COMPARATIVE GENETIC ANALYSIS AND FINE MAPPING OF A MAJOR PREHARVEST SPROUTING QTL INTERVAL IN WHITE WINTER WHEAT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Suthasinee Somyong

May 2010

COMPARATIVE GENETIC ANALYSIS AND FINE MAPPING OF A MAJOR PREHARVEST SPROUTING QTL INTERVAL IN WHITE WINTER WHEAT

Suthasinee Somyong, Ph.D.

Cornell University 2010

Wheat preharvest sprouting (PHS) occurs when grain germinates on the plant before harvest, resulting in reduced grain quality. Previous mapping of quantitative trait loci (QTL) revealed a major PHS QTL located on chromosome 2B.1 that was significant in 16 environments and explained from 5 to 31% of the phenotypic variation. The objective of this project was to fine map the PHS QTL interval on 2B.1. For fine mapping the QTL interval, ESTs (expressed sequence tags) and comparative mapping were used to design 278 primer pairs, of which 22 produced polymorphic amplicons that mapped to the group 2 chromosomes. Fourteen mapped to chromosome 2B but only 10 were located in the QTL interval. Recombinant backcross populations (BC1F4 and BC1F5) were developed by backcrossing selected double haploids to a recurrent parent and selfing to the F4 and F5 generations. In each generation, three markers in the PHS QTL interval were used to screen for recombinants. Comparative analysis revealed good macrocollinearity between the PHS interval and a 3 million base pair (mb) region in rice chromosomes 7 and 3, and a 2.5 mb region in *Brachypodium* Super_0. Fine mapping revealed that the 2B.1 PHS QTL interval contained 2 PHS QTLs. The first PHS QTL, located between Wmc453c and Barc55, contributed one third of phenotypic variation and collocated with the seed dormancy QTL. The second PHS QTL, between Wmc474 and rCaPK, contributed two thirds of the variation. The PHS resistance alleles were contributed from Cayuga parent. One of the PHS Cayuga resistance alleles originated in Golden Chief, a parent of Clark's Cream. One of the

candidate genes, Calmodulin/Ca^{2+} dependent protein kinase, linked with one PHS QTL. Although many recombinant families were identified, the lack of polymorphism for markers in the QTL interval prevented the localization of the recombination breakpoints and identification of the gene underlying the phenotype.

BIOGRAPHICAL SKETCH

Suthasinee Somyong was born on April 26, 1980 in Nakhonratchasima province, in the Isan region of Thailand. She grew up in a farming family that grows cassava and coffee. She accompanied her parents and siblings on the farm during her early years but she has been studying away from home since high school. Early on, she did not realize her interest in science. Instead she thought that she wanted to become a doctor or nurse. But later on at Burirum Pitayakom High School, she received a scholarship to study biology at Khonkaen University. She did pummelo tissue culture and transformation techniques in her Bachelor's project. She continued with a Master's program at Mahidol University in the Institute of Molecular Biology and Genetics. She did her small advanced project about infectivity of the papaya ringspot virus. However, her main research was changed to work with, tiger prawn. She studied the serotonin receptor in prawn. But her interest in plants never subsided. In 2005, she received a Thai government scholarship to do a Ph.D program at Cornell University, in the Department of Plant Breeding. She did fine-mapping in white wheat at the Small Grain Research Project Laboratories. After she completed her Ph.D program in 2010, she fulfilled a requirement of her scholarship by working at the National Center for Genetic Engineering and Biotechnology in Thailand. Her focus will be on plant physiology and biochemistry. She may also incorporate plant molecular breeding into her research.

I would like to dedicate this effort to my parents, Mr. and Mrs. Somyong, whose simple life lessons and encouragement helped me to push forward during difficult times.

ACKNOWLEDGMENTS

I am wholeheartedly grateful and appreciative for the helpfulness, encouragement and advice from my advisor and special committee chair, Prof. Mark E. Sorrells throughout my Ph.D. study and research work. Also, I am deeply grateful to my committee members, Prof. Ralph L. Obendorf and Prof. James J. Giovannoni for their helpfulness and advice. This fine-mapping project has been an accomplishment due to the helpfulness and advice from our lab members and field crews. I appreciate all of their help, especially David Bencher, James Tanaka, Jesse Munkvold, Mahmoud Zeid and Prof. Yong Gu Cho. I would like to thank my friends, all staff and faculty of the Plant Breeding Department, especially, Cynda Farnham, Tee Havananda, Sarah Collier, Michael Kovach and Robert Bode.

I would like to thank the Thailand government for this opportunity and their financial support. I thank the Office of Educational Affairs in Washington D.C., for taking care of many aspects of my Ph.D study.

Finally, I would like to express my gratitude to my parents, brother and sisters for their helpfulness and encouragement and my love, Singh who always stayed beside me in both sadness and happiness.

Suthasinee Somyong

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

min = minute

cM = centimorgan

cm = centimeter

rpm = engine speed

M = molar

µl = microliter

ml = milliliter

hrs = hours

PCR = Polymerase Chain Reaction

kb = kilobase pair

chro. = chromosome

CHAPTER 1

LITERATURE REVIEW

Preharvest sprouting

Preharvest sprouting (PHS) is a condition in which germination of grain occurs after seed maturation but before harvesting. Preharvest sprouting is a problem in cereals such as wheat, rice, barley and maize (Bewley and Black 1994). In wheat, the main effects of PHS are a lower yield due to harvest losses and, more importantly, a reduction in end-product quality such as low volume bread, compact interior and dark crust. For example, sprouting damage leads to sticky dough, which causes gummy crumb leading to problems with bread slicing (Dexter 1993) (Figure 1.1).

