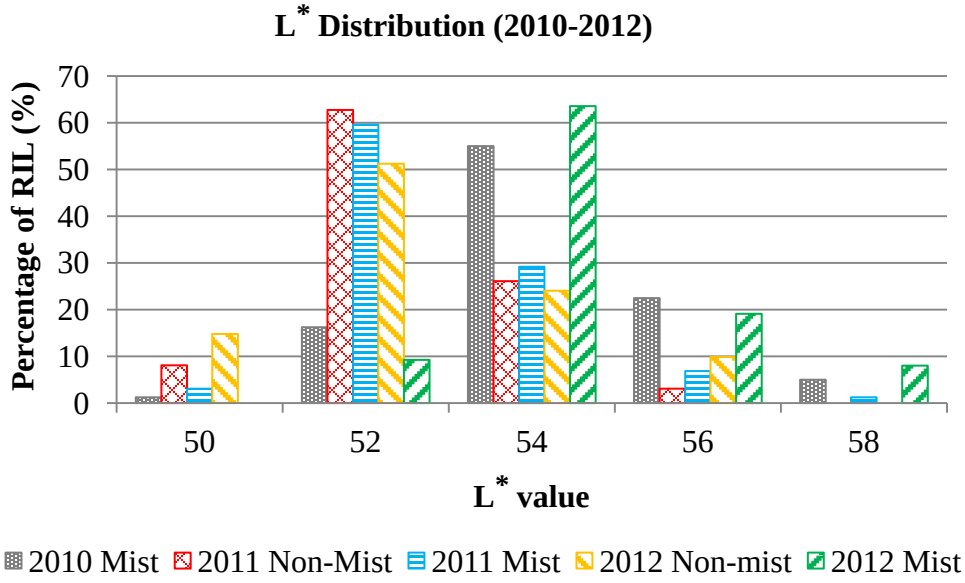
A genetic map, based upon 1692 SNP markers and three SSR markers, was constructed for all 21 linkage groups with a total map size of 1992.6 cM (Table 2.2; Table A.3). The chromosome assignment of mapped SNPs was based on the consensus map released by the 9K SNP array design team (Cacanagh et al. 2013). The A genome had the largest subgenome size (931.5 cM) out of the three sub-genomes with a relatively even distribution of chromosome map size and marker density. The B genome had a linkage map size of 749 cM while Chromosome 3B was potentially under-represented with 26 markers spanning on a 42.1cM map. On average, the A genome had a SNP density of 0.95 SNPs per cM, while B genome had an average of 1.01 SNPs per cM. Both A and B genomes were significantly better represented than D genome, which had the smallest linkage map size (312.1 cM) and fewest markers (81). However, due to a lack of linkage, all subgenomes had some linkage groups were segmented. However, with the chromosome assignment information released by the 9k SNP chip design team, we were able to retain all the linkage groups for later QTL study.

Phenotype distribution of seed color components and α-amylase activity

Based on the SSR markers linked to the red seed color locus, the population was further divided into eight genotype groups (χ^2= 0.7145) according to the combination of three homeologs (Table 2.1). Figure 2.1 shows the distribution of color components (L^*, a^*, b) and Box-Cox transformed α-amylase activity (EnzBC) for three years with two post-harvest

Table 2.2 Summary of linkage map constructed with 1695 markers for ‘Vida’ × MTHW0471 population

Chromosome Group	A genome		B genome		D genome	
	SNP	cM	SNP	cM	SNP	cM
1	134	128.4	109	140.7	17	75.3
2	97	123.7	156	113.7	15	81.5
3	133	124.1	26	42.1	9	33.5
4	141	104.3	59	114	9	39.5
5	132	148.1	214	153.6	12	41.5
6	87	127	130	83.2	4	0.3
7	127	175.9	69	101.7	15	40.5
Sum	851	931.5	763	749	81	312.1



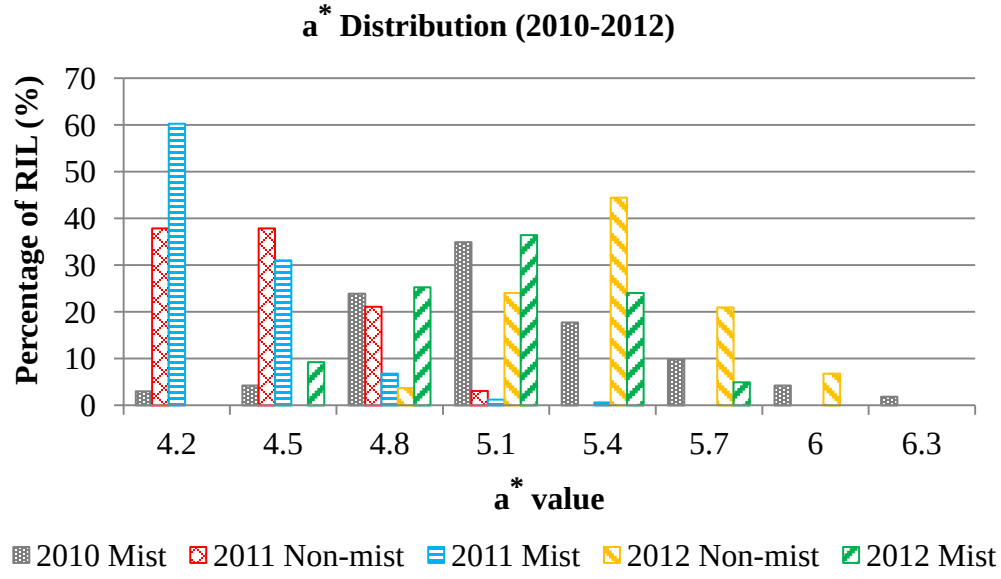
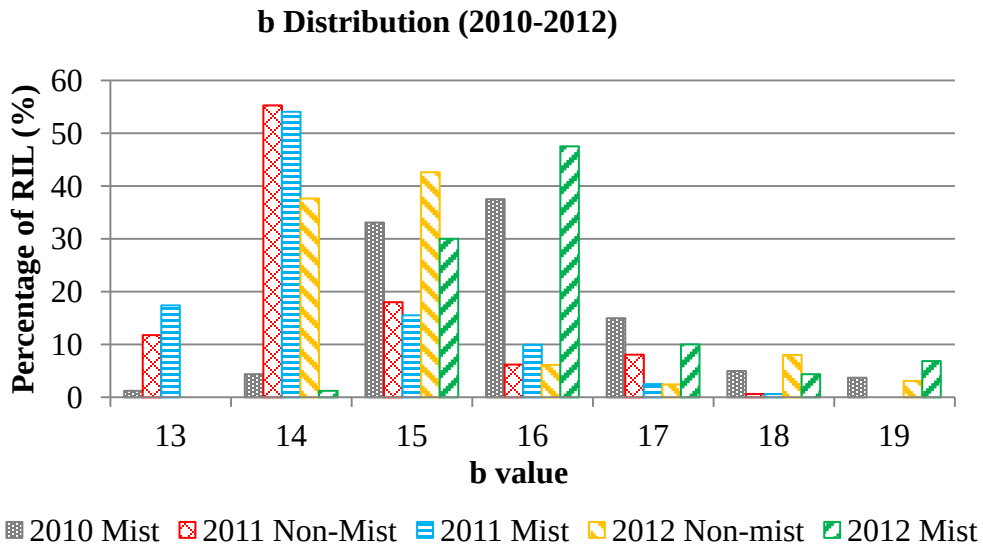
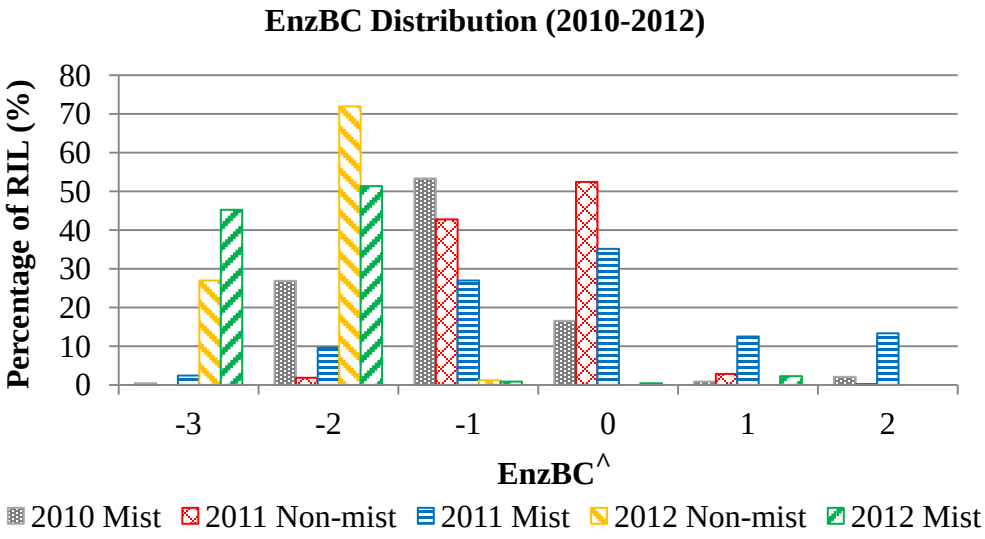


Figure 2.1 Distribution of color components and α-amylase activity for 'Vida' × MTHW0471 population from 2010 to 2012.
For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.
^ EnzBC is the Box-Cox transformed value of α-amylase activity.

Figure 2.1 (cont'd)





treatments. All the traits are quantitatively distributed and the year effect was significant to affect the value of all three color components and α-amylase activity ($p< 0.05$). Moreover, across the three years, a trend was shown that the more precipitation the plant received during the harvest period, the higher α-amylase activity of the population would have for that year (Table A.1). This trend corresponds well with the fact that in a humid year, the plants are more susceptible to PHS damage.

Correlation between seed coat color and α-amylase activity

Color measurements and α-amylase activity were conducted on both mist and non-mist material and correlations of these traits within each environment were summarized in Table 2.3. For all five conditions, the highest correlation were identified between L^* and b, ranging from 78.1-85.8%. Both L^* and a^* were highly correlated with EnzBC in four out of five conditions. While the white color, represented by L^*, was positively correlated with the α-amylase activity; red color, represented by a^*, was in negative correlation ($p < 0.01$). On the other hand, when considering a single trait across environments, EnzBC in 2012 mist condition was significantly correlated with EnzBC from all other environments, while strong correlations between mist and non-mist materials were also found between traits across environments (Table A.4).

Genotypic effects on color components and α-amylase activity

Within each environment, genotypic effects on color components and α-amylase activity were measured at genotype group, dosage group, and allele group levels. Table 2.4 summarized the ANOVA results for all five conditions. The genotype group, or the eight different allele combinations, showed a significant difference ($p<0.001$) for seed whiteness (L*) and yellowness

Table 2.3 Phenotypic correlations of traits within each environment

2010 Mist[^]	L*	a*	b				
a*	0.251**						
b	0.855***	0.450***					
EnzBC[#]	0.460***	-0.092	0.449***				
2011 Mist	L*	a*	b	2011 Non-mist	L*	a*	b
a*	-0.079			a*	-0.082		
b	0.840***	0.125		b	0.781***	0.326***	
EnzBC	0.422***	-0.398***	0.539***	EnzBC	0.291**	-0.256**	0.299**
2012 Mist	L*	a*	b	2012 Non-mist	L*	a*	b
a*	-0.100			a*	-0.181		
b	0.858***	0.280**		b	0.856***	0.229	
EnzBC	0.280**	-0.402***	0.132	EnzBC	0.234	-0.301***	0.052

[^] Mist/Non-mist represents the treatment each group received; 2010/2011/2012 represented the year experiment conducted;

[#] EnzBC is the Box-Cox transformed value of α-amylase activity;

, * significant at 0.01 level and 0.001 level respectively, all the p-values were Tukey-adjusted.

Table 2.4 Analysis of variance of color components and α-amylase activity evaluated in three years with two post-harvest treatments in Mason, MI for 'Vida'× MTHW0471population

	Source	Environment				
		2010 Mist	2011 Non-mist	2011 Mist	2012 Non-mist	2012 Mist
L^*	Genotype group	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Dosage group	0.028	<0.0001	<0.0001	<0.0001	<0.0001
	R-3A	<0.01	<0.01	<0.01	<0.01	<0.01
	R-3B	<0.01	<0.01	<0.01	<0.01	<0.01
	R-3D	<0.01	<0.01	<0.01	<0.01	<0.01
a^*	Genotype group	NS^	NS	0.038	0.002	NS
	Dosage group	NS	0.043	0.028	<0.0001	NS
	R-3A	NS	NS	NS	<0.05	NS
	R-3B	<0.05	<0.05	NS	<0.05	NS
	R-3D	NS	NS	NS	<0.05	NS
b	Genotype group	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Dosage group	0.006	<0.0001	<0.0001	<0.0001	<0.0001
	R-3A	<0.01	<0.01	<0.01	<0.01	<0.01
	R-3B	<0.01	<0.01	<0.01	<0.01	<0.01
	R-3D	<0.01	<0.01	<0.01	<0.01	<0.01
$EnzBC^{\#}$	Genotype group	<0.0001	<0.0001	<0.0001	<0.0001	0.003
	Dosage group	<0.0001	<0.0001	<0.0001	NS	0.016
	R-3A	<0.01	<0.01	<0.01	<0.01	<0.01
	R-3B	<0.01	<0.01	<0.01	<0.01	<0.01
	R-3D	<0.01	NS	<0.01	NS	NS

^ NS: non-significant, $p > 0.05$;

EnzBC: the Box-Cox transformed value of α-amylase activity.

(b), while seed redness (a^*) only differed in two out of five conditions ($p<0.05$). At dosage group level, L^* and b, again showed significant differences for all five conditions; while color component a^* was significantly impacted in three out of five conditions. For each specific allele (R-3A, R-3B, R-3D), when comparing individuals containing a specific allele as dominant and individuals containing that same allele as recessive, both whiteness (L^*) and yellowness (b) of the seed coat showed significant differences for all three alleles. For grain redness (a^*), groups contained R-3B as dominant showed a significant difference with groups containing R-3B as recessive for three out of five conditions. While the R-3A and R-3D allele only showed significant effect in 2012 non-mist condition.

For α-amylase activity, the genotypic effect was significant for all environments at genotype group level ($p < 0.01$). At dosage group level, the different dosage can significantly impact the α-amylase activity in all environments except for 2012 mist condition ($p < 0.02$). At allele group level, a dominant red allele at R-3A or R-3B can significantly reduce α-amylase activity when compared with a recessive allele at the same loci in all environments, while R-3D red allele were only effective in 2010 and 2011 mist conditions in terms of reducing α-amylase activity when compared between individuals contained R-3D allele as dominant versus recessive.

On the other hand, α-amylase activity was significantly different between white wheat (0R) and red wheat (2R, 4R, 6R) except for 2012 non-mist environment, while within red wheat groups, the dosage difference didn't show a significant effect except for 2011 mist condition, which significant differences for all different dosages (Table 2.5).

Table 2.5 Least square means of α-amylase activity for different dosage groups across different environments

EnzBC*	2010 Mist#	2011 Mist	2011 Non-mist	2012 Mist	2012 Non-mist
0R	-0.536a^	0.722a	-0.611a	-2.742a	-2.772a
2R	-1.593b	-0.855b	-0.949b	-3.001b	-2.761a
4R	-1.668b	-1.297c	-0.943b	-3.045b	-2.783a
6R	-1.676b	-2.017d	-0.905b	-3.035b	-2.685a

* EnzBC is the Box-Cox transformed value of α-amylase activity;

Mist/Non-mist represents the Mist or Non-mist treatment each group received; 2010, 2011, 2012 represents the year the experiment conducted;

^ LS Means followed by different letters are significantly different according to Fisher's protected LSD ($p = 0.05$).

QTL analysis for seed color measurements and α-amylase activity

A total of 38 QTL were identified from three years with two post-harvest treatments for color components and α-amylase activity using MQM method in MapQTL6. All the significant QTL were confirmed by CIM and MIM methods in WinQTLCartographer V2.5 (Table 2.6).

Three QTL on Chromosome 3A, 3B and 3D were co-located with the three SSR markers representing color loci. They were shown consistently in multiple environments for different color components and α-amylase activity traits which indicated a potential pleiotropic effects. The QTL on Chromosome 3A explains 11-13.4% of phenotypic variation for L^*, 12.9-15.1% for b, and 10.4-14.7% for EnzBC, respectively. However, it was not identified as a significant QTL for a^* in any environment. The QTL on Chromosome 3B was always shown together with QTL on 3A while it explains a range of 8.0-8.8% for L^* and 9.5-16.0% for b.

The QTL directly related to grain redness, a^*, were only detected in 2012. There were three QTL on Chromosome 5A, 5D, and 6B that were identified in both mist and non-mist conditions with similar QTL effects, while a QTL on Chromosome 2A was only detected in non-mist environment. When all the QTL related to a^* were added together, they can explained a total of 24.6 and 33.8% of phenotypic variation in non-mist and mist conditions, respectively.

The QTL related to α-amylase activity (EnzBC) were detected in all five conditions. In 2012 non-mist condition, one major QTL on 2B can explains up to 37.6% phenotypic variation for α-amylase activity. In most environments, except for 2011 mist condition, multiple QTL were identified for EnzBC in each environment. When added together, these QTL explain 32.2-61.3% of phenotypic variation within a specific environment. Furthermore, all the QTL

Table 2.6 Summary of QTL identified in 'Vida' × MTHW0471 from 2010 to 2012

Environment	Trait	Chromosome	Location (cM)	LOD	PT[#]	Variance %	Additive effects
2010	L[*]	3A2[^]	0.0	5.2	3.4	13.2	-0.57
Mist	b	3A2	0.0	5.8	3.2	13.2	-0.47
		3B3	11.3	4.2		9.5	-0.40
	EnzBC	1B	41.6	6.9	3.2	17.5	-0.57
		3A2	0.0	5.7		14.7	-0.52
2011	L[*]	3A2	0.0	5.0	3.4	12.2	-0.45
Non-mist	b	3A2	2.0	7.1	3.3	15.1	-0.43
		3B3	3.7	6.4		13.6	-0.39
	EnzBC	1B	41.2	9.5	3.2	23.8	-0.25
		2B1	73.1	4.2		10.8	0.17
2011	L[*]	3A2	0.0	5.5	3.3	13.4	-0.47
Mist		3B3	3.7	3.4		8.0	-0.37
	b	3A2	0.0	6.3	3.3	13.5	-0.36
		3B3	3.7	5.6		11.8	-0.34
	EnzBC	3A2	0.0	3.8	3.3	10.4	-0.42
2012	L[*]	3A2	0.0	5.1	3.3	12.5	-0.56
Non-mist	a[*]	5A1	46.6	5.9	3.2	11.9	0.10
		5D2	0.0	4.1		8.0	-0.08
		2A	103.6	3.8		7.4	-0.08
		6B1	62.8	3.3		6.5	0.07
	b	3A2	0.0	7.6	3.2	14.1	-0.53
		3B3	3.7	5.5		10.0	-0.42
		3D1	25.5	4.3		7.5	-0.47
		1B	41.2	3.5		6.2	-0.33
	EnzBC	2B1	66.2	16.5	3.3	37.6	0.02
		4B	43.1	4.4		11.9	0.01
		7B2	33.8	4.4		11.8	-0.01
2012	L[*]	3B3	11.3	4.2	3.2	8.8	-0.38
Mist		3A2	0.0	5.2		11	-0.46
	a[*]	5D2	4.0	4.0	3.3	9.0	-0.10
		6B1	62.8	3.5		7.9	0.09
		5A1	47.6	3.4		7.7	0.09
	b	3B3	4.0	7.6	3.3	16.0	-0.50
		3A2	0.0	6.2		12.9	-0.44
		3D1	25.5	3.6		7.3	-0.40
	EnzBC	1B	40.3	3.4	3.3	9.2	-0.35
		2B1	66.2	11.2		27.4	0.63
		7B2	33.8	4.4		11.7	-0.41

Table 2.6 (cont'd)

[#] Permutation threshold, empirical likelihood of odds (LOD) threshold based on 1000 permutations at 5% significance level;

[^] The number behind letter represented the specific linkage group assigned to that chromosome based on Cavanagh et al. (2013).

were identified more than once in both mist and non-mist conditions, except for one QTL on Chromosome 4B which identified only in 2012 non-mist condition.

The QTL related to α-amylase activity (EnzBC) were detected in all five conditions. In 2012 non-mist condition, one major QTL on 2B can explains up to 37.6% phenotypic variation for α-amylase activity. In most environments, except for 2011 mist condition, multiple QTL were identified for EnzBC in each environment. When added together, these QTL explain 32.2-61.3% of phenotypic variation within a specific environment. Furthermore, all the QTL were identified more than once in both mist and non-mist conditions, except for one QTL on Chromosome 4B which identified only in 2012 non-mist condition.

All the QTL identified by CIM method were also validated by MIM method in WinQTLCartographer. In addition, novel QTL were identified by MIM also with minor effects (not listed here). Additive by additive interactions (A × A) were found for multiple traits across environments (Table 2.7). It can occur between QTL identified only in CIM, QTL identified only in MIM, or one from each. For each trait, these interactions can explain 1.5-22% of phenotypic variation. One of the hotspots containing Additive × Additive interactions is the QTL on 3A, which interacted with not only its homeolog (3B, 3D), but also QTL on other chromosomes (5B and 7A). The major QTL identified on 2B for EnzBC were also found to interact with 3B and 7B, which explains an additional 5% of the phenotypic variation. The broad-sense heritability was calculated for each trait when considering all the QTL and their interactions. The heritability for L^*, a^*, b ranged from 0.14-0.39, 0.13-0.33, 0.33-0.64, while EnzBC was only calculated in 2012 mist condition as 0.4.

Table 2.7 QTL identified by MIM having Additive × Additive interaction and the heritability of the trait

Environment	Trait	Additive × Additive interaction		Variance	$h^{2\#}$
		QTL-CIM	QTL-MIM	%	
2010 Mist	L*		Chr 5B and Chr 7D	1.5	0.14
	b	Chr^ 3A2	Chr 3D1	1.9	0.48
		Chr 3A2 and Chr 3B3		7.2	
2011 Non-mist	a*	Chr 3B3	Chr 1B	2.0	0.22
	b	Chr 3A2 and Chr 3B3		7.4	0.38
		Chr 3A2	Chr 7A1	3.0	
2011 Mist	L*	Chr 3A2 and Chr 3B3		3.1	0.29
	a*	Chr 7B2	Chr 4A1	4.2	0.13
	b	Chr 3A2	Chr 6B2	4.2	0.33
		Chr 3A2 and Chr 3B3		1.2	
2012 Non-mist	L*	Chr 3A2	Chr 7A1	2.2	0.29
	b	Chr 3A2 and Chr 3B3		9.5	
		Chr 3A2 and Chr 3D1		6.4	0.64
		Chr 3A2 and Chr 5B		6.1	
2012 Mist	L*	Chr 3A2 and Chr 3B3		2.8	
		Chr 3A2 and Chr 7A1		2.8	0.39
		Chr 7A1	Chr 3D1	0.2	
	a*	Chr 5D2 and Chr 6B1		1.9	0.33
	b	Chr 3A2 and Chr 3D1		9.7	0.39
	EnzBC	Chr 2B1 and Chr 7B2		3.6	0.4
		Chr 2B1 and Chr 3B3		1.2	

h^2 stands for broad-sense heritability of the trait;

^ Chr stands for chromosome.

Discussions

Map construction and QTL mapping with high-density markers

In this study, Infinium 9k SNP array was used to develop a high density genetic map. Due to large linkage gaps in current RIL population, more than 21 linkage groups were initially formed. Based on the consensus map released by Cavanagh et al. (2013), we were able to merge and assign groups into specific chromosomes and the orders of the markers were also validated. Due to the lack of genetic diversity in D genome (Cavanagh et al. 2013), SNP markers for D genome were under-represented in current map. Furthermore, the fact that only two QTL from D genome were mapped on the current map with a relatively large marker interval may also due to this low map resolution for D genome. More D genome markers are needed to map D genome QTL at a better resolution for future studies.

Recent years, new technologies such as single nucleotide polymorphism (SNP), Diversity Arrays Technology (DaRT, (Jing et al., 2009), Insertion site based polymorphism (ISBP, Paux et al., 2010) and Genotyping-by-Sequencing (GBS, Elshire et al., 2011) make marker discovery a high-throughput process. However, the use of DaRT markers was limited due to the business model of this technology while only 2000 markers with sequences information are available at this time (http://www.diversityarrays.com/sequences.html, accessed on July 10th, 2013). Compared with DaRT's low, ISBP and GBS had a much higher marker density but most genetic markers were identified from repeats or intergenic region, while the low coverage caused genotype calling issued in GBS make the downstream analysis even more complicated in polyploid species. On the contrary, Infinium SNP array technology, which was recently made available to the wheat community, was designed to capture major genetic diversity at exonic regions (Cavanagh et al., 2013). The SNP genotype data output is easily managed and the

marker quality is confirmed. The high-throughput genotyping protocol of Infinium 9k SNP array allows 192 samples to be genotyped in a 3-day period along with user-friendly genotype calling process makes it an attractive marker set for breeding programs. At the same time, the second generation of Infinum wheat array with around 90,000 SNPs was under development by the same group, which is aimed to significantly enrich the marker density (Dr. Akhunov, *pers. comm.*). Moreover, a core set of exonic SNPs might shorten the candidate gene mapping process for QTL mapping projects when QTL were identified in a marker dense region. In this study, most QTL regions defined by two flanking SNP markers were less than 2 cM. The SNPs that were developed from exonic regions will also help us identify potential candidate genes more effectively. Furthermore, the Infinium wheat arrays are a standard marker set that makes genetic studies across different populations more comparable.

Correlations of phenotypic traits

Various strong correlations between α-amylase activity and color components within each environment were identified (Table 2.3). The positive correlation between color component L^* and α-amylase activity (EnzBC) and negative correlation between color component a^* and EnzBC fitted perfectly with the observation that white wheat are generally more PHS susceptible, which was represented by an elevated level of α-amylase activity.

The strong correlations between non-mist and mist groups in all the environments suggested the possibility of breeding for PHS resistance through selection under various environments (Table A.3). EnzBC in 2012 mist condition was significantly correlated with EnzBC from all other environments, which indicated the potential of breeding for PHS resistance using artificial misting treatment, especially in a dry year.

Dosage effect of homeologs for grain color and α-amylase activity

In this study, dosage grouping of color loci was based on the SSR genotyping results shared by Dr. Sherman (*pers. comm.*). L^* and b showed a significant difference between white (0R) and red (2R, 4R, 6R) wheat across all five conditions. However, a^*, the direct indicator of grain redness, has not shown a consistent dosage effect across environments. This might be due to the complexity of seed coat color, which is a combination of red (phlobaphenes) and brown (oxidized proanthocyanidin). Similar results were presented in Groos et al. (2002), who used a ratio of a^*/L^* to represent grain redness. Historically, there are some debates about using colorimeter to measure grain color (Matus-Cadiz et al., 2003; Peterson et al., 2001). This might be explained by the genotype x environment effect of seed color (Wu et al., 1999). In current study, our samples were grown in the same location which might reduce this effect to some extent.

Moreover, the significant differences on grain redness between groups having R-3B allele as dominant or recessive suggested an unequal contribution of specific alleles to the seed coat color might also contribute to the inconsistency. The potentially larger contribution by R-3B allele was also supported by Groos et al. (2002) who reported a stronger influence coming from QTL on Chromosome 3B. The additive × additive interactions among Chromosome Group 3 identified within or between traits also suggest a interaction between color loci homeologs (Table 2.7). Recent studies on polyploidy evolution also provide evidence about an unbalanced usage of alleles among subgenomes (Feldman and Levy, 2012; Page et al., 2013).

QTL identified for PHS related traits

Most QTL identified in this study were also found by previous studies related to pre-harvest sprouting. In this study, all SSR markers, representing color loci, were positioned on the SNP map at expected location, the long arm of Chromosome Group 3. Hence, we divided the QTL identified into two groups: QTL linked to color loci on Chromosome Group 3 and QTL identified on other chromosome. The majority of the first group of QTL were QTL for color components while the 3A QTL were also found to explain 10-15% of the phenotypic variations of α-amylase activity in 2010 and 2011 mist conditions. A close linkage between color loci and PHS related QTL were previously documented by Groos et al. (2002). Gu et al. (2011) also claimed that the relationship between pericarp color and seed dormancy in weedy rice was due to pleiotropy. Rasul et al. (2009) also identified QTL on 3A and 3D that are related to sprouting index. Moreover, *Viviparous 1* (*VP1*) is also mapped on the long arm of Chromosome Group 3 but in loose linkage with R gene loci (Bailey et al., 1999). Recently, Liu et al. (2013) cloned a gene, *TaPHS1*, proposed to controlling PHS on Chrmosome 3AS, which is independent from grain color. They proposed that while *TaPHS1* controlled PHS resistance qualitatively, seed color may modify PHS resistance in a quantitative manner. The accumulation of PHS related genes on Chromosome 3A and color homeologs on Chromosome Group 3 may help to explain the close relationship between seed coat color and PHS resistance while further study is required to explore this relationship at a higher resolution. The recent study using genotyping-by-sequencing technology provided massive amount of markers near a color loci region, which might help us to dissect that region in the near future (Saintenac et al. 2013). On the other hand, the Chromosome 3R of rye is syntenic to wheat Chromosome Group 3. Masojć and Milczarski

(2009) found two QTL loci on Chromosome 3R had overlapping effects for both α-amylase activity and PHS.

For QTL not mapped to Chromosome Group 3, there were various types of candidate genes related to them. In a recent genome wide association study conducted in white winter wheat association panel, Kulwal et al. (2012) identified a significant correlation between a marker on Chromosome 1B and PHS related traits. This marker may be related to the QTL we identified on Chromosome 1B for EnzBC under 2011 non-mist and 2012 mist conditions. Groos et al. (2002) also found a QTL on Chromosome 5A which was in close linkage to grain color and this was also validated in our population for color component, a^*, in 2012 environments. A recent study in tetraploid wheat found genes related to seed dormancy and PHS resistance on Chromosomes 2A, 2B, 3A, 4A and 7B may help explain several QTL identified in current study for α-amylase activity (Chao et al., 2010). A QTL identified on Chromosome 4B for EnzBC may be co-located with QTL for seed dormancy previously located on wheat Chromosome Group 4 (Kato et al. 2001). Furthermore, this QTL may be orthologous to the cloned seed dormancy gene, *sdr4* in weedy rice (Gu et al. 2011). The 4B QTL was also identified by Rasul et al. (2009, 2012) controlling falling number, germination index and sprouting index. The QTL identified on Chromosome 6B for color component, a^*, in both conditions of 2012 may related to a QTL identified for sprouting score described by Roy et al. (1999). In the same paper, a QTL on Chromosome 7D was also related to PHS resistance, which might be homeolog to the 7B QTL we identified for EnzBC under 2012 conditions. Chromosome Group 6, especially 6B, and group 7 were also known to be where wheat α-amylase gene mapped (Emebiri et al., 2010; Gale et al., 1983; Mrva and Mares, 1999).

QTL by QTL interaction

Groos et al. (2002) claimed no significant interaction among QTLs that were identified; hence the effects of the red dominant alleles seem to act essentially as additive factors. However, several additive by additive interactions were found in current study using MIM method for both color traits and α-amylase activity (Table 2.8). As mentioned earlier, Chromosome Group 3 was confirmed to be a hotspot for the interactions among Chromosomes 1B, 2B, 5B and 7A. However, a combined effect for grain dormancy between 3A and group 4, which was found by Mori et al. (2005), were not identified by current study.

As for QTL effects, the multiple QTL identified for each trait, such as a^*, usually contain conflicting additive effects and the sum is around zero. This might indicate the need of pyramiding correct QTL alleles for future PHS resistance breeding. In a recent study evaluating epistatic effects of QTL controlling α-amylase activity (Yang and Ham, 2012), a strong QTL by environment interaction was identified. The epistatic effect and various types of QTL interactions can really complicate the expression of PHS resistance even within red wheat and might be able to explain why it is common to identify different QTL across environments for EnzBC.

In conclusion, by using a high density genotyping platform and a population segregating with all combinations of seed color alleles helped us to dissect the dosage and allelic effect of each color allele and enabled us to identify QTL × QTL interactions between traits. However, a dissection of color loci and PHS related traits was not accomplished even with improved mapping resolution, while the identification of Chromosome Group 3 as a hotspot for QTL interactions might complicate the dissection.

APPENDIX

Table A.1 Growth stage of 'Vida' × MTHW0471 population and precipitation from physiological maturity to the end of harvest from 2010 to 2012

Year	Planting Date	50% Flowering Interval	Physiological Maturity Interval	Harvest Interval	Precipitation (cm)*
2010	April 12th	June 13 - July 3	July 16 - July 29	July 19 - Aug.1	2.1
2011	May 9th	July 1 – July 21	July 28 - Aug.9	July 31 - Aug. 12	11.3
2012	March 29th	June 10 - June 17	July 7 - July 15	July 10 –18-Jul	0.5

* Precipitation was recorded from July 16 to August 1 for 2010, July 28 to August 12 for 2011, July 7 to July 18 for 2012.