Preharvest sprouting resistance is quantitative trait (Zanetti et al. 2000). Genetic studies have revealed the potential genes controlling the PHS mechanism (Li et al. 2004; Bentsink et al. 2006; Gu et al. 2009; Sugimoto et al. 2010). Most reports explain PHS resistance in terms of a seed dormancy mechanism (Li et al. 2004; Ogbonnaya et al. 2008), however, PHS resistance results from many factors such as water uptake, drying rate of the ear, seed dormancy and storage reserve mobilization (King and Richards 1984a; King and von Wettstein-Knowles 2000; Nielsen et al. 1984).

Types of seed dormancy

Seed dormancy is the absence of germination of viable seeds under conditions that are favorable for germination. There are two types of seed dormancy; seed coat-imposed dormancy and embryo dormancy (Bewley and Black 1994). For coat-imposed dormancy, lemma, palea, pericarp, testa and endosperm are the barrier to prevent water uptake, gas exchange, and hormone penetration (Adkins et al. 2002).

A B C

Figure 1.1. Baked bread from sprouted wheat flour affecting bread slicing.

Loaves were prepared from control (A), 3% germinated (B) and 5% germinated (C) wheat, respectively (Dexter 1993)

For embryo dormancy, abscisic acid (ABA) plays the major role (McCrate et al. 1982). Also, the types of seed dormancy can be distinguished according to the timing of dormancy. There are two types of dormancy; primary and secondary dormancy. The seeds that are released from a mother plant in a dormant state exhibit primary dormancy. The seed released from a mother in a non-dormant state but then become dormant state due to unfavorable conditions exhibit secondary dormancy.

For coat-imposed dormancy, the surrounding seed coat acts in the following ways. The seed coat can act as a permeability barrier that prevents water uptake and gaseous exchange. This kind of dormancy can be overcome by weakening and breaking down the seed coat and pericarp tissue of the caryopsis, which is called scarification. It can also create a mechanical barrier that prevents embryo expansion. This may be overcome by enzymes produced by the embryo or by physical parameters such as temperature fluctuation. Animal digestion or fire may also weaken this barrier. Seed coat can serve as a germination inhibitor, preventing water uptake. This coat-imposed dormancy is common in dicots but not common in grain crops. ABA and coumarin are common inhibitors. Prolonged leaching out by rainfall and decay of coat tissues will lessen this form of dormancy (Adkins et al. 2002).

Embryo dormancy is believed to be an important mechanism in grasses (Simpson 1990). It involves a balance between ABA and GA (Gibberellins), expression of some genes, activity of respiratory pathways and mobilization and utilization of food reserves. In wheat, dormancy variation among cultivars was primarily caused by differential response of their embryos to endogenous inhibitors that counteracted with GA (McCrate et al. 1982). Also, excised embryos from preharvest resistant genotype germinated slowly in the presence of exogenous inhibitors, including ABA (Stoy and Sundin 1976; Walker-Simmons 1987) and embryo sensitivity is affected from both endogenous inhibitors and exogenous ABA

and is countered by GA (Upadhyay et al. 1988). Also, embryos of wheat mutants, which lacked dormancy at seed maturation, rapidly lost the sensitivity to ABA (Kawakami et al. 1997). Embryo dormancy can be released by temperature treatment such as temperature fluctuation and chilling or by light requirement. For example, in many warm season grasses, light may operate by altering between germination promotion and inhibition and released by chemical stimulant: GA, ABA, cytokinins and ethylene that are believed to play major roles in overcoming seed dormancy (Adkins et al. 2002)

Seed dormancy mechanisms

Seed dormancy is the absence of germination of viable seeds under conditions that are favorable for germination. A lack of seed dormancy is one mechanism that may contribute to PHS damage (Bewley and Black 1994). In general, ABA promotes and maintains dormancy whereas GA promotes the release of dormancy though germination (Finkelstein et al. 2008; Finch-Savage and Leubner-Metzger 2006; Kucera et al. 2005; Finch-Savage et al. 2007). Environmental signals regulate this balance by modifying the expression of biosynthetic and catabolic enzymes (Finkelstein et al. 2008). Seed dormancy induction takes place during seed development until seed maturation (Finkelstein et al. 2008; Amen 1968; Bentsink and Koornneef 2008). Seed dormancy maintenance takes place during the desiccation tolerance phase after seed maturation (Finkelstein et al. 2008; Kermode 2005).

Many genes have been reported to be involved in seed dormancy and seed germination (Finch-Savage et al. 2007; Gualberti et al. 2002; Oh et al. 2004; Utsugi et al. 2008; Yamauchi et al. 2004). Most of those genes were studied in Arabidopsis but some genes or ortholog genes were studied in cereals. Factors from both the mother plant and the seed embryo affect the seed dormancy mechanism. The overview of

some genes that were reported to involve in seed dormancy during seed maturation until germination was shown on Figure 1.2. Most of these genes were studied in Arabidopsis. However, those genes may be useful to identify the orthologous genes in cereals.

During dormancy induction, which took place at seed maturation stage, ABA plays the major role, with many negative and positive regulators involved. ABA affects embryonic maturation, prevents seed germination and controls nutrient reserve accumulation. In this stage, dormancy induction is also controlled from both maternal factors such as zinc-finger factors *DAG1* and *DAG2* and embryonic factors such as *LEC1* (*LEAFY COTYLEDON*), *LEC2* and *FUS3* (*FUSCA3*). *DAG1* activates maternal genes that promote dormancy in embryo whereas *DAG2* represses the action by binding with the *DAG1* factor but the specific targets of the DAG have not yet identified (Gualberti et al. 2002) and inactivation of *DAG1* affected the response to light and testa integrity (Papi et al. 2002). In Arabidopsis, *LEC1*, *LEC2* and *FUS3* controlled most aspects of seed maturation such as accumulation of storage compound, desiccation tolerance and dormancy as well as embryogenesis (To et al. 2006; Tanaka et al. 2008). Also, *LEC1* and *LEC2* were reported to up-regulate *FUS3* and *ABI3*, *Vp1* ortholog (Gonzalez-Garcia et al. 2003).

For ABA-dependent signaling, upstream regulation involves both ABA biosynthesis and catabolism. For ABA biosynthesis, 9-cis-epoxycarotenoid dioxygenases (*NCEDs*) and zeaxanthin epoxidase (*ZEP*) are involved. Cytochrome P450 707A (*CYP 707As*) controls ABA catabolism (Frey et al. 1999; Lefebvre et al. 2006; Okamoto et al. 2006). In downstream ABA regulation, protein phosphatase 2C (*PP2Cs*), *ERA1* (*ENHANCED RESPONSE TO ABA*), *ABI3* (*ABA INSENSITIVE*), Vp1 (Viviparous *ABI3* ortholog), *DOG* (*DELAY OF GERMINATION*),

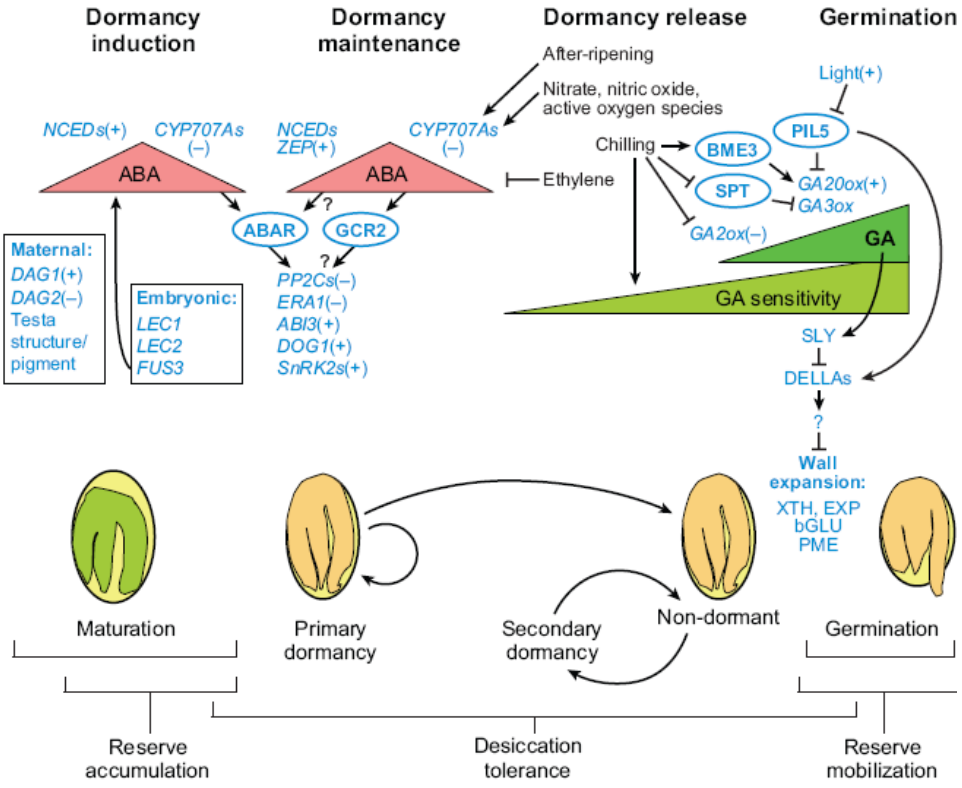


Figure 1.2. Summary of some molecular factors that were involved in seed dormancy mechanism from seed maturation to seed germination (Finkelstein et al. 2008).

Positive regulation is indicated by (+) and negative regulation is indicated by (-).

SnPK (SNF1-related protein kinase), ABAR (ABA receptor), GCR (G-protein coupled receptor) are involved (Finkelstein et al. 2008). It was suggested that protein phosphatase 2C (PP2Cs) regulated ABA signaling by promoting the transition from seed dormancy to germination during stratification in beech (*Fagus sylvatica*). The GCR gene (Arabidopsis G protein-coupled receptor gene) involved inhibition of ABA response. Over expression of this gene abolished seed dormancy and reduced time to flowering (Colucci et al. 2002; Liu et al. 2007) whereas *ABI3*, *Vp1* and *DOG* positively responded to ABA. The Arabidopsis *DOG* locus has been identified based on positional cloning and mutant analysis revealed it was required for seed dormancy induction (Bentsink et al. 2006).

For dormancy release, environmental factors such as after-ripening, nitrate, nitric oxide, active oxygen species and chilling mediated by phytohormones promote the changing balance between ABA and GA. For example, *PIL5* (Phytochrome-Interacting Factor 3-Like 5) gene is the basic helix-loop-helix (bHLH) transcription factor (Oh et al. 2004). It prevented seed germination in the dark by repression of the expression of *GA20ox1* (Gibberrellin 20 oxidase) and *GA3ox2* (Gibberrellin 3 oxidase). Light activates the degradation of PIL5 protein to promote seed germination though gibberellins in Arabidopsis (Oh et al. 2006). Fine mapping in rice chromosome 12 revealed that *PIL5* gene was linked to a seed dormancy QTL (Gu et al. 2009).