Table A.2 SNP filtering pipeline (from raw genotype calling to Joinmap input)

SNP		Filtering criteria
Kept	Removed	
8632	0	Began with 8632 SNP from the Infinium SNP array
7921	711	Removed SNPs deemed bad or null (Feb 2012, Dr. Chao)
7304	617	Removed SNPs with greater than 20% (29 progeny) no-call
7013	291	Removed SNPs having genotype calling error
6801	212	Removed SNPs with low-confidence genotype-calls
1849	4952	Removed monomorphic SNPs
1852		Added 3 SSR markers

Table A.3 1695 markers mapped by ‘Vida’ x MTHW0471population

SNP-ID	Chromosome	cM
4163	1A1	0
4008	1A1	0.047
4164	1A1	0.131
1387	1A1	0.631
4754	1A1	18.997
268	1A1	19.428
459	1A1	19.518
4126	1A1	19.923
7021	1A1	21.2
4327	1A1	21.226
4326	1A1	21.305
6338	1A1	23.496
5268	1A1	23.881
2982	1A1	23.892
7779	1A1	24.033
7557	1A1	24.101
4661	1A1	24.168
3533	1A1	24.292
7945	1A1	24.297
5083	1A1	24.356
4265	1A1	24.401
5084	1A1	24.447
3528	1A1	24.47
5009	1A1	24.477
3957	1A1	24.539
3955	1A1	24.547
7104	1A1	24.548
3666	1A1	24.589
3665	1A1	24.597
6972	1A1	24.618
5031	1A1	24.619
3695	1A1	24.621
2981	1A1	24.623
6971	1A1	24.624
6031	1A1	24.633
1279	1A1	24.638
42	1A1	24.638
4577	1A1	24.641
3956	1A1	24.696

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
2744	1A1	24.861
8198	1A1	24.946
6984	1A1	24.948
6985	1A1	24.948
2762	1A1	24.95
1806	1A1	24.952
2630	1A1	24.953
3538	1A1	24.959
1807	1A1	24.961
3536	1A1	24.963
1952	1A1	24.967
4797	1A1	24.975
1785	1A1	24.985
2629	1A1	25.01
4852	1A1	25.016
3144	1A1	25.074
3867	1A1	25.129
3472	1A1	25.222
3882	1A1	25.247
3473	1A1	25.313
5140	1A1	25.334
3883	1A1	25.341
3499	1A1	25.382
3451	1A1	25.415
4302	1A1	25.442
6044	1A1	25.497
3160	1A1	25.531
4301	1A1	25.56
6636	1A1	25.716
498	1A1	26.053
3884	1A1	26.218
4283	1A1	32.569
7868	1A1	34.25
7869	1A1	34.446
605	1A1	36.834
8101	1A1	37.092
2490	1A1	41.774
5776	1A1	41.866
2314	1A1	45.997

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
3980	1A1	49.496
2541	1A1	50.327
3340	1A1	50.38
5125	1A1	50.763
530	1A1	50.942
4080	1A1	51.108
7639	1A1	51.632
6934	1A1	55.061
8334	1A1	62.784
3060	1A1	62.941
691	1A1	63.62
1195	1A1	63.651
5754	1A1	69.02
3146	1A1	70.809
3145	1A1	70.809
4931	1A1	73.071
5505	1A1	80.913
475	1A1	81.695
4934	1A1	81.74
8135	1A1	83.439
7924	1A1	83.501
1618	1A1	83.513
3486	1A1	83.529
1619	1A1	83.575
1119	1A1	83.661
1118	1A1	83.805
5910	1A1	86.815
2405	1A1	87.008
2450	1A1	98.465
3409	1A1	98.669
1368	1A1	100.196
6530	1A1	107.914
1560	1A1	110.603
3089	1A1	110.879
2035	1A1	111.173
4271	1A1	114.804
4123	1A1	115.666
3977	1A1	115.838
3661	1A1	116.055

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
4122	1A1	116.149
4120	1A1	116.156
3799	1A1	116.476
7290	1A1	116.71
5734	1A1	117.463
7488	1A2	0
4119	1A2	0.615
2326	1A2	0.636
2325	1A2	0.641
8351	1A2	0.665
8213	1A2	0.676
8214	1A2	0.677
5483	1A2	0.694
5484	1A2	0.697
7034	1A2	0.698
7477	1A2	0.768
1063	1A2	10.924
7067	1B	0
1480	1B	1.584
5976	1B	1.599
6449	1B	2.462
1578	1B	2.473
361	1B	2.501
7480	1B	2.51
4504	1B	2.56
2998	1B	2.57
4975	1B	4.496
7737	1B	4.892
5301	1B	37.928
6448	1B	38.775
7703	1B	38.78
6062	1B	39.289
6611	1B	39.607
6610	1B	39.77
4093	1B	39.811
1566	1B	39.943
8392	1B	40.279
5681	1B	40.597
7219	1B	40.614

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
5592	1B	40.652
8338	1B	40.717
6259	1B	41.166
6891	1B	41.483
3631	1B	41.642
6890	1B	41.642
7343	1B	42.106
139	1B	42.166
4402	1B	42.166
7234	1B	42.206
6581	1B	42.213
5546	1B	42.219
5304	1B	42.226
3295	1B	42.549
107	1B	44.419
106	1B	44.663
44	1B	44.808
5779	1B	47.7
3057	1B	48.441
4556	1B	48.46
188	1B	48.539
4198	1B	48.604
141	1B	48.628
189	1B	48.769
5278	1B	48.887
890	1B	48.907
6073	1B	48.91
4197	1B	48.937
7594	1B	48.959
6674	1B	48.987
2554	1B	49.016
4987	1B	49.033
3307	1B	49.198
7017	1B	49.27
4557	1B	49.472
3945	1B	50.284
270	1B	50.485
4316	1B	50.8
491	1B	50.883

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
7700	1B	51.001
7836	1B	51.032
269	1B	51.063
573	1B	52.509
2315	1B	56.129
2040	1B	63.215
734	1B	63.333
2790	1B	63.551
1729	1B	63.609
2788	1B	63.615
4939	1B	63.719
2588	1B	63.73
5159	1B	63.739
4940	1B	63.897
2586	1B	63.904
2041	1B	63.925
2890	1B	64.04
7040	1B	64.04
4702	1B	64.289
4703	1B	64.325
1313	1B	64.361
7527	1B	64.502
7811	1B	64.542
2889	1B	64.644
6479	1B	64.69
4488	1B	67.514
146	1B	70.402
7178	1B	73.737
7179	1B	74.306
3341	1B	74.374
8379	1B	74.402
255	1B	74.528
367	1B	74.633
2989	1B	74.853
5186	1B	75.314
5749	1B	82.782
696	1B	97.355
695	1B	98.551
1092	1B	112.798

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
3998	1B	117.106
1791	1B	130.584
724	1B	132.138
6512	1B	139.458
6647	1B	139.64
2928	1B	140.282
2077	1B	140.284
1504	1B	140.708
4935	1B	140.752
3125	1D	0
1397	1D	7.95
7797	1D	7.967
1787	1D	7.993
8551	1D	31.133
5020	1D	56.888
5018	1D	56.929
5019	1D	56.936
57	1D	57.522
362	1D	57.704
830	1D	59.045
1192	1D	59.236
1193	1D	59.267
642	1D	59.427
7425	1D	59.693
165	1D	61.593
7154	1D	75.291
1512	2A	0
6922	2A	0.275
1562	2A	0.364
1563	2A	0.365
4989	2A	0.424
2425	2A	0.662
5423	2A	1.17
681	2A	13.231
3047	2A	26.872
2696	2A	33.918
2059	2A	34.235
3235	2A	35.179
5762	2A	36.571

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
841	2A	36.705
1242	2A	36.713
1166	2A	36.736
4441	2A	37.02
4830	2A	48.684
5495	2A	62.17
901	2A	62.235
991	2A	62.335
5824	2A	62.338
562	2A	62.371
5793	2A	62.412
2007	2A	62.422
2005	2A	62.425
2006	2A	62.459
2948	2A	67.766
887	2A	68.783
411	2A	68.827
72	2A	68.88
32	2A	69.661
70	2A	69.708
33	2A	69.753
71	2A	69.786
69	2A	69.867
812	2A	69.945
120	2A	70.187
3151	2A	70.273
111	2A	70.88
7947	2A	73.059
488	2A	73.207
1174	2A	73.212
5243	2A	82.394
1960	2A	83.682
8157	2A	84.772
7864	2A	85.164
4373	2A	85.308
4375	2A	85.386
5215	2A	86.347
5733	2A	86.48
5463	2A	86.509

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
7433	2A	86.686
6499	2A	86.7
5214	2A	86.762
5216	2A	86.768
7593	2A	86.846
6307	2A	86.873
2884	2A	86.984
6844	2A	87.049
6503	2A	87.084
7540	2A	87.131
5856	2A	87.309
5855	2A	87.461
3576	2A	87.972
5244	2A	89.716
7998	2A	89.966
7876	2A	91.35
2612	2A	92.008
8040	2A	103.218
5686	2A	103.241
8041	2A	103.284
5685	2A	103.326
6549	2A	103.573
5840	2A	112.146
227	2A	112.556
7761	2A	112.974
7412	2A	114.296
4870	2A	114.818
4228	2A	115.08
6840	2A	115.082
7142	2A	115.086
4229	2A	115.099
1347	2A	115.105
6841	2A	115.114
1351	2A	115.125
6839	2A	115.127
6620	2A	115.159
5588	2A	115.162
3743	2A	115.168
702	2A	115.189

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
1350	2A	115.326
5072	2A	115.343
2601	2A	116.024
5759	2A	117.464
4491	2A	123.473
4493	2A	123.734
2846	2B1	0
7936	2B1	0.423
5137	2B1	0.809
2407	2B1	1.375
749	2B1	1.707
6767	2B1	3.354
6768	2B1	3.565
7697	2B1	19.265
1929	2B1	25.602
5708	2B1	25.758
6085	2B1	26.014
2117	2B1	26.078
7799	2B1	26.292
1930	2B1	26.61
888	2B1	26.616
2116	2B1	26.626
2115	2B1	26.657
2440	2B1	26.935
7120	2B1	26.974
5721	2B1	26.995
2442	2B1	26.995
2443	2B1	27.185
2441	2B1	27.286
4285	2B1	27.742
4284	2B1	28.307
4554	2B1	41.709
6739	2B1	41.74
7069	2B1	41.866
6740	2B1	41.875
4421	2B1	41.927
8083	2B1	41.943
4420	2B1	41.964
5392	2B1	52.326

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
8381	2B1	53.276
6364	2B1	53.836
607	2B1	54.785
7030	2B1	54.899
7029	2B1	54.957
8182	2B1	55.259
7567	2B1	55.26
608	2B1	55.272
5753	2B1	55.275
5560	2B1	55.28
3081	2B1	55.284
2557	2B1	55.285
762	2B1	55.286
2887	2B1	55.291
3329	2B1	55.294
5916	2B1	55.304
3080	2B1	55.309
295	2B1	55.396
3824	2B1	55.922
4532	2B1	56.284
1114	2B1	56.338
4531	2B1	56.366
5246	2B1	64.324
6554	2B1	66.244
7263	2B1	66.688
5038	2B1	67.016
5464	2B1	67.107
6462	2B1	67.111
4102	2B1	67.113
3428	2B1	67.117
5811	2B1	67.12
6136	2B1	67.136
1059	2B1	67.338
5149	2B1	67.397
3657	2B1	67.4
439	2B1	67.409
3213	2B1	67.438
2977	2B1	67.453
7951	2B1	67.456

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
6476	2B1	67.473
1938	2B1	67.477
4984	2B1	67.484
4983	2B1	67.485
6664	2B1	67.486
6830	2B1	67.499
5059	2B1	67.61
6875	2B1	67.719
3656	2B1	67.949
5724	2B1	68.434
8517	2B1	71.723
1656	2B1	71.95
6430	2B1	72.112
2290	2B1	72.134
6209	2B1	73.881
4541	2B1	78.432
5128	2B1	81.007
1036	2B1	81.219
6969	2B1	81.3
8359	2B1	81.321
5547	2B1	81.341
4853	2B1	81.343
1690	2B1	81.359
5525	2B1	81.404
5008	2B1	81.944
6175	2B1	82.04
2924	2B1	94.431
4956	2B1	94.604
6093	2B1	94.676
2131	2B1	94.782
2903	2B1	94.986
2261	2B1	95.191
5414	2B1	95.236
469	2B1	95.248
5415	2B1	95.482
1389	2B1	95.969
3935	2B1	96.072
4890	2B1	96.325
4948	2B1	96.335

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
3395	2B1	97.713
4357	2B1	97.752
4358	2B1	97.783
4356	2B1	97.861
933	2B1	98.319
8195	2B1	99.91
1273	2B1	102.619
6122	2B1	102.889
8362	2B1	103.096
6121	2B1	103.1
7371	2B1	103.191
2701	2B1	103.233
4909	2B1	103.477
4097	2B1	103.49
2678	2B1	103.502
2676	2B1	103.509
4095	2B1	103.625
4098	2B1	103.781
2158	2B1	103.927
3176	2B1	105.302
7955	2B1	106.072
1765	2B1	106.145
2459	2B1	106.18
2502	2B1	106.634
4130	2B1	106.936
3010	2B1	107.599
5460	2B1	107.613
8555	2B1	107.617
8406	2B1	107.617
8449	2B1	107.668
1599	2B1	107.671
1822	2B1	107.674
5810	2B1	108.286
5081	2B1	109.856
5694	2B2	0
6852	2B2	0.238
4619	2B2	1.589
2046	2B2	1.701
1667	2B2	1.821

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
1668	2B2	1.875
2551	2B2	2.126
3252	2B2	3.069
3773	2B2	3.408
4118	2B2	3.441
5442	2B2	3.847
7273	2D	0
4496	2D	0.544
2722	2D	24.947
5637	2D	34.951
5252	2D	49.347
2631	2D	74.037
4666	2D	76.347
3849	2D	76.429
4108	2D	76.536
230	2D	76.588
229	2D	76.652
3976	2D	76.928
1440	2D	77.225
552	2D	81.151
6813	2D	81.495
2174	3A1	0
6387	3A1	1.003
447	3A1	2.136
3939	3A1	24.57
8105	3A1	24.735
8106	3A1	24.785
5050	3A1	29.771
4257	3A1	29.833
2048	3A1	37.22
2047	3A1	37.345
7086	3A1	37.435
2049	3A1	37.596
7085	3A1	37.744
5641	3A1	44.685
3448	3A1	44.686
443	3A1	53.6
6413	3A1	54.261
4676	3A1	54.273

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
1972	3A1	55.096
2019	3A1	57.025
2985	3A1	57.412
7662	3A1	57.722
3166	3A1	57.964
1812	3A1	58.009
7501	3A1	58.415
5399	3A1	59.241
4009	3A1	59.254
335	3A1	59.825
720	3A1	61.713
6306	3A1	62.252
3498	3A1	63.705
1614	3A1	63.89
143	3A1	64.297
5006	3A1	64.55
4913	3A1	64.909
2156	3A1	64.921
5005	3A1	64.921
1922	3A1	64.924
4912	3A1	64.925
1422	3A1	64.93
4381	3A1	64.93
3156	3A1	64.934
8577	3A1	64.944
5786	3A1	64.944
3794	3A1	66.159
1699	3A1	66.475
1019	3A1	68.715
7114	3A1	69.002
7355	3A1	69.194
5332	3A1	69.222
1887	3A1	69.224
5632	3A1	69.26
4883	3A1	69.332
249	3A1	69.342
234	3A1	70.418
133	3A1	70.805
743	3A1	70.805

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
8465	3A1	70.857
5124	3A1	70.86
7891	3A1	71.043
7970	3A1	71.049
3600	3A1	71.065
1507	3A1	71.087
7817	3A1	71.13
3512	3A1	71.141
3772	3A1	71.149
1701	3A1	71.152
2944	3A1	71.156
5994	3A1	71.156
6907	3A1	71.164
8621	3A1	71.176
4001	3A1	71.212
4930	3A1	71.215
8558	3A1	71.249
4110	3A1	71.257
1983	3A1	71.305
1605	3A1	71.306
3250	3A1	71.314
2925	3A1	71.315
5579	3A1	71.318
1700	3A1	71.323
4707	3A1	71.323
1762	3A1	71.367
2332	3A1	71.388
7150	3A1	71.392
5315	3A1	71.398
3771	3A1	71.482
7541	3A1	71.482
5286	3A1	72.028
1678	3A1	72.03
5285	3A1	72.035
5284	3A1	72.067
7159	3A1	72.075
5650	3A1	72.079
7877	3A1	72.081
5651	3A1	72.084

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
6750	3A1	72.131
3198	3A1	72.259
3093	3A1	72.398
5649	3A1	72.524
7073	3A1	80.018
2649	3A1	84.492
2291	3A1	84.619
1462	3A1	84.734
2774	3A1	84.884
1463	3A1	84.895
4851	3A1	90.115
6652	3A1	90.564
R	3A2	0
1207	3A2	5.39
7297	3A2	5.667
2372	3A2	6.119
4259	3A2	20.275
2949	3A2	20.387
6173	3A2	20.521
4258	3A2	20.545
7835	3A2	20.675
3559	3A2	21.262
5111	3A2	21.483
2396	3A2	24.921
6716	3A2	25.093
8000	3A2	25.324
2397	3A2	25.52
7157	3A2	32.819
1457	3A2	32.864
3178	3A2	32.877
2870	3A2	32.944
602	3A2	32.952
4850	3A2	33.092
3177	3A2	33.164
7158	3A2	33.219
3949	3A2	33.543
4796	3B1	0
6471	3B1	0.264
3103	3B1	0.91

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
5202	3B1	1.708
2177	3B1	2.145
4801	3B1	2.145
5106	3B1	5.915
4800	3B1	5.921
5426	3B1	6.069
7342	3B2	0
280	3B2	2.081
8460	3B2	5.743
6464	3B2	6.916
2360	3B3	0
6056	3B3	3.165
4324	3B3	3.281
5013	3B3	3.714
2462	3B3	4.053
8053	3B3	5.659
3331	3B3	5.847
8054	3B3	5.891
R	3B3	11.25
7542	3B3	21.994
939	3B3	28.662
8058	3B3	29.054
1814	3D1	0
R	3D1	25.517
4559	3D2	0
1321	3D2	0.395
7468	3D2	0.399
5695	3D2	1.49
5224	3D3	0
6485	3D3	6.504
811	4A1	0
3061	4A1	2.495
3698	4A1	2.502
5968	4A1	3.113
3068	4A1	4.154
7322	4A1	7.769
7653	4A1	18.864
2761	4A1	19.701
3774	4A1	21.289

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
3756	4A1	25.957
6035	4A1	29.489
6906	4A1	33.051
6733	4A1	36.151
5200	4A1	42.283
1720	4A1	42.316
2123	4A1	46.367
1505	4A1	46.762
4651	4A1	50.121
7264	4A1	50.266
7265	4A1	50.361
6696	4A1	52.854
6884	4A1	57.334
6882	4A1	57.372
6883	4A1	57.5
7442	4A1	61.493
4023	4A1	62.135
2460	4A1	62.467
7077	4A1	62.608
2606	4A1	62.872
6193	4A1	63.033
5123	4A1	70.567
2900	4A1	72.638
3191	4A1	72.801
7699	4A1	72.908
2585	4A1	73.06
2901	4A1	73.155
7604	4A1	74.089
730	4A1	74.117
1400	4A1	74.161
4319	4A1	74.164
4709	4A1	74.17
1727	4A1	74.362
385	4A1	75.15
4067	4A1	75.327
1639	4A1	75.437
3826	4A1	75.463
6743	4A1	75.464
1329	4A1	75.465

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
40	4A1	75.466
493	4A1	75.525
6025	4A1	75.559
157	4A1	75.6
142	4A1	75.687
160	4A1	75.743
1141	4A1	75.748
4876	4A1	75.783
1327	4A1	75.825
192	4A1	75.925
6103	4A1	76.178
4079	4A1	77.524
112	4A1	77.57
2334	4A1	77.676
1060	4A1	77.76
750	4A1	77.899
7082	4A1	77.932
6659	4A1	77.991
6392	4A1	78.147
1824	4A1	78.15
5652	4A1	78.159
5069	4A1	78.2
7657	4A1	78.204
3541	4A1	78.213
7271	4A1	78.215
115	4A1	78.216
5975	4A1	78.217
3311	4A1	78.266
7134	4A1	78.274
8296	4A1	78.278
8220	4A1	78.284
6678	4A1	78.308
5196	4A1	78.326
912	4A1	78.327
3818	4A1	78.331
1969	4A1	78.331
7081	4A1	78.333
8265	4A1	78.339
5127	4A1	78.339

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
3119	4A1	78.352
826	4A1	78.352
1178	4A1	78.352
5705	4A1	78.356
3763	4A1	78.357
3671	4A1	78.357
6702	4A1	78.367
5237	4A1	78.368
5897	4A1	78.37
7859	4A1	78.441
1341	4A1	78.441
4772	4A1	78.442
3845	4A1	78.448
3028	4A1	78.449
7270	4A1	78.45
5865	4A1	78.455
4405	4A1	78.457
4700	4A1	78.457
6540	4A1	78.457
3582	4A1	78.458
2000	4A1	78.459
3565	4A1	78.46
110	4A1	78.463
6873	4A1	78.464
2781	4A1	78.464
8414	4A1	78.465
6597	4A1	78.465
7133	4A1	78.471
3027	4A1	78.474
1919	4A1	78.479
109	4A1	78.481
3326	4A1	78.489
3088	4A1	78.499
1694	4A1	78.502
7395	4A1	78.53
7092	4A1	78.537
3361	4A1	78.588
172	4A1	78.599
3542	4A1	78.612

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
4560	4A1	78.621
1693	4A1	78.628
6911	4A1	78.635
3029	4A1	78.671
7522	4A1	78.674
1320	4A1	78.858
7382	4A1	79.013
8150	4A1	79.08
7124	4A1	79.973
7632	4A1	103.551
108	4A1	103.978
3749	4A2	0
3751	4A2	0.317
3747	4A2	0.317
8425	4A2	0.318
6457	4B	0
2298	4B	19.565
8108	4B	19.822
4569	4B	23.725
3290	4B	26.148
102	4B	36.391
4854	4B	43.066
8263	4B	43.546
75	4B	43.648
76	4B	43.697
453	4B	43.918
327	4B	44.302
2591	4B	45.364
113	4B	46.193
6011	4B	51.855
1405	4B	52.002
3846	4B	52.057
4070	4B	52.53
1344	4B	52.602
7437	4B	52.681
1007	4B	52.747
1818	4B	52.772
1006	4B	52.774
2732	4B	52.814

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
4330	4B	52.819
1105	4B	52.824
7167	4B	52.826
2532	4B	52.829
2155	4B	52.842
6898	4B	52.861
2745	4B	52.867
5195	4B	52.884
8019	4B	53.064
1961	4B	53.172
2666	4B	53.215
2683	4B	54.231
7752	4B	54.538
1035	4B	54.58
3396	4B	54.592
6635	4B	54.679
7313	4B	55.324
2754	4B	57.703
2755	4B	57.937
1382	4B	58.035
2171	4B	58.214
1028	4B	59.74
908	4B	59.764
892	4B	61.588
3038	4B	69.53
3042	4B	69.654
3039	4B	69.918
3697	4B	85.423
2031	4B	86.441
5358	4B	92.637
564	4B	95.045
7299	4B	113.556
4618	4B	113.725
1798	4B	113.876
2087	4B	114.044
752	4D	0
5381	4D	15.651
4044	4D	21.853
161	4D	22.478

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
430	4D	23.448
465	4D	33.126
3555	4D	34.75
3815	4D	35.851
7482	4D	39.452
2558	5A1	0
7880	5A1	6.022
5002	5A1	11.248
5003	5A1	11.612
3334	5A1	12.197
649	5A1	12.698
648	5A1	13.11
3083	5A1	14.527
2802	5A1	17.383
2828	5A1	20.556
7162	5A1	20.564
1670	5A1	21.342
2897	5A1	22.372
6641	5A1	22.437
3323	5A1	28.734
5154	5A1	28.83
2282	5A1	46.602
2163	5A1	50.528
3623	5A1	54.343
2959	5A1	55.169
675	5A1	57.305
674	5A1	57.411
7044	5A1	57.413
4744	5A1	60.355
4468	5A2	0
6683	5A2	0.001
6682	5A2	0.046
3625	5A2	0.092
5838	5A2	0.188
1439	5A2	0.188
7553	5A2	0.799
687	5A3	0
8093	5A3	0.156
1120	5A3	0.284

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
1201	5A3	0.417
4687	5A3	0.439
2429	5A3	0.587
1236	5A3	3.828
4299	5A3	5.538
4051	5A3	5.684
5118	5A3	5.742
4052	5A3	5.777
7529	5A3	5.811
6573	5A3	7.096
6949	5A3	8.465
4629	5A3	10.558
8559	5A3	10.668
3313	5A3	11.489
6515	5A3	11.522
66	5A3	11.882
3413	5A3	15.254
740	5A3	15.485
738	5A3	15.516
6255	5A3	15.844
6523	5A3	15.859
7436	5A3	17.464
7220	5A3	23.892
2447	5A3	24.114
5329	5A3	25.433
5567	5A3	31.394
121	5A3	33.299
122	5A3	33.456
4734	5A3	33.69
300	5A3	33.88
3873	5A3	35.873
5529	5A3	37.05
5528	5A3	37.132
5884	5A3	37.709
5032	5A3	38.675
5034	5A3	38.886
5033	5A3	39.313
7130	5A3	43.657
7565	5A3	44.313

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
3099	5A3	44.547
1301	5A3	44.552
4149	5A3	44.86
4454	5A3	45.071
5395	5A3	45.093
3100	5A3	45.275
2480	5A3	45.471
1988	5A3	45.747
315	5A3	45.937
278	5A3	45.973
3445	5A3	46.031
5496	5A3	46.112
7351	5A3	46.195
1951	5A3	46.318
5521	5A3	46.416
3349	5A3	46.49
4736	5A3	46.558
2120	5A3	46.564
8563	5A3	46.709
2467	5A3	46.75
87	5A3	46.763
6415	5A3	46.765
1950	5A3	47.079
7961	5A3	47.111
7960	5A3	47.133
1943	5A3	47.182
7926	5A3	47.226
7925	5A3	47.241
7691	5A3	47.303
7690	5A3	47.315
114	5A3	47.395
1253	5A3	47.431
5105	5A3	47.574
7129	5A3	47.874
7109	5A3	48.445
3263	5A3	49.24
3190	5A3	49.878
8154	5A3	50.103
7061	5A3	53.682

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
3530	5A3	53.733
4465	5A3	62.518
5728	5A3	62.566
2378	5A3	63.751
1062	5A3	63.754
3811	5A3	63.779
4767	5A3	63.834
4765	5A3	63.841
5614	5A3	63.868
4069	5A3	63.868
330	5A3	64.025
331	5A3	64.1
3365	5A3	65.277
6859	5A3	65.925
5368	5A3	69.001
3196	5A3	83.04
3197	5A3	83.119
14	5A3	83.64
4932	5A3	83.98
7361	5A3	86.906
23	5B	0
4635	5B	1.25
4748	5B	1.265
5454	5B	1.492
6211	5B	1.568
22	5B	2.648
197	5B	6.187
1564	5B	9.555
7340	5B	11.67
3972	5B	12.089
3214	5B	12.138
8031	5B	12.142
6393	5B	12.148
3984	5B	12.294
6416	5B	12.719
8444	5B	17.059
7020	5B	17.093
8433	5B	17.225
6097	5B	17.229

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
2827	5B	17.369
7585	5B	27.969
2388	5B	28.326
7965	5B	28.386
6782	5B	30.169
6148	5B	30.693
6779	5B	30.775
6147	5B	30.929
7494	5B	30.93
1433	5B	31.009
7493	5B	31.371
4184	5B	31.528
7963	5B	41.823
8262	5B	41.889
7668	5B	41.892
5179	5B	41.893
3800	5B	42.546
7910	5B	52.823
1577	5B	53.37
7733	5B	53.4
4057	5B	55.442
5836	5B	56.266
5835	5B	56.609
3226	5B	56.651
3432	5B	57.343
2565	5B	57.359
2255	5B	57.41
3479	5B	57.457
6024	5B	57.505
1775	5B	57.84
1780	5B	59.342
8508	5B	60.469
3002	5B	61.413
6894	5B	70.674
6895	5B	70.775
1755	5B	71.536
6905	5B	79.79
6112	5B	80.652
6111	5B	81.082

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
7815	5B	81.172
4280	5B	81.637
6516	5B	81.638
8069	5B	81.887
7175	5B	82.062
337	5B	82.194
5488	5B	82.312
4632	5B	82.354
8187	5B	82.399
2694	5B	82.414
2335	5B	82.417
4571	5B	82.424
7471	5B	82.427
3009	5B	82.441
2336	5B	82.459
952	5B	82.474
7470	5B	82.475
6171	5B	82.475
3044	5B	82.481
3008	5B	82.481
7776	5B	82.482
5139	5B	82.511
2697	5B	82.538
2430	5B	82.539
7844	5B	82.545
5482	5B	82.546
3719	5B	82.548
951	5B	82.554
3964	5B	82.56
5947	5B	82.56
6366	5B	82.578
6125	5B	82.607
6766	5B	82.699
7446	5B	82.712
5487	5B	82.726
6235	5B	82.732
5795	5B	82.74
1446	5B	82.75
987	5B	82.869

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
2133	5B	82.882
338	5B	82.905
1772	5B	82.992
3165	5B	82.993
4832	5B	83.103
1445	5B	83.137
265	5B	83.842
4074	5B	86.28
3985	5B	86.336
6383	5B	86.758
6867	5B	86.95
5283	5B	87.632
1018	5B	89.07
5217	5B	89.558
6291	5B	89.658
4622	5B	89.793
2934	5B	89.901
2432	5B	89.975
2536	5B	91.391
1401	5B	92.295
1402	5B	92.307
6638	5B	93.996
5331	5B	94.2
3436	5B	94.636
1471	5B	102.834
6992	5B	102.865
7944	5B	103.248
5742	5B	105.367
8603	5B	106.902
5289	5B	110.921
1584	5B	111.165
301	5B	111.4
7227	5B	112.28
1776	5B	112.344
6344	5B	112.527
1777	5B	112.712
5280	5B	112.845
6526	5B	112.97
4158	5B	114.251

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
4414	5B	114.354
8005	5B	114.531
3706	5B	114.732
3106	5B	114.868
2596	5B	114.929
4300	5B	115.096
396	5B	115.224
7272	5B	115.295
4494	5B	115.344
5166	5B	115.454
3630	5B	115.575
2597	5B	115.622
1705	5B	115.783
1706	5B	115.864
1994	5B	115.873
2071	5B	115.876
5764	5B	116.524
2742	5B	116.607
5497	5B	116.849
2032	5B	116.978
1057	5B	118.616
3209	5B	119.599
7608	5B	119.621
1965	5B	119.797
1176	5B	119.814
5079	5B	119.848
7609	5B	119.923
1394	5B	119.995
4377	5B	120.031
4281	5B	120.07
7300	5B	120.081
4282	5B	120.082
1342	5B	120.1
2910	5B	120.204
6568	5B	120.223
6447	5B	120.229
7217	5B	120.233
6909	5B	120.234
6910	5B	120.239

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
894	5B	120.241
5494	5B	120.275
3870	5B	120.328
7988	5B	120.352
6567	5B	120.364
279	5B	120.419
5126	5B	120.44
1461	5B	120.48
5334	5B	120.506
620	5B	120.508
3101	5B	120.584
6817	5B	120.611
2930	5B	120.613
1781	5B	120.629
2563	5B	120.645
6555	5B	120.652
6065	5B	120.697
6521	5B	120.721
4379	5B	120.779
5078	5B	120.781
5108	5B	120.782
6556	5B	120.941
6030	5B	121.126
2320	5B	121.48
2610	5B	123.16
2609	5B	123.459
7211	5B	136.536
4355	5B	142.463
7223	5B	143.574
3514	5B	143.765
7153	5B	148.557
5176	5B	148.691
1709	5B	149.077
4416	5B	152.397
6402	5B	152.443
7400	5B	152.478
4313	5B	152.66
5537	5B	152.774
4415	5B	153.666

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
5366	5D1	0
7517	5D1	0.331
4550	5D1	0.658
2803	5D1	24.844
3429	5D1	25.193
1681	5D2	0
4087	5D2	14.085
1427	5D3	0
1428	5D3	0.204
1431	5D3	1.422
2878	5D3	1.422
2877	5D3	2.189
8160	6A	0
1033	6A	22.498
6601	6A	26.222
1205	6A	26.941
680	6A	41.436
8510	6A	48.321
612	6A	48.452
3321	6A	49.049
2018	6A	53.9
5930	6A	54.285
2017	6A	54.291
6337	6A	54.77
233	6A	57.778
522	6A	68.062
6986	6A	71.61
7940	6A	72.035
5441	6A	72.145
1423	6A	72.28
1285	6A	72.38
2187	6A	72.389
7847	6A	72.397
4371	6A	72.398
4370	6A	72.416
2366	6A	72.421
5421	6A	72.423
7438	6A	72.424
1498	6A	72.443

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
2421	6A	72.444
4607	6A	72.482
7492	6A	72.485
7458	6A	72.489
5656	6A	72.492
3529	6A	72.493
2805	6A	72.535
6095	6A	72.612
5057	6A	72.615
2186	6A	72.615
5712	6A	72.644
6560	6A	72.685
416	6A	73.027
7052	6A	77.222
741	6A	77.471
651	6A	77.488
879	6A	77.516
6699	6A	77.605
4737	6A	77.619
7563	6A	77.63
2192	6A	77.72
650	6A	77.799
6596	6A	77.907
5074	6A	77.938
6508	6A	77.965
5073	6A	78.101
3767	6A	78.221
1097	6A	87.375
4699	6A	87.392
1868	6A	87.751
1867	6A	87.752
7497	6A	88.026
2639	6A	89.025
7894	6A	89.073
1510	6A	89.134
2055	6A	89.45
2054	6A	89.5
5655	6A	89.663
4691	6A	90.029

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
6537	6A	90.032
3067	6A	90.039
7874	6A	90.398
6316	6A	90.405
7747	6A	90.414
2580	6A	91.487
442	6A	91.9
1000	6A	91.961
1497	6A	92.77
3954	6A	95.203
4278	6A	103.242
214	6A	104.278
8568	6A	105.331
3488	6A	108.066
2705	6A	108.171
3487	6A	108.249
383	6A	109.042
6355	6A	111.334
259	6A	126.68
7366	6A	126.945
4950	6A	126.965
4246	6B1	0
824	6B1	0.749
5605	6B1	2.051
7116	6B1	15.362
4717	6B1	28.969
1816	6B1	29.054
7056	6B1	33.588
1666	6B1	37.46
2474	6B1	38.853
1268	6B1	49.831
219	6B1	55.757
4337	6B1	55.866
5148	6B1	56
283	6B1	56.318
3327	6B1	56.455
2830	6B1	56.755
1473	6B1	56.775
3636	6B1	56.782

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
4338	6B1	56.789
4339	6B1	56.825
297	6B1	56.951
3967	6B1	56.965
4959	6B1	56.994
1472	6B1	57.013
221	6B1	57.167
2773	6B1	57.339
8611	6B1	58.432
387	6B1	58.786
4500	6B1	60.185
4503	6B1	60.461
5044	6B1	60.659
971	6B1	60.951
755	6B1	60.955
3030	6B1	61.971
6825	6B1	62.483
3450	6B1	62.528
7380	6B1	62.536
1743	6B1	62.542
3133	6B1	62.591
5104	6B1	62.592
4487	6B1	62.612
4086	6B1	62.637
2843	6B1	62.642
1839	6B1	62.715
6855	6B1	62.719
3878	6B1	62.741
5102	6B1	62.756
1838	6B1	62.768
5241	6B1	62.773
4599	6B1	62.784
2780	6B1	62.828
7935	6B1	62.844
2342	6B1	62.848
5157	6B1	62.867
8144	6B1	62.903
3632	6B1	62.906
3167	6B1	62.908

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
6142	6B1	62.912
8175	6B1	62.919
5785	6B1	62.933
3131	6B1	62.96
1151	6B1	62.963
5095	6B1	62.971
1742	6B1	62.977
4169	6B1	62.981
5748	6B1	62.984
5029	6B1	62.995
4924	6B1	63
2135	6B1	63.057
941	6B1	63.067
8189	6B1	63.068
617	6B1	63.089
8192	6B1	63.095
185	6B1	63.129
3132	6B1	63.145
3652	6B1	63.146
4848	6B1	63.153
3917	6B1	63.163
1545	6B1	63.197
5242	6B1	63.198
5966	6B1	63.202
2090	6B1	63.205
613	6B1	63.205
6628	6B1	63.25
6153	6B1	63.258
434	6B1	63.305
5225	6B1	63.326
7995	6B1	63.427
5531	6B1	63.599
3459	6B1	63.846
4440	6B1	63.847
2109	6B1	63.987
3677	6B1	67.072
5042	6B1	67.087
2975	6B1	67.157
5043	6B1	67.231

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
5056	6B1	68.563
6494	6B1	68.568
1905	6B1	68.606
6953	6B1	68.607
7689	6B1	68.613
2439	6B1	68.664
5055	6B1	68.694
8284	6B1	68.711
7807	6B1	68.835
7809	6B1	68.875
7618	6B1	68.972
3300	6B1	68.998
3501	6B1	69.313
3923	6B1	69.366
1657	6B1	69.416
1640	6B1	69.623
6466	6B1	69.636
2219	6B1	69.794
7937	6B1	69.945
7810	6B1	70.066
1660	6B1	71.933
4408	6B1	72.139
7240	6B1	72.482
7257	6B1	74.449
4010	6B2	0
4290	6B2	3.399
7725	6B2	4.19
2098	6B2	6.856
969	6B2	7.061
6759	6B2	7.197
666	6B2	7.672
1254	6B2	7.737
7070	6B2	8.212
860	6B2	8.759
203	6D1	0
4592	6D1	0.323
2808	6D2	0
3624	6D2	0
3979	7A1	0