During germination, GA plays a major role. In Arabidopsis, GA biosynthesis involves *GA20ox* and *GA3ox* (Finch-Savage et al. 2007; Yamauchi et al. 2004). There was a report that GA20-oxidase, on barley 5HL, was identified as a candidate gene within the QTL that controlled seed dormancy and PHS. GA20-oxidase was collinear with the terminal end of the long arm of rice chromosome 3 and collinear with wheat 4AL (Li et al. 2004). Downstream GA regulation involves GA sensitivity such as an interaction of SLY (SLEEPY) and DELLA affecting seed wall expansion leading to

germination (Finkelstein et al. 2008). However unlike dicots, mutations in the rice GA receptor (OsGID1), which interacted with a rice DELLA protein, decreased α-amylase but still allowed germination (Ueguchi-Tanaka et al. 2005). This suggests that some factors and pathways contribute to germination in cereals differently from Arabidopsis. GA responses were repressed by DELLA because it contained GA-dependent proteosomal degradation domain and DELLA mutant was unable to interact with the GA receptor (Willige et al. 2007). In addition to GA, other hormones such as Brassinosteroid and Ethylene play a role in seed germination (Leubner-Metzger 2001).

Vivipary occurs before seed maturation so it involves seed dormancy induction and upstream ABA regulation. Mutated PHS genes that were involved in ABA biosynthesis were studied in rice (Fang and Chu 2008). Enzymes encoded by these PHS genes were phytoene desaturase (*PDS*), carotene desaturase (*ZDS*), lycopene beta-cyclase (*beta-LCY*) and zeaxanthin epoxidase (*ZEP*). They are involved in the synthesis of a carotenoid precursor of ABA. These mutated PHS genes resulted in ABA-deficiency phenotype that might contribute to the PHS trait in rice (Fang et al. 2008). This may indicate that both upstream and downstream ABA-dependent regulation affects PHS. There were many studies in wheat and cereals about ABA downstream regulation genes that affect seed dormancy (Gomez-Cadenas et al. 1999; Gomez-Cadenas et al. 2001; Anderberg and Walker-Simmons 1992; Johnson et al. 2002). For example, PKABA1-protein serine/ threonine kinase acts as the key intermediate in the signal transduction pathway mediating ABA induction (Gomez-Cadenas et al. 1999; Gomez-Cadenas et al. 2001; Anderberg and Walker-Simmons 1992). ABA induction results in suppression of GA-inducible gene expression in cereal aleurone layers (Gomez-Cadenas et al. 1999). One example is the suppression of α-amylase expression (Gomez-Cadenas et al. 2001).

Viviparous (*Vp1*), also affects suppression of α-amylase expression. Hexaploid wheat possesses 3 Vp1 homoeologous genes (*TaVp1*) on *TaVp-A1, TaVp-B1* and *TaVp-D1* on chromosome group 3. From cDNA sequence analysis, *TaVp-A1* and *TaVp-D1* transcripts were spliced incorrectly, resulting in truncated and deleted protein (McKibbin et al. 2002). *TaVp-B1*, however, was spliced correctly. *TaVp-B1* may function as the key transcription factor for seed dormancy in wheat. From transient expression, *TaVp-B1* might activate *Em* expression and repress α-amylase expression (Utsugi et al. 2008).Viviparous gene (*Vp1*) regulates seed dormancy by activating genes which are responsive to ABA. The activation involves ABA-responsive elements (ABREs), such as *Em* gene promoter. However, *Vp1* doesn't bind with ABREs directly. Instead, a bZIP factor containing a basic region leucine zipper such as TRAB1 interacts with both *Vp1* and ABREs (Hobo et al. 1999). Several bZIP factors bind with ABREs, such as wheat EmBP1 (Guiltinan et al. 1990), tobacco TAF-1 (Oeda et al. 1991), riceOSBZ8 (Nakagawa et al. 1996) and rice osZIP-1a (Nantel and Quatrano 1996). The activation of the bZIP factors is controlled via their phosphorylation states by kinases (Finkelstein et al. 2008). Those kinases are MAP (mitogene activated protein) kinase, CDPK (calcium-dependent protein kinase), SnPK (SNF-related protein kinase). For example, the PKABA1 gene was found to control wheat bZIP factor TaABF (Johnson et al. 2002) leading to repression of GA-induced genes and activation of ABA induced genes(Gomez-Cadenas et al. 1999; Gomez-Cadenas et al. 2001).

Wheat cDNA PKABA1 contains features of serine / threonine protein kinase, with all 12 conserved regions of catalytic domain highly homologous to other serine / threonine protein kinases (Anderberg and Walker-Simmons 1992). Similar kinases include rat calcium/ calmodulin dependent protein kinase II, soybean calcium dependent protein kinase, yeast SNF1 and yeast nim1+ protein kinase, from 43-52%

amino acid identity. The relationship between endogenous ABA and PKABA1 was assessed in developing wheat. PKABA transcript level was highest in mature dry seed embryos although this is the stage in which ABA declined (Anderberg and Walker-Simmons 1992). Calcium-dependent protein kinase (CDPK) is one kind of kinase that involves ABA regulation. In wheat, 20 CDPKs were identified locating in wheat chromosomes group 1, 2, 4, 5 and 6, and they responded to various stimuli including cold, hydrogen peroxide, salt, drought, powdery mildew, ABA and GA (Li et al. 2008).