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
6160	7A1	0.393
6519	7A1	0.441
7321	7A1	0.67
4558	7A1	0.693
1921	7A1	0.881
7196	7A1	0.895
7978	7A1	0.907
6642	7A1	1.444
2570	7A1	1.565
5336	7A1	1.719
7197	7A1	1.858
3850	7A1	2.554
1735	7A1	8.905
6127	7A1	13.689
557	7A1	17.071
7093	7A1	21.619
930	7A1	22.52
834	7A1	22.616
6475	7A1	42.414
4614	7A1	43.234
8390	7A1	47.705
472	7A1	47.774
473	7A1	47.869
1805	7A1	49.269
8161	7A1	62.274
7121	7A1	62.305
4386	7A1	62.381
3351	7A1	87.68
947	7A1	88.332
796	7A1	88.848
486	7A1	88.861
222	7A1	88.866
5369	7A1	89.714
8066	7A1	89.743
3863	7A1	89.853
2786	7A1	89.95
5132	7A1	90.201
1456	7A1	90.25
3693	7A1	90.383

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
3727	7A1	90.413
3694	7A1	90.424
6183	7A1	90.432
8171	7A1	90.444
2954	7A1	90.451
2381	7A1	90.453
7990	7A1	90.453
3557	7A1	90.462
4818	7A1	90.529
4072	7A1	90.542
4573	7A1	90.559
3054	7A1	90.653
7975	7A1	90.69
7554	7A1	90.777
1278	7A1	90.825
7855	7A1	90.912
4574	7A1	91.058
4082	7A1	91.47
4638	7A1	93.988
2301	7A1	94.592
3062	7A1	94.812
7193	7A1	94.998
4637	7A1	95.091
5119	7A1	95.091
1833	7A1	95.103
3090	7A1	95.19
6629	7A1	95.197
3091	7A1	95.286
2082	7A1	95.402
1834	7A1	95.418
7798	7A1	101.489
4846	7A1	101.861
7756	7A1	102.093
7755	7A1	102.133
4109	7A1	102.737
6004	7A1	103.648
2011	7A1	104.42
2010	7A1	104.496
4288	7A1	104.553

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
8076	7A1	108.646
7933	7A1	109.019
6562	7A1	109.816
4196	7A1	119.237
2270	7A1	120.024
3562	7A1	121.039
5912	7A1	122.46
5913	7A1	122.545
4910	7A1	123.131
4911	7A1	123.163
1032	7A1	124.005
1031	7A1	124.192
7045	7A1	126.469
7325	7A1	135.132
4483	7A1	136.217
7884	7A1	138.633
3367	7A1	139.012
2725	7A1	142.245
6892	7A1	142.27
1519	7A1	159.258
1515	7A1	159.365
1271	7A1	159.417
4176	7A1	159.463
4137	7A1	159.588
4364	7A1	159.605
4594	7A1	159.99
7184	7A1	159.999
4595	7A1	160.027
3371	7A1	160.245
4175	7A1	160.35
2929	7A1	160.535
5873	7A1	162.706
178	7A1	163.298
179	7A1	164.248
6576	7A1	167.979
5904	7A1	174.51
2506	7A2	0
679	7A2	0.054
6331	7A2	0.127

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
2735	7A2	0.16
2507	7A2	0.163
7205	7A2	0.893
2820	7A2	1.053
4770	7A2	1.139
7460	7A2	1.24
275	7A2	1.286
7161	7A2	1.365
1156	7A2	1.421
783	7B1	0
1181	7B1	0.18
1526	7B1	1.097
2894	7B2	0
2893	7B2	0.001
1241	7B2	1.71
4977	7B2	1.896
4968	7B2	2.049
1437	7B2	2.156
6901	7B2	2.164
1314	7B2	2.17
1436	7B2	2.173
5389	7B2	2.183
1315	7B2	2.261
3915	7B2	6.02
4092	7B2	6.022
3958	7B2	6.52
3572	7B2	8.012
3508	7B2	8.709
3507	7B2	8.973
7232	7B2	9.426
7233	7B2	9.457
6700	7B2	18.433
355	7B2	27.413
8469	7B2	28.088
4249	7B2	28.759
1963	7B2	29.083
8387	7B2	29.201
5071	7B2	29.274
4190	7B2	29.334

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
1420	7B2	29.341
3807	7B2	29.348
5171	7B2	29.361
4191	7B2	29.363
832	7B2	29.367
7450	7B2	29.387
3810	7B2	29.446
1964	7B2	29.455
3691	7B2	29.479
507	7B2	29.484
1419	7B2	29.503
3655	7B2	29.611
8021	7B2	29.957
8022	7B2	29.977
449	7B2	30.626
495	7B2	32.892
594	7B2	33.803
5706	7B2	40.889
4160	7B2	40.914
6619	7B2	41.419
6618	7B2	41.448
3928	7B2	45.525
1339	7B2	45.612
6608	7B2	45.65
3423	7B2	45.676
2767	7B2	48.413
7403	7B2	60.335
4305	7B2	60.771
7402	7B2	60.809
4306	7B2	60.954
5129	7B2	61.236
4393	7B2	73.874
3387	7B2	73.877
2193	7B2	74.767
6246	7B2	74.812
4749	7B2	74.836
4864	7B2	75
431	7B2	75.031
4750	7B2	75.611

Table A.3 (cont'd)

SNP-ID	Chromosome	cM
432	7B2	76.107
2770	7B2	87.653
8448	7B2	101.714
1902	7D	0
1537	7D	7.149
604	7D	10.988
1257	7D	11.338
5557	7D	11.414
2772	7D	11.558
2208	7D	12.069
266	7D	33.886
2522	7D	36.929
3970	7D	38.386
2521	7D	38.468
7610	7D	38.565
2523	7D	38.746
7827	7D	40.34
7828	7D	40.498

Table A.4 Significant correlations between phenotypic traits within and across environments.

V1	V2	correlation	lower 95%	upper 95%	significance
A-MS10	L-MS10	0.2509	0.1011	0.3896	0.0012
B-MS10	L-MS10	0.8549	0.8072	0.8914	<.0001
B-MS10	A-MS10	0.4496	0.3179	0.5643	<.0001
EnzBCMS10	L-MS10	0.4605	0.3301	0.5736	<.0001
EnzBCMS10	B-MS10	0.4485	0.3166	0.5634	<.0001
L-CK11	L-MS10	0.4432	0.3089	0.5602	<.0001
L-CK11	A-MS10	-0.1853	-0.3314	-0.0305	0.0194
L-CK11	B-MS10	0.4409	0.3062	0.5582	<.0001
L-CK11	EnzBCMS10	0.4988	0.3729	0.6067	<.0001
B-CK11	L-MS10	0.5259	0.4032	0.63	<.0001
B-CK11	B-MS10	0.6164	0.5097	0.7045	<.0001
B-CK11	EnzBCMS10	0.5731	0.4591	0.6685	<.0001
B-CK11	L-CK11	0.7814	0.7129	0.8351	<.0001
B-CK11	A-CK11	0.3258	0.1802	0.4574	<.0001
EnzBCCK11	L-MS10	0.2559	0.1044	0.3958	0.0011
EnzBCCK11	B-MS10	0.2454	0.0933	0.3863	0.0018
EnzBCCK11	EnzBCMS10	0.5278	0.4063	0.631	<.0001
EnzBCCK11	L-CK11	0.2909	0.1426	0.4264	0.0002
EnzBCCK11	A-CK11	-0.2556	-0.3947	-0.1051	0.0011
EnzBCCK11	B-CK11	0.299	0.1514	0.4336	0.0001
L-MS11	L-MS10	0.4679	0.3368	0.5812	<.0001
L-MS11	B-MS10	0.5016	0.3752	0.6096	<.0001
L-MS11	EnzBCMS10	0.5326	0.4119	0.635	<.0001
L-MS11	L-CK11	0.6568	0.5589	0.7366	<.0001
L-MS11	B-CK11	0.7339	0.6534	0.798	<.0001
L-MS11	EnzBCCK11	0.3063	0.1591	0.4401	<.0001
A-MS11	L-MS10	-0.1721	-0.3192	-0.0169	0.0301
A-MS11	A-MS10	0.2907	0.1414	0.427	0.0002
A-MS11	EnzBCMS10	-0.2821	-0.4185	-0.1333	0.0003
A-MS11	L-CK11	-0.3609	-0.4883	-0.2184	<.0001
A-MS11	A-CK11	0.2567	0.1062	0.3956	0.001
A-MS11	B-CK11	-0.265	-0.4032	-0.1151	0.0007
A-MS11	EnzBCCK11	-0.2982	-0.4329	-0.1504	0.0001
B-MS11	L-MS10	0.4615	0.3296	0.5758	<.0001
B-MS11	B-MS10	0.5765	0.4623	0.6718	<.0001
B-MS11	EnzBCMS10	0.5176	0.3946	0.6225	<.0001
B-MS11	L-CK11	0.5624	0.4466	0.6597	<.0001

Table A.4 (cont'd)

B-MS11	A-CK11	0.2072	0.0542	0.3506	0.0084
B-MS11	EnzBCCK11	0.2381	0.0866	0.3788	0.0024
B-MS11	L-MS11	0.8399	0.7875	0.8802	<.0001
EnzBCMS11	L-MS10	0.3358	0.1891	0.4678	<.0001
EnzBCMS11	B-MS10	0.4258	0.2884	0.546	<.0001
EnzBCMS11	EnzBCMS10	0.6017	0.4922	0.6925	<.0001
EnzBCMS11	L-CK11	0.5277	0.4053	0.6314	<.0001
EnzBCMS11	B-CK11	0.5577	0.4402	0.6563	<.0001
EnzBCMS11	EnzBCCK11	0.4166	0.279	0.5374	<.0001
EnzBCMS11	L-MS11	0.4219	0.285	0.542	<.0001
EnzBCMS11	A-MS11	-0.3981	-0.5214	-0.2584	<.0001
EnzBCMS11	B-MS11	0.5386	0.418	0.6405	<.0001
LCK12	L-MS10	0.4775	0.3481	0.589	<.0001
LCK12	B-MS10	0.5084	0.3835	0.615	<.0001
LCK12	EnzBCMS10	0.5198	0.3974	0.624	<.0001
LCK12	L-CK11	0.5407	0.4212	0.6417	<.0001
LCK12	B-CK11	0.6537	0.5551	0.7341	<.0001
LCK12	EnzBCCK11	0.3748	0.2336	0.5004	<.0001
LCK12	L-MS11	0.6425	0.5417	0.7251	<.0001
LCK12	A-MS11	-0.3519	-0.4804	-0.2085	<.0001
LCK12	B-MS11	0.5849	0.473	0.6782	<.0001
LCK12	EnzBCMS11	0.5663	0.4504	0.6635	<.0001
ACK12	L-MS10	0.159	0.004	0.3066	0.0446
ACK12	B-MS10	0.2227	0.0699	0.3652	0.0046
ACK12	A-CK11	0.4259	0.2903	0.5447	<.0001
ACK12	B-CK11	0.2541	0.1035	0.3933	0.0011
ACK12	A-MS11	0.2126	0.0599	0.3556	0.0068
ACK12	B-MS11	0.3087	0.1618	0.4423	<.0001
ACK12	LCK12	-0.181	-0.3261	-0.0275	0.0212
BCK12	L-MS10	0.5622	0.4459	0.6598	<.0001
BCK12	B-MS10	0.6556	0.5571	0.7359	<.0001
BCK12	EnzBCMS10	0.6364	0.5346	0.7199	<.0001
BCK12	L-CK11	0.6395	0.538	0.7227	<.0001
BCK12	B-CK11	0.8295	0.7742	0.8723	<.0001
BCK12	EnzBCCK11	0.3651	0.223	0.492	<.0001
BCK12	L-MS11	0.7074	0.6206	0.7771	<.0001
BCK12	A-MS11	-0.3134	-0.4465	-0.1669	<.0001
BCK12	B-MS11	0.7524	0.6764	0.8125	<.0001
BCK12	EnzBCMS11	0.6554	0.5565	0.7359	<.0001
BCK12	LCK12	0.8558	0.8083	0.8923	<.0001

Table A.4 (cont'd)

EnzBCCK12	EnzBCMS10	0.2131	0.0609	0.3556	0.0065
EnzBCCK12	A-CK11	-0.1883	-0.3333	-0.0347	0.0167
EnzBCCK12	EnzBCCK11	0.4164	0.2798	0.5366	<.0001
EnzBCCK12	LCK12	0.2343	0.0831	0.375	0.0027
EnzBCCK12	ACK12	-0.3005	-0.4346	-0.1535	0.0001
LMS12	L-MS10	0.5291	0.4074	0.6324	<.0001
LMS12	B-MS10	0.5673	0.4519	0.664	<.0001
LMS12	EnzBCMS10	0.5178	0.3952	0.6223	<.0001
LMS12	L-CK11	0.6059	0.4979	0.6954	<.0001
LMS12	B-CK11	0.6751	0.5811	0.7513	<.0001
LMS12	EnzBCCK11	0.243	0.0917	0.3832	0.0019
LMS12	L-MS11	0.5941	0.4839	0.6858	<.0001
LMS12	A-MS11	-0.2832	-0.4195	-0.1344	0.0003
LMS12	B-MS11	0.6121	0.5052	0.7004	<.0001
LMS12	EnzBCMS11	0.5591	0.442	0.6576	<.0001
LMS12	LCK12	0.7453	0.6679	0.8068	<.0001
LMS12	BCK12	0.7789	0.7099	0.833	<.0001
AMS12	A-CK11	0.3611	0.2186	0.4885	<.0001
AMS12	EnzBCCK11	-0.2153	-0.358	-0.0627	0.0061
AMS12	A-MS11	0.2645	0.1145	0.4027	0.0007
AMS12	LCK12	-0.2987	-0.433	-0.1515	0.0001
AMS12	ACK12	0.5522	0.4351	0.651	<.0001
AMS12	EnzBCCK12	-0.3221	-0.4537	-0.1766	<.0001
BMS12	L-MS10	0.5498	0.4314	0.6495	<.0001
BMS12	B-MS10	0.64	0.5384	0.7234	<.0001
BMS12	EnzBCMS10	0.5179	0.3953	0.6224	<.0001
BMS12	L-CK11	0.593	0.4826	0.6848	<.0001
BMS12	A-CK11	0.179	0.025	0.3246	0.0231
BMS12	B-CK11	0.7782	0.7089	0.8327	<.0001
BMS12	EnzBCCK11	0.1822	0.0283	0.3276	0.0207
BMS12	L-MS11	0.6171	0.5113	0.7046	<.0001
BMS12	A-MS11	-0.1899	-0.3348	-0.0363	0.0158
BMS12	B-MS11	0.7192	0.6352	0.7864	<.0001
BMS12	EnzBCMS11	0.5642	0.4479	0.6617	<.0001
BMS12	LCK12	0.6193	0.5142	0.7061	<.0001
BMS12	ACK12	0.3758	0.2353	0.501	<.0001
BMS12	BCK12	0.8312	0.7766	0.8735	<.0001
BMS12	LMS12	0.8582	0.8114	0.8941	<.0001
BMS12	AMS12	0.2803	0.1318	0.4165	0.0003
EnzBCMS12	B-MS10	0.2341	0.0819	0.3756	0.0029

Table A.4 (cont'd)

EnzBCMS12	EnzBCMS10	0.2743	0.1254	0.4111	0.0004
EnzBCMS12	L-CK11	0.2752	0.1259	0.4124	0.0004
EnzBCMS12	B-CK11	0.258	0.1076	0.3968	0.001
EnzBCMS12	EnzBCCK11	0.3315	0.1864	0.4625	<.0001
EnzBCMS12	L-MS11	0.2317	0.0799	0.373	0.0031
EnzBCMS12	A-MS11	-0.1892	-0.3341	-0.0356	0.0162
EnzBCMS12	B-MS11	0.1974	0.044	0.3416	0.0121
EnzBCMS12	EnzBCMS11	0.3262	0.1797	0.4586	<.0001
EnzBCMS12	LCK12	0.3235	0.1782	0.455	<.0001
EnzBCMS12	ACK12	-0.2021	-0.3455	-0.0494	0.0099
EnzBCMS12	BCK12	0.263	0.1134	0.4009	0.0007
EnzBCMS12	EnzBCCK12	0.6089	0.5019	0.6976	<.0001
EnzBCMS12	LMS12	0.2799	0.1314	0.4161	0.0003
EnzBCMS12	AMS12	-0.4017	-0.5234	-0.2638	<.0001

REFERENCES

REFERENCES

American Association of Cereal Chemists, International. (2002). Approved method of AACCI,10th edition. St Paul, MN, USA

Bewley J.D. (1997) Seed germination and dormancy. Plant Cell 9: 1055-1066.

Black M., Bewley J.D., Halmer P. (2006) The encyclopedia of seeds science, technology and uses. CABI Publishing, Wallingford, Oxfordshire, p 528.

Box G.E.P., Cox D.R. (1964) An analysis of transformations, Journal of the Royal Statistical Society, Series B, 26, 211-252.

Cavanagh C.R., Chao S., Wang S., Huang B.E., Stephen S., Kiani S., Forrest K., Saintenac C., Brown-Guedira G.L., Akhunova A., See D., Bai G., Pumphrey M., Tomar L., Wong D., Kong S., Reynolds M., da Silva M.L., Bockelman H., Talbert L., Anderson J.A., Dreisigacker S., Baenziger S., Carter A., Korzun V., Morrell P.L., Dubcovsky J., Morell M.K., Sorrells M.E., Hayden M.J., Akhunov E. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences 110:8057-8062.

Chao S., Xu S.S., Elias E.M., Faris J.D., Sorrells M.E. (2010) Identification of chromosome locations of genes affecting preharvest sprouting and seed dormancy using chromosome substitution lines in tetraploid wheat (*Triticum turgidum* L.) Crop Science 50:1180-1187.

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

Derera N.F. (Ed.) (1989) Preharvest field sprouting in cereals., CRC Press Inc., Boca Raton, Florida.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

Emebiri L.C., Oliver J.R., Mrva K., Mares D. (2010) Association mapping of late maturity alpha-amylase (LMA) activity in a collection of synthetic hexaploid wheat. Molecular Breeding 26:39-49.

Feldman M., and Levy A.A. (2012) Genome evolution due to allopolyploidization in wheat. Genetics 192:763-774.

Gale M.D., Law C.N., Chojecki A.J., Kempton R.A. (1983) Genetic control of α-amylase production in wheat. Theoretical and Applied Genetics 64:309-316.

Gale M.D., Lenton J.R. (1987) Pre-harvest sprouting in wheat: a complex genetic and physiological problem affecting bread making quality in UK wheat. Aspects Appl. Biol 15:115-124.

Gale M.D. (1989) The genetics of preharvest sprouting in cereals, particularly in wheat. In: Derera N.F. (Ed.) Preharvest field sprouting in cereals. CRC Press Inc., Boca Raton, Florida.

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

Groos C., Gay G., Perretant M.R., Gervais L., Bernard M., Dedryver F., Charmet D. (2002) Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white × red grain bread-wheat cross. Theoretical and Applied Genetics 104:39-47.

Gu X., Foley M.E., Horvath D.P., Anderson J.V., Feng J., Zhang L., Mowry C.R., Ye H., Suttle J.C., Kadowaki K., Chen Z. (2011) Association between seed dormancy and pericarp color is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice. Genetics 189:1515-1524.

Himi E., Nisar A., Noda K. (2005) Colour genes (*R* and *Rc*) for grain and coleoptile upregulate flavonoid biosynthesis genes in wheat. Genome 48:747-754.

Himi E., Noda K. (2005) Red grain colour gene (*R*) of wheat is a *Myb*-type transcription factor. Euphytica 143:239-242.

Imtiaz M., Ogbonnaya F.C., Oman J., van Ginkel M. (2008) Characterization of Quantitative Trait Loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. Genetics 178:1725-1736.

Jing H.C., Bayon C., Kanyuka K., Berry S., Wenzl P., Huttner E., Kilian A.E., Hammond-Kosack K. (2009) DArT markers: diversity analyses, genomes comparison, mapping and integration with SSR markers in *Triticum monococcum*. BMC Genomics 10:458.

Kato K, Nakamura W, Tabiki T, Miura H, Sawada S (2001) Detection of loci controlling seed dormancy on group 4 chromosomes of wheat and comparative mapping with rice and barley genomes. Theoretical and Applied Genetics 102:980-985.

Kulwal P., Ishikawa G., Benscher D., Feng Z.Y., Yu L.X., Jadhav A., Mehetre S., Sorrells M.E. (2012) Association mapping for pre-harvest sprouting resistance in white winter wheat. Theoretical and Applied Genetics 125:793-805.

Mares D., Mrva K., Cheong J., Williams K., Watson B., Storlie E, Sutherland M, Zou Y. (2005) A QTL located on Chromosome 4A associated with dormancy in white- and red-grained wheats of diverse origin. Theoretical and Applied Genetics 111: 1357-1364.

Masojć P., Milczarski P. (2009) Relationship between QTLs for preharvest sprouting and alpha-amylase activity in rye grain. Molecular Breeding 23:75-84.

Matus-Cadiz M.A., Hucl P., Perron C.E., Tyler R.T. (2003) Genotype X environment interaction for grain color in hard white spring wheat. Crop Science 43:219-226.

Mori M., Uchino N., Chono M., Kato K., Miura H. (2005) Mapping QTLs for grain dormancy on wheat Chromosome 3A and the group 4 chromosomes, and their combined effect. Theoretical and Applied Genetics 110:1315-1323.

Mrva K., Mares D.J. (1999) Regulation of high pI alpha-amylase synthesis in wheat aleurone by a gene(s) located on Chromosome 6B. Euphytica 109:17-23.

National Agricultural Statistics Service. (2012) http://www.nass.usda.gov/Statistics_by_Subject/index.php?sector=CROPS, Accessed on July 10, 2013

Page J.T., Gingle A.R., Udall J.A. (2013) PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. G3: Genes Genomes Genetics 3:517-525.

Paux E., Faure S., Choulet F., Roger D., Gauthier V., Martinant J.-P., Sourdille P., Balfourier F., Le Paslier M.-C., Chauveau A., Cakir M., Gandon B., Feuillet C. (2010) Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnology Journal 8:196-210.

Peng J., Sun D., Nevo E. (2011) Domestication evolution, genetics and genomics in wheat. Molecular Breeding 28:281-301.

Peterson C.J., Shelton D.R., Martin T.J., Sears R.G., Williams E., Graybosch, R.A. (2001) Grain color stability and classification of hard white wheat in the US. Euphytica 119: 101-106.

Rasul G., Humphreys D.G., Brule-Babel A., McCartney C.A., Knox R.E., DePauw R.M., Somers D.J. (2009) Mapping QTLs for pre-harvest sprouting traits in the spring wheat cross 'RL4452/AC Domain'. Euphytica 168:363-378.

Rasul G., Humphreys D.G., Wu J.X., Brule-Babel A., Fofana B., Glover K.D. (2012) Evaluation of preharvest sprouting traits in a collection of spring wheat germplasm using genotype and genotype environment interaction model. Plant Breeding 131:244-251.

Roy JK, Prasad M, Varshney RK, Balyan HS, Blake TK, Dhaliwal HS, Singh H, Edwards KJ, Gupta PK (1999) Identification of a microsatellite on Chromosome 6B and a STS on 7D of

bread wheat showing an association with preharvest sprouting tolerance. Theoretical and Applied Genetics 99:336-340.

Saintenac C., Jiang D., Wang S., Akhunov E. (2013) Sequence-based mapping of the polyploid wheat genome. G3 3:1105-1114.

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

Smith, T., Guild, J. (1931). The C.I.E. colorimetric standards and their use. Transactions of the Optical Society 33 (3): 73-134.

Van Ooijen J. W. (2004) MapQTL®5, Software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Van Ooijen J. W. (2006) JoinMap 4.0: Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Wang S., C. J. Basten, and Z.-B. Zeng (2012). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Winkel-Shirley B. (2001) Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiology: 126(2): 485-493.

Wu J.M., Carver B.F., Goad C.L. (1999) Kernel color variability of hard white and hard red winter wheat. Crop Science 39:634-638.

Yang R.C. and Ham B.J. (2012) Stability of genome-wide QTL effects on malt α-amylase activity in a barley doubled-haploid population. Euphytica 188:131-139.

CHAPTER 3 RNA-seq analysis of wheat transcriptome during pre-harvest sprouting induction

Introduction

Pre-harvest sprouting (PHS) is the precocious germination of seed in the head before harvesting. It can be induced by prolonged wet conditions. Sprouting can significantly downgrade the storage and processing quality of the seed. Direct annual losses caused by PHS can reach up to US $1 billion worldwide (Black et al., 2006). Due to this reason, the percentage of sprouted kernel has become an important criterion in grain grading and has been closely inspected by processors. Sprouting resistance is a critical quality trait for wheat breeding programs, especially for white wheat, which is more susceptible to PHS than red wheat.

PHS is a process that can be affected by both internal factors such as dormancy, hormone balance at different developmental stages, and external environments, such as humidity during physiological maturity or cold shock during seed development (Barrero et al., 2013; Derera, 1989; DePauw et al., 2012). Given its complexity, PHS resistance has been examined from different approaches ranging from physiology to biochemistry and genetics (Derera, 1989). For example, QTL analysis and association mapping have been utilized to identify QTL regions or significant markers associated with various traits related to PHS, such as α-amylase activity and falling number test of milled flour, germination rate of detached seeds, or sprouting index of intact spikes (Imtiaz et al., 2008; Kulwal et al., 2005, 2012; Mares et al., 2005; Zanetti et al., 2000). These studies indicated a complex genetic structure of PHS. Some QTL were identified in close linkage with color loci on Chromosome Group 3 while other QTL were scattered in diverse regions of wheat genome (Flintham et al., 2000). The latter group might offer breeders an opportunity to pyramid different PHS resistance sources into elite germplasm, especially for white wheat.

Red wheat is categorically more resistant to PHS compared with white wheat (Gale and Lenton, 1987). Pigments in wheat seed coat are mainly phlobaphene and proanthocynidin (Debeaujon et al., 2007). Both compounds were found to contribute to seed dormancy by either serving as a germination inhibitor or interacting with hormone pathways (Miyamoto and Everson, 1958; Winkel-Shirley, 2001). In wheat, red grain color is controlled by three major loci on chromosome group 3 and is influenced by minor genes (Freed, 1976). Red is dominant to white, and the degree of redness can be affected by the environment (Wu et al. 1999). The phenotypic correlation between seed coat color and PHS resistance were reported by Groos et al. (2002), but either this relationship is caused by a close linkage or a pleiotrophy still need more study.

Previous studies on PHS mainly focused on various phenotypic screening methods, such as germination rate or sprouting count, to measure the sprouting resistance based on the downstream events during germination process. However, these methods do not capture the initial change of PHS, which usually happens in the absence of visual symptoms. Therefore, the measurement at transcription level would be critical to identify potential candidates involved in the induction of PHS. Transcriptional events in seeds are arrested after seeds reach physiological maturity and will resume when environmental conditions are suitable for germination (Dobrzanska et al., 1973). During dry-down process, a small amount of RNA is carried over during seed development and may be expressed at a low level (Leubner-Metzger, 2005). During germination, seed RNA expression is activated in a cascading manner in the order of mRNA, rRNA, and tRNA (Dobrzanska et al., 1973). Messenger RNAs are quickly mobilized at the beginning of imbibition to support *de novo* protein synthesis and prepare for germination when large scale transcription initiates (Nonogaki et al. 2010). However, little information has been reported for the expression patterns during PHS process for which the seed started dry-down

process after physiological maturity. Therefore, information on expression levels at the initial stage of PHS may help us to generate a more complete picture of PHS process, and enable us to compare PHS with germination, for which seed had gone through dry-down. Moreover, a comparison of the expression pattern between red wheat and white wheat at the initial stage of PHS can also help us to understand how red and white wheat respond to PHS induction and what the relationship is between seed coat color and PHS resistance.

Traditionally, microarrays have been used for measuring expression pattern differences at a global scale. The probe design requires previous knowledge about the transcripts sequences (Dalma-Weiszhausz, 2006). So far, the most popular microarray platform for wheat, Affymetrix GeneChip® (http://www.affymetrix.com/estore/browse/products.jsp?productId=131517), is based on wheat unigene dataset built in 2004, which is outdated with limited gene representation. Extra issues related to microarray technology, such as high background noise caused by non-specific hybridization and data repeatability (Wang et al., 2009), also made it an obsolete option. Recent years, RNA-seq, the next generation sequencing (NGS) of cDNA sequences, has become faster, more accurate, and less expensive (Wang et al., 2009). In a single assay, it can identify novel genes and splice variants and quantify transcriptome-wide expression levels. Thus, it has been widely used to identify potential candidate genes underlying various biological processes in plants, such as development (Kyndt et al., 2012; Teoh et al., 2013), disease infection (Savory et al., 2012), and different types of abiotic stresses (Zhang et al., 2013; Miller et al. 2010). The transcripts assembled from the RNA-seq study can also be a valuable resource for single nucleotide polymorphism (SNP) mining, which can be applied in breeding programs (Trick et al., 2012).

The transcript abundance in RNA-seq is measured by counting the reads mapped to specific loci instead of measuring hybridization signal intensity as seen in microarray studies. The absolute measurement of transcripts makes RNA-seq more accurate. The measuring range is also more dynamic when compared to microarray technology which measures differential expression (DE) based on the relative intensity of florescent hybridization signals (Wang et al., 2009). RNA-seq based DE would be more user-friendly if a well annotated reference genome is available. However, with the development of *de novo* assembly methods, DE estimates based on a *de novo* assembled transcriptome is now possible (Mutasa-Gottgens et al., 2012). Assemblers using *de novo* methods leverage high sequencing depth and uses a *de Bruijn* graph method to construct a reference transcriptome without a draft genome (Martin and Wang, 2011). Several assemblers have been proven efficient in *de novo* assembly of complex transcriptomes and the accuracy of downstream analysis using these assemblies has been confirmed (Grabherr et al., 2011; Robertson et al., 2010). Another alternative for transcriptome assembly is to align reads to a phylogenetically close model species (Mayer et al., 2011) or a comprehensive set of ESTs (Trick et al. 2012). This strategy, called Genome Guided assembly (GG), leverages the available genomic resources to filter out contamination and sequencing artifacts. The GG strategy can recover full-length transcripts more easily with available genomic sequences, but can also limit the gene discovery to the available genomic regions. On the other hand, the DV strategy may have more fragmented sequences but can recover novel transcripts from regions missing in the genome assembly (Martin and Wang, 2011).

Due to a large genome size and polyploidy nature, hexploid wheat genome has not been fully sequenced yet. However, its D genome progenitor, *Aegilops tauschii*, was sequenced with NGS technology recently (Jia et al., 2013). Its availability enabled us to evaluate the DV and GG

assembly strategies for wheat transcriptome for the first time. Then we assumed three scenarios in terms of different amount of input reads (single sample, combining biological replicates of one condition/time or all samples) in order to evaluate transcripts recovery rate. The availability of annotated *Aegilops tauschii* genome also simplified the transciptome level comparison. Four comparisons were made between red wheat and white wheat before and after a 48 hr misting treatment, in order to identify the effect of seed coat color and misting to PHS induction. This work demonstrated the feasibility of adapting RNA-seq technology in wheat PHS research and provided candidates for future dissection of gene networks involved in the PHS process.

Materials and methods

Plant materials

Eight F5:7 recombinant inbred lines were selected from a spring wheat population ('Vida' × MTHW0471) segregating for seed coat color (Sherman et al., 2008). Individuals were selected based on their color loci genotype and α-amylase activity. Out of the eight lines, four lines from the white wheat group had all three color loci as recessive and consistently high α-amylase activity in comparison with the rest of the population. Four lines from the red wheat group had all three color loci as dominant with a consistently low α-amylase activity in comparison with the rest of the population. The eight lines were grown in the greenhouse until physiological maturity (loss of green color at peduncle). Plants were transferred to the greenhouse with a misting schedule of 45 minutes misting every 6 hr for a period of 48 hr (see Chapter 2 for misting treatment details). Plants were kept in pots and held upright to mimic field conditions. Four lines within each group (red/white) were considered as biological replicates. Seeds were collected from single plant at 0 hr, and 48 hr under both mist and non-mist treatments (Table 3.1). Seeds

Table 3.1 Summary for RNA-seq comparisons

Comparisons	A	B	C	D
Wheat group by color*	White (3) vs Red (4)	White (3) vs Red (2)	Red (2) vs Red (2)	White (3) vs White (3)
Treatment	Non-mist	Mist	Non mist vs Mist	Non-mist vs Mist
Duration	0 hr	48 hr	0 hr vs 48 hr	0 hr vs 48 hr

* Number within parenthesis is the biological replicates actually used for differential expression analysis within that group.

were flash frozen in liquid nitrogen, then transferred to a -80°C freezer for long term storage until RNA extraction.

RNA extraction, library construction and Illumina sequencing

For RNA extraction, a single seed was used as a sample for each recombinant inbred line to minimize heterogeneity within a sample. Total RNA was extracted using TRIzol LS reagent (Invitrogen, 10296-010) following manufacture's instruction. The quality and quantity of RNA were inspected on Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) before library preparation. Sample library preparation and Illumina sequencing were performed by the Research Technology Supporting Facility at Michigan State University (http://rtsf.msu.edu/). Samples were sequenced in two rounds: first set of samples were run on the Illumina Genome Analyzer II system (Illumina, San Diego, CA, USA) generating 75 bp paired-end (PE) reads, while the second set were sequenced on Illumina HiSeq 2000 system, which generated 100 bp PE reads. Due to the improved throughput, samples from second set usually had 2-3 times the number of reads compared with samples in the first set.

Quality trimming of short reads

Reads used for transcriptome assembly were trimmed using Cutadapt version 1.1 (Martin, 2011) with the parameters '-f fastq -e 0.01 -m 20', which trimmed reads to a minimum length of 20 bp and a quality cutoff of 20. Then reads were mapped against the Triticeae Repeat Sequence Database (TREP, release 10, http://wheat.pw.usda.gov/ITMI/Repeats/), ribosomal RNA database (SILVA, release 109, www.arb-silva.de), wheat chloroplast genome sequence and wheat mitochondria genome sequence, sequentially using Tophat 1.4.1 with default parameters (Trapnell et al., 2009). The reads that mapped onto any of these databases were removed. Only paired reads were used in transcriptome assembly.

Reads used for differential expression were trimmed by Trimmomatic-0.27 (Lohse et al., 2012) with parameters 'PE -phred 33 ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 LEADING:13 TRAILING:13 SLIDINGWINDOW:4:20 MINLEN:25', which removed adapters and scanned the read within a 4-base sliding window, reads were removed when the average quality per base dropped below 20, and a minimum length of 25.

Short read assembly and quality evaluation

Trinity v. 20130225 was used for *de novo* assembly (DV) with paired reads in fasta format (-seqType fa, –kmer_method meryl) and minimum contig size was set to be 201bp (-min_contig_length 200). Genome Guided assembly (GG) was done by first aligning trimmed paired-end and single-end RNA-seq reads using GMAP/20130331 (Wu and Nacu, 2010) to the *Aegilops tauschii* draft genome (Jia et al., 2013). The parameters used for alignment are -N 1 -Q -B 5 -t 2 and Sequence Alignment/Map (SAM) format was used. The alignment results were then used by Trinity v. 20130225 to generate GG. Both assemblies were performed with single k-mer (k=21) due to software configurations. CD-HIT-EST v.4.6.1c (Fu et al. 2012) was used to measure the similarities of two assembly results. Sequence similarity search were conducted by Blast+ v. 2.2.25 (BLASTn, BLASTp, BLASTx, MegaBLAST) to identify common transcripts in different datasets and to assign gene functions to *de novo* assembled transcripts by identifying their top match in public databases. Databases used included brachypodium, v1.0 (ftp://ftpmips.helmholtz-muenchen.de/plants/brachypodium/v1.0/brachy1.0_wholegenome_unmasked.mfa.gz), barley, cultivar 'Bowman', (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/sequences/assembly5_WGSBowman34x_renamed_blastable_carma.zip) and rice, v 7.0

(ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/all.seq). The parameter used for defining a top match is based on a BLAST Expect-value (E-value) cutoff of $1\times e^{-20}$, if not specified otherwise.

Comparison of assembly strategy

In this study eight RNA-seq datasets with 75 bp PE reads (Table 3.2) were used to generate *de novo* assemblies in three ways: 1. Single sample assembly; 2. Bulk assembly with merged reads from four biological replicates; 3. Assembly with reads from all eight samples. Read alignment rate, via Tophat 1.4.1, as well as transcript representation via reciprocal BLASTn search, were compared between different assemblies to identify the strength and weakness of each assembly strategy.

Analysis of differential gene expression

Differential expression was conducted between samples collected under four conditions (Table 3.1). RNA-seq reads were aligned to *Aegilops tauschii* draft genome (Jia et al. 2013) using Tophat 2.0.8b, while Cufflinks 2.1.1 were used to estimate transcripts abundance in fragments per kilo base pair of target transcript length per million reads mapped (FPKM) and for differential expression analysis with red and white wheat with different treatment conditions following the protocol described by Trapnell et al. (2012). Parameters of both software were adjusted (-I 200000 --min-intron-length 30 -b *A.tauschii*.fasta) based on annotation files provided by Dr. Chi Zhang, BGI (pers. comm.). Enrichment of gene ontology (GO) category was determined by following methods described by Miller et al. (2010). P values from Fisher's exact tests were calculated with statistical package R/2.14.1 (R Development Core Team, 2008) and adjusted with the Benjamini and Hochberg (1995) method to avoid multiple-testing. False discovery rate (FDR) of 1% was used as the threshold for calling differentially expressed genes.

Table 3.2 Summary of sample raw reads (million pairs) for comparison of assembly strategy

Sample*	1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8
Reads amount	7.9	8.0	16.1	10.5	12.3	8.8	8.6	12.5

* White wheat: 1-1, 1-2, 1-7, 1-8; Red wheat: 1-3, 1-4, 1-5, 1-6.

Results and discussions

Global characteristics of transcript assemblies

An RNA-seq dataset with 24 million pairs of 100 bp PE reads was used to assess the quality of Trinity assemblies using *de novo* (DV) and Genome Guided (GG) methods. The library was prepared from a single seed of a white wheat line at 4 days past physiological maturity and the seed was undergone a 48 hr mist treatment.

Both assemblies contained multiple splicing forms for each locus, which were indicated by the number after "seq" of the transcript name (Haas et al. 2013). In total, DV assembly contained 105,634 transcripts representing 59,628 gene loci. GG assembly contained 68,365 transcripts representing 44,898 gene loci. To get an overview, two assemblies were compared between each other and against four wheat transcripts collections: (1) UK454, a *de novo* wheat cDNA assembly using 454 reads (Brenchley et al., 2012); (2) Unigene_uniq_v63 with the longest transcripts of wheat Unigene build 63 clusters (ftp://ftp.ncbi.nih.gov/repository/UniGene/Triticum_aestivum/); (3) CAAS, a *de novo* wheat transcriptome assembly using Illumina reads (Duan et al., 2012), and (4) TriflDB, a collection of wheat full-length cDNA (Mochida et al., 2009). Key statistics were summarized in Table 3.3.

Unigene_uniq_v63 and CAAS had the largest number of transcripts, which may be due to their comprehensive collection of treatment conditions and tissue types. GG assembly, DV assembly and UK454 had similar amount of transcripts while TriflDB, which contained only full length cDNA, had the least amount of transcripts. N50 size, the length sum of all the contigs that had at least N50 contig size (see below), is strongly correlated with the number of transcripts (R2=0.951). N50 contig size is defined as the maximum length whereby at least 50% of the total assembled sequences reside in contigs of at least that length. Therefore, a length distribution of

Table 3.3 Summary of Trinity assembled transcripts compared with public databases

Database	*de novo*	Genome Guided	UK454	Unigene_ uniq_v63	CAAS	TriFLDB
Num_of_transcripts	105,634	68,365	90,174	178,464	223,082	6,137
Num of loci	59,628	44,898	NA[^]	NA	NA	NA
N50 size (Mb)	44.6	28.5	46.1	54.9	67.5	5.4
N50 contig (bp)	1,408	1,392	1,335	821	812	1,998
Max_len_trans[*] (bp)	16,351	16,357	10,382	11,803	10,027	8,930
Min_len_trans[#] (bp)	201	201	201	103	201	134
References	NA	NA	(Brenchley et al., 2012)	NCBI Unigene	(Duan et al., 2012)	(Mochida et al., 2009)

[^] NA: Not available;

[*] Maximum length of transcripts in the dataset;

[#] Minimum length of transcripts in the dataset.

transcript size was used to explain the differences in N50 contig size (Figure. 3.1). Unigene_uniq_v63 and CAAS had the smallest N50 contig size. Alternately, TriflDB with flcDNA had the largest N50 contig size. GG assembly, DV assembly and UK454 had nearly the same N50 contig size while their contig length distribution was also not significantly different from each other (χ^2=0.001, p value =1). The contig length distribution also indicated that the length was normally distributed for the DV and GG assembled transcripts when compared with other datasets (Figure 3.1). However, there are 28 transcripts in GG and DV assembly that are above 10,000 bp (upper limit of public database transcript size). These transcripts were searched against the NCBI refseq_rna database to determine their identity. All the transcripts were hit by authentic transcripts from related species (Table B.1).

For both GG and DV assemblies, BLASTn search was conducted against wheat unigene dataset, Unigene_uniq_v63. 93.4% of GG transcripts and 85.1% of DV transcripts were confirmed to be putative wheat transcripts based on their BLASTn matches. For transcripts that didn't return any matches, BLASTx search was conducted against *A. tauschii* proteome to check if they matched protein coding sequences from phylogenetically related species including brachypodium, barley and rice. No new GG transcripts were further identified from these databases, while 2,589 DV transcripts had matches. For the transcripts that don't have a BLAST match, second round BLASTn search were performed against whole genome sequences of brachypodium, barley, and rice to check if the assembled transcripts were part of non-coding genomic sequences. About 40-50% of the remaining sequences from both assemblies were aligned to the barley genome while only 1-2% of the sequences aligned to the brachypodium and rice genomes. In total, only 2,092, or 3.1%, of GG transcripts and 7,855, or 7.4%, of DV

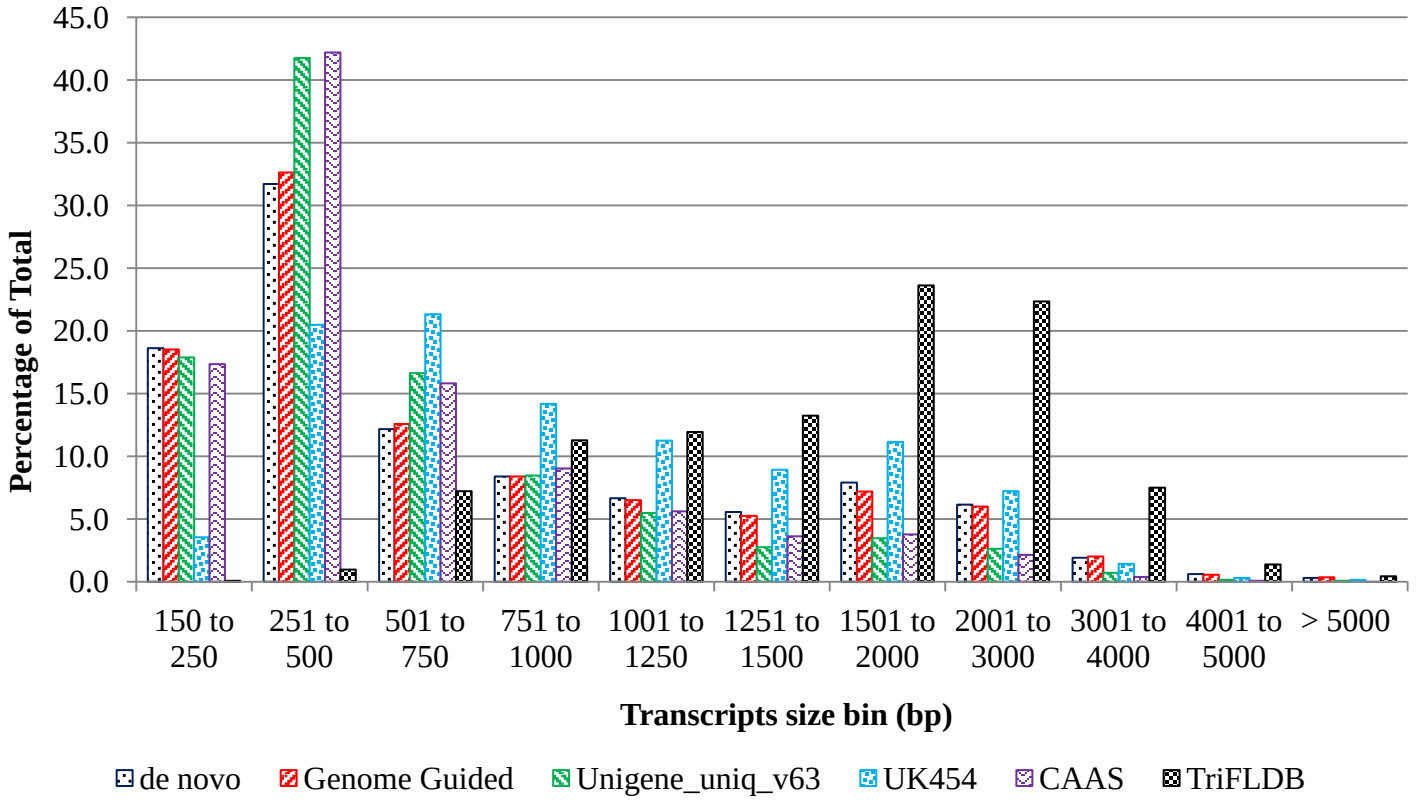


Figure 3.1 Length distribution of Trinity assembly compared against public databases.

transcripts had no hit, which might be novel protein coding transcripts that was not covered by current databases or sequenced poly-adenylated non-coding mRNA.

Concordance analysis of *de novo* and Genome Guided assemblies

In order to check the transcripts similarity of the two assemblies, reciprocal BLASTn search were performed to check for concordance between GG and DV assemblies. In total, 62,535, or 91.5% of GG assembled transcripts and 73,001, or 69.1% of DV assembled transcripts were shown to have a match in their counterpart's assemblies. As mentioned earlier, Trinity assembly can have multiple splicing forms for some loci. DV assembly contained 47,512 singleton loci, which indicated only one splicing form were available for that locus, and 12,116 loci with an average of 4.8 splicing forms while GG assembly contained 35,660 singleton loci and 9,238 loci with an average of 3.5 splicing forms. The imbalance of transcript match happened again when compared at loci level: 39,000, or 87% of GG loci matched 31,000, or 52.0% of DV loci. This imbalance between the two assemblies may be due to the potential mis-assembly of DV transcripts, which might cause one DV transcript to match with multiple GG transcripts at the same time while one GG transcripts generally hit one DV transcripts. By reviewing the BLAST results, most loci with splicing forms were covered by both assemblies, while singleton loci were more specific to their own assemblies. For GG assembly, singleton contain more novel transcripts might be limited due to the reference genome chosen, while DV singleton transcripts may either be novel transcripts that do not overlap with current sequence databases or a potential mis-assembly.

In order to measure the number of transcripts that are common in both assemblies, reciprocal best BLAST hit (RBBH) were defined by the hits that were common in both counterparts. 24,055 DV assembled loci (77.5% of DV loci that have hit in GG) and 27,358 GG

assembled loci (70.1% of GG loci that have hit in DV) were identified which indicating a roughly 50% concordance rate based on the number of common transcripts shared by both methods. These transcripts can be considered as a set of transcripts with high confidence.

In summary, the two methods were sharing approximately 50% of the assembled transcripts. The difference in the number of assembled transcripts can be due to following reasons. The align-then-assemble mechanism used by GG only assembled reads that mapped to a genome. Thus, GG assembly's completeness and quality can be heavily impacted by the reference genome. As for DV assembly, its relatively large transcriptome size with more transcripts can be explained by the fact that the *de novo* algorithm may split a potential long transcript into two fragmented ones due to the lack of reads connecting the read parts in between. This is consistent with the observation that 19144, or 18.1% of DV transcripts had hits from two or more DV loci. The GG assembly can avoid the fragmentation issue using a draft genome. Moreover, the GG assembled transcripts may have a higher sequence similarity based on the fact that fewer reads were mapped uniquely in GG than DV. The use of diploid progenitor genome as reference may collapse the homeologs of common wheat into one transcript in GG assembly while the DV assembly can be more accurate in recovering homeologs which maybe indicated by more novel transcripts shown in the DV assembly. However, further investigation is needed to confirm current assumptions.

CD-HIT-EST was used to measure the redundancy, which was defined as the number of transcripts got merged into other transcripts, of the GG and DV assemblies by clustering analysis based on protein sequence similarities (Fu et al. 2012). A range of 80% to 100% identities were tested. At 100% identity, the DV assembly had 49 transcripts (0.05%) clustered into larger transcripts while the GG assembly had 436 transcripts (0.64%) clustered (80% identity). These

clustering results were closely related to the clustering of transcripts with splicing forms (Table 3.4).

Transcripts representation and assembly completeness

Transcript completeness is defined in Martin and Wang (2011) as percentage of expressed transcripts in reference databases covered by the assembled transcripts. The higher the percentage, the more complete the assembled transcriptome is. The ideal scenario would be 100%, which means all the transcripts in reference database were represented in assembled transcripts. It can be further divided into two levels: the total number of transcripts represented by assembled transcripts and the length percentage of each transcript covered by assembled transcripts. For the first level, GG and DV assemblies were compared with three cDNA libraries deposited in wheat Unigene databases with similar growth stage or treatment (Table 3.5). From 62.2 to 75.4% of the ESTs in the three datasets were covered by both assemblies. For the second level, the ABA-treated cDNA library was used to calculate the percentage length coverage of the ESTs covered by Trinity transcripts, which ranged from 20 to100%. Approximately 50% of the ESTs overlapped with a single Trinity transcript for more than 80% of their lengths (Figure 3.2).

In addition to the single condition representation, an understanding of the percentage of transcripts expressed under current condition versus total expressed wheat transcripts may help us to compare current biological process with other biological processes. Two types of similarity searches were performed to evaluate this percentage. At nucleotide level, MegaBLAST search against three comprehensive wheat EST databases (Unigene_uniq_v63, UK454, and CAAS) returned a result ranging from 16 to 24% of transcripts in reference databases were shown in GG assembly and 20 to 27% were shown in DV assembly (Table 3.6). At protein level, BLASTx against *A. tauschii* protein coding sequences was performed (Jia et al. 2013) against both GG and

Table 3.4 Clustering results of Trinity assembled transcripts based on CD-HIT-EST at different sequence similarities

Sequence similarity	80%	85%	90%	95%	99%	100%
Genome Guided	45,112 (66.0)*	47,555 (69.6)	18,111 (73.5)	13,884 (79.2)	6,763 (89.8)	49 (99.9)
de novo	68,343 (64.7)	72,250 (68.4)	28,918 (72.6)	22,008 (79.7)	10,757 (90.1)	436 (99.4)

* Number in parenthesis is the percentage of transcripts that cluster when compared with original assemblies: Genome Guided assembly contains 68,366 transcripts, *de novo* assembly contains 105,564.

Table 3.5 MegaBLAST results of Trinity assembled transcripts vs. Unigene cDNA library

Unigene Library ID	EST	Brief description	Hit by GG*	Hit by DV#
5552	2,173	Mature seed treated with 25 mM ABA for 12h at 22°C	1,547 (71.2%)	1,639 (75.4%)
12171	1,000	Seeds malted for 55 hr at 22°C	633 (63.3%)	663 (66.3%)
8825	2,954	Embryo from mature dormant seed	1,837 (62.2%)	1,934 (65.5%)

* GG: Genome Guided assembly;

DV: *de novo* assembly.

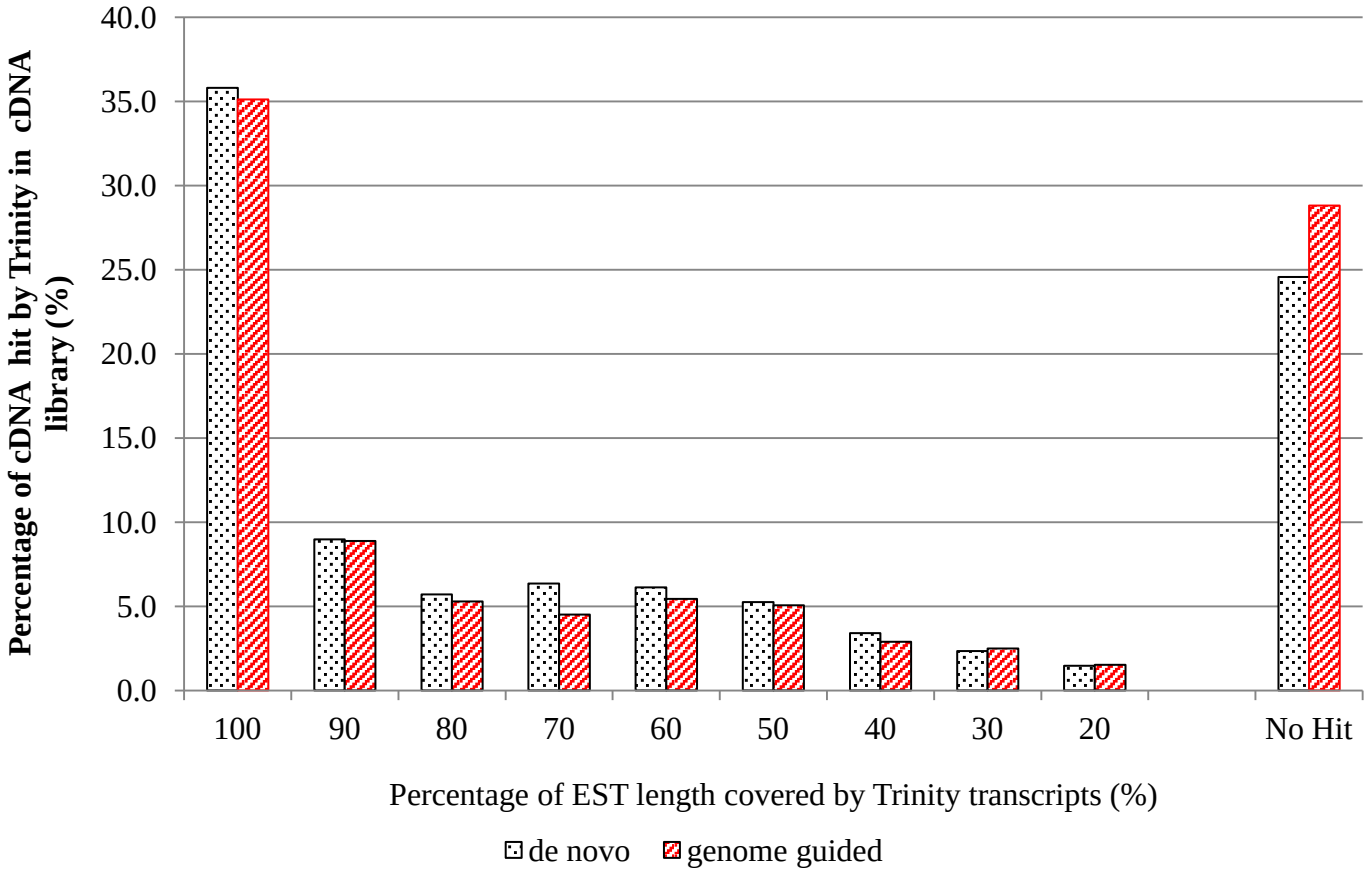


Figure 3.2 The distribution of length coverage of EST by the Trinity assembled transcripts in cDNA library made from ABA-treated seed.

Table 3.6 Percentage of Trinity assembled transcripts in public wheat EST databases

MegaBLAST	Unigene_uniq_v63	UK454	CAAS
Genome Guided (GG)	28,623 (16.0)[#]	23,060 (23.7)	35,480 (15.9)
de novo (DV)	37,803 (21.2)	26,592 (27.3)	45,172 (20.2)
References	NCBI	(Brenchley et al., 2012)	(Duan et al., 2012)

[#] The number in parenthesis is the percentage of the transcripts in target databases with homology.

DV assemblies. A total of 24,900 (55.5%) GG assembled loci, represented by 41,245 GG transcripts, hit 16,688 (38.7%) predicted protein coding genes in *Aegilops tauschii* proteome, while 27,682 (40.5%) DV assembled loci, represented by 53,583 DV transcripts, hit 18,008 (41.7%) predicted protein coding genes. This percentage was in line with a recent transcriptome study conducted in Arabidopsis, which recorded a range of 7,000 to 14,000 genes, or 28-56% (assume 25,000 genes in Arabidopsis), expressed during germination process (Dekkers et al. 2013)

In summary, in terms of transcript completeness of Trinity assembly, both GG and DV assemblies showed a relatively good coverage for this biological stage based on the fact that 75% of the ESTs in cDNA library were represented in both assemblies while approximately 50% of these EST were covered by Trinity transcripts for 80% or more of their length (Table 3.5; Figure 3.2). However, when it compared with general wheat EST collections, the relatively low rate of transcript representation from the general databases showed the diversity and comprehensiveness of wheat transcriptome (Table 3.6). Therefore, in order to build a comprehensive wheat transcriptome, materials from multiple biological conditions and different tissues at different growth stages are required.

Transcripts contiguity

Due to the nature of short read sequencing, transcripts contiguity is another performance metric for transcriptome assembly. The transcript contiguity is defined in Martin and Wang (2011) as the percentage of transcripts in reference databases covered by a single, longest-assembled transcript. Both assemblies were aligned to TriflDB, the full-length cDNA set, using BLASTn. The distributions of the percentage of full-length cDNA covered by single Trinity transcripts were summarized in Figure 3.3. In general, more than 40% of the full length cDNA

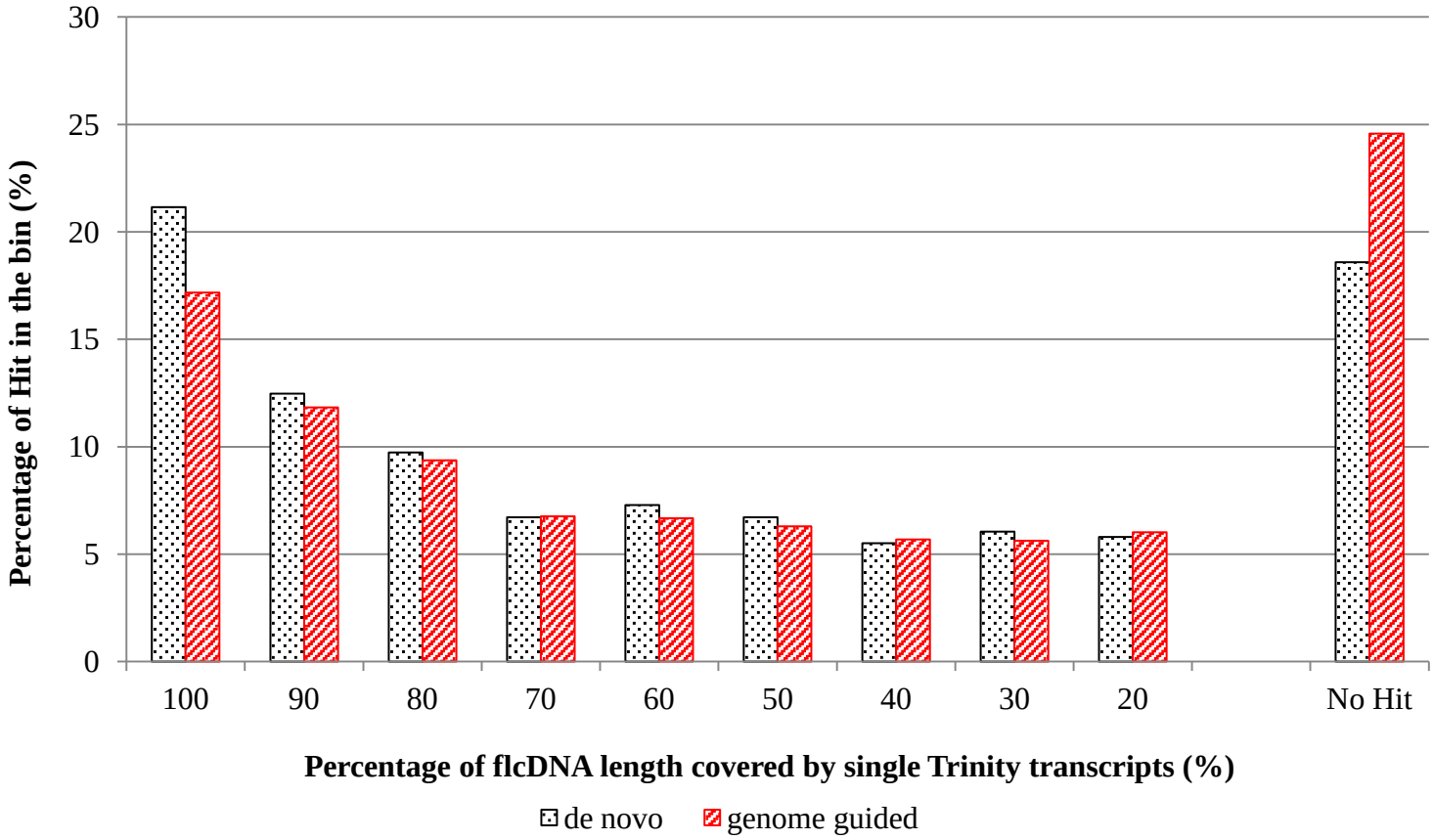


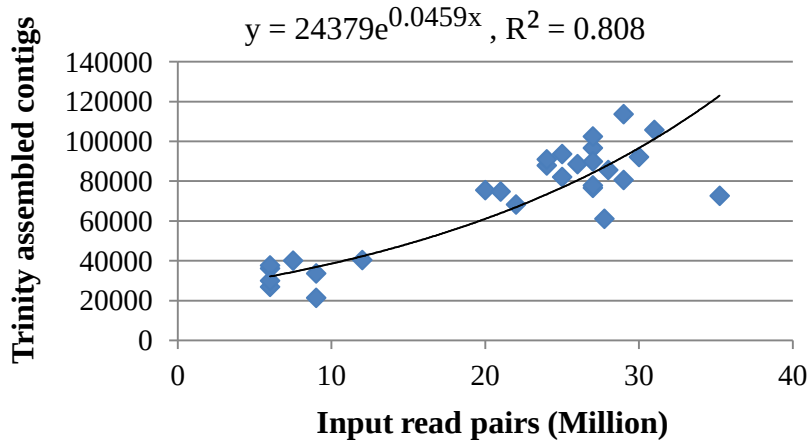
Figure 3.3 The distribution of length coverage of full length cDNA (flcDNA) by the Trinity transcripts.

were overlapped with both assemblies for more than 80% of their lengths. The DV assembly had more full length cDNA covered in total and in all categories compared with GG assembly. In conclusion, the similar quality of the two assemblies and their comparable quality with the public wheat ESTs made us confident about the *de novo* algorithms used by Trinity to assemble full length transcripts in complex transcriptome like wheat. This is in accordance with previous results on the efficiency of Trinity to recover full-length transcripts and spliced isoforms in other species (Grabherr et al., 2011).

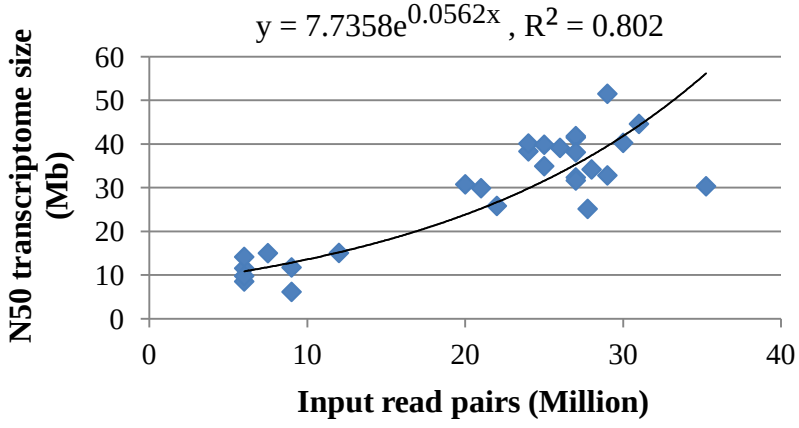
Evaluation of *de novo* assembly strategy

In this study, eight RNA-seq datasets with 75 bp PE reads (Table 3.2) was used to generate *de novo* assemblies in three ways: 1) single sample assembly, 2) bulk assembly with merged reads from four biological replicates, and 3) assembly with reads from all eight samples together. For single sample assembly, there was an exponential increase in numbers of transcripts assembled (R^2=0.81) and N50 size (R^2=0.80) and a logarithmic increase for N50 contig size (R^2=0.70) (Figure 3.4, (a-c)) as more reads were included. Similar trends were shown in merged assemblies. When analyzing all three types together, a linear correlation is shown between input reads number and assembly statistics (Figure 3.4, (d-f)).

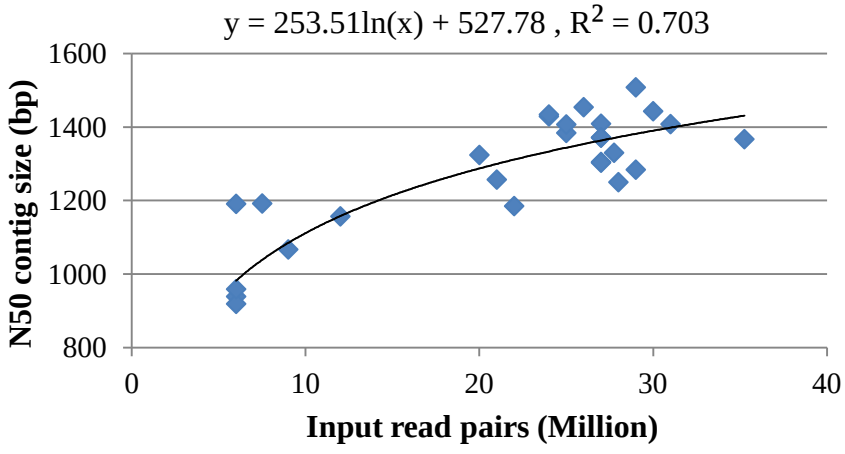
Based on the fact that the more reads gave the larger the *de novo* transcriptome size, we would like to determine if the size increase was due to the novel transcripts represented in different samples or just redundancy. The redundancy of each assembly was measured by CD-HIT-EST based on 100% sequence similarity. Each single sample assembly yielded less than 0.3% redundancy, while the merged assembly was around 0.6%. Two samples with similar input read numbers and assembly statistics were chosen for further comparisons.



(a)



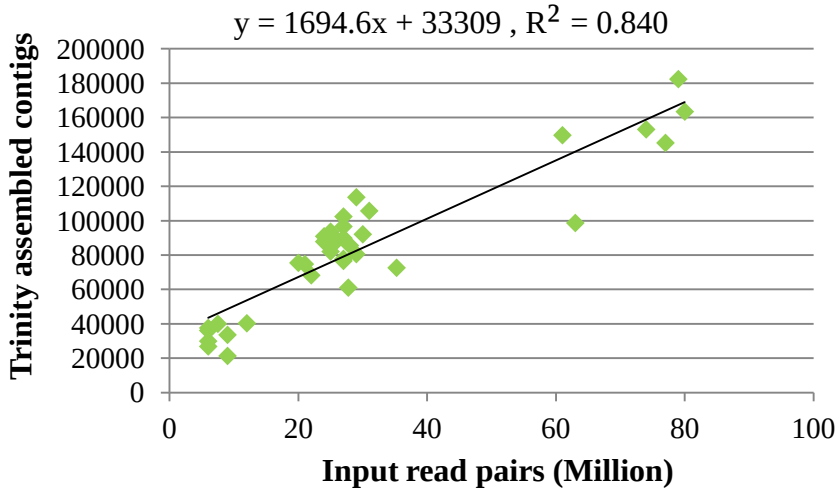
(b)



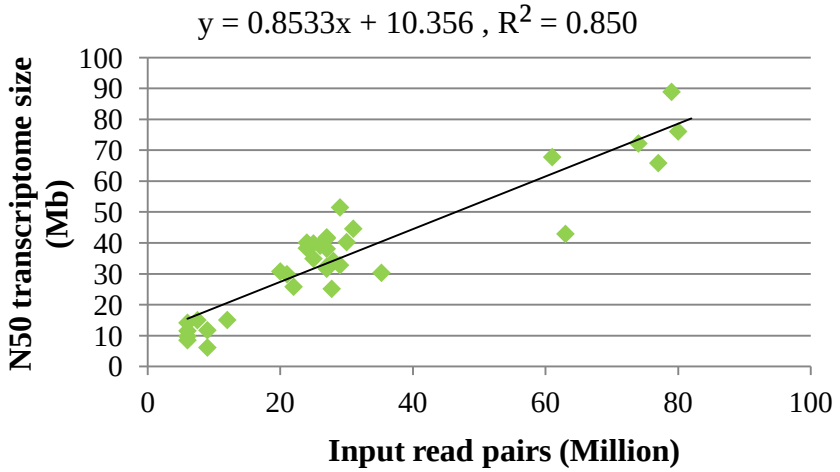
(c)

Figure 3.4 Correlations between input read pairs and other assembly statistics.
Figures (a), (b), (c) were based on single sample assembly. Figures (d), (e), (f) were based on single and merged assemblies.