During seed germination, reserve mobilization is regulated by GA. Conversion of oil to sugar is one part that provides precursors for germination. ICL (isocitrate lyase) is an enzyme involved in gluconeogenesis. Studies show that GA-dependent induction of ICL activity is mediated by the gene *SLENDER1* and requires cGMP, but does not involve the transcription factor *GAMYB*. Gibberellins and ABA therefore act antagonistically to regulate gluconeogenesis in the aleurone layer (Eastmond and Jones 2005). GA treatment on cereal aleurone increased the levels of several factors, such as Ca^{2+}, CaM (Calmodulin), cGMP, GA-Myb, and changes pH. All of these factors, in turn, cause an increase in α-amylase (Bethke et al. 1997).

Seed dormancy in red- and white- grained wheat

Preharvest sprouting resistance in some white-grain wheat varieties likely involves embryo dormancy (Bailey et al. 1999). In red-grain wheat varieties, the red testa color in maternal tissue and R alleles contribute to coat-imposed dormancy. Enhanced dormancy and red testa color are pleiotropic effects of dominant R alleles at triplicate loci in hexaploid wheat (Flintham 2000; Warner et al. 2000; Himi et al. 2002). Differences in dormancy between two red wheat lines carrying all three dominant R alleles might involve embryo dormancy in the background as well

(Flintham 2000). White-grain wheat contains homozygous recessive R genes on all three genomes but red-grain wheat contains at least one dominant allele. These R-genes are located on 3A (Metzger and Silbaugh 1970), 3B (Metzger and Silbaugh 1970; Allan and Vogel 1965) and 3D (Sears 1944). The R-gene is some 60 cM from the centromere on the long arm of group 3 (Metzger and Silbaugh 1970). Wheat gene *taVP1* is orthologous to *Vp1* (maize VIVIPAROUS 1) and *ABI3* and contributes to embryo dormancy. Gene *taVP1* was mapped on 3AL, 3BL and 3DL about 30 cM from centromere and some 30 cM proximal to the R-gene. There was no difference in ABA accumulation in red and white-grained lines (Himi et al. 2002). These authors suggested that the R-gene may influence grain dormancy by increasing ABA sensitivity in red-grained lines rather than increasing the ABA level. One QTL study reported that grain color (GC) might associate with PHS because four QTLs of GC and PHS were colocalized on 2B.1, 2B.2, 3B, 6B (Kumar et al. 2009a).

The R gene was reported to be a transcription factor of the flavonoid synthesis pathway (Himi and Noda 2004). The red pigment of wheat grain has been reported to be a polyphenolic compound, phlobaphene, synthesized at the early step of flavonoid synthesis (Miyamoto and Everson 1958). Dihydroflavonol 4-reductase (*DFR*) and Chalcone synthase (*CHS*) are genes encoding enzymes for the flavonoid pathway. *DFR* was shown to be upregulated in the grain-coated tissue in Chinese spring red wheat than white wheat whereas *CHS* was expressed in red grain wheat but not in white wheat (Himi and Noda 2004). However, different red-grain and white-grain wheat lines may either exhibit dormancy or non-dormancy (Mares et al. 2005). But, red-grained lines are usually more dormant (Ogbonnaya et al. 2007). The higher dormancy of red-grained lines results from both embryonic and maternal factors (Flintham 2000; Warner et al. 2000). In white-grain wheat, the seed dormancy or PHS QTL have been reported to be located on group 3 chromosomes (Liu et al. 2008).

Other chromosomes throughout the genome are involved, such as 4A (Mares et al. 2005; Tan et al. 2006), 5B (Tan et al. 2006), 2B (Anderson et al. 1993; Munkvold et al. 2009), 2A, and 2D (Mares et al. 2002).

Seed dormancy and preharvest sprouting QTL on wheat group 2 and orthologous regions in other species

For previous work in white winter wheat, approximately 15 PHS QTLs were located across the genome (Munkvold et al. 2009). The population consisted of doubled haploids (DH) from a cross between white-grained Cayuga (PHS resistant) and Caledonia (PHS susceptible). A major QTL, on chromosome 2BS, was significant in all 16 environments, and explained 5-31% of phenotypic variation (average = 24% of the variation). The QTL was not located specifically at the same interval in all environments. For the mean of all environments the QTL peak was at Barc55, surrounded by the flanking SSR markers; wmc474 and gwm429. The QTL was contributed by PHS resistance parent, Cayuga. QTLs were also detected on other chromosomes, including 2D, 3D and 6D. The QTLs were significant in 7, 4 and 10 environments, respectively, out of 16 environments. Consequently, seed dormancy and PHS studies between wheat group 2 and orthologous regions would help us better understand the relationship of this region with other wheat populations and other grass species.

Wheat group 2 chromosomes show high similarity with rice chromosomes 4 and 7 (La Rota and Sorrells 2004; Salse et al. 2008). The SSR marker Barc55 amplifies a locus in the deletion Bin 2BS-0.53-0.75 on wheat chromosome 2BS. This bin shows high similarity with rice chromosome 7 as well (La Rota and Sorrells 2004). Munkvold et al (2009) reported that the 2B QTL region controlling PHS also controlled seed dormancy (SD). The SD and PHS QTL overlapped but the QTL

confidence interval distance for the SD QTL was wider than that of the PHS QTL. SD and PHS QTLs also co-localized on other chromosomes such as 4A (Ogbonnaya et al. 2008; Mares et al. 2005; Tan et al. 2006).

Many studies have reported that the PHS QTLs are located on 2B in white-grain wheat on both short and long arms of chromosome 2B ((Kumar et al. 2009a; Liu et al. 2008; Anderson et al. 1993; Munkvold et al. 2009; Mares et al. 2002). Anderson et al (1993) mapped PHS QTL in two populations, one from a cross between NY18 (moderate PHS resistance) and Clark's Cream (high PHS resistance) and the other between NY18 (moderate PHS resistance) and NY10 (PHS susceptible). PHS QTL from NY18 were located on 2BS ($r^2 = 8.3$) and 2BL (2 QTL - $r^2 = 5.9$-8.8). Collinearity among homoeologous group 2 chromosomes of wheat, rye and barley is conserved except in the distal region of the short arm of wheat chromosome 2B and rye 2R. These two areas have interchromosomal translocations. Chromosomes 2BS, 6AS and 6DS were found to be homologous by using a probe designed from ABA-induced cDNA indicating 2BS/6BS reciprocal translocation occurred (Devos et al. 1993). Interestingly, major QTLs for PHS were detected on 2BS and 6D in winter white-grained wheat (Munkvold et al. 2009). If these QTLs were located at homoeologous regions, it may suggest that the duplicated gene contributes to PHS. However this needs further verification. The PHS QTL 2B was found in both white and red-grained wheat, but in the red-grain wheat population it was from the resistant parent ($r^2 = 6.12$-14.61) (Kumar et al. 2009a; Liu et al. 2008). In a second study, in a population of hard white winter wheat, grown in 3 greenhouses environments and 1 field experiment a major QTL on 3AS explained 41% of phenotypic variation (Liu et al. 2008). They also reported two minor QTLs on 2B but in different linkage groups. The 2B.1 QTL was significant in 2 greenhouse environments and all mean greenhouse environments, explained 5-6.4% of phenotypic variation. The second minor QTL on

2B.2 was significant in only one greenhouse experiment and explained 4.5% of phenotypic variation. These two QTLs on 2B.1 and 2B.2 were located at Marker barc50 and barc334, respectively. Using Wmc 474 as a landmark, they were located outside of the Cayuga QTL on 2B.1, reported by Munkvold et al (2009) and Kumar et al (2009a). In another study in white-grained wheat, the PHS QTL was mapped in a doubled haploid population that was a cross between Cranbrook (PHS extreme susceptible) and Halberd (moderate PHS resistant). The significant PHS QTLs were on chromosomes 2AL, 2DL and 4AL (Mares et al. 2002).

Homologous regions in related species, such as rice, might be useful to identify the regions contributing to PHS as well as additional markers. PHS or SD QTLs on orthologous regions between wheat and rice could come from the same gene(s). Wheat group 2 chromosomes show high similarity to rice chromosomes 4 and 7 (La Rota and Sorrells 2004; Salse et al. 2008). Two linked dormancy QTLs were detected in weedy rice chromosome 7 (Gu et al. 2004). This could suggest that multiple closely linked loci are involved in seed dormancy. Other studies also detected SD QTLs on rice chromosome 7 (Wan et al. 2006; Miura et al. 2002; Lin et al. 1998; Dong et al. 2003; Cai and Morishima 2002; Ishimaru et al. 2001). Those studies suggest that there is more than one gene or region controlling seed dormancy and PHS on wheat chromosome 2B and the orthologous region in rice chromosome 7.

Factors contributing to seed dormancy and preharvest sprouting

There are many characteristics and environmental factors contributing to SD and PHS such as spike length, heading date, spike characteristics, seed shattering, grain color, temperature and rainfall.

Spike length

Spike morphology may affect PHS. One study reported a correlation between

spike length and PHS (Zanetti et al. 2000). Three of six QTLs with major effects (r^2 > 15%) on PHS resistance coincided with QTLs for spike length. The QTL with the largest effect was located on 5AL at the q locus, responsible for the spike morphology of spelt. However, a close linkage between a QTL for spike length and PHS might contribute to the correlation.

Heading date (HD)

Heading date is the date of wheat spike emergence from the leaf sheath. It is controlled by multiple genes and environmental conditions such as temperature and day length. Genes involved include vernalization response (*Vrn*), photoperiod response (*Ppd*) and earliness per se (*Eps*). Zhang et al. (2009) reported HD QTLs were on 1B, 2B, 5D, 6D, 7A and 7D. Heading date QTL was reported to co-localize with SD on rice chromosome 3 (Lin et al. 1998; Takeuchi et al. 2003). The PHS QTL on wheat chromosome 2B was reported to be close to a HD QTL on 2B but was not the same locus in a doubled haploid population (Munkvold et al. 2009) . A correlation between SD and HD has been reported in rice (Jing et al. 2008).

Spike characteristics

The wheat cultivars without awns and with club heads have reduced spike water uptake (King and Richards 1984). Epicuticular waxes were found to reduce sprouting by increasing water repellency of mature ears in wheat and barley. Spikes of the glaucous lines showed a clear reduction of wetting (20–30% less) and, after 72 hours of wetting, their in-ear sprouting was reduced by 50 to 65% (King and von Wettstein-Knowles 2000).

Seed shattering

Seed shattering is the dispersion of seeds immediately after maturity. Non-shattering of the seed and reduced seed dormancy were traits selected during the domestication of cereals. Seed shattering and SD QTLs were found to be linked in rice

because three out of four shattering QTLs were accompanied by seed dormancy QTLs on chromosomes 1, 8 and 11 (Cai and Morishima 2000). Also, they proposed that seed dormancy QTLs on rice chromosome 3, 5, 8 and 11 control coat-imposed dormancy because the QTL loci were significant in intact seeds but not significant in de-hulled seeds. The other loci on chromosomes 9 and 11 might affect embryo-imposed dormancy because both the QTL loci were significant in intact seeds and de-hulled seeds. Later in rice, seed shattering locus *sh-h* was tightly linked with *Rc* locus conferring pericarp color as well seed dormancy QTL on chromosome 7 implying inheritance as a domestication block in the evolution in rice (Ji et al. 2006).