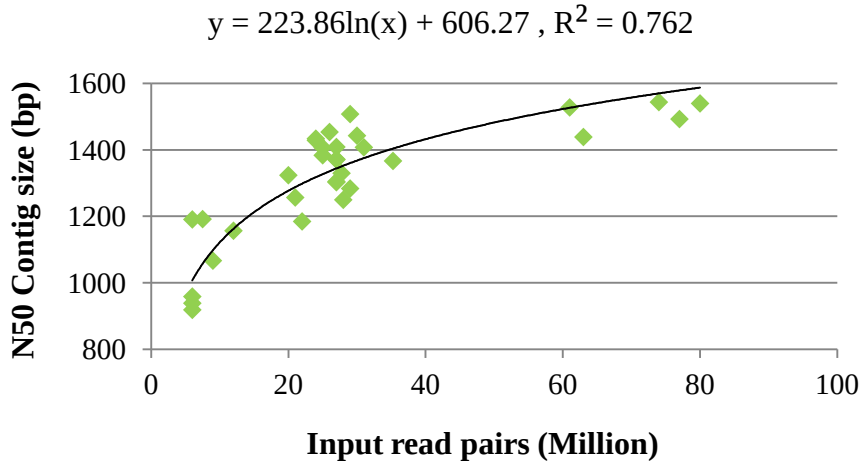
Figure 3.4 (cont'd)



(d)



(e)



(f)

Sample 1-1 was from white wheat. Sample 1-6 are derived from red wheat (Table 3.2). Bowtie2 (Langmead and Salzberg, 2012) was used to map reads back to their own assembly (1-1-1, 1-6-6), to each other's assembly (1-1-6, 1-6-1) and to the bulk or total assembly (1-1-rrr, 1-1-RRR, 1-1-All; 1-6-rrr, 1-6-RRR, 1-6-All) to see if there are any differences in alignment rate (Table 3.7). For both samples, the alignment rate increased when mapped to the merged assembly compared with single assemblies. The smaller alignment difference of sample 1-1 and 1-6 when aligned to assembly of 1-6 versus aligned to assembly of 1-1 may also be explained by this. Moreover, each sample had a higher alignment rate when aligned to its own bulk (1-1-rrr, 1-6-RRR), compared with their counterpart's bulk (1-1-RRR, 1-6-rrr), while no differences were shown when aligned to the total assembly (1-1-All, 1-6-All). In general, the transcriptome assembled from a single sample were comparable with the merged assembly in terms of alignment rates but the merged assembly may capture a broader range of transcripts which is indicated by the increased alignment rate.

Bowtie2 alignment indicated that around 90% of the reads can be aligned to assembly without significant bias to a specific sample. This alignment may be biased by highly expressed transcripts while the sample-specific transcripts may have a lower abundance. In order to understand what kinds of transcripts tend to be specifically aligned to one sample, the single sample assemblies were searched against the bulk and total assemblies using BLASTn (Table 3.8). When single sample assembly was searched against total assembly (1-All), around 99% of their transcripts had a hit in the total assembly. However, not all the transcripts had a unique hit in the total assembly. This might be explained by the different splicing forms or a potential mis-assembly of transcripts from different samples. For a single sample, the percentage of its BLAST matches in total assembly was highly correlated to the number of

Table 3.7 Bowtie2 alignment for single samples to Trinity assemblies with different assembly strategies

Sample	Overall mapped (%)	Properly paired (%)	Sample	Overall mapped (%)	Properly paired (%)
1-1-1	91.3	70.0	1-6-6	92.1	68.0
1-6-1	89.7	68.0	1-1-6	91.3	68.0
1-1-rrr*	93.6	66.6	1-1-RRR#	93.4	66.2
1-6-rrr	92.9	65.5	1-6-RRR	93.8	66.0
1-1-All^	94.4	65.1	1-6-All	94.3	64.1

* rrr was assembled using reads from 1-1, 1-2, 1-7, 1-8;

RRR was assembled using reads from 1-3, 1-4, 1-5, 1-6;

^ All was assembled using reads from all 8 samples.

Table 3.8 MegaBLAST results for single sample assembly against bulk (1-rrr, 1-RRR) and total assembly (1-All)

Sample*	No. of Contigs	Had Hit In 1-All#	Targets in 1-All^ (98,059)	Had hit In 1-rrr	Targets in 1-rrr (60,745)	Had hit in 1-RRR	Targets in 1-RRR (72,163)
1-1	36109	99.0	22.1	98.8	36.5	97.8	29.5
1-2	26830	99.4	17.4	99.2	28.2	98.7	23.4
1-3	40316	99.1	26.7	94.4	41.4	98.9	37.0
1-4	39999	99.0	26.8	93.7	41.5	98.9	37.2
1-5	21323	99.4	13.8	98.9	21.8	99.3	18.7
1-6	37595	99.1	24.9	94.8	39.0	99.0	34.4
1-7	29919	99.3	19.7	99.1	31.8	98.3	26.1
1-8	33576	99.2	22.1	99.0	36.5	97.7	29.3
1-rrr	61036	98.7	42.4	NA	NA	89.2	49.6
1-RRR	72583	98.4	52.5	92.7	49.4	NA	NA

* Sample 1-1 to 1-8 represented single sample data set; 1-rrr was assembled using reads from 1-1, 1-2, 1-7, 1-8; 1-RRR was assembled using reads from 1-3, 1-4, 1-5, 1-6;

The number shown below is the percentage of unique transcripts in single sample assembly had a hit in 1-All, which was assembled from all 8 samples;

^ The number shown below is the percentage of unique targets in 1-All.

contigs within each single sample assembly (R^2=0.98). This indicated the effectiveness of total assembly in capturing sample specific transcripts. The same correlation was shown when the bulk assemblies (1-rrr, 1-RRR) were searched against the total assembly (1-all).

The low percentage of BLAST matches of a single sample assembly against total assembly may indicate two scenarios: 1. each single sample had similar hits in the total assembly, which can be a core set of highly expressed transcripts that expressed in all samples while the rest transcripts are sample specific, or 2. each single sample hit different transcripts in the total assembly, which might indicate a very diverse expression patterns across biological replicates. Based on previous bowtie2 alignment result, the first scenario was more reasonable. To confirm its validity, all the single samples were searched against each other by BLASTn to measure the overlap of transcripts between single samples. The results showed a high percentage (60-95%) of overlap between samples. Moreover, a frequency count of transcripts in total assembly hit by single samples showed that around 13,500 transcripts were hit by six or more samples, which represented 50-99% of transcripts shared by all eight samples. Therefore, the low BLAST match by single sample in total assembly was mainly due to the diversity of samples in the total assembly, which might be due to biological reason or insufficient sequencing depth.

In conclusion, when sequencing depth is sparse, a single sample assembly will only cover the highly expressed genes. The merge-and-assemble strategy can be an effective strategy for *de novo* assembly to include most transcripts expressed. However, if read depths are sufficient, a bulk merge of all the biological replicates or individuals having similar phenotypes may gave a more accurate representation of bulk-specific transcripts by reducing the risk of false-merging of bulk-specific transcripts, especially in de novo assembly, and reducing the demand for computation resources.

RNA-seq transcriptome profiles and DE calling

The "Tuxedo" pipeline, Tophat and Cufflinks, was used in this study with *Aegilops tauschii* draft genome as reference (Jia et al., 2013; Trapnell et al., 2012). Samples were sequenced on paired end module with 75bp and 100bp run lengths. Genes that had FPKM values larger than the lower bound of the 95 percentile of FPKM distributions were considered as expressed. Correlation between input read amounts and genes expressed was shown in both datasets: R^2 for 75bp PE reads set equals 0.68 and R^2 for 100 bp PE reads set equals 0.81. To identify DE transcripts, Cuffdiff2 was used based on the log-fold-change in transcripts expression against the null hypothesis of no change accounting for read mapping and assignment uncertainty (Trapnell et al. 2013). Furthermore, when sequencing depth is low, the lowly expressed transcripts may only showed in one sample but not the other (based on the FPKM cutoff). Thus the statistical test for DE calling can be biased. Thus, in order to reduce the false positives, DE was called not only when it met software's test statistic threshold, but also required FPKM in both samples to be above the threshold expression level. Simulation results showed that increasing biological replicates was more effective in terms of gaining statistical power for calling DE transcripts when compared with increasing library/technical replicates or sequencing depth. Sequencing depth could be reduced as low as 15% of the original depth without substantial impacts on false positive or true positive rates (Robles et al., 2012). Hence, the use of a more conservative DE calling criterion and including multiple biological replicates in comparison can help reduce the false positive rate and the impact of a potential lack of depth.

Transcriptome profiling during PHS induction stage

Red wheat is known to be more resistant to PHS compared with white wheat (Gale and Lenton, 1987). Previous QTL mapping results (Chapter 2) identified a QTL for α-amylase

activity co-located with the seed coat color loci on chromosome 3A. In current experiment, individuals from the same QTL mapping population were selected at physiological maturity to explore the relationship between seed coat color and PHS resistance at the transcriptome level. PHS resistance was measured in terms of the expression of genes involved in germination process. The correlation of gene expression pattern between biological replicates for both the white wheat group (0 hr: 4 replicates; 48 hr: 3 replicates) and the red wheat group (0 hr: 4 replicates, 48 hr: 2 replicates) during the misting process were summarized in Table 3.9. The highest correlation was found between biological replicates within the same seed color and same misting treatment. The correlations between red wheat and white wheat at 0 hr were generally higher than the correlation between them at 48 hr. This is intuitive since the base genetics without treatment should be expected to have a higher correlation.

Four comparisons were conducted in this study to explore the common and differential expression patterns of red wheat versus white wheat in response to a 48 hr misting treatment (Table 3.1). Comparison A was aimed to draw the baseline of the genetic background differences between red wheat and white wheat before misting. Comparison B was to identify the DE transcripts between red wheat and white wheat after 48 hr misting treatment. This treatment was shown to be effective in inducing PHS in greenhouse based on elevated α-amylase activity (Chapter 2). In theory, there are two major categories of transcripts can be the potential targets in response to misting treatment: 1. the transcripts showed as non-DE in comparison A, but showed as DE in comparison B; 2. the DE transcripts shown in both comparisons but with opposite regulation patterns. There were three unique transcripts in the category 1 while none in the category 2 (Table B.2).

Table 3.9 Sample correlations between biological replicates used for differential expression

Sample correlation^	W-0 hr-1*	W-0 hr-2	W-0 hr-3	W-0 hr-4	W-48hr-1**	W-48hr-2	W-48hr-3	R-0 hr-1	R-0 hr-2	R-0 hr-3	R-0 hr-4	R-48hr-1	R-48hr-2
W-0 hr-1	1.00												
W-0 hr-2	0.99	1.00											
W-0 hr-3	0.99	1.00	1.00										
W-0 hr-4	0.99	1.00	1.00	1.00									
W-48hr-1	0.99	0.97	0.97	0.97	1.00								
W-48hr-2	0.95	0.92	0.90	0.92	0.97	1.00							
W-48hr-3	0.99	0.99	0.98	0.99	0.98	0.96	1.00						
R-0 hr-1	0.99	1.00	1.00	1.00	0.97	0.91	0.98	1.00					
R-0 hr-2	1.00	0.99	0.99	1.00	0.98	0.93	0.99	1.00	1.00				
R-0 hr-3	0.96	0.98	0.97	0.98	0.93	0.90	0.97	0.97	0.97	1.00			
R-0 hr-4	0.99	1.00	0.99	1.00	0.97	0.92	0.98	1.00	1.00	0.97	1.00		
R-48hr-1	0.99	0.97	0.96	0.97	0.99	0.98	0.99	0.96	0.98	0.95	0.97	1.00	
R-48hr-2	0.91	0.86	0.83	0.86	0.94	0.98	0.92	0.84	0.88	0.82	0.86	0.95	1.00

* W(R)-0 hr-1: White (Red) wheat, 0 hr misting, 1-4 is biological replicates;

** W(R)-48hr-1: White (Red) wheat, 48 hr after misting, 1-3 is biological replicates.

On the other hand, the DE transcripts uniquely shown in A, and DE transcripts that were common in comparisons A and B with same regulation pattern can be related to seed coat color differences. Transcripts in this category included WRKY and MADS-box transcription factor. For comparisons C and D, the genetic background effects were minimized since the comparisons were done among red wheat samples (comparison C) and among white wheat samples (comparison D). The only difference within each comparison was the misting treatment. There are seventy-two DE transcripts that were common between comparisons C and D can be attributed to how wheat respond to misting treatment while DE transcripts specific within each group can be considered as their own response to misting (seven in red, five in white) (Table B.2). The low amount of common DE transcripts between comparison C and D may indicate the different gene networks involved in red and white wheat during their germination process.

Annotation of differentially expressed genes

The DE transcripts were assessed by testing the observed log-fold-change in its expression against the null hypothesis of no significant change. To adjust for multiple-testing, an FDR of 0.05 was used as the threshold. The DE transcripts can be further divided into two categories depended on the gene expression was up or down regulated between the two groups involved in comparisons. In this study, only about 1% of the transcripts were reported as differentially expressed at 0 hr when compared between red and white wheat. However, the comparison of the seed transcripts before and after misting treatment (Comparison C and D) showed an approximately 10% of the genes differentially expressed (Table B.3), which was comparable with a similar study conducted in pear, a 2-7% DE transcripts during dormancy break process (Liu et al., 2012). Another study conducted using microarray to compare Arabidopsis seeds with and without priming reported a 20% of DE genes (Ligternink et al. 2007).

This difference might be due to the stronger effects of priming treatment and the seed sensitivity. The differentially expressed genes from comparison A and B were further traced back to their biological functions based on the annotation files of *Aegilops tauschii* genome (Dr. Chi Zhang, BGI, *pers. comm.*, Table B.4). A total of 15,229, 73.2%, expressed loci, were assigned gene ontology (GO) terms. The increased amount of DE transcripts when compared between 0 hr and 48 hr indicated the initiation of biological processes in response to either misting treatment or germination. Many DE transcripts were involved in iron transport and DNA/protein binding process, which indicated the activation of major biological events. The absence of candidate genes involved in hormone balance pathway as differentially expressed genes may be due to the sampling stage. In comparison A, a transcript (TCONS_00081538) coding for late embryogenesis abundant (LEA) protein was found to have a differential expression for its multiple isoforms. Its expression level in white wheat is about 6 times higher than red wheat under non-mist condition. After misting for 48 hr, the expression level became similar between red and white wheat. LEA protein is known to be involved in water-binding, to maintain protein and membrane structure. It accumulates during late embryogenesis and then disappears during germination (Gallardo et al., 2001). Therefore, the significant reduction of LEA expression in comparison A may indicate that the germination process started in white wheat but not in red wheat. In comparison B, genes related to a hydrolase family was shown as differentially expressed. Hydrolase activity is important in seed germination by weakening the cell wall of endosperm (Leubner-Metzger et al., 1995). White wheat showed a higher hydrolase activity, which may indicate a more advanced stage into germination compared with red wheat.

There were also several transcription factors (TFs) found differentially expressed in comparisons A and B (Table 3.1). The gene of one DE transcript was assigned to the CCCH-

Type Zinc Finger Family, which was previously identified to be involved in ABA and drought stress in maize (Peng et al., 2012). It was up-regulated in red wheat more significantly than white wheat after misting treatment. Other TFs identified with similar function are basic leucine zipper (bZIP) and myeloblastosis (MYB). The earlier one was known to be involved in *Arabidopsis* abscisic acid signal transduction pathway under drought stress (Uno et al. 2000); while the later one was identified in rice panicle under drought stress (Gorantla et al. 2007). Another TFs found were WRKY, which is known as the TF for multiple stress tolerance process and regulated in a ABA-dependent manner (Zhu et al. 2013). These TFs can be the connecting dots to other transcription networks involved in germination process.

GO enrichment of differentially expressed genes

GO enrichment of differentially expressed transcripts was conducted following the method described in Miller et al. (2010). A total of 95 GO terms from four comparisons were identified as significantly enriched (Fisher's exact test, adjusted p-value < 0.01; Table 3.10 and Table 3.11). In general, major categories of molecular function terms enriched in DE genes including different types of binding proteins, transmembrane proteins and protein kinases were identified. All of them are major players in signal transduction and processes like germination. Proteins involved in oxidation reduction process, such as thioredoxin, were more germination specific and had been reported as key regulators in germination. Several studies had been done to explore its function in improving PHS resistance in cereals (Guo et al. 2013; Ren et al. 2007). In this study multiple GO terms were shown to be both up and down regulated, however they are usually related to different proteins.

Table 3.10 Gene Ontology (GO) categories significantly enriched in Up-regulated wheat genes in four comparisons

GO		Annotation	Comparison	DE loci		Non-DE loci		P-value[e]	FDR[f]	Trend[g]
				GO[a]	Non-GO[b]	GO[c]	Non-GO[d]			
GO:0003676	mf[h]	nucleic acid binding	A	2	27	1	6374	5.92E-05	1.89E-04	Up
GO:0005515	mf	protein binding	A	2	27	1	6374	5.92E-05	1.89E-04	Up
GO:0005524	mf	ATP binding	A	2	27	1	6374	5.92E-05	1.89E-04	Up
GO:0008270	mf	zinc ion binding	A	2	27	1	6374	5.92E-05	1.89E-04	Up
GO:0006468	bp[i]	protein phosphorylation	B	4	315	1	12454	1.87E-06	9.10E-06	Up
GO:0055085	bp	transmembrane transport	B	3	316	1	12454	6.06E-05	1.90E-04	Up
GO:0055114	bp	oxidation-reduction process	B	7	312	1	12454	4.45E-11	5.15E-10	Up
GO:0016020	cc[j]	membrane	B	5	314	1	12454	5.54E-08	4.46E-07	Up
GO:0003677	mf	DNA binding	B	4	315	1	12454	1.87E-06	9.10E-06	Up
GO:0004672	mf	protein kinase activity	B	4	315	1	12454	1.87E-06	9.10E-06	Up
GO:0004713	mf	protein tyrosine kinase activity	B	4	315	1	12454	1.87E-06	9.10E-06	Up
GO:0005506	mf	iron ion binding	B	4	315	1	12454	1.87E-06	9.10E-06	Up
GO:0005515	mf	protein binding	B	6	313	2	12453	6.22E-09	5.48E-08	Up
GO:0005524	mf	ATP binding	B	5	314	1	12454	5.54E-08	4.46E-07	Up
GO:0009055	mf	electron carrier activity	B	2	317	1	12454	0.001834	3.69E-03	Up
GO:0016705	mf	oxidoreductase activity	B	2	317	1	12454	0.001834	3.69E-03	Up
GO:0020037	mf	heme binding	B	4	315	1	12454	1.87E-06	9.10E-06	Up
GO:0003676	mf	nucleic acid binding	C	11	2911	1	10053	7.09E-07	4.67E-06	Up
GO:0003677	mf	DNA binding	C	18	2904	1	10053	3.19E-11	4.03E-10	Up

Table 3.10 (cont'd)

GO:0003700	mf	sequence-specific DNA binding transcription factor activity	C	19	2903	1	10053	7.51E-12	1.16E-10	Up
GO:0003824	mf	catalytic activity	C	10	2912	1	10053	2.90E-06	1.34E-05	Up
GO:0004553	mf	hydrolase activity, hydrolyzing O-glycosyl compounds	C	6	2916	1	10053	0.000734	1.86E-03	Up
GO:0004672	mf	protein kinase activity	C	7	2915	1	10053	0.000188	5.69E-04	Up
GO:0004713	mf	protein tyrosine kinase activity	C	6	2916	1	10053	0.000734	1.86E-03	Up
GO:0005506	mf	iron ion binding	C	18	2904	1	10053	3.19E-11	4.03E-10	Up
GO:0005515	mf	protein binding	C	25	2897	2	10052	1.30E-14	3.44E-13	Up
GO:0006355	bp	regulation of transcription, DNA-dependent	D	12	1552	137	7122	0.001067	2.63E-03	Up
GO:0006468	bp	protein phosphorylation	D	2	1562	177	7082	6.08E-13	1.25E-11	Up
GO:0006508	bp	proteolysis	D	2	1562	64	7195	0.000515	1.38E-03	Up
GO:0008152	bp	metabolic process	D	2	1562	115	7144	6.18E-08	4.76E-07	Up
GO:0055085	bp	transmembrane transport	D	4	1560	85	7174	0.000378	1.03E-03	Up
GO:0055114	bp	oxidation-reduction process	D	12	1552	184	7075	2.60E-06	1.23E-05	Up
GO:0005622	cc	intracellular	D	4	1560	99	7160	3.95E-05	1.43E-04	Up
GO:0005634	cc	nucleus	D	5	1559	123	7136	3.43E-06	1.55E-05	Up
GO:0016020	cc	membrane	D	5	1559	171	7088	1.43E-09	1.47E-08	Up
GO:0016021	cc	integral to membrane	D	1	1563	92	7167	3.79E-07	2.60E-06	Up
GO:0003676	mf	nucleic acid binding	D	4	1560	104	7155	1.87E-05	7.21E-05	Up
GO:0003677	mf	DNA binding	D	6	1558	176	7083	2.95E-09	2.87E-08	Up

Table 3.10 (cont'd)

GO:0003723	mf	RNA binding	D	2	1562	51	7208	0.003649	7.26E-03	Up
GO:0003824	mf	catalytic activity	D	6	1558	94	7165	0.000885	2.21E-03	Up
GO:0004672	mf	protein kinase activity	D	2	1562	177	7082	6.08E-13	1.25E-11	Up
GO:0004713	mf	protein tyrosine kinase activity	D	2	1562	146	7113	2.66E-10	2.89E-09	Up
GO:0005515	mf	protein binding	D	14	1550	431	6828	3.46E-22	2.13E-20	Up
GO:0005524	mf	ATP binding	D	6	1558	337	6922	1.04E-21	4.81E-20	Up
GO:0008270	mf	zinc ion binding	D	9	1555	168	7091	7.32E-07	4.67E-06	Up
GO:0009055	mf	electron carrier activity	D	4	1560	82	7177	0.000534	1.41E-03	Up
GO:0016491	mf	oxidoreductase activity	D	5	1559	83	7176	0.00173	3.69E-03	Up
GO:0020037	mf	heme binding	D	4	1560	75	7184	0.001594	3.60E-03	Up

[a]GO, Number of significantly up-regulated genes with the GO annotation in question;

[b]Non-GO, Number of significantly up-regulated genes without the GO annotation;

[c]GO, Number of genes without significant expression change with the GO annotation;

[d]Non-GO, Number of genes with no significant expression change that do not have the GO annotation;

[e]Fisher's exact test P value. [f]The FDR value is calculated with R package with Benjamini & Hochberg (1995) method;

[g]Trend: FPKM value comparison. For A (0 hr), B (48 hr): White < Red; for comparison C (Red), D (White): 0 hr <48 hr;

[h]mf, Molecular function;

[i]cc, Cellular component;

[j]bp, Biological process.

Table 3.11 Gene Ontology (GO) categories significantly enriched in Down-regulated wheat genes in four comparisons

GO		Annotation	Comparison	DE loci		Non-DE loci		P-value[e]	FDR[f]	Trend[g]
				GO[a]	Non-GO[b]	GO[c]	Non-GO[d]			
GO:0005515	mf[h]	protein binding	A	2	174	546	8050	0.001461	3.34E-03	Down
GO:0055114	bp[i]	oxidation-reduction process	B	3	210	1	10348	3.19E-05	1.20E-04	Down
GO:0005506	mf	iron ion binding	B	2	211	1	10348	0.001198	2.77E-03	Down
GO:0005524	mf	ATP binding	B	2	211	1	10348	0.001198	2.77E-03	Down
GO:0009055	mf	electron carrier activity	B	2	211	1	10348	0.001198	2.77E-03	Down
GO:0016705	mf	oxidoreductase activity	B	2	211	1	10348	0.001198	2.77E-03	Down
GO:0020037	mf	heme binding	B	2	211	1	10348	0.001198	2.77E-03	Down
GO:0016020	cc[j]	membrane	C	3	1119	1	9132	0.004798	9.34E-03	Down
GO:0003676	mf	nucleic acid binding	C	14	1108	2	9131	3.18E-12	5.35E-11	Down
GO:0003677	mf	DNA binding	C	7	1115	1	9132	1.34E-06	7.75E-06	Down
GO:0003723	mf	RNA binding	C	12	1110	1	9132	3.27E-11	4.03E-10	Down
GO:0003735	mf	structural constituent of ribosome	C	7	1115	1	9132	1.34E-06	7.75E-06	Down
GO:0003824	mf	catalytic activity	C	6	1116	1	9132	1.08E-05	4.25E-05	Down
GO:0004672	mf	protein kinase activity	C	7	1115	1	9132	1.34E-06	7.75E-06	Down
GO:0004713	mf	protein tyrosine kinase activity	C	6	1116	2	9131	3.90E-05	1.43E-04	Down
GO:0005515	mf	protein binding	C	35	1087	4	9129	7.67E-30	7.09E-28	Down
GO:0005524	mf	ATP binding	C	5	1117	1	9132	8.48E-05	2.61E-04	Down
GO:0008270	mf	zinc ion binding	C	4	1118	1	9132	0.000651	1.70E-03	Down
GO:0043565	mf	sequence-specific DNA binding	C	3	1119	1	9132	0.004798	9.34E-03	Down

Table 3.11 (cont'd)

GO:0006355	bp	regulation of transcription, DNA-dependent	D	7	1556	1	9402	8.28E-06	3.33E-05	Down
GO:0006468	bp	protein phosphorylation	D	7	1556	1	9402	8.28E-06	3.33E-05	Down
GO:0006508	bp	proteolysis	D	5	1558	1	9402	0.000309	8.54E-04	Down
GO:0006511	bp	ubiquitin-dependent protein catabolic process	D	4	1559	1	9402	0.001823	3.69E-03	Down
GO:0006886	bp	intracellular protein transport	D	7	1556	1	9402	8.28E-06	3.33E-05	Down
GO:0006952	bp	defense response	D	6	1557	1	9402	5.11E-05	1.75E-04	Down
GO:0008152	bp	metabolic process	D	4	1559	1	9402	0.001823	3.69E-03	Down
GO:0009058	bp	biosynthetic process	D	5	1558	1	9402	0.000309	8.54E-04	Down
GO:0016192	bp	vesicle-mediated transport	D	4	1559	1	9402	0.001823	3.69E-03	Down
GO:0005622	cc	intracellular	D	6	1557	1	9402	5.11E-05	1.75E-04	Down
GO:0005634	cc	nucleus	D	9	1554	1	9402	2.08E-07	1.48E-06	Down
GO:0016020	cc	membrane	D	4	1559	1	9402	0.001823	3.69E-03	Down
GO:0016021	cc	integral to membrane	D	4	1559	1	9402	0.001823	3.69E-03	Down
GO:0003676	mf	nucleic acid binding	D	16	1547	2	9401	3.12E-12	5.35E-11	Down
GO:0003677	mf	DNA binding	D	9	1554	1	9402	2.08E-07	1.48E-06	Down
GO:0003700	mf	sequence-specific DNA binding transcription factor activity	D	4	1559	1	9402	0.001823	3.69E-03	Down
GO:0003723	mf	RNA binding	D	5	1558	1	9402	0.000309	8.54E-04	Down
GO:0003824	mf	catalytic activity	D	11	1552	1	9402	5.00E-09	4.63E-08	Down
GO:0004672	mf	protein kinase activity	D	7	1556	1	9402	8.28E-06	3.33E-05	Down

Table 3.11 (cont'd)

GO:0004713	mf	protein tyrosine kinase activity	D	7	1556	1	9402	8.28E-06	3.33E-05	Down
GO:0005515	mf	protein binding	D	45	1518	4	9399	5.75E-34	1.06E-31	Down
GO:0005524	mf	ATP binding	D	19	1544	1	9402	1.32E-15	4.07E-14	Down
GO:0005525	mf	GTP binding	D	5	1558	1	9402	0.000309	8.54E-04	Down
GO:0008270	mf	zinc ion binding	D	22	1541	1	9402	4.26E-18	1.58E-16	Down
GO:0009055	mf	electron carrier activity	D	5	1558	1	9402	0.000309	8.54E-04	Down
GO:0016787	mf	hydrolase activity	D	4	1559	1	9402	0.001823	3.69E-03	Down
GO:0030170	mf	pyridoxal phosphate binding	D	5	1558	1	9402	0.000309	8.54E-04	Down
GO:0043531	mf	ADP binding	D	6	1557	1	9402	5.11E-05	1.75E-04	Down

[a]GO, Number of significantly down-regulated genes with the GO annotation in question;

[b]Non-GO, Number of significantly down-regulated genes without the GO annotation;

[c]GO, Number of genes without significant expression change with the GO annotation;

[d]Non-GO, Number of genes with no significant expression change that do not have the GO annotation;

[e]Fisher's exact test P value;

[f]The FDR value is calculated with R package with Benjamini & Hochberg (1995) method;

[g]Trend: : FPKM value comparison. For A (0 hr), B (48 hr): White > Red; for comparison C (Red), D (White): 0 hr >48 hr;

[h]mf, Molecular function;

[i]cc, Cellular component;

[j]bp, Biological process.

For example in comparison A, the two that were up-regulated, one was a Zinc-Finger CCCH protein and the other was Lysine-specific demethylase 3B, which was responsible for histone demethylation and epigenetic transcriptional regulation. The BLASTp search showed the down-regulated protein belongs to thioredoxin like superfamily. Thioredoxin (Trx) was reported as a catalyst for germination and α-amylase activation (Wong et al. 2002). Several transgenic studies manipulated Trx gene expression level proved its direct impact to the germination speed (Guo et al. 2013; Ren et al. 2007). The higher the Trx content, the easier the seed to germinate. Therefore, the higher Trx in white wheat may partially explain why white wheat is more susceptible to PHS. In general, comparison A had the smallest percentage (0.6%) of differentially expressed genes (Table B.3). But the differentially expressed "binding" related loci indicated the induction of biological processesmight be able to explain the potential physiological difference between red and white wheat at the starting point.

In comparison B, "membrane" is the only GO term that belonged to "cellular components" that is up-regulated, which indicate a more drastic change in white wheat than red wheat after 48 hr of misting. By reviewing the BLASTp hit of the loci matching this GO term, multiple hits were shown to be closely related to germination process. Out of the five loci with the same GO, three are transporter related proteins. The other two are pentatricopeptide repeat (PPR) and transparent testa 12 (tt12) proteins. PPR was recently shown to be enriched in Arabidopsis germination specific gene set and only transiently expressed during germination process (Narsai et al. 2011), while *tt12* is known to be related to seed dormancy (Debeaujon et al. 2000). The lower the *tt12*, like here in white wheat, the lower dormancy it would have when compared with red wheat.

When compared between comparison C (red wheat) and D (white wheat), white wheat had a more comprehensive set of GO terms enriched, which indicated a more advanced stage in germination due to the various biological pathways activated. Further pathway analysis will be beneficial to understand the genetic network underlying the current induction process, while the proteins underlying each enriched GO term will be valuable for mining of PHS resistance gene.

DE transcripts without annotations

About 50% of DE transcripts identified from current study had putative biological function related to germination process (Figure 3.3). However, the rest of the DE transcripts were mapped to the nearby regions of annotated genes. These genes were ignored in current analysis but maybe useful due to several reasons. First, Cufflinks has the ability to assemble novel transcripts and the novel transcripts might contribute to some part of these mapped DE transcripts. Secondly, the reference genome used was a draft release that was likely to miss a number of genes. Thus, our results can even be used to improve the annotation. Lastly, a diploid progenitor was used as the reference genome. The genetic distance between common wheat genome and *Aegilops tauschii* genome might also affect the final mapping results, which is indicated by a relatively low Tophat mapping rates. This diploid genome can be considered as a simplified version of wheat genome but the complexity will be much less than the hexaploid wheat. Therefore, there is also a chance that the transcripts mapped to nearby exonic regions are putative splicing forms in common wheat but not captured in the current genome annotation. Similar situations have been reported in model species *C. reinhardtii* as “intergenic transcriptional unit” by Miller et al. (2010). Further analysis is needed to understand impact of using diploid genome as reference for polyploidy study and the origin of the “missing” transcripts.

Conclusions

Wheat as a commodity crop has been bred for fast and uniform germination unintentionally since its domestication. Seed with low dormancy is more susceptible to problems such as pre-harvest sprouting. Previous studies focused on evaluating the sprout damage which could not resolve the molecular events leading to the initiation of the PHS process. Thus, an understanding of the initial stage through expression profiling may be critical to understand how plants interact with this environmental change and ultimately help us understand the network involved in the process. In this study, NGS technology was used to characterize the transcriptome of red wheat and white wheat during a PHS induction treatment (misting). In order to build a set of high confidence transcripts, Genome Guided and *de novo* assembly strategies were evaluated with other public databases and compared between each other for the transcript concordance, completeness, and contiguity. The results showed that both assemblies had more than 90% of transcripts were authentic wheat transcripts while a good concordance was shown between the two methods for around 50% of their transcripts assembled. When the transcripts were compared with cDNA libraries with similar growth stage or treatments, more than 60% of the cDNA were shown in both assembled transcripts. When the two assemblies were compared with a comprehensive set of wheat ESTs, both assemblies showed a lower representation but were still in line with previous studies done in seed germination field. In current study, the GG assembly was done with the *A. tauschii* genome instead of hexploid wheat genome, which is not available at current stage. This approximation may affect the quality of GG to some extent and the non-specific read alignment rate may be one of the phenomenon. Further investigation is required to understand the use of diploid genome for polyploidy genome assembly.

Due to the fast growth of NGS, genotyping by sequencing at population level has been tested in multiple crops, such as maize (Elshire et al. 2011) and complex polyploid like switchgrass, but wheat has not been tested. With current technology, it may still not be feasible to produce sequencing depth that is high enough to identify most lowly expressed transcripts or rare alleles. Therefore, current study compared the assemblies from three different assembly strategies using different amount of reads for sequence redundancy and completeness. The comparison had a direct application for researchers want to do multiplexing to reduce experimental cost and increase throughput. Our results confirmed the fact that the more input reads, the more transcripts you will get from a in a *de novo* assembly. The comparison between single sample assembly and merged sample assembly suggested that when sequencing depth is shallow, the assembly can only cover highly expressed transcripts, so it is important to merge samples in order to get a better coverage of the transcripts expressed during this stage. When sequencing depth is sufficient, merged sample assembly didn't significantly increase the redundancy when compared with a single sample assembly, while there is a potential to falsely merge fragmented transcripts together. Moreover, with the more reads feeding into the assembler, the computation power required to assemble it can quickly become the next challenge. The assembly strategy that merged biological replicates or bulk individuals with similar phenotype prior to assembly was favored here. This strategy can provide a good balance between keeping sample-specific transcripts and eliminating unnecessary fusion transcripts while leveraging more reads to generate a better assembly covering most transcripts expressed in the merged bulk.

In the end, differential expression and GO enrichment of DE transcripts were performed. Comparison of common or unique transcripts between comparisons confirmed multiple candidate genes involved in germination process, most of which were identified by previous

studies. However, further validation of DE transcripts is needed before making final conclusion. The series of studies conducted here showed the efficiency of using RNA-seq to identify potential gene network behind complex phenotypes such as PHS. The set of transcripts assembled with Trinity can be incorporated to public wheat transcripts while DE transcripts can be used for mining of candidate genes and gene networks for PHS resistance.

APPENDIX

Table B.1 BLASTn results of 28 Trinity assembled transcripts longer than 10,000bp

query	Origin	subject	%id	q len	mismatch	gap	qs	qe	ss	se	E	bit
comp46213_c0_seq1	DV	gi514709750	85.67	9768	1322	54	442	10185	9914	201	0	10211
comp46213_c0_seq2	DV	gi514709750	85.67	9768	1322	54	624	10367	9914	201	0	10211
comp46213_c0_seq4	DV	gi514709750	85.67	9768	1322	54	560	10303	9914	201	0	10211
comp46213_c0_seq5	DV	gi514709754	85.89	5585	752	25	624	6199	8946	3389	0	5915
comp46213_c0_seq6	DV	gi514709754	85.89	5585	752	25	560	6135	8946	3389	0	5915
comp46258_c0_seq2	DV	gi357166713	90.33	10847	1019	22	2	10830	739	11573	0	14192
comp46260_c0_seq1	DV	gi357150792	90.65	10540	920	44	112	10632	13316	2823	0	13947
comp46260_c0_seq2	DV	gi357150792	91.14	8390	714	18	1	8376	11197	2823	0	11350
comp46260_c0_seq3	DV	gi357150792	91.14	8390	714	18	1	8376	11197	2823	0	11350
comp46260_c0_seq4	DV	gi357150792	90.65	10540	920	44	112	10632	13316	2823	0	13947
comp46261_c0_seq1	DV	gi357150722	89.96	11006	1044	21	238	11214	1	10974	0	14146
comp46261_c0_seq2	DV	gi357150722	90.09	10988	1046	14	238	11196	1	10974	0	14218
comp46277_c0_seq1	DV	gi357140567	87.13	7071	856	26	9136	16177	8361	15406	0	7967
comp46277_c0_seq3	DV	gi357140567	87.13	7071	856	26	9050	16091	8361	15406	0	7967
s16289-comp49_c0_seq1	GG	gi357150722	90.09	10988	1049	13	221	11182	1	10974	0	14222
s49346-comp37_c0_seq2	GG	gi357150792	90.95	10423	899	29	197	10604	13216	2823	0	13983
s49472-comp63_c0_seq2	GG	gi357118204	86.02	5522	615	76	3286	8745	6934	1508	0	5775
s50393-comp64_c0_seq1	GG	gi514817801	83.82	6217	948	37	493	6668	11299	5100	0	5854
s55518-comp12_c0_seq11	GG	gi514742213	85.22	9965	1409	46	4	9937	4476	14407	0	10183
s56197-comp29_c0_seq2	GG	gi357116227	92.14	9791	717	24	883	10632	1914	11692	0	13766
s56199-comp29_c0_seq4	GG	gi357116227	92.21	10496	761	28	265	10717	1211	11692	0	14798
s56201-comp29_c0_seq6	GG	gi357116227	92.21	10496	761	28	265	10717	1211	11692	0	14798
s56203-comp29_c0_seq8	GG	gi357116227	92.14	9791	717	24	1963	11712	1914	11692	0	13766
s56204-comp29_c0_seq9	GG	gi357116227	92.14	9791	717	24	1963	11712	1914	11692	0	13766
s56205-comp29_c0_seq10	GG	gi357116227	92.06	11707	870	30	135	11797	1	11692	0	16417
s56206-comp29_c0_seq11	GG	gi357116227	92.14	9791	717	24	883	10632	1914	11692	0	13766
s56207-comp29_c0_seq12	GG	gi357116227	92.06	11707	870	30	135	11797	1	11692	0	16417
s58116-comp170_c0_seq1	GG	gi357140567	87.22	7073	852	28	9107	16154	8361	15406	0	8006

Table B.2 Differentially expressed transcripts with BLASTp and KEGG annotations
Transcripts involved in biological process in response to misting
(Criteria: Common DE in Comparison C and D (yes, |log2| >1), same regulation trend)

Gene loci	Protein	Regulation	BLASTp, KEGG hit
XLOC_001947	AEGTA27601	Down	Os07g0530600; K09773 hypothetical protein
XLOC_002637	AEGTA29317	Down	Os03g0278200; K12450 UDP-glucose 4,6-dehydratase [EC:4.2.1.76]
XLOC_003835	AEGTA13659	Down	hypothetical protein LOC100261060; K13148 integrator complex subunit 11
XLOC_003945	AEGTA26638	Down	hypothetical protein; K03234 elongation factor EF-2 [EC:3.6.5.3]
XLOC_004384	AEGTA33193	Down	Disease resistance protein RPM1, putative (EC:3.1.3.16); K13457 h
XLOC_004857	AEGTA03473	Down	Os02g0709200; K00817 histidinol-phosphate aminotransferase [EC:2.6.1.9]
XLOC_007331	AEGTA32170	Down	hypothetical protein; K03240 translation initiation factor eIF-2B epsilon subunit
XLOC_008042	AEGTA21328	Down	#N/A
XLOC_009705	AEGTA06240	Down	similar to DNA-J; K09518 DnaJ homolog subfamily B member 12
XLOC_010154	AEGTA04881	Down	#N/A
XLOC_017331	AEGTA27691	Down	hypothetical protein ; K03671 thioredoxin 1
XLOC_023567	AEGTA20698	Down	Os05g0353400; K14293 importin subunit beta-1
XLOC_024149	AEGTA15798	Down	MRPR1; NBS-LRR type disease resistance protein; K13457
XLOC_026354	AEGTA09072	Down	60S ribosomal protein L44; K02929 large subunit ribosomal protein L44e
XLOC_031084	AEGTA06575	Down	#N/A
XLOC_031986	AEGTA15563	Down	hypothetical protein; K01952 phosphoribosylformylglycinamidine synthase
XLOC_034101	AEGTA10192	Down	hypothetical protein LOC100252764; K13457 disease resistance protein RPM1
XLOC_035455	AEGTA30415	Down	hypothetical protein; K14301 nuclear pore complex protein Nup107
XLOC_037451	AEGTA15768	Down	nucleolar GTP-binding protein, putative; K06943 nucleolar GTP-binding protein
XLOC_043418	AEGTA27224	Down	hypothetical protein; K12382 saposin
XLOC_043506	AEGTA22028	Down	Os03g0117200; K11752
XLOC_044135	AEGTA20589	Down	similar to predicted protein; K09667 polypeptide N-acetylglucosaminyltransferase
XLOC_044528	AEGTA08085	Down	protein binding; K12821 pre-mRNA-processing factor 40
XLOC_045838	AEGTA00465	Down	#N/A

Table B.2 (cont'd)

Gene loci	Protein	Regulation	BLASTp, KEGG hit
XLOC_001266	AEGTA21145	Up	Os11g0456300; K03094 S-phase kinase-associated protein 1
XLOC_002814	AEGTA30004	Up	peroxidase, putative; K00430 peroxidase [EC:1.11.1.7]
XLOC_003561	AEGTA27889	Up	hypothetical protein; K14497 protein phosphatase 2C [EC:3.1.3.16]
XLOC_006066	AEGTA16028	Up	hypothetical protein LOC100383693; K09843 (+)-abscisic acid 8'-hydroxylase
XLOC_006511	AEGTA13579	Up	Glycosyltransferase QUASIMODO1, putative; K13648
XLOC_006868	AEGTA13757	Up	EPHX2; epoxide hydrolase 2, cytoplasmic; K08726 soluble epoxide hydrolase
XLOC_007857	AEGTA11337	Up	Os11g0456300; K03094 S-phase kinase-associated protein 1
XLOC_008762	AEGTA32945	Up	#N/A
XLOC_009320	AEGTA22368	Up	Os07g0629000; K09422 myb proto-oncogene protein, plant
XLOC_011302	AEGTA24565	Up	hypothetical protein; K01187 alpha-glucosidase [EC:3.2.1.20]
XLOC_012837	AEGTA06614	Up	Os02g0672200; K11596 argonaute
XLOC_013890	AEGTA06050	Up	#N/A
XLOC_014787	AEGTA29360	Up	#N/A
XLOC_016339	AEGTA01779	Up	DREB2A; DNA binding / transcription activator/ transcription factor; K09286
XLOC_017794	AEGTA31244	Up	ATHB-1 (ARABIDOPSIS HOMEOBOX 1); DNA binding / protein homodimerization; K09338 homeobox-leucine zipper
XLOC_019735	AEGTA26202	Up	hypothetical protein; K00487 trans-cinnamate 4-monooxygenase
XLOC_019806	AEGTA24649	Up	hypothetical protein LOC100253371; K03164 DNA topoisomerase II
XLOC_021127	AEGTA12305	Up	#N/A
XLOC_021706	AEGTA02683	Up	Os04g0494100; K01183 chitinase [EC:3.2.1.14]
XLOC_021742	AEGTA26495	Up	#N/A
XLOC_022405	AEGTA28595	Up	Os03g0700400; K00454 lipoxygenase [EC:1.13.11.12]
XLOC_023590	AEGTA07926	Up	CYP78A8; electron carrier/ heme binding / iron ion binding / monooxygenase/ oxygen binding; K00517 [EC:1.14.-.-]
XLOC_024570	AEGTA43357	Up	Os09g0400000; K00083 cinnamyl-alcohol dehydrogenase [EC:1.1.1.195]
XLOC_025915	AEGTA15810	Up	#N/A
XLOC_027523	AEGTA27951	Up	#N/A

Table B.2 (cont'd)

Gene loci	Protein	Regulation	BLASTp, KEGG hit
XLOC_028765	AEGTA20358	Up	hypothetical protein; K09285 AP2-like factor, ANT lineage
XLOC_029950	AEGTA03883	Up	hypothetical protein ; K01175 [EC:3.1.-.-]
XLOC_030124	AEGTA06401	Up	AG; agamous; K09264 MADS-box transcription factor, plant
XLOC_030694	AEGTA09239	Up	GE13493 gene product from transcript GE13493-RA; K14572 midasin
XLOC_031756	AEGTA04393	Up	ISU3; ISU3 (ISCU-LIKE 3); structural molecule; K04488 nitrogen fixation protein
XLOC_031834	AEGTA14386	Up	hypothetical protein; K11982 E3 ubiquitin-protein ligase RNF115/126
XLOC_032234	AEGTA28681	Up	hypothetical protein; K09753 cinnamoyl-CoA reductase [EC:1.2.1.44]
XLOC_034661	AEGTA14505	Up	hypothetical protein; K13091 RNA-binding protein 39
XLOC_034895	AEGTA10808	Up	APK2A; APK2A (PROTEIN KINASE 2A); ATP binding / kinase; K00924
XLOC_035449	AEGTA27654	Up	hypothetical protein LOC100267258; K08081 tropine dehydrogenase
XLOC_036512	AEGTA12433	Up	#N/A
XLOC_037454	AEGTA27979	Up	Os03g0700400; K00454 lipoxygenase [EC:1.13.11.12]
XLOC_039379	AEGTA29279	Up	mybH; SWIRM domain-containing protein; K11865 protein MYSM1
XLOC_039818	AEGTA24810	Up	INT3; INT3 (NOSITOL TRANSPORTER 3); carbohydrate transmembrane transporter/ sugar:hydrogen symporter; K08150 MFS transporter, SP family
XLOC_039862	AEGTA30807	Up	#N/A
XLOC_040636	AEGTA14662	Up	#N/A
XLOC_040851	AEGTA08681	Up	hypothetical protein; K01620 threonine aldolase [EC:4.1.2.5]
XLOC_044228	AEGTA45260	Up	hypothetical protein; K02021 putative ABC transport system ATP-binding protein
XLOC_044263	AEGTA17702	Up	#N/A
XLOC_044525	AEGTA13133	Up	BT2; BT2 (BTB AND TAZ DOMAIN PROTEIN 2); protein binding / transcription factor/ transcription regulator; K00517 [EC:1.14.-.-]
XLOC_045499	AEGTA18150	Up	#N/A
XLOC_045907	AEGTA31210	Up	hypothetical protein; K01090 protein phosphatase [EC:3.1.3.16]
XLOC_046079	AEGTA17028	Up	#N/A

Table B.2 (cont'd)
Transcripts involved in seed color difference
(Criteria: 1-DE (yes, |log2| >1)/2-nonDE)

Gene loci	Protein	Regulation	BLASTp, KEGG hit
XLOC_005537	AEGTA03848	Up	CHD3; chromodomain helicase DNA binding protein 3; K11642
XLOC_024894	AEGTA19657	Down	WRKY transcription factor, putative; K13424 WRKY transcription factor 33
XLOC_037700	AEGTA27908	Down	hypothetical protein; K09264 MADS-box transcription factor, plant

Transcripts involved in red wheat's response to misting
Criteria: DE loci only shown in comparison C not D.

Gene loci	Protein	Regulation	BLASTp, KEGG hit
XLOC_019174	AEGTA21896	Up	Ammonium transporter 2 member 1 [Aegilops tauschii]
XLOC_020144	AEGTA10711	Up	LOX3; LOX3; electron carrier/ iron ion binding / lipoxygenase/ metal ion binding / oxidoreductase, incorporation of two atoms of oxygen; K00454 lipoxygenase
XLOC_022250	AEGTA11275	Up	Os01g0718300; K13415 protein brassinosteroid insensitive 1 [EC:2.7.10.1 2.7.11.1]
XLOC_032818	AEGTA03520	Up	ERF1 (ETHYLENE RESPONSE FACTOR 1); DNA binding / transcription activator/ transcription factor; K14516 ethylene-responsive transcription factor 1
XLOC_038715	AEGTA19454	Up	hypothetical protein; K13027 tyrosine N-monooxygenase [EC:1.14.13.41]
XLOC_041562	AEGTA19458	Up	hypothetical protein LOC100260645; K01179 endoglucanase [EC:3.2.1.4]
XLOC_040917	AEGTA00880	Down	Serine/threonine-protein kinase PBS1 [Triticum urartu]

Transcripts involved in white wheat's response to misting
(Criteria: DE loci only shown in comparison D not C)

Gene loci	Protein	Regulation	BLASTp, KEGG hit
XLOC_010277	AEGTA23816	Up	IDP65; LOC100193874; K02894 large subunit ribosomal protein L23e
XLOC_018723	AEGTA25569	Down	alpha gliadin [Triticum aestivum]
XLOC_019390	AEGTA13087	Down	atp1-1; ATPase subunit 1; K02132 F-type H+-transporting ATPase subunit alpha
XLOC_040814	AEGTA26248	Down	alpha-gliadin [Triticum aestivum]
XLOC_044316	AEGTA29379	Down	alpha-gliadin protein [Aegilops tauschii × Secale cereale]

Table B.3 Differential expressed transfrags based on Cuffdiff categorization

Comparisons	A*	B	C	D
gene	89 (0.6)#	375 (1.6)	2046 (9.5)	1690 (9.2)
isoform	73 (0.4)	230 (0.6)	1171 (3.6)	838 (3.3)
tss^	75 (0.5)	301 (1.0)	1633 (6.2)	1161 (5.2)
cds&	12 (0.3)	15 (0.2)	236 (3.5)	116 (2.1)
promoter	0	0	0	129 (2.5)
splicing	0	0	0	222 (5.8)

* Comparison A: white vs red, 0h non-mist; B: white vs red, 48h mist; C. White, 0h vs 48h; D. Red, 0h vs 48h.

The number in the parenthesis represents the percentage of DE transcripts in the total number of transcripts that are testable. DE were called with the threshold of FDR=0.05

^ tss: transcription start site

& coding sequences

Table B.4 GO terms that match proteins at differentially expressed loci

GO:0000015; phosphopyruvate hydratase complex; Cellular Component
GO:0000036; acyl carrier activity; Molecular Function
GO:0000045; autophagic vacuole assembly; Biological Process
GO:0000049; tRNA binding; Molecular Function
GO:0000062; fatty-acyl-CoA binding; Molecular Function
GO:0000079; regulation of cyclin-dependent protein kinase activity; Biological Process
GO:0000086; G2/M transition of mitotic cell cycle; Biological Process
GO:0000105; histidine biosynthetic process; Biological Process
GO:0000139; Golgi membrane; Cellular Component
GO:0000145; exocyst; Cellular Component
GO:0000148; 1,3-beta-D-glucan synthase complex; Cellular Component
GO:0000151; ubiquitin ligase complex; Cellular Component
GO:0000154; rRNA modification; Biological Process
GO:0000155; two-component sensor activity; Molecular Function
GO:0000156; two-component response regulator activity; Molecular Function
GO:0000159; protein phosphatase type 2A complex; Cellular Component
GO:0000160; two-component signal transduction system (phosphorelay); Biological Process
GO:0000162; tryptophan biosynthetic process; Biological Process
GO:0000166; nucleotide binding; Molecular Function
GO:0000172; ribonuclease MRP complex; Cellular Component
GO:0000175; 3'-5'-exoribonuclease activity; Molecular Function
GO:0000179; rRNA (adenine-N6,N6-)-dimethyltransferase activity; Molecular Function
GO:0000184; nuclear-transcribed mRNA catabolic process, nonsense-mediated decay; Biological Process
GO:0000213; tRNA-intron endonuclease activity; Molecular Function
GO:0000226; microtubule cytoskeleton organization; Biological Process
GO:0000228; nuclear chromosome; Cellular Component
GO:0000247; C-8 sterol isomerase activity; Molecular Function
GO:0000272; polysaccharide catabolic process; Biological Process
GO:0000275; mitochondrial proton-transporting ATP synthase complex, catalytic core F(1); Cellular Component
GO:0000276; mitochondrial proton-transporting ATP synthase complex, coupling factor F(o); Cellular Component
GO:0000287; magnesium ion binding; Molecular Function
GO:0000398; nuclear mRNA splicing, via spliceosome; Biological Process
GO:0000439; core TFIIH complex; Cellular Component
GO:0000502; proteasome complex; Cellular Component
GO:0000724; double-strand break repair via homologous recombination; Biological Process

Table B.4 (cont'd)

GO:0000774; adenyl-nucleotide exchange factor activity; Molecular Function
GO:0000775; chromosome, centromeric region; Cellular Component
GO:0000785; chromatin; Cellular Component
GO:0000786; nucleosome; Cellular Component
GO:0000808; origin recognition complex; Cellular Component
GO:0000922; spindle pole; Cellular Component
GO:0000976; transcription regulatory region sequence-specific DNA binding; Molecular Function
GO:0001104; RNA polymerase II transcription cofactor activity; Molecular Function
GO:0001510; RNA methylation; Biological Process
GO:0001522; pseudouridine synthesis; Biological Process
GO:0001671; ATPase activator activity; Molecular Function
GO:0001682; tRNA 5'-leader removal; Biological Process
GO:0001882; nucleoside binding; Molecular Function
GO:0003333; amino acid transmembrane transport; Biological Process
GO:0003676; nucleic acid binding; Molecular Function
GO:0003677; DNA binding; Molecular Function
GO:0003678; DNA helicase activity; Molecular Function
GO:0003684; damaged DNA binding; Molecular Function
GO:0003689; DNA clamp loader activity; Molecular Function
GO:0003697; single-stranded DNA binding; Molecular Function
GO:0003700; sequence-specific DNA binding transcription factor activity; Molecular Function
GO:0003712; transcription cofactor activity; Molecular Function
GO:0003713; transcription coactivator activity; Molecular Function
GO:0003714; transcription corepressor activity; Molecular Function
GO:0003723; RNA binding; Molecular Function
GO:0003725; double-stranded RNA binding; Molecular Function
GO:0003735; structural constituent of ribosome; Molecular Function
GO:0003743; translation initiation factor activity; Molecular Function
GO:0003746; translation elongation factor activity; Molecular Function
GO:0003747; translation release factor activity; Molecular Function
GO:0003755; peptidyl-prolyl cis-trans isomerase activity; Molecular Function
GO:0003774; motor activity; Molecular Function
GO:0003777; microtubule motor activity; Molecular Function
GO:0003779; actin binding; Molecular Function
GO:0003824; catalytic activity; Molecular Function
GO:0003827; alpha-1,3-mannosylglycoprotein 2-beta-N-acetylglucosaminyltransferase activity; Molecular Function
GO:0003830; beta-1,4-mannosylglycoprotein 4-beta-N-acetylglucosaminyltransferase activity; Molecular Function

Table B.4 (cont'd)

GO:0003840; gamma-glutamyltransferase activity; Molecular Function
GO:0003843; 1,3-beta-D-glucan synthase activity; Molecular Function
GO:0003852; 2-isopropylmalate synthase activity; Molecular Function
GO:0003854; 3-beta-hydroxy-delta5-steroid dehydrogenase activity; Molecular Function
GO:0003857; 3-hydroxyacyl-CoA dehydrogenase activity; Molecular Function
GO:0003868; 4-hydroxyphenylpyruvate dioxygenase activity; Molecular Function
GO:0003871; 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase activity; Molecular Function
GO:0003872; 6-phosphofructokinase activity; Molecular Function
GO:0003879; ATP phosphoribosyltransferase activity; Molecular Function
GO:0003883; CTP synthase activity; Molecular Function
GO:0003885; D-arabinono-1,4-lactone oxidase activity; Molecular Function
GO:0003887; DNA-directed DNA polymerase activity; Molecular Function
GO:0003896; DNA primase activity; Molecular Function
GO:0003899; DNA-directed RNA polymerase activity; Molecular Function
GO:0003905; alkylbase DNA N-glycosylase activity; Molecular Function
GO:0003910; DNA ligase (ATP) activity; Molecular Function
GO:0003913; DNA photolyase activity; Molecular Function
GO:0003916; DNA topoisomerase activity; Molecular Function
GO:0003917; DNA topoisomerase type I activity; Molecular Function
GO:0003918; DNA topoisomerase (ATP-hydrolyzing) activity; Molecular Function
GO:0003919; FMN adenylyltransferase activity; Molecular Function
GO:0003924; GTPase activity; Molecular Function
GO:0003935; GTP cyclohydrolase II activity; Molecular Function
GO:0003950; NAD+ ADP-ribosyltransferase activity; Molecular Function
GO:0003951; NAD+ kinase activity; Molecular Function
GO:0003964; RNA-directed DNA polymerase activity; Molecular Function
GO:0003968; RNA-directed RNA polymerase activity; Molecular Function
GO:0003978; UDP-glucose 4-epimerase activity; Molecular Function
GO:0003980; UDP-glucose:glycoprotein glucosyltransferase activity; Molecular Function
GO:0003984; acetolactate synthase activity; Molecular Function
GO:0003987; acetate-CoA ligase activity; Molecular Function
GO:0003989; acetyl-CoA carboxylase activity; Molecular Function
GO:0003993; acid phosphatase activity; Molecular Function
GO:0003995; acyl-CoA dehydrogenase activity; Molecular Function
GO:0003997; acyl-CoA oxidase activity; Molecular Function
GO:0004003; ATP-dependent DNA helicase activity; Molecular Function
GO:0004013; adenosylhomocysteinase activity; Molecular Function
GO:0004014; adenosylmethionine decarboxylase activity; Molecular Function

Table B.4 (cont'd)

GO:0004017; adenylate kinase activity; Molecular Function
GO:0004019; adenylosuccinate synthase activity; Molecular Function
GO:0004030; aldehyde dehydrogenase [NAD(P)+] activity; Molecular Function
GO:0004044; amidophosphoribosyltransferase activity; Molecular Function
GO:0004045; aminoacyl-tRNA hydrolase activity; Molecular Function
GO:0004055; argininosuccinate synthase activity; Molecular Function
GO:0004066; asparagine synthase (glutamine-hydrolyzing) activity; Molecular Function
GO:0004070; aspartate carbamoyltransferase activity; Molecular Function
GO:0004075; biotin carboxylase activity; Molecular Function
GO:0004089; carbonate dehydratase activity; Molecular Function
GO:0004096; catalase activity; Molecular Function
GO:0004097; catechol oxidase activity; Molecular Function
GO:0004106; chorismate mutase activity; Molecular Function
GO:0004109; coproporphyrinogen oxidase activity; Molecular Function
GO:0004112; cyclic-nucleotide phosphodiesterase activity; Molecular Function
GO:0004121; cystathionine beta-lyase activity; Molecular Function
GO:0004129; cytochrome-c oxidase activity; Molecular Function
GO:0004134; 4-alpha-glucanotransferase activity; Molecular Function
GO:0004143; diacylglycerol kinase activity; Molecular Function
GO:0004144; diacylglycerol O-acyltransferase activity; Molecular Function
GO:0004148; dihydrolipoyl dehydrogenase activity; Molecular Function
GO:0004161; dimethylallyltranstransferase activity; Molecular Function
GO:0004163; diphosphomevalonate decarboxylase activity; Molecular Function
GO:0004170; dUTP diphosphatase activity; Molecular Function
GO:0004175; endopeptidase activity; Molecular Function
GO:0004176; ATP-dependent peptidase activity; Molecular Function
GO:0004177; aminopeptidase activity; Molecular Function
GO:0004181; metallocarboxypeptidase activity; Molecular Function
GO:0004185; serine-type carboxypeptidase activity; Molecular Function
GO:0004190; aspartic-type endopeptidase activity; Molecular Function
GO:0004197; cysteine-type endopeptidase activity; Molecular Function
GO:0004198; calcium-dependent cysteine-type endopeptidase activity; Molecular Function
GO:0004221; ubiquitin thiolesterase activity; Molecular Function
GO:0004222; metalloendopeptidase activity; Molecular Function
GO:0004252; serine-type endopeptidase activity; Molecular Function
GO:0004298; threonine-type endopeptidase activity; Molecular Function
GO:0004315; 3-oxoacyl-[acyl-carrier-protein] synthase activity; Molecular Function
GO:0004325; ferrochelatase activity; Molecular Function
GO:0004326; tetrahydrofolylpolyglutamate synthase activity; Molecular Function

Table B.4 (cont'd)

GO:0004329; formate-tetrahydrofolate ligase activity; Molecular Function
GO:0004332; fructose-bisphosphate aldolase activity; Molecular Function
GO:0004345; glucose-6-phosphate dehydrogenase activity; Molecular Function
GO:0004348; glucosylceramidase activity; Molecular Function
GO:0004351; glutamate decarboxylase activity; Molecular Function
GO:0004356; glutamate-ammonia ligase activity; Molecular Function
GO:0004357; glutamate-cysteine ligase activity; Molecular Function
GO:0004358; glutamate N-acetyltransferase activity; Molecular Function
GO:0004363; glutathione synthase activity; Molecular Function
GO:0004367; glycerol-3-phosphate dehydrogenase [NAD+] activity; Molecular Function
GO:0004368; glycerol-3-phosphate dehydrogenase activity; Molecular Function
GO:0004370; glycerol kinase activity; Molecular Function
GO:0004372; glycine hydroxymethyltransferase activity; Molecular Function
GO:0004386; helicase activity; Molecular Function
GO:0004392; heme oxygenase (decyclizing) activity; Molecular Function
GO:0004399; histidinol dehydrogenase activity; Molecular Function
GO:0004402; histone acetyltransferase activity; Molecular Function
GO:0004407; histone deacetylase activity; Molecular Function
GO:0004420; hydroxymethylglutaryl-CoA reductase (NADPH) activity; Molecular Function
GO:0004421; hydroxymethylglutaryl-CoA synthase activity; Molecular Function
GO:0004425; indole-3-glycerol-phosphate synthase activity; Molecular Function
GO:0004427; inorganic diphosphatase activity; Molecular Function
GO:0004435; phosphatidylinositol phospholipase C activity; Molecular Function
GO:0004449; isocitrate dehydrogenase (NAD+) activity; Molecular Function
GO:0004450; isocitrate dehydrogenase (NADP+) activity; Molecular Function
GO:0004452; isopentenyl-diphosphate delta-isomerase activity; Molecular Function
GO:0004459; L-lactate dehydrogenase activity; Molecular Function
GO:0004470; malic enzyme activity; Molecular Function
GO:0004476; mannose-6-phosphate isomerase activity; Molecular Function
GO:0004478; methionine adenosyltransferase activity; Molecular Function
GO:0004497; monooxygenase activity; Molecular Function
GO:0004499; N,N-dimethylaniline monooxygenase activity; Molecular Function
GO:0004506; squalene monooxygenase activity; Molecular Function
GO:0004512; inositol-3-phosphate synthase activity; Molecular Function
GO:0004514; nicotinate-nucleotide diphosphorylase (carboxylating) activity; Molecular Function
GO:0004516; nicotinate phosphoribosyltransferase activity; Molecular Function
GO:0004518; nuclease activity; Molecular Function
GO:0004519; endonuclease activity; Molecular Function

Table B.4 (cont'd)

GO:0004523; ribonuclease H activity; Molecular Function
GO:0004525; ribonuclease III activity; Molecular Function
GO:0004526; ribonuclease P activity; Molecular Function
GO:0004527; exonuclease activity; Molecular Function
GO:0004540; ribonuclease activity; Molecular Function
GO:0004550; nucleoside diphosphate kinase activity; Molecular Function
GO:0004553; hydrolase activity, hydrolyzing O-glycosyl compounds; Molecular Function
GO:0004556; alpha-amylase activity; Molecular Function
GO:0004560; alpha-L-fucosidase activity; Molecular Function
GO:0004563; beta-N-acetylhexosaminidase activity; Molecular Function
GO:0004568; chitinase activity; Molecular Function
GO:0004571; mannosyl-oligosaccharide 1,2-alpha-mannosidase activity; Molecular Function
GO:0004573; mannosyl-oligosaccharide glucosidase activity; Molecular Function
GO:0004576; oligosaccharyl transferase activity; Molecular Function
GO:0004579; dolichyl-diphosphooligosaccharide-protein glycotransferase activity; Molecular Function
GO:0004594; pantothenate kinase activity; Molecular Function
GO:0004601; peroxidase activity; Molecular Function
GO:0004609; phosphatidylserine decarboxylase activity; Molecular Function
GO:0004610; phosphoacetylglucosamine mutase activity; Molecular Function
GO:0004612; phosphoenolpyruvate carboxykinase (ATP) activity; Molecular Function
GO:0004615; phosphomannomutase activity; Molecular Function
GO:0004616; phosphogluconate dehydrogenase (decarboxylating) activity; Molecular Function
GO:0004617; phosphoglycerate dehydrogenase activity; Molecular Function
GO:0004618; phosphoglycerate kinase activity; Molecular Function
GO:0004619; phosphoglycerate mutase activity; Molecular Function
GO:0004629; phospholipase C activity; Molecular Function
GO:0004634; phosphopyruvate hydratase activity; Molecular Function
GO:0004640; phosphoribosylanthranilate isomerase activity; Molecular Function
GO:0004645; phosphorylase activity; Molecular Function
GO:0004649; poly(ADP-ribose) glycohydrolase activity; Molecular Function
GO:0004650; polygalacturonase activity; Molecular Function
GO:0004652; polynucleotide adenylyltransferase activity; Molecular Function
GO:0004654; polyribonucleotide nucleotidyltransferase activity; Molecular Function
GO:0004655; porphobilinogen synthase activity; Molecular Function
GO:0004657; proline dehydrogenase activity; Molecular Function
GO:0004659; prenyltransferase activity; Molecular Function
GO:0004664; prephenate dehydratase activity; Molecular Function

Table B.4 (cont'd)

GO:0004665; prephenate dehydrogenase (NADP+) activity; Molecular Function
GO:0004672; protein kinase activity; Molecular Function
GO:0004673; protein histidine kinase activity; Molecular Function
GO:0004674; protein serine/threonine kinase activity; Molecular Function
GO:0004707; MAP kinase activity; Molecular Function
GO:0004713; protein tyrosine kinase activity; Molecular Function
GO:0004719; protein-L-isoaspartate (D-aspartate) O-methyltransferase activity; Molecular Function
GO:0004721; phosphoprotein phosphatase activity; Molecular Function
GO:0004725; protein tyrosine phosphatase activity; Molecular Function
GO:0004735; pyrroline-5-carboxylate reductase activity; Molecular Function
GO:0004743; pyruvate kinase activity; Molecular Function
GO:0004746; riboflavin synthase activity; Molecular Function
GO:0004747; ribokinase activity; Molecular Function
GO:0004748; ribonucleoside-diphosphate reductase activity; Molecular Function
GO:0004750; ribulose-phosphate 3-epimerase activity; Molecular Function
GO:0004751; ribose-5-phosphate isomerase activity; Molecular Function
GO:0004765; shikimate kinase activity; Molecular Function
GO:0004781; sulfate adenylyltransferase (ATP) activity; Molecular Function
GO:0004784; superoxide dismutase activity; Molecular Function
GO:0004788; thiamine diphosphokinase activity; Molecular Function
GO:0004789; thiamine-phosphate diphosphorylase activity; Molecular Function
GO:0004794; L-threonine ammonia-lyase activity; Molecular Function
GO:0004797; thymidine kinase activity; Molecular Function
GO:0004806; triglyceride lipase activity; Molecular Function
GO:0004807; triose-phosphate isomerase activity; Molecular Function
GO:0004809; tRNA (guanine-N2-)-methyltransferase activity; Molecular Function
GO:0004812; aminoacyl-tRNA ligase activity; Molecular Function
GO:0004813; alanine-tRNA ligase activity; Molecular Function
GO:0004815; aspartate-tRNA ligase activity; Molecular Function
GO:0004816; asparagine-tRNA ligase activity; Molecular Function
GO:0004819; glutamine-tRNA ligase activity; Molecular Function
GO:0004820; glycine-tRNA ligase activity; Molecular Function
GO:0004821; histidine-tRNA ligase activity; Molecular Function
GO:0004822; isoleucine-tRNA ligase activity; Molecular Function
GO:0004824; lysine-tRNA ligase activity; Molecular Function
GO:0004825; methionine-tRNA ligase activity; Molecular Function
GO:0004827; proline-tRNA ligase activity; Molecular Function
GO:0004828; serine-tRNA ligase activity; Molecular Function
GO:0004829; threonine-tRNA ligase activity; Molecular Function

Table B.4 (cont'd)

GO:0004830; tryptophan-tRNA ligase activity; Molecular Function
GO:0004831; tyrosine-tRNA ligase activity; Molecular Function
GO:0004832; valine-tRNA ligase activity; Molecular Function
GO:0004834; tryptophan synthase activity; Molecular Function
GO:0004835; tubulin-tyrosine ligase activity; Molecular Function
GO:0004842; ubiquitin-protein ligase activity; Molecular Function
GO:0004852; uroporphyrinogen-III synthase activity; Molecular Function
GO:0004857; enzyme inhibitor activity; Molecular Function
GO:0004861; cyclin-dependent protein kinase inhibitor activity; Molecular Function
GO:0004864; protein phosphatase inhibitor activity; Molecular Function
GO:0004866; endopeptidase inhibitor activity; Molecular Function
GO:0004867; serine-type endopeptidase inhibitor activity; Molecular Function
GO:0004869; cysteine-type endopeptidase inhibitor activity; Molecular Function
GO:0004871; signal transducer activity; Molecular Function
GO:0004930; G-protein coupled receptor activity; Molecular Function
GO:0004965; G-protein coupled GABA receptor activity; Molecular Function
GO:0004970; ionotropic glutamate receptor activity; Molecular Function
GO:0005053; peroxisome matrix targeting signal-2 binding; Molecular Function
GO:0005083; small GTPase regulator activity; Molecular Function
GO:0005085; guanyl-nucleotide exchange factor activity; Molecular Function
GO:0005086; ARF guanyl-nucleotide exchange factor activity; Molecular Function
GO:0005089; Rho guanyl-nucleotide exchange factor activity; Molecular Function
GO:0005093; Rab GDP-dissociation inhibitor activity; Molecular Function
GO:0005094; Rho GDP-dissociation inhibitor activity; Molecular Function
GO:0005097; Rab GTPase activator activity; Molecular Function
GO:0005198; structural molecule activity; Molecular Function
GO:0005215; transporter activity; Molecular Function
GO:0005216; ion channel activity; Molecular Function
GO:0005234; extracellular-glutamate-gated ion channel activity; Molecular Function
GO:0005247; voltage-gated chloride channel activity; Molecular Function
GO:0005249; voltage-gated potassium channel activity; Molecular Function
GO:0005267; potassium channel activity; Molecular Function
GO:0005315; inorganic phosphate transmembrane transporter activity; Molecular Function
GO:0005337; nucleoside transmembrane transporter activity; Molecular Function
GO:0005351; sugar:hydrogen symporter activity; Molecular Function
GO:0005375; copper ion transmembrane transporter activity; Molecular Function
GO:0005381; iron ion transmembrane transporter activity; Molecular Function
GO:0005452; inorganic anion exchanger activity; Molecular Function
GO:0005471; ATP:ADP antiporter activity; Molecular Function