Temperature and rainfall

Many studies have been shown that temperature and rainfall have a major influence on the expression of seed dormancy between anthesis and maturation (Nielsen et al. 1984; Mares 1993; Nyachiro et al. 2002; Wiesner and Grabe 1972). The effects of the environment after maturation and ripening were important as well (Auld and Paulsen 2003). Low temperature during seed development can induce high dormancy but low temperature during germination break dormancy (Nyachiro et al. 2002; Gu et al. 2006). Low temperatures induced greater seed dormancy in long day plants such as wheat and barley during seed development while in short day plants such as rice high temperature induced greater seed dormancy during seed development (Stand 1980). In wheat, lower temperatures from 8.5 °C-26.5 °C during the 20 days before harvest produced slightly greater PHS resistance than did high temperatures from 17.5 °C to 34 °C (Mares 1993). High temperatures two weeks before maturation in the field also reduced PHS tolerance in hard white winter wheat (Nielsen et al. 1984). In grain grown at low temperature (15°C), embryonic ABA steadily increased during development, reaching a maximum before grain desiccation. However, at high temperature (25°C), embryonic ABA levels were very high at early stages and

dropped rapidly during seed maturation (Walker-Simmons and Sesing 1990).

Different varieties of wheat responded to low temperatures differently. In South Africa, the climatic effect was studied using eighteen cultivars of winter wheat over four years (Barnard and Smith 2009). The climatic effect was determined for various stages of grain development, in three distinct groups of cultivars, from PHS susceptible to PHS resistant. They found a strong positive correlation ($r = 0.715$, $p = 0.008$) between PHS rates and minimum temperature during grain filling. Eight of 18 cultivars were more sensitive to temperature effects than others and PHS rates differed between sites and years for individual cultivars. Lower dormancy is generally associated with high temperatures, short days, red light, drought and high nitrogen levels during seed development (Fenner 1991). Other studies, however, reported that high temperatures induce dormancy. High temperatures during grain filling were associated with induced dormancy in certain cultivars of non-dormant genotypes but didn't affect dormancy in dormant genotypes (Biddulph et al. 2007). Also, three wheat cultivars differing in PHS susceptibility were grown in the pacific northwest of the US, where cool temperatures during grain development and high temperatures after ripening promoted rapid loss of seed dormancy and led to severe PHS (Hagemann and Ciha 1987). The effects of environment and the interaction between genotype and environment are significant. However, the interactions did not account for a large proportion (<6%variation) of the variation for sprouting tolerance in the study by Biddulph et al. (2008). The greatest differences in seed germination tended to be between 15 °C and 20 °C. Therefore, the level of seed dormancy depends on the genotype, sites and germination temperature (Nyachiro et al. 2002).

Even though the temperature plays the major role in seed dormancy and preharvest resistance, rain might play a minor role. High relative humidity during grain filling reduces grain dormancy (King 1993). Most of the seasonal variation in

PHS tolerance could be explained in terms of the amount of rain during the 20 day period before grain harvesting. Rainfall over the 20 days accounted for 84% of variation whereas rainfall over a 10 day period before harvest accounted for 57% of variation (Mares 1993).

Dormancy after maturation is a necessary characteristic in wheat varieties, protecting against sprouting in the head during the rainy period of late summer. Temperatures are also a major factor affecting PHS after seed maturation. Low temperature during grain maturation and high temperature after maturation induce seed dormancy. High temperatures (30°C) after maturation induce deep secondary dormancy more in dormant seeds when compared to non-dormant seeds. (George 1967). Storage temperature is also an important consideration for secondary dormancy. Storage of harvested ripe wheat seeds at -15°C delayed the loss of dormancy up to 270 days after the ripening period (Noll and Czarnecki 1980). High temperature storage accelerates loss of dormancy (Mares 1983).

Comparative studies of the relationship of grass species

Grass originated 55-75 million years ago (Kellogg 2001). Three major cereals; rice, maize and wheat diverged from a common ancestor around 40 million years ago (Figure 1.3). Humans and wheat share a remarkably parallel evolutionary history. Three million years ago, humans diverged from apes and diploid A, B and D progenitor species of wheat diverged from a common ancestor (Huang et al. 2002). About 200,000 years ago at nearly the same time that modern humans originated in Africa, two diploid grass species hybridized to form polyploid wheat in the middle east. Later, around 15,000 years ago, wheat was domesticated by humans (Gill et al. 2004). *Triticum urartu* was confirmed as the A-genome donor of tetraploid and