Table B.4 (cont'd)

GO:0005488; binding; Molecular Function
GO:0005506; iron ion binding; Molecular Function
GO:0005507; copper ion binding; Molecular Function
GO:0005509; calcium ion binding; Molecular Function
GO:0005515; protein binding; Molecular Function
GO:0005516; calmodulin binding; Molecular Function
GO:0005524; ATP binding; Molecular Function
GO:0005525; GTP binding; Molecular Function
GO:0005529; sugar binding; Molecular Function
GO:0005542; folic acid binding; Molecular Function
GO:0005543; phospholipid binding; Molecular Function
GO:0005544; calcium-dependent phospholipid binding; Molecular Function
GO:0005576; extracellular region; Cellular Component
GO:0005618; cell wall; Cellular Component
GO:0005622; intracellular; Cellular Component
GO:0005634; nucleus; Cellular Component
GO:0005643; nuclear pore; Cellular Component
GO:0005663; DNA replication factor C complex; Cellular Component
GO:0005664; nuclear origin of replication recognition complex; Cellular Component
GO:0005665; DNA-directed RNA polymerase II, core complex; Cellular Component
GO:0005666; DNA-directed RNA polymerase III complex; Cellular Component
GO:0005667; transcription factor complex; Cellular Component
GO:0005669; transcription factor TFIID complex; Cellular Component
GO:0005672; transcription factor TFIIA complex; Cellular Component
GO:0005674; transcription factor TFIIF complex; Cellular Component
GO:0005675; holo TFIIH complex; Cellular Component
GO:0005680; anaphase-promoting complex; Cellular Component
GO:0005681; spliceosomal complex; Cellular Component
GO:0005694; chromosome; Cellular Component
GO:0005730; nucleolus; Cellular Component
GO:0005737; cytoplasm; Cellular Component
GO:0005739; mitochondrion; Cellular Component
GO:0005740; mitochondrial envelope; Cellular Component
GO:0005741; mitochondrial outer membrane; Cellular Component
GO:0005742; mitochondrial outer membrane translocase complex; Cellular Component
GO:0005743; mitochondrial inner membrane; Cellular Component
GO:0005744; mitochondrial inner membrane presequence translocase complex; Cellular Component
GO:0005746; mitochondrial respiratory chain; Cellular Component
GO:0005759; mitochondrial matrix; Cellular Component

Table B.4 (cont'd)

GO:0005777; peroxisome; Cellular Component
GO:0005778; peroxisomal membrane; Cellular Component
GO:0005779; integral to peroxisomal membrane; Cellular Component
GO:0005783; endoplasmic reticulum; Cellular Component
GO:0005787; signal peptidase complex; Cellular Component
GO:0005789; endoplasmic reticulum membrane; Cellular Component
GO:0005794; Golgi apparatus; Cellular Component
GO:0005795; Golgi stack; Cellular Component
GO:0005798; Golgi-associated vesicle; Cellular Component
GO:0005801; cis-Golgi network; Cellular Component
GO:0005815; microtubule organizing center; Cellular Component
GO:0005838; proteasome regulatory particle; Cellular Component
GO:0005839; proteasome core complex; Cellular Component
GO:0005840; ribosome; Cellular Component
GO:0005853; eukaryotic translation elongation factor 1 complex; Cellular Component
GO:0005856; cytoskeleton; Cellular Component
GO:0005874; microtubule; Cellular Component
GO:0005875; microtubule associated complex; Cellular Component
GO:0005938; cell cortex; Cellular Component
GO:0005942; phosphatidylinositol 3-kinase complex; Cellular Component
GO:0005945; 6-phosphofructokinase complex; Cellular Component
GO:0005956; protein kinase CK2 complex; Cellular Component
GO:0005971; ribonucleoside-diphosphate reductase complex; Cellular Component
GO:0005975; carbohydrate metabolic process; Biological Process
GO:0005985; sucrose metabolic process; Biological Process
GO:0005986; sucrose biosynthetic process; Biological Process
GO:0005992; trehalose biosynthetic process; Biological Process
GO:0006006; glucose metabolic process; Biological Process
GO:0006007; glucose catabolic process; Biological Process
GO:0006012; galactose metabolic process; Biological Process
GO:0006014; D-ribose metabolic process; Biological Process
GO:0006021; inositol biosynthetic process; Biological Process
GO:0006032; chitin catabolic process; Biological Process
GO:0006071; glycerol metabolic process; Biological Process
GO:0006072; glycerol-3-phosphate metabolic process; Biological Process
GO:0006073; cellular glucan metabolic process; Biological Process
GO:0006075; (1->3)-beta-D-glucan biosynthetic process; Biological Process
GO:0006081; cellular aldehyde metabolic process; Biological Process
GO:0006094; gluconeogenesis; Biological Process
GO:0006096; glycolysis; Biological Process

Table B.4 (cont'd)

GO:0006098; pentose-phosphate shunt; Biological Process
GO:0006099; tricarboxylic acid cycle; Biological Process
GO:0006102; isocitrate metabolic process; Biological Process
GO:0006108; malate metabolic process; Biological Process
GO:0006122; mitochondrial electron transport, ubiquinol to cytochrome c; Biological Process
GO:0006139; nucleobase-containing compound metabolic process; Biological Process
GO:0006164; purine nucleotide biosynthetic process; Biological Process
GO:0006165; nucleoside diphosphate phosphorylation; Biological Process
GO:0006183; GTP biosynthetic process; Biological Process
GO:0006184; GTP catabolic process; Biological Process
GO:0006200; ATP catabolic process; Biological Process
GO:0006207; 'de novo' pyrimidine base biosynthetic process; Biological Process
GO:0006221; pyrimidine nucleotide biosynthetic process; Biological Process
GO:0006228; UTP biosynthetic process; Biological Process
GO:0006241; CTP biosynthetic process; Biological Process
GO:0006259; DNA metabolic process; Biological Process
GO:0006260; DNA replication; Biological Process
GO:0006265; DNA topological change; Biological Process
GO:0006269; DNA replication, synthesis of RNA primer; Biological Process
GO:0006270; DNA-dependent DNA replication initiation; Biological Process
GO:0006278; RNA-dependent DNA replication; Biological Process
GO:0006281; DNA repair; Biological Process
GO:0006282; regulation of DNA repair; Biological Process
GO:0006284; base-excision repair; Biological Process
GO:0006289; nucleotide-excision repair; Biological Process
GO:0006298; mismatch repair; Biological Process
GO:0006302; double-strand break repair; Biological Process
GO:0006303; double-strand break repair via nonhomologous end joining; Biological Process
GO:0006306; DNA methylation; Biological Process
GO:0006308; DNA catabolic process; Biological Process
GO:0006310; DNA recombination; Biological Process
GO:0006333; chromatin assembly or disassembly; Biological Process
GO:0006334; nucleosome assembly; Biological Process
GO:0006338; chromatin remodeling; Biological Process
GO:0006351; transcription, DNA-dependent; Biological Process
GO:0006352; transcription initiation, DNA-dependent; Biological Process
GO:0006353; transcription termination, DNA-dependent; Biological Process
GO:0006355; regulation of transcription, DNA-dependent; Biological Process
GO:0006357; regulation of transcription from RNA polymerase II promoter; Biological Process

Table B.4 (cont'd)

GO:0006364; rRNA processing; Biological Process
GO:0006366; transcription from RNA polymerase II promoter; Biological Process
GO:0006367; transcription initiation from RNA polymerase II promoter; Biological Process
GO:0006370; mRNA capping; Biological Process
GO:0006379; mRNA cleavage; Biological Process
GO:0006383; transcription from RNA polymerase III promoter; Biological Process
GO:0006388; tRNA splicing, via endonucleolytic cleavage and ligation; Biological Process
GO:0006396; RNA processing; Biological Process
GO:0006397; mRNA processing; Biological Process
GO:0006400; tRNA modification; Biological Process
GO:0006402; mRNA catabolic process; Biological Process
GO:0006412; translation; Biological Process
GO:0006413; translational initiation; Biological Process
GO:0006414; translational elongation; Biological Process
GO:0006415; translational termination; Biological Process
GO:0006418; tRNA aminoacylation for protein translation; Biological Process
GO:0006419; alanyl-tRNA aminoacylation; Biological Process
GO:0006421; asparaginyl-tRNA aminoacylation; Biological Process
GO:0006422; aspartyl-tRNA aminoacylation; Biological Process
GO:0006425; glutaminyl-tRNA aminoacylation; Biological Process
GO:0006426; glycyl-tRNA aminoacylation; Biological Process
GO:0006427; histidyl-tRNA aminoacylation; Biological Process
GO:0006428; isoleucyl-tRNA aminoacylation; Biological Process
GO:0006430; lysyl-tRNA aminoacylation; Biological Process
GO:0006431; methionyl-tRNA aminoacylation; Biological Process
GO:0006433; prolyl-tRNA aminoacylation; Biological Process
GO:0006434; seryl-tRNA aminoacylation; Biological Process
GO:0006435; threonyl-tRNA aminoacylation; Biological Process
GO:0006436; tryptophanyl-tRNA aminoacylation; Biological Process
GO:0006437; tyrosyl-tRNA aminoacylation; Biological Process
GO:0006438; valyl-tRNA aminoacylation; Biological Process
GO:0006452; translational frameshifting; Biological Process
GO:0006457; protein folding; Biological Process
GO:0006461; protein complex assembly; Biological Process
GO:0006464; protein modification process; Biological Process
GO:0006465; signal peptide processing; Biological Process
GO:0006468; protein phosphorylation; Biological Process
GO:0006470; protein dephosphorylation; Biological Process
GO:0006471; protein ADP-ribosylation; Biological Process

Table B.4 (cont'd)

GO:0006476; protein deacetylation; Biological Process
GO:0006479; protein methylation; Biological Process
GO:0006486; protein glycosylation; Biological Process
GO:0006487; protein N-linked glycosylation; Biological Process
GO:0006505; GPI anchor metabolic process; Biological Process
GO:0006506; GPI anchor biosynthetic process; Biological Process
GO:0006508; proteolysis; Biological Process
GO:0006511; ubiquitin-dependent protein catabolic process; Biological Process
GO:0006520; cellular amino acid metabolic process; Biological Process
GO:0006526; arginine biosynthetic process; Biological Process
GO:0006527; arginine catabolic process; Biological Process
GO:0006529; asparagine biosynthetic process; Biological Process
GO:0006536; glutamate metabolic process; Biological Process
GO:0006537; glutamate biosynthetic process; Biological Process
GO:0006542; glutamine biosynthetic process; Biological Process
GO:0006544; glycine metabolic process; Biological Process
GO:0006556; S-adenosylmethionine biosynthetic process; Biological Process
GO:0006561; proline biosynthetic process; Biological Process
GO:0006562; proline catabolic process; Biological Process
GO:0006563; L-serine metabolic process; Biological Process
GO:0006564; L-serine biosynthetic process; Biological Process
GO:0006568; tryptophan metabolic process; Biological Process
GO:0006571; tyrosine biosynthetic process; Biological Process
GO:0006597; spermine biosynthetic process; Biological Process
GO:0006605; protein targeting; Biological Process
GO:0006614; SRP-dependent cotranslational protein targeting to membrane; Biological Process
GO:0006621; protein retention in ER lumen; Biological Process
GO:0006625; protein targeting to peroxisome; Biological Process
GO:0006626; protein targeting to mitochondrion; Biological Process
GO:0006629; lipid metabolic process; Biological Process
GO:0006631; fatty acid metabolic process; Biological Process
GO:0006633; fatty acid biosynthetic process; Biological Process
GO:0006635; fatty acid beta-oxidation; Biological Process
GO:0006637; acyl-CoA metabolic process; Biological Process
GO:0006644; phospholipid metabolic process; Biological Process
GO:0006659; phosphatidylserine biosynthetic process; Biological Process
GO:0006662; glycerol ether metabolic process; Biological Process
GO:0006665; sphingolipid metabolic process; Biological Process

Table B.4 (cont'd)

GO:0006694; steroid biosynthetic process; Biological Process
GO:0006696; ergosterol biosynthetic process; Biological Process
GO:0006725; cellular aromatic compound metabolic process; Biological Process
GO:0006730; one-carbon metabolic process; Biological Process
GO:0006750; glutathione biosynthetic process; Biological Process
GO:0006754; ATP biosynthetic process; Biological Process
GO:0006777; Mo-molybdopterin cofactor biosynthetic process; Biological Process
GO:0006779; porphyrin-containing compound biosynthetic process; Biological Process
GO:0006783; heme biosynthetic process; Biological Process
GO:0006788; heme oxidation; Biological Process
GO:0006796; phosphate-containing compound metabolic process; Biological Process
GO:0006801; superoxide metabolic process; Biological Process
GO:0006807; nitrogen compound metabolic process; Biological Process
GO:0006808; regulation of nitrogen utilization; Biological Process
GO:0006810; transport; Biological Process
GO:0006811; ion transport; Biological Process
GO:0006812; cation transport; Biological Process
GO:0006813; potassium ion transport; Biological Process
GO:0006814; sodium ion transport; Biological Process
GO:0006820; anion transport; Biological Process
GO:0006821; chloride transport; Biological Process
GO:0006826; iron ion transport; Biological Process
GO:0006839; mitochondrial transport; Biological Process
GO:0006855; drug transmembrane transport; Biological Process
GO:0006857; oligopeptide transport; Biological Process
GO:0006869; lipid transport; Biological Process
GO:0006879; cellular iron ion homeostasis; Biological Process
GO:0006885; regulation of pH; Biological Process
GO:0006886; intracellular protein transport; Biological Process
GO:0006887; exocytosis; Biological Process
GO:0006888; ER to Golgi vesicle-mediated transport; Biological Process
GO:0006891; intra-Golgi vesicle-mediated transport; Biological Process
GO:0006897; endocytosis; Biological Process
GO:0006904; vesicle docking involved in exocytosis; Biological Process
GO:0006909; phagocytosis; Biological Process
GO:0006913; nucleocytoplasmic transport; Biological Process
GO:0006915; apoptotic process; Biological Process
GO:0006950; response to stress; Biological Process
GO:0006952; defense response; Biological Process

Table B.4 (cont'd)

GO:0006974; response to DNA damage stimulus; Biological Process
GO:0006979; response to oxidative stress; Biological Process
GO:0007010; cytoskeleton organization; Biological Process
GO:0007017; microtubule-based process; Biological Process
GO:0007018; microtubule-based movement; Biological Process
GO:0007021; tubulin complex assembly; Biological Process
GO:0007030; Golgi organization; Biological Process
GO:0007031; peroxisome organization; Biological Process
GO:0007034; vacuolar transport; Biological Process
GO:0007047; cellular cell wall organization; Biological Process
GO:0007049; cell cycle; Biological Process
GO:0007050; cell cycle arrest; Biological Process
GO:0007067; mitosis; Biological Process
GO:0007090; regulation of S phase of mitotic cell cycle; Biological Process
GO:0007154; cell communication; Biological Process
GO:0007155; cell adhesion; Biological Process
GO:0007165; signal transduction; Biological Process
GO:0007186; G-protein coupled receptor signaling pathway; Biological Process
GO:0007205; activation of protein kinase C activity by G-protein coupled receptor protein signaling pathway; Biological Process
GO:0007264; small GTPase mediated signal transduction; Biological Process
GO:0007275; multicellular organismal development; Biological Process
GO:0007585; respiratory gaseous exchange; Biological Process
GO:0008017; microtubule binding; Molecular Function
GO:0008020; G-protein coupled photoreceptor activity; Molecular Function
GO:0008026; ATP-dependent helicase activity; Molecular Function
GO:0008033; tRNA processing; Biological Process
GO:0008060; ARF GTPase activator activity; Molecular Function
GO:0008061; chitin binding; Molecular Function
GO:0008080; N-acetyltransferase activity; Molecular Function
GO:0008083; growth factor activity; Molecular Function
GO:0008094; DNA-dependent ATPase activity; Molecular Function
GO:0008097; 5S rRNA binding; Molecular Function
GO:0008104; protein localization; Biological Process
GO:0008107; galactoside 2-alpha-L-fucosyltransferase activity; Molecular Function
GO:0008108; UDP-glucose:hexose-1-phosphate uridylyltransferase activity; Molecular Function
GO:0008121; ubiquinol-cytochrome-c reductase activity; Molecular Function
GO:0008131; primary amine oxidase activity; Molecular Function
GO:0008138; protein tyrosine/serine/threonine phosphatase activity; Molecular Function

Table B.4 (cont'd)

GO:0008146; sulfotransferase activity; Molecular Function
GO:0008152; metabolic process; Biological Process
GO:0008168; methyltransferase activity; Molecular Function
GO:0008170; N-methyltransferase activity; Molecular Function
GO:0008171; O-methyltransferase activity; Molecular Function
GO:0008173; RNA methyltransferase activity; Molecular Function
GO:0008198; ferrous iron binding; Molecular Function
GO:0008199; ferric iron binding; Molecular Function
GO:0008219; cell death; Biological Process
GO:0008233; peptidase activity; Molecular Function
GO:0008234; cysteine-type peptidase activity; Molecular Function
GO:0008235; metalloexopeptidase activity; Molecular Function
GO:0008236; serine-type peptidase activity; Molecular Function
GO:0008237; metallopeptidase activity; Molecular Function
GO:0008242; omega peptidase activity; Molecular Function
GO:0008270; zinc ion binding; Molecular Function
GO:0008276; protein methyltransferase activity; Molecular Function
GO:0008283; cell proliferation; Biological Process
GO:0008289; lipid binding; Molecular Function
GO:0008290; F-actin capping protein complex; Cellular Component
GO:0008295; spermidine biosynthetic process; Biological Process
GO:0008299; isoprenoid biosynthetic process; Biological Process
GO:0008308; voltage-gated anion channel activity; Molecular Function
GO:0008312; 7S RNA binding; Molecular Function
GO:0008318; protein prenyltransferase activity; Molecular Function
GO:0008324; cation transmembrane transporter activity; Molecular Function
GO:0008373; sialyltransferase activity; Molecular Function
GO:0008374; O-acyltransferase activity; Molecular Function
GO:0008375; acetylglucosaminyltransferase activity; Molecular Function
GO:0008378; galactosyltransferase activity; Molecular Function
GO:0008380; RNA splicing; Biological Process
GO:0008408; 3'-5' exonuclease activity; Molecular Function
GO:0008409; 5'-3' exonuclease activity; Molecular Function
GO:0008417; fucosyltransferase activity; Molecular Function
GO:0008430; selenium binding; Molecular Function
GO:0008440; inositol-1,4,5-trisphosphate 3-kinase activity; Molecular Function
GO:0008455; alpha-1,6-mannosylglycoprotein 2-beta-N-acetylglucosaminyltransferase activity; Molecular Function
GO:0008466; glycogenin glucosyltransferase activity; Molecular Function
GO:0008474; palmitoyl-(protein) hydrolase activity; Molecular Function

Table B.4 (cont'd)

GO:0008478; pyridoxal kinase activity; Molecular Function
GO:0008479; queuine tRNA-ribosyltransferase activity; Molecular Function
GO:0008483; transaminase activity; Molecular Function
GO:0008508; bile acid:sodium symporter activity; Molecular Function
GO:0008519; ammonium transmembrane transporter activity; Molecular Function
GO:0008531; riboflavin kinase activity; Molecular Function
GO:0008535; respiratory chain complex IV assembly; Biological Process
GO:0008553; hydrogen-exporting ATPase activity, phosphorylative mechanism; Molecular Function
GO:0008565; protein transporter activity; Molecular Function
GO:0008601; protein phosphatase type 2A regulator activity; Molecular Function
GO:0008610; lipid biosynthetic process; Biological Process
GO:0008612; peptidyl-lysine modification to hypusine; Biological Process
GO:0008616; queuosine biosynthetic process; Biological Process
GO:0008641; small protein activating enzyme activity; Molecular Function
GO:0008643; carbohydrate transport; Biological Process
GO:0008649; rRNA methyltransferase activity; Molecular Function
GO:0008652; cellular amino acid biosynthetic process; Biological Process
GO:0008654; phospholipid biosynthetic process; Biological Process
GO:0008676; 3-deoxy-8-phosphooctulonate synthase activity; Molecular Function
GO:0008686; 3,4-dihydroxy-2-butanone-4-phosphate synthase activity; Molecular Function
GO:0008716; D-alanine-D-alanine ligase activity; Molecular Function
GO:0008725; DNA-3-methyladenine glycosylase activity; Molecular Function
GO:0008762; UDP-N-acetylmuramate dehydrogenase activity; Molecular Function
GO:0008792; arginine decarboxylase activity; Molecular Function
GO:0008831; dTDP-4-dehydrorhamnose reductase activity; Molecular Function
GO:0008835; diaminohydroxyphosphoribosylaminopyrimidine deaminase activity; Molecular Function
GO:0008836; diaminopimelate decarboxylase activity; Molecular Function
GO:0008839; dihydrodipicolinate reductase activity; Molecular Function
GO:0008853; exodeoxyribonuclease III activity; Molecular Function
GO:0008883; glutamyl-tRNA reductase activity; Molecular Function
GO:0008889; glycerophosphodiester phosphodiesterase activity; Molecular Function
GO:0008897; holo-[acyl-carrier-protein] synthase activity; Molecular Function
GO:0008898; homocysteine S-methyltransferase activity; Molecular Function
GO:0008915; lipid-A-disaccharide synthase activity; Molecular Function
GO:0008942; nitrite reductase [NAD(P)H] activity; Molecular Function
GO:0008963; phospho-N-acetylmuramoyl-pentapeptide-transferase activity; Molecular Function
GO:0008964; phosphoenolpyruvate carboxylase activity; Molecular Function

Table B.4 (cont'd)

GO:0008977; prephenate dehydrogenase activity; Molecular Function
GO:0008987; quinolinate synthetase A activity; Molecular Function
GO:0009039; urease activity; Molecular Function
GO:0009041; uridylate kinase activity; Molecular Function
GO:0009052; pentose-phosphate shunt, non-oxidative branch; Biological Process
GO:0009055; electron carrier activity; Molecular Function
GO:0009058; biosynthetic process; Biological Process
GO:0009059; macromolecule biosynthetic process; Biological Process
GO:0009060; aerobic respiration; Biological Process
GO:0009072; aromatic amino acid family metabolic process; Biological Process
GO:0009073; aromatic amino acid family biosynthetic process; Biological Process
GO:0009082; branched chain family amino acid biosynthetic process; Biological Process
GO:0009086; methionine biosynthetic process; Biological Process
GO:0009089; lysine biosynthetic process via diaminopimelate; Biological Process
GO:0009094; L-phenylalanine biosynthetic process; Biological Process
GO:0009097; isoleucine biosynthetic process; Biological Process
GO:0009098; leucine biosynthetic process; Biological Process
GO:0009107; lipoate biosynthetic process; Biological Process
GO:0009113; purine base biosynthetic process; Biological Process
GO:0009116; nucleoside metabolic process; Biological Process
GO:0009168; purine ribonucleoside monophosphate biosynthetic process; Biological Process
GO:0009186; deoxyribonucleoside diphosphate metabolic process; Biological Process
GO:0009228; thiamine biosynthetic process; Biological Process
GO:0009229; thiamine diphosphate biosynthetic process; Biological Process
GO:0009231; riboflavin biosynthetic process; Biological Process
GO:0009236; cobalamin biosynthetic process; Biological Process
GO:0009245; lipid A biosynthetic process; Biological Process
GO:0009247; glycolipid biosynthetic process; Biological Process
GO:0009252; peptidoglycan biosynthetic process; Biological Process
GO:0009269; response to desiccation; Biological Process
GO:0009306; protein secretion; Biological Process
GO:0009308; amine metabolic process; Biological Process
GO:0009311; oligosaccharide metabolic process; Biological Process
GO:0009312; oligosaccharide biosynthetic process; Biological Process
GO:0009331; glycerol-3-phosphate dehydrogenase complex; Cellular Component
GO:0009396; folic acid-containing compound biosynthetic process; Biological Process
GO:0009405; pathogenesis; Biological Process
GO:0009408; response to heat; Biological Process
GO:0009415; response to water; Biological Process

Table B.4 (cont'd)

GO:0009416; response to light stimulus; Biological Process
GO:0009432; SOS response; Biological Process
GO:0009435; NAD biosynthetic process; Biological Process
GO:0009443; pyridoxal 5'-phosphate salvage; Biological Process
GO:0009451; RNA modification; Biological Process
GO:0009452; RNA capping; Biological Process
GO:0009507; chloroplast; Cellular Component
GO:0009512; cytochrome b6f complex; Cellular Component
GO:0009521; photosystem; Cellular Component
GO:0009522; photosystem I; Cellular Component
GO:0009523; photosystem II; Cellular Component
GO:0009536; plastid; Cellular Component
GO:0009538; photosystem I reaction center; Cellular Component
GO:0009584; detection of visible light; Biological Process
GO:0009607; response to biotic stimulus; Biological Process
GO:0009611; response to wounding; Biological Process
GO:0009654; oxygen evolving complex; Cellular Component
GO:0009664; plant-type cell wall organization; Biological Process
GO:0009678; hydrogen-translocating pyrophosphatase activity; Molecular Function
GO:0009690; cytokinin metabolic process; Biological Process
GO:0009725; response to hormone stimulus; Biological Process
GO:0009765; photosynthesis, light harvesting; Biological Process
GO:0009767; photosynthetic electron transport chain; Biological Process
GO:0009772; photosynthetic electron transport in photosystem II; Biological Process
GO:0009790; embryo development; Biological Process
GO:0009934; regulation of meristem structural organization; Biological Process
GO:0009966; regulation of signal transduction; Biological Process
GO:0009982; pseudouridine synthase activity; Molecular Function
GO:0009987; cellular process; Biological Process
GO:0010024; phytochromobilin biosynthetic process; Biological Process
GO:0010038; response to metal ion; Biological Process
GO:0010044; response to aluminum ion; Biological Process
GO:0010181; FMN binding; Molecular Function
GO:0010277; chlorophyllide a oxygenase [overall] activity; Molecular Function
GO:0010285; L,L-diaminopimelate aminotransferase activity; Molecular Function
GO:0010309; acireductone dioxygenase [iron(II)-requiring] activity; Molecular Function
GO:0010333; terpene synthase activity; Molecular Function
GO:0010380; regulation of chlorophyll biosynthetic process; Biological Process
GO:0015002; heme-copper terminal oxidase activity; Molecular Function

Table B.4 (cont'd)

GO:0015018; galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase activity; Molecular Function
GO:0015031; protein transport; Biological Process
GO:0015035; protein disulfide oxidoreductase activity; Molecular Function
GO:0015074; DNA integration; Biological Process
GO:0015078; hydrogen ion transmembrane transporter activity; Molecular Function
GO:0015079; potassium ion transmembrane transporter activity; Molecular Function
GO:0015137; citrate transmembrane transporter activity; Molecular Function
GO:0015171; amino acid transmembrane transporter activity; Molecular Function
GO:0015205; nucleobase transmembrane transporter activity; Molecular Function
GO:0015232; heme transporter activity; Molecular Function
GO:0015238; drug transmembrane transporter activity; Molecular Function
GO:0015297; antiporter activity; Molecular Function
GO:0015299; solute:hydrogen antiporter activity; Molecular Function
GO:0015385; sodium:hydrogen antiporter activity; Molecular Function
GO:0015450; P-P-bond-hydrolysis-driven protein transmembrane transporter activity; Molecular Function
GO:0015629; actin cytoskeleton; Cellular Component
GO:0015662; ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism; Molecular Function
GO:0015746; citrate transport; Biological Process
GO:0015851; nucleobase transport; Biological Process
GO:0015886; heme transport; Biological Process
GO:0015930; glutamate synthase activity; Molecular Function
GO:0015934; large ribosomal subunit; Cellular Component
GO:0015935; small ribosomal subunit; Cellular Component
GO:0015936; coenzyme A metabolic process; Biological Process
GO:0015937; coenzyme A biosynthetic process; Biological Process
GO:0015969; guanosine tetraphosphate metabolic process; Biological Process
GO:0015977; carbon fixation; Biological Process
GO:0015979; photosynthesis; Biological Process
GO:0015986; ATP synthesis coupled proton transport; Biological Process
GO:0015991; ATP hydrolysis coupled proton transport; Biological Process
GO:0015992; proton transport; Biological Process
GO:0016020; membrane; Cellular Component
GO:0016021; integral to membrane; Cellular Component
GO:0016043; cellular component organization; Biological Process
GO:0016068; type I hypersensitivity; Biological Process
GO:0016070; RNA metabolic process; Biological Process
GO:0016075; rRNA catabolic process; Biological Process
GO:0016125; sterol metabolic process; Biological Process

Table B.4 (cont'd)

GO:0016149; translation release factor activity, codon specific; Molecular Function
GO:0016151; nickel cation binding; Molecular Function
GO:0016157; sucrose synthase activity; Molecular Function
GO:0016161; beta-amylase activity; Molecular Function
GO:0016165; lipoxygenase activity; Molecular Function
GO:0016168; chlorophyll binding; Molecular Function
GO:0016192; vesicle-mediated transport; Biological Process
GO:0016208; AMP binding; Molecular Function
GO:0016209; antioxidant activity; Molecular Function
GO:0016226; iron-sulfur cluster assembly; Biological Process
GO:0016272; prefoldin complex; Cellular Component
GO:0016301; kinase activity; Molecular Function
GO:0016303; 1-phosphatidylinositol-3-kinase activity; Molecular Function
GO:0016307; phosphatidylinositol phosphate kinase activity; Molecular Function
GO:0016310; phosphorylation; Biological Process
GO:0016311; dephosphorylation; Biological Process
GO:0016428; tRNA (cytosine-5-)-methyltransferase activity; Molecular Function
GO:0016429; tRNA (adenine-N1-)-methyltransferase activity; Molecular Function
GO:0016459; myosin complex; Cellular Component
GO:0016469; proton-transporting two-sector ATPase complex; Cellular Component
GO:0016485; protein processing; Biological Process
GO:0016491; oxidoreductase activity; Molecular Function
GO:0016558; protein import into peroxisome matrix; Biological Process
GO:0016567; protein ubiquitination; Biological Process
GO:0016568; chromatin modification; Biological Process
GO:0016570; histone modification; Biological Process
GO:0016575; histone deacetylation; Biological Process
GO:0016592; mediator complex; Cellular Component
GO:0016597; amino acid binding; Molecular Function
GO:0016614; oxidoreductase activity, acting on CH-OH group of donors; Molecular Function
GO:0016615; malate dehydrogenase activity; Molecular Function
GO:0016616; oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; Molecular Function
GO:0016619; malate dehydrogenase (oxaloacetate-decarboxylating) activity; Molecular Function
GO:0016620; oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor; Molecular Function
GO:0016624; oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor; Molecular Function
GO:0016627; oxidoreductase activity, acting on the CH-CH group of donors; Molecular Function

Table B.4 (cont'd)

GO:0016630; protochlorophyllide reductase activity; Molecular Function
GO:0016636; oxidoreductase activity, acting on the CH-CH group of donors, iron-sulfur protein as acceptor; Molecular Function
GO:0016638; oxidoreductase activity, acting on the CH-NH2 group of donors; Molecular Function
GO:0016651; oxidoreductase activity, acting on NADH or NADPH; Molecular Function
GO:0016671; oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor; Molecular Function
GO:0016679; oxidoreductase activity, acting on diphenols and related substances as donors; Molecular Function
GO:0016701; oxidoreductase activity, acting on single donors with incorporation of molecular oxygen; Molecular Function
GO:0016702; oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen; Molecular Function
GO:0016705; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; Molecular Function
GO:0016706; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors; Molecular Function
GO:0016708; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NADH or NADPH as one donor, and incorporation of two atoms of oxygen into one donor; Molecular Function
GO:0016717; oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water; Molecular Function
GO:0016740; transferase activity; Molecular Function
GO:0016742; hydroxymethyl-, formyl- and related transferase activity; Molecular Function
GO:0016743; carboxyl- or carbamoyltransferase activity; Molecular Function
GO:0016746; transferase activity, transferring acyl groups; Molecular Function
GO:0016747; transferase activity, transferring acyl groups other than amino-acyl groups; Molecular Function
GO:0016756; glutathione gamma-glutamylcysteinyltransferase activity; Molecular Function
GO:0016757; transferase activity, transferring glycosyl groups; Molecular Function
GO:0016758; transferase activity, transferring hexosyl groups; Molecular Function
GO:0016760; cellulose synthase (UDP-forming) activity; Molecular Function
GO:0016762; xyloglucan:xyloglucosyl transferase activity; Molecular Function
GO:0016763; transferase activity, transferring pentosyl groups; Molecular Function
GO:0016765; transferase activity, transferring alkyl or aryl (other than methyl) groups; Molecular Function
GO:0016769; transferase activity, transferring nitrogenous groups; Molecular Function
GO:0016772; transferase activity, transferring phosphorus-containing groups; Molecular Function

Table B.4 (cont'd)

GO:0016773; phosphotransferase activity, alcohol group as acceptor; Molecular Function
GO:0016779; nucleotidyltransferase activity; Molecular Function
GO:0016780; phosphotransferase activity, for other substituted phosphate groups; Molecular Function
GO:0016787; hydrolase activity; Molecular Function
GO:0016788; hydrolase activity, acting on ester bonds; Molecular Function
GO:0016790; thiolester hydrolase activity; Molecular Function
GO:0016791; phosphatase activity; Molecular Function
GO:0016798; hydrolase activity, acting on glycosyl bonds; Molecular Function
GO:0016810; hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds; Molecular Function
GO:0016811; hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides; Molecular Function
GO:0016812; hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amides; Molecular Function
GO:0016813; hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amidines; Molecular Function
GO:0016817; hydrolase activity, acting on acid anhydrides; Molecular Function
GO:0016818; hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides; Molecular Function
GO:0016820; hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances; Molecular Function
GO:0016829; lyase activity; Molecular Function
GO:0016831; carboxy-lyase activity; Molecular Function
GO:0016832; aldehyde-lyase activity; Molecular Function
GO:0016841; ammonia-lyase activity; Molecular Function
GO:0016844; strictosidine synthase activity; Molecular Function
GO:0016846; carbon-sulfur lyase activity; Molecular Function
GO:0016847; 1-aminocyclopropane-1-carboxylate synthase activity; Molecular Function
GO:0016853; isomerase activity; Molecular Function
GO:0016857; racemase and epimerase activity, acting on carbohydrates and derivatives; Molecular Function
GO:0016868; intramolecular transferase activity, phosphotransferases; Molecular Function
GO:0016872; intramolecular lyase activity; Molecular Function
GO:0016874; ligase activity; Molecular Function
GO:0016876; ligase activity, forming aminoacyl-tRNA and related compounds; Molecular Function
GO:0016881; acid-amino acid ligase activity; Molecular Function
GO:0016884; carbon-nitrogen ligase activity, with glutamine as amido-N-donor; Molecular Function
GO:0016887; ATPase activity; Molecular Function

Table B.4 (cont'd)

GO:0016891; endoribonuclease activity, producing 5'-phosphomonoesters; Molecular Function
GO:0016901; oxidoreductase activity, acting on the CH-OH group of donors, quinone or similar compound as acceptor; Molecular Function
GO:0016972; thiol oxidase activity; Molecular Function
GO:0016984; ribulose-bisphosphate carboxylase activity; Molecular Function
GO:0016987; sigma factor activity; Molecular Function
GO:0016992; lipoate synthase activity; Molecular Function
GO:0016998; cell wall macromolecule catabolic process; Biological Process
GO:0017004; cytochrome complex assembly; Biological Process
GO:0017038; protein import; Biological Process
GO:0017056; structural constituent of nuclear pore; Molecular Function
GO:0017069; snRNA binding; Molecular Function
GO:0017089; glycolipid transporter activity; Molecular Function
GO:0017111; nucleoside-triphosphatase activity; Molecular Function
GO:0017148; negative regulation of translation; Biological Process
GO:0017150; tRNA dihydrouridine synthase activity; Molecular Function
GO:0017176; phosphatidylinositol N-acetylglucosaminyltransferase activity; Molecular Function
GO:0018024; histone-lysine N-methyltransferase activity; Molecular Function
GO:0018106; peptidyl-histidine phosphorylation; Biological Process
GO:0018279; protein N-linked glycosylation via asparagine; Biological Process
GO:0018298; protein-chromophore linkage; Biological Process
GO:0018342; protein prenylation; Biological Process
GO:0019001; guanyl nucleotide binding; Molecular Function
GO:0019008; molybdopterin synthase complex; Cellular Component
GO:0019139; cytokinin dehydrogenase activity; Molecular Function
GO:0019205; nucleobase-containing compound kinase activity; Molecular Function
GO:0019211; phosphatase activator activity; Molecular Function
GO:0019239; deaminase activity; Molecular Function
GO:0019277; UDP-N-acetylgalactosamine biosynthetic process; Biological Process
GO:0019288; isopentenyl diphosphate biosynthetic process, mevalonate-independent pathway; Biological Process
GO:0019295; coenzyme M biosynthetic process; Biological Process
GO:0019307; mannose biosynthetic process; Biological Process
GO:0019318; hexose metabolic process; Biological Process
GO:0019363; pyridine nucleotide biosynthetic process; Biological Process
GO:0019538; protein metabolic process; Biological Process
GO:0019439; aromatic compound catabolic process; Biological Process
GO:0019556; histidine catabolic process to glutamate and formamide; Biological Process

Table B.4 (cont'd)

GO:0019646; aerobic electron transport chain; Biological Process
GO:0019684; photosynthesis, light reaction; Biological Process
GO:0019752; carboxylic acid metabolic process; Biological Process
GO:0019773; proteasome core complex, alpha-subunit complex; Cellular Component
GO:0019843; rRNA binding; Molecular Function
GO:0019867; outer membrane; Cellular Component
GO:0019887; protein kinase regulator activity; Molecular Function
GO:0019898; extrinsic to membrane; Cellular Component
GO:0019901; protein kinase binding; Molecular Function
GO:0019904; protein domain specific binding; Molecular Function
GO:0019953; sexual reproduction; Biological Process
GO:0020037; heme binding; Molecular Function
GO:0022857; transmembrane transporter activity; Molecular Function
GO:0022891; substrate-specific transmembrane transporter activity; Molecular Function
GO:0022900; electron transport chain; Biological Process
GO:0030001; metal ion transport; Biological Process
GO:0030036; actin cytoskeleton organization; Biological Process
GO:0030054; cell junction; Cellular Component
GO:0030060; L-malate dehydrogenase activity; Molecular Function
GO:0030071; regulation of mitotic metaphase/anaphase transition; Biological Process
GO:0030117; membrane coat; Cellular Component
GO:0030126; COPI vesicle coat; Cellular Component
GO:0030127; COPII vesicle coat; Cellular Component
GO:0030131; clathrin adaptor complex; Cellular Component
GO:0030132; clathrin coat of coated pit; Cellular Component
GO:0030145; manganese ion binding; Molecular Function
GO:0030151; molybdenum ion binding; Molecular Function
GO:0030163; protein catabolic process; Biological Process
GO:0030170; pyridoxal phosphate binding; Molecular Function
GO:0030173; integral to Golgi membrane; Cellular Component
GO:0030234; enzyme regulator activity; Molecular Function
GO:0030244; cellulose biosynthetic process; Biological Process
GO:0030246; carbohydrate binding; Molecular Function
GO:0030247; polysaccharide binding; Molecular Function
GO:0030259; lipid glycosylation; Biological Process
GO:0030288; outer membrane-bounded periplasmic space; Cellular Component
GO:0030410; nicotianamine synthase activity; Molecular Function
GO:0030418; nicotianamine biosynthetic process; Biological Process
GO:0030488; tRNA methylation; Biological Process

Table B.4 (cont'd)

GO:0030515; snoRNA binding; Molecular Function
GO:0030529; ribonucleoprotein complex; Cellular Component
GO:0030598; rRNA N-glycosylase activity; Molecular Function
GO:0030599; pectinesterase activity; Molecular Function
GO:0030677; ribonuclease P complex; Cellular Component
GO:0030688; preribosome, small subunit precursor; Cellular Component
GO:0030833; regulation of actin filament polymerization; Biological Process
GO:0030904; retromer complex; Cellular Component
GO:0030955; potassium ion binding; Molecular Function
GO:0030976; thiamine pyrophosphate binding; Molecular Function
GO:0030983; mismatched DNA binding; Molecular Function
GO:0031011; Ino80 complex; Cellular Component
GO:0031072; heat shock protein binding; Molecular Function
GO:0031120; snRNA pseudouridine synthesis; Biological Process
GO:0031145; anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process; Biological Process
GO:0031167; rRNA methylation; Biological Process
GO:0031227; intrinsic to endoplasmic reticulum membrane; Cellular Component
GO:0031305; integral to mitochondrial inner membrane; Cellular Component
GO:0031361; integral to thylakoid membrane; Cellular Component
GO:0031418; L-ascorbic acid binding; Molecular Function
GO:0031461; cullin-RING ubiquitin ligase complex; Cellular Component
GO:0031625; ubiquitin protein ligase binding; Molecular Function
GO:0031966; mitochondrial membrane; Cellular Component
GO:0032012; regulation of ARF protein signal transduction; Biological Process
GO:0032040; small-subunit processome; Cellular Component
GO:0032065; cortical protein anchoring; Biological Process
GO:0032259; methylation; Biological Process
GO:0032312; regulation of ARF GTPase activity; Biological Process
GO:0032313; regulation of Rab GTPase activity; Biological Process
GO:0032324; molybdopterin cofactor biosynthetic process; Biological Process
GO:0032549; ribonucleoside binding; Molecular Function
GO:0032955; regulation of barrier septum assembly; Biological Process
GO:0032957; inositol trisphosphate metabolic process; Biological Process
GO:0032968; positive regulation of transcription elongation from RNA polymerase II promoter; Biological Process
GO:0033014; tetrapyrrole biosynthetic process; Biological Process
GO:0033177; proton-transporting two-sector ATPase complex, proton-transporting domain; Cellular Component
GO:0033178; proton-transporting two-sector ATPase complex, catalytic domain; Cellular Component

Table B.4 (cont'd)

GO:0033897; ribonuclease T2 activity; Molecular Function
GO:0033903; endo-1,3(4)-beta-glucanase activity; Molecular Function
GO:0033925; mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase activity; Molecular Function
GO:0033926; glycopeptide alpha-N-acetylgalactosaminidase activity; Molecular Function
GO:0034755; iron ion transmembrane transport; Biological Process
GO:0034968; histone lysine methylation; Biological Process
GO:0035004; phosphatidylinositol 3-kinase activity; Molecular Function
GO:0035091; phosphatidylinositol binding; Molecular Function
GO:0035434; copper ion transmembrane transport; Biological Process
GO:0035435; phosphate ion transmembrane transport; Biological Process
GO:0035556; intracellular signal transduction; Biological Process
GO:0042026; protein refolding; Biological Process
GO:0042147; retrograde transport, endosome to Golgi; Biological Process
GO:0042176; regulation of protein catabolic process; Biological Process
GO:0042218; 1-aminocyclopropane-1-carboxylate biosynthetic process; Biological Process
GO:0042254; ribosome biogenesis; Biological Process
GO:0042256; mature ribosome assembly; Biological Process
GO:0042309; homoiothermy; Biological Process
GO:0042318; penicillin biosynthetic process; Biological Process
GO:0042393; histone binding; Molecular Function
GO:0042398; cellular modified amino acid biosynthetic process; Biological Process
GO:0042545; cell wall modification; Biological Process
GO:0042546; cell wall biogenesis; Biological Process
GO:0042549; photosystem II stabilization; Biological Process
GO:0042578; phosphoric ester hydrolase activity; Molecular Function
GO:0042586; peptide deformylase activity; Molecular Function
GO:0042623; ATPase activity, coupled; Molecular Function
GO:0042626; ATPase activity, coupled to transmembrane movement of substances; Molecular Function
GO:0042651; thylakoid membrane; Cellular Component
GO:0042719; mitochondrial intermembrane space protein transporter complex; Cellular Component
GO:0042742; defense response to bacterium; Biological Process
GO:0042765; GPI-anchor transamidase complex; Cellular Component
GO:0042802; identical protein binding; Molecular Function
GO:0042803; protein homodimerization activity; Molecular Function
GO:0043022; ribosome binding; Molecular Function
GO:0043039; tRNA aminoacylation; Biological Process
GO:0043043; peptide biosynthetic process; Biological Process

Table B.4 (cont'd)

GO:0043085; positive regulation of catalytic activity; Biological Process
GO:0043086; negative regulation of catalytic activity; Biological Process
GO:0043140; ATP-dependent 3'-5' DNA helicase activity; Molecular Function
GO:0043161; proteasomal ubiquitin-dependent protein catabolic process; Biological Process
GO:0043169; cation binding; Molecular Function
GO:0043234; protein complex; Cellular Component
GO:0043461; proton-transporting ATP synthase complex assembly; Biological Process
GO:0043531; ADP binding; Molecular Function
GO:0043565; sequence-specific DNA binding; Molecular Function
GO:0043631; RNA polyadenylation; Biological Process
GO:0043666; regulation of phosphoprotein phosphatase activity; Biological Process
GO:0043682; copper-transporting ATPase activity; Molecular Function
GO:0043754; dihydrolipoyllysine-residue (2-methylpropanoyl)transferase activity; Molecular Function
GO:0044070; regulation of anion transport; Biological Process
GO:0044237; cellular metabolic process; Biological Process
GO:0044262; cellular carbohydrate metabolic process; Biological Process
GO:0044267; cellular protein metabolic process; Biological Process
GO:0044431; Golgi apparatus part; Cellular Component
GO:0045039; protein import into mitochondrial inner membrane; Biological Process
GO:0045040; protein import into mitochondrial outer membrane; Biological Process
GO:0045116; protein neddylation; Biological Process
GO:0045156; electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity; Molecular Function
GO:0045226; extracellular polysaccharide biosynthetic process; Biological Process
GO:0045261; proton-transporting ATP synthase complex, catalytic core F(1); Cellular Component
GO:0045263; proton-transporting ATP synthase complex, coupling factor F(o); Cellular Component
GO:0045300; acyl-[acyl-carrier-protein] desaturase activity; Molecular Function
GO:0045454; cell redox homeostasis; Biological Process
GO:0045735; nutrient reservoir activity; Molecular Function
GO:0045892; negative regulation of transcription, DNA-dependent; Biological Process
GO:0045893; positive regulation of transcription, DNA-dependent; Biological Process
GO:0045901; positive regulation of translational elongation; Biological Process
GO:0045905; positive regulation of translational termination; Biological Process
GO:0045980; negative regulation of nucleotide metabolic process; Biological Process
GO:0046034; ATP metabolic process; Biological Process
GO:0046080; dUTP metabolic process; Biological Process
GO:0046168; glycerol-3-phosphate catabolic process; Biological Process

Table B.4 (cont'd)

GO:0046274; lignin catabolic process; Biological Process
GO:0046373; L-arabinose metabolic process; Biological Process
GO:0046422; violaxanthin de-epoxidase activity; Molecular Function
GO:0046488; phosphatidylinositol metabolic process; Biological Process
GO:0046556; alpha-N-arabinofuranosidase activity; Molecular Function
GO:0046836; glycolipid transport; Biological Process
GO:0046854; phosphatidylinositol phosphorylation; Biological Process
GO:0046872; metal ion binding; Molecular Function
GO:0046873; metal ion transmembrane transporter activity; Molecular Function
GO:0046907; intracellular transport; Biological Process
GO:0046912; transferase activity, transferring acyl groups, acyl groups converted into alkyl on transfer; Molecular Function
GO:0046923; ER retention sequence binding; Molecular Function
GO:0046933; hydrogen ion transporting ATP synthase activity, rotational mechanism; Molecular Function
GO:0046938; phytochelatin biosynthetic process; Biological Process
GO:0046949; fatty-acyl-CoA biosynthetic process; Biological Process
GO:0046961; proton-transporting ATPase activity, rotational mechanism; Molecular Function
GO:0046983; protein dimerization activity; Molecular Function
GO:0047134; protein-disulfide reductase activity; Molecular Function
GO:0047325; inositol tetrakisphosphate 1-kinase activity; Molecular Function
GO:0047617; acyl-CoA hydrolase activity; Molecular Function
GO:0047750; cholestenol delta-isomerase activity; Molecular Function
GO:0047800; cysteamine dioxygenase activity; Molecular Function
GO:0048015; phosphatidylinositol-mediated signaling; Biological Process
GO:0048037; cofactor binding; Molecular Function
GO:0048038; quinone binding; Molecular Function
GO:0048046; apoplast; Cellular Component
GO:0048193; Golgi vesicle transport; Biological Process
GO:0048278; vesicle docking; Biological Process
GO:0048280; vesicle fusion with Golgi apparatus; Biological Process
GO:0048478; replication fork protection; Biological Process
GO:0048500; signal recognition particle; Cellular Component
GO:0048544; recognition of pollen; Biological Process
GO:0050080; malonyl-CoA decarboxylase activity; Molecular Function
GO:0050242; pyruvate, phosphate dikinase activity; Molecular Function
GO:0050307; sucrose-phosphate phosphatase activity; Molecular Function
GO:0050511; undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase activity; Molecular Function
GO:0050660; flavin adenine dinucleotide binding; Molecular Function

Table B.4 (cont'd)

GO:0050661; NADP binding; Molecular Function
GO:0050662; coenzyme binding; Molecular Function
GO:0050790; regulation of catalytic activity; Biological Process
GO:0050825; ice binding; Molecular Function
GO:0050826; response to freezing; Biological Process
GO:0050832; defense response to fungus; Biological Process
GO:0050897; cobalt ion binding; Molecular Function
GO:0051020; GTPase binding; Molecular Function
GO:0051082; unfolded protein binding; Molecular Function
GO:0051087; chaperone binding; Molecular Function
GO:0051186; cofactor metabolic process; Biological Process
GO:0051205; protein insertion into membrane; Biological Process
GO:0051258; protein polymerization; Biological Process
GO:0051276; chromosome organization; Biological Process
GO:0051287; NAD binding; Molecular Function
GO:0051301; cell division; Biological Process
GO:0051536; iron-sulfur cluster binding; Molecular Function
GO:0051537; 2 iron, 2 sulfur cluster binding; Molecular Function
GO:0051539; 4 iron, 4 sulfur cluster binding; Molecular Function
GO:0051603; proteolysis involved in cellular protein catabolic process; Biological Process
GO:0051726; regulation of cell cycle; Biological Process
GO:0051861; glycolipid binding; Molecular Function
GO:0051920; peroxiredoxin activity; Molecular Function
GO:0052716; hydroquinone:oxygen oxidoreductase activity; Molecular Function
GO:0052725; inositol-1,3,4-trisphosphate 6-kinase activity; Molecular Function
GO:0052726; inositol-1,3,4-trisphosphate 5-kinase activity; Molecular Function
GO:0055085; transmembrane transport; Biological Process
GO:0055114; oxidation-reduction process; Biological Process
GO:0070402; NADPH binding; Molecular Function
GO:0070403; NAD+ binding; Molecular Function
GO:0070461; SAGA-type complex; Cellular Component
GO:0071266; 'de novo' L-methionine biosynthetic process; Biological Process
GO:0071805; potassium ion transmembrane transport; Biological Process
GO:0072488; ammonium transmembrane transport; Biological Process
GO:0097157; pre-mRNA intronic binding; Molecular Function
GO:2001070; starch binding; Molecular Function

REFERENCES

REFERENCES

Barrero, J. M., Mrva, K., Talbot, M. J., White, R. G., Taylor, J., Gubler, F., Mares, D. J. (2013) Genetic, hormonal, and physiological analysis of late maturity alpha-amylase in wheat. Plant Physiology 161: 1265-1277.

Black, M., J. D. Bewley, P. Halmer. (2006) The encyclopedia of seeds science, technology and uses. CABI Publishing, Wallingford, Oxfordshire, p 528.

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

Brenchley R., Spannagl M., Pfeifer M., Barker G.L., D'Amore R., Allen A.M., McKenzie N., Kramer M., Kerhornou A., Bolser D., Kay S., Waite D., Trick M., Bancroft I., Gu Y., Huo N., Luo M.C., Sehgal S., Gill B., Kianian S., Anderson O., Kersey P., Dvorak J., McCombie W.R., Hall A., Mayer K.F., Edwards K.J., Bevan M.W., Hall N. (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705-710.

Dalma-Weiszhausz D.D., Warrington J., Tanimoto E.Y., Miyada C.G.(2006) The Affymetrix GeneChip® Platform: An Overview, In: Alan Kimmel, and Brian Oliver, Editor(s), Methods in Enzymology, Academic Press, 2006, Volume 410, pp. 3-28.

Debeaujon I., Lepiniec L., Pourcel L., Routaboul J.-M. (2007) Seed coat development and dormancy. Annual Plant Reviews Volume 27: Seed Development, Dormancy and Germination, Blackwell Publishing Ltd. pp. 25-49.

Debeaujon I, Léon-Kloosterziel K.M., Koornneef M. (2000) Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. Plant Physiol.122: 403-414.

Dekkers B.J., Pearce S., Bolderen-Veldkamp R.P., Marshall A., Widera P., Gilbert J., Drost H-G, Bassel G., Müller K., King J., Wood A., Grosse I., Quint M., Krasnogor N., Leubner-Metzger G., Holdsworth M., Bentsink L. (2013) Transcriptional dynamics of two seed compartments with opposing roles in Arabidopsis seed germination. Plant Physiol.pp113

DePauw R. M., Knox R. E., Singh A. K., Fox S. L., Humphreys D. G., Hucl P. (2012) Developing standardized methods for breeding preharvest sprouting resistant wheat, challenges and successes in Canadian wheat. Euphytica 188(1): 7-14.

Derera N. F. (Ed.)(1989) Preharvest field sprouting in cereals., CRC Press Inc., Boca Raton, Florida.

Dobrzanska M., Tomaszewski M., Grzelczak Z., Rejman E., Buchowicz J. (1973) Cascade activation of genome transcription in wheat. Nature, 244:507-509.

Duan J., Xia C., Zhao G., Jia J., Kong X. (2012) Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. BMC Genomics 13:392.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

Flintham J.E. (2000) Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. Seed Science Research 10:43-50.

Freed R.D., Everson E.H., Ringlund K., Gullord M. (1976) Seed coat in wheat and the relationship to seed dormancy at maturity. Cereal Res. Comm. 4:147-148.

Fu L., Niu B., Zhu Z., Li W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28: 3150-3152.

Gale M.D., Lenton J.R. (1987) Pre-harvest sprouting in wheat: a complex genetic and physiological problem affecting bread making quality in UK wheat. Aspects Appl. Biol 15:115-124.

Guo HX, Wang SX, Xu FF, Li YC, Ren JP, Wang X, Niu HB, Yin J. (2013) The role of thioredoxin h in protein metabolism during wheat (*Triticum aestivum* L.) seed germination. Plant Physiology and Biochemistry 67: 137-143.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29:644-652.

Groos C., Gay G., Perretant M.R., Gervais L., Bernard M., Dedryver F., Charmet D. (2002) Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white × red grain bread-wheat cross. Theoretical and Applied Genetics 104:39-47.

Haas BJ., Papanicolaou A. ,Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., MacManes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., LeDuc R.D., Friedman N., Regev A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protocols: 8: 1494-1512.

Hanft J.M., Wych R.D. (1982) Visual indicators of physiological maturity of hard red spring wheat. Crop Science 22:584-588.

Imtiaz M., Ogbonnaya F.C., Oman J., van Ginkel M. (2008) Characterization of Quantitative Trait Loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. Genetics 178:1725-1736.

Jia J., Zhao S., Kong X., Li Y., Zhao G., He W., Appels R., Pfeifer M., Tao Y., Zhang X., Jing R., Zhang C., Ma Y., Gao L., Gao C., Spannagl M., Mayer K.F.X., Li D., Pan S., Zheng F., Hu Q., Xia X., Li J., Liang Q., Chen J., Wicker T., Gou C., Kuang H., He G., Luo Y., Keller B., Xia Q., Lu P., Wang J., Zou H., Zhang R., Xu J., Gao J., Middleton C., Quan Z., Liu G., Wang J., Yang H., Liu X., He Z., Mao L., Wang J. (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91-95.

Kulwal P., Ishikawa G., Benscher D., Feng Z.Y., Yu L.X., Jadhav A., Mehetre S., Sorrells M.E. (2012) Association mapping for pre-harvest sprouting resistance in white winter wheat. Theoretical and Applied Genetics 125:793-805.

Kulwal P.L., Kumar N., Gaur A., Khurana P., Khurana J.P., Tyagi A.K., Balyan H.S., Gupta P.K. (2005) Mapping of a major QTL for pre-harvest sprouting tolerance on chromosome 3A in bread wheat. Theoretical and Applied Genetics 111:1052-1059.

Kyndt T., Denil S., Haegeman A., Trooskens G., De Meyer T., Van Criekinge W., Gheysen G. (2012) Transcriptome analysis of rice mature root tissue and root tips in early development by massive parallel sequencing. Journal of Experimental Botany 63:2141-2157.

Langmead B and Salzberg S. Fast gapped-read alignment with Bowtie 2. (2012) Nature Methods. 9:357-359.

Leubner-Metzger G., Frundt C., Vogeli-Lange R., Meins Jr F. (1995) Class I ß-1,3-Glucanases in the endosperm of tobacco during germination. Plant Physiology 109:751-759.

Leubner-Metzger G. (2005) ß-1,3-Glucanase gene expression in low-hydrated seeds as a mechanism for dormancy release during tobacco after-ripening. The Plant Journal 41: 133-145.

Ligternink, W., J. Kodde, M. Lammers, H. Dassen, A. H. M. van der Geest, R. A. de Maagd, and H. W. M. Hilhorst. 2007. Stress-inducible gene expression and its impact on seed and plant performance: a microarray approach. In: S. Adkins, S. Ashmore, and S. C. Navie, eds. Seeds: Biology, development and ecology. Wallingford, UK: CAB International pp. 139-148.

Liu GQ, Li WS, Zheng PH, Xu T, Chen LJ, Liu DF, Hussain S, Teng YW. (2012) Transcriptomic analysis of ‘Suli’ pear (*Pyrus pyrifolia* white pear group) buds during the dormancy by RNA-Seq. BMC Genomics 13:700.

Lohse M., Bolger A.M., Nagel A., Fernie A.R., Lunn J.E., Stitt M., Usadel B. (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Research 40:W622-W627.

Mares D., Mrva K., Cheong J., Williams K., Watson B., Storlie E., Sutherland M., Zou Y. (2005) A QTL located on chromosome 4A associated with dormancy in white- and red-grained wheats of diverse origin. Theoretical and Applied Genetics 111:1357-1364.

Martin J.A., and Wang Z. (2011) Next-generation transcriptome assembly. Nature Reviews Genetics 12:671-682.

Martin M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, North America, 17, may. 2011. Available at: http://journal.embnet.org/index.php/embnetjournal/article/view/200. Date accessed: 15 July. 2013.

Mayer KFX, Martis M, Hedley PE, Simkova H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubalakova M, Suchankova P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249-1263.

Miller R., Wu. G., Deshpande,RR., Vieler A., Gartner K, Li X.,Moellering ER, Zauner S., Cornish A.J., BS, Bullard , Sears B.B.,Kuo MH, Hegg EL, Shachar-Hill Y., Shiu S-H, and Benning C. (2010) Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. Plant Physiology: 154: 1737-1752.

Miyamoto T. and Everson E.H. (1958) Biochemical and physiological studies of wheat seed pigmentation. Agron Jour 50:733-734.

Mochida K., Yoshida T., Sakurai T., Ogihara Y., Shinozaki K. (2009) TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. Plant Physiology 150:1135-1146.

Mutasa-Gottgens E., Joshi A., Holmes H., Hedden P., Gottgens B. (2012) A new RNAseq-based reference transcriptome for sugar beet and its application in transcriptome-scale analysis of vernalization and gibberellin responses. BMC Genomics 13:99.

Narsai R., Law S.R., Carrie C., Xu L., Whelan J. (2011) In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in arabidopsis. Plant Physiol. 157: 1342-1362.

Nonogaki H., Bassel G.W., Bewley J.D. (2010) Germination-Still a mystery, Plant Science 179 (6): 574-581.

Peng X., Zhao Y., Cao J., Zhang W., Jiang H., Li X., Ma Q., Zhu S., Cheng B. (2012) CCCH-Type Zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. PLoS ONE 7:e40120.

R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Ren J.P., Yin J., Niu H.B., Wang X.G., Li Y.C. (2007) Effects of antisense-thioredoxin s gene on expression of endogenous thioredoxin h gene in transgenic wheat seed. J. Plant Physiol. Mol. Biol., 33: 325–332 (in Chinese).

Robertson G., Schein J., Chiu R., Corbett R., Field M., Jackman S.D., Mungall K., Lee S., Okada H.M., Qian J.Q., Griffith M., Raymond A., Thiessen N., Cezard T., Butterfield Y.S., Newsome R., Chan S.K., She R., Varhol R., Kamoh B., Prabhu A.L., Tam A., Zhao Y.J., Moore R.A., Hirst M., Marra M.A., Jones S.J.M., Hoodless P.A., Birol I. (2010) De novo assembly and analysis of RNA-seq data. Nature Methods 7:909-U62.

Robles J., Qureshi S.E., Stephen S.J., Wilson S.R., Burden C.J., Taylor J.M. (2012) Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics*13:484.

Savory E.A., Adhikari B.N., Hamilton J.P., Vaillancourt B., Buell C.R., Day B. (2012) mRNA-Seq analysis of the *Pseudoperonospora cubensis* transcriptome during cucumber (*Cucumis sativus* L.) infection. PLoS ONE 7.

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

Teoh K.T., Requesens D.V., Devaiah S.P., Johnson D., Huang X.Z., Howard J.A., Hood E.E. (2013) Transcriptome analysis of embryo maturation in maize. BMC Plant Biology: 13.

Trapnell C., Pachter L., Salzberg S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105-1111.

Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L., Pachter L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protocols 7:562-578.

Trapnell C., Hendrickson D.G., Sauvageau M., Goff L., Rinn J.L., Pachter L. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotech 7:562-578.

Trick M., Adamski N., Mugford S., Jiang C.-C., Febrer M., Uauy C. (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biology 12:14.

Uno Y., Furihata T., Abe H., Yoshida R., Shinozaki K., Yamaguchi-Shinozaki K. (2000) Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proceedings of the National Academy of Sciences 97: 11632-11637.

Wang Z., Gerstein M., Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57-63.

Winkel-Shirley B. (2001) Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiology 126:485-493.

Wu J.M., Carver B.F., Goad C.L. (1999) Kernel color variability of hard white and hard red winter wheat. Crop Science 39:634-638.

Wu T., Nacu S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 26: 873-881.

Zanetti S., Winzeler M., Keller M., Keller B., Messmer M. (2000) Genetic analysis of pre-harvest sprouting resistance in a wheat x spelt cross. Crop Science 40:1406-1417.

Zhang X.Y., Yao D.X., Wang Q.H., Xu W.Y., Wei Q., Wang C.C., Liu C.L., Zhang C.J., Yan H., Ling Y., Su Z., Li F.G. (2013) mRNA-seq Analysis of the *Gossypium arboreum* transcriptome Reveals Tissue Selective Signaling in Response to Water Stress during Seedling Stage. PLoS ONE 8.

Zhu X., Liu S., Meng C., Qin L., Kong L., Xia G. (2013). WRKY transcription Factors in Wheat and their induction by biotic and abiotic stress. Plant Mol Biol Rep, 31:1053-1067.

CHAPTER 4 Future directions

The resources established from this study included wheat lines that captured multiple QTL, a high-density genetic map, seed-specific transcript assembly. All together, they set up the stage for leveraging next generation sequencing (NGS) technology for PHS study in wheat. There are several potential areas can be explored based on the results from current research.

The development of heterogeneous inbred family (HIF) (Tuinstra et al., 1997) from lines containing recombination around QTL regions in the 'Vida' × MTHW0471 population would be a good way to validate QTL identified from current study. HIF focused on the residual heterozygosity around identified QTL regions. Recombinants selected for HIF have the potential to segregate within QTL regions. Multiple HIF can be developed to cover different combinations of multiple QTLs. With the high density SNP genotyping results, several QTL identified in current study were located between markers separated in less than 2 cM. With the help of HIF, fine mapping of these QTL will be relatively faster when compared with fine-mapping using neo-isogenic lines. The QTL that is of most interest is the major QTL on Chromosome 2B which explained nearly 40% of phenotypic variation of α-amylase activity. Other ones that had QTL × QTL interactions can also be further assessed by comparing between HIF families containing different combinations of these QTL. Moreover, Liu et al. (2013) cloned a gene on the short arm of Chromosome 3A from a white wheat population, which was claimed to be able to improve PHS resistance significantly. The SNP markers reported from that paper can be used to screen our population for recombinants around that region. If yes, a potential fine-mapping study would be able to explore the allelic variation between red and white wheat lines within that region.

SNP identification from NGS data was proven feasible in autopolyploid species recently (Krasileva et al., 2013; Trick et al. 2012). Thus, an attempt to identify novel SNPs from the sequenced individuals could be valuable for map saturation or future fine-mapping. On the other

hand, a fine-mapping using bulk segregant analysis (BSA) with RNA-seq reads had been done in durum wheat, which showed the potential of using next-generation sequencing data for marker-trait associations. Reads by comparing red wheat and white wheat bulks can help to identify SNPs within genic regions. When combining NGS with HIF, the SNP identification can be narrowed down to genic regions with a potential of collocating with differentially expressed genes.

During the design of wheat 9K SNP array, the non-genome specific SNPs, and especially D-genome markers, were the major limiting factors for the generation of a high-density maps. The genotyping results need to be re-clustered based on population segregation and all the heterozygotes information were lost during the re-cluster process (Cavanagh et al. 2013). Therefore, the development of genome-specific SNP is critical when considering for future map enrichment.

A recent paper utilizing progenitor genome SNPs to categorize genomic origin in cotton can also be adapted to wheat system with some modifications (Page et al., 2013). The recently published wheat A and D progenitor genome sequence can be used to help categorize genome-specific reads and call SNP within each sub-genome (Jia et al., 2013; Ling et al., 2013). Two other alternative strategies currently adopted by other groups. These resources can also be used for the validation of our categorization method. First is to use flow-sorted genome-specific chromosomes and identify chromosome specific SNP from there. This strategy is currently adopted by the International Wheat Genome Sequencing Consortium (IWGSC) and survey sequence of chromosome specific sequences will be made available for public use soon (Kellye Eversole, IWGSC, pers. comm.). The second strategy used in durum is similar to the SNP-categorization method mentioned above but using phasing information instead. (Krasileva et al.,

2013). Except for making the SNP calling easier and more accurate, collection of genome specific reads is also critical to build sub-genome specific transcriptome, which should reduce the chance of mis-assembly due to homeolog issue. As seen in a couple studies in polyploidy, the expression bias between sub-genome can be huge. Thus, the third potential application of categorized sub-genome specific reads is to capture the sub-genome level differential expression and even when DE at whole genome level is not significant. On the other hand, the extracted sub-genome can be more amenable to using diploid wheat progenitors as reference genome than using hexaploid wheat genome, which is not available at this stage. As shown in current study, the use of *A. tauschii* genome as reference can cause potential loss of sub-genome specific loci, which might bias the downstream analysis.

To the best of author's knowledge, this study is one of the first studies systematically evaluate the 9k-SNP array and NGS technology in hexaploid wheat genetic study. The analysis revealed previous identified and novel QTL and potential candidate genes involved in PHS process. The pipeline developed here can also be used in other trait discovery pipelines. With these power tools, the dissection of complex trait in polyploidy species like wheat will become more accurate and efficient.

REFERENCES

REFERENCES

Cavanagh C.R., Chao S., Wang S., Huang B.E., Stephen S., Kiani S., Forrest K., Saintenac C., Brown-Guedira G.L., Akhunova A., See D., Bai G., Pumphrey M., Tomar L., Wong D., Kong S., Reynolds M., da Silva M.L., Bockelman H., Talbert L., Anderson J.A., Dreisigacker S., Baenziger S., Carter A., Korzun V., Morrell P.L., Dubcovsky J., Morell M.K., Sorrells M.E., Hayden M.J., Akhunov E. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences 110:8057-8062.

Jia J., Zhao S., Kong X., Li Y., Zhao G., He W., Appels R., Pfeifer M., Tao Y., Zhang X., Jing R., Zhang C., Ma Y., Gao L., Gao C., Spannagl M., Mayer K.F.X., Li D., Pan S., Zheng F., Hu Q., Xia X., Li J., Liang Q., Chen J., Wicker T., Gou C., Kuang H., He G., Luo Y., Keller B., Xia Q., Lu P., Wang J., Zou H., Zhang R., Xu J., Gao J., Middleton C., Quan Z., Liu G., Wang J., Yang H., Liu X., He Z., Mao L., Wang J. (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91-95.

Krasileva K., Buffalo V., Bailey P., Pearce S., Ayling S., Tabbita F., Soria M., Wang S., Consortium I., Akhunov E., Uauy C., Dubcovsky J. (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. Genome Biology 14:R66.

Ling H.-Q., Zhao S., Liu D., Wang J., Sun H., Zhang C., Fan H., Li D., Dong L., Tao Y., Gao C., Wu H., Li Y., Cui Y., Guo X., Zheng S., Wang B., Yu K., Liang Q., Yang W., Lou X., Chen J., Feng M., Jian J., Zhang X., Luo G., Jiang Y., Liu J., Wang Z., Sha Y., Zhang B., Wu H., Tang D., Shen Q., Xue P., Zou S., Wang X., Liu X., Wang F., Yang Y., An X., Dong Z., Zhang K., Zhang X., Luo M.-C., Dvorak J., Tong Y., Wang J., Yang H., Li Z., Wang D., Zhang A., Wang J. (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87-90.

Liu S., Sehgal S.K., Li J., Lin M., Trick H.N., Yu J., Gill B.S., Bai G. (2013) Cloning and characterization of a critical regulator for pre-harvest sprouting in wheat. Genetics 195:263-273.

Page J.T., Gingle A.R., Udall J.A. (2013) PolyCat: A Resource for Genome Categorization of Sequencing Reads From Allopolyploid Organisms. G3: Genes|Genomes|Genetics 3:517-525.

Tuinstra M.R., Ejeta G., Goldsbrough P.B. (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theoretical and Applied Genetics 95:1005-1011.

Trick M., Adamski N., Mugford S., Jiang C.C., Febrer M., Uauy C. (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biology 12:14.