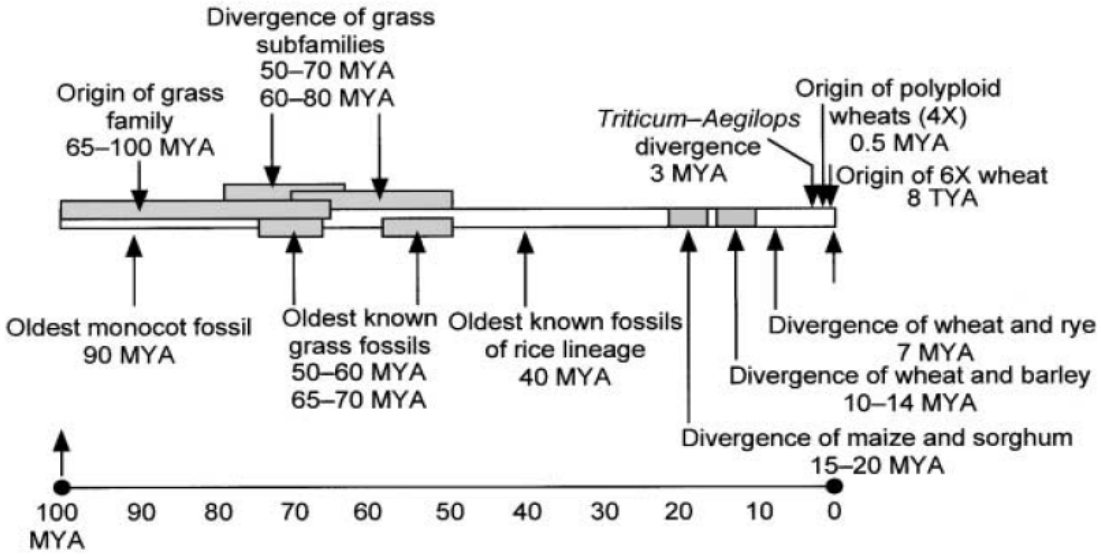


Figure 1.3. Evolutionary timeline of grass species (Gill et al. 2004).

hexaploid wheat, which diverged from *Triticum urartu* half a million ago and the D genome sequences of *T. aestivum* and *Aegilops tauschii* are similar, confirming that *T. aestivum* arose from hybridization of *T. turgidum* and *Ae. tauschii* only 8,000 years ago (Huang et al. 2002). *Aegilops* genome analysis did not conclusively identify the B-genome donor so the diploid progenitor of B genome still remained unknown. Later, analysis of progenitors using AFLP loci and haplotypes of chloroplast and nuclear loci strongly implicated that the origin of the B-genome was *Ae. speltoides* (Kilian et al. 2007). Examples of spikes of ancestors of hexaploid wheat are shown in Figure 1.4.

Genomes of maize, rice, sorghum, barley and wheat are highly variable and evolve quickly (Bennetzen and Ma 2003). Genic regions have undergone many more rearrangements than has been revealed by recombinational mapping studies in rice, barley and foxtail millets (Leister et al. 1998). Comparative DNA sequence analysis of mapped wheat Expressed Sequence Tags (ESTs) revealed the complexity of genome relationships between rice and wheat (La Rota and Sorrells 2004; Salse et al. 2008). Wheat group 1 chromosomes show high similarity with rice chromosomes 5 and 10, wheat group 2 with rice chromosomes 4 and 7, wheat group 3 with rice chromosome 1, wheat group 4 with rice chromosomes 3 and 11, wheat group 5 with rice chromosomes 9 and 12 and wheat group 6 with rice chromosome 2. Rice gene sequences were compared with 6426 mapped wheat ESTs, identifying 29 interchromosomal duplications. The duplicated regions covered 72% of the rice genome and 67.5% of the wheat genome (Salse et al. 2008). Thirteen blocks of orthologous regions between rice and wheat genomes were detected covering 81.3% and 90.4% of rice and wheat genomes, respectively. Seven duplicated regions were shared between rice and wheat. It has also been proposed that grass, sorghum, maize, rice and wheat, have evolved from a common ancestor in Figure 1.5 with a basic

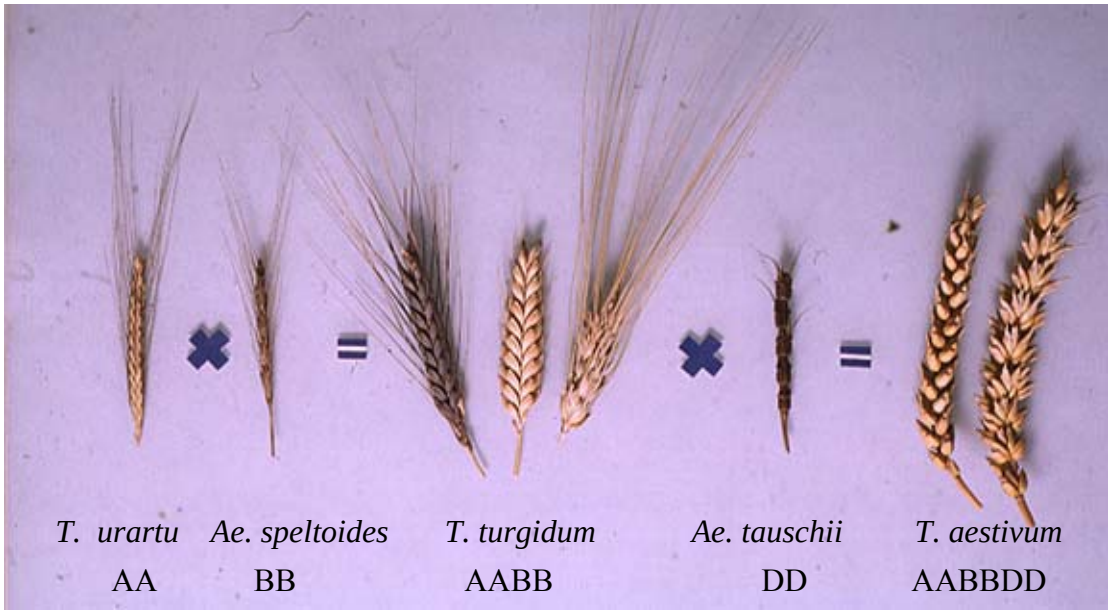


Figure 1.4 Examples of spikes of hexaploid wheat (*Triticum aestivum*) and progenitors

(http://www2.mpiz-koeln.mpg.de/pr/garten/schau/Triticumaestivum/wheat.html).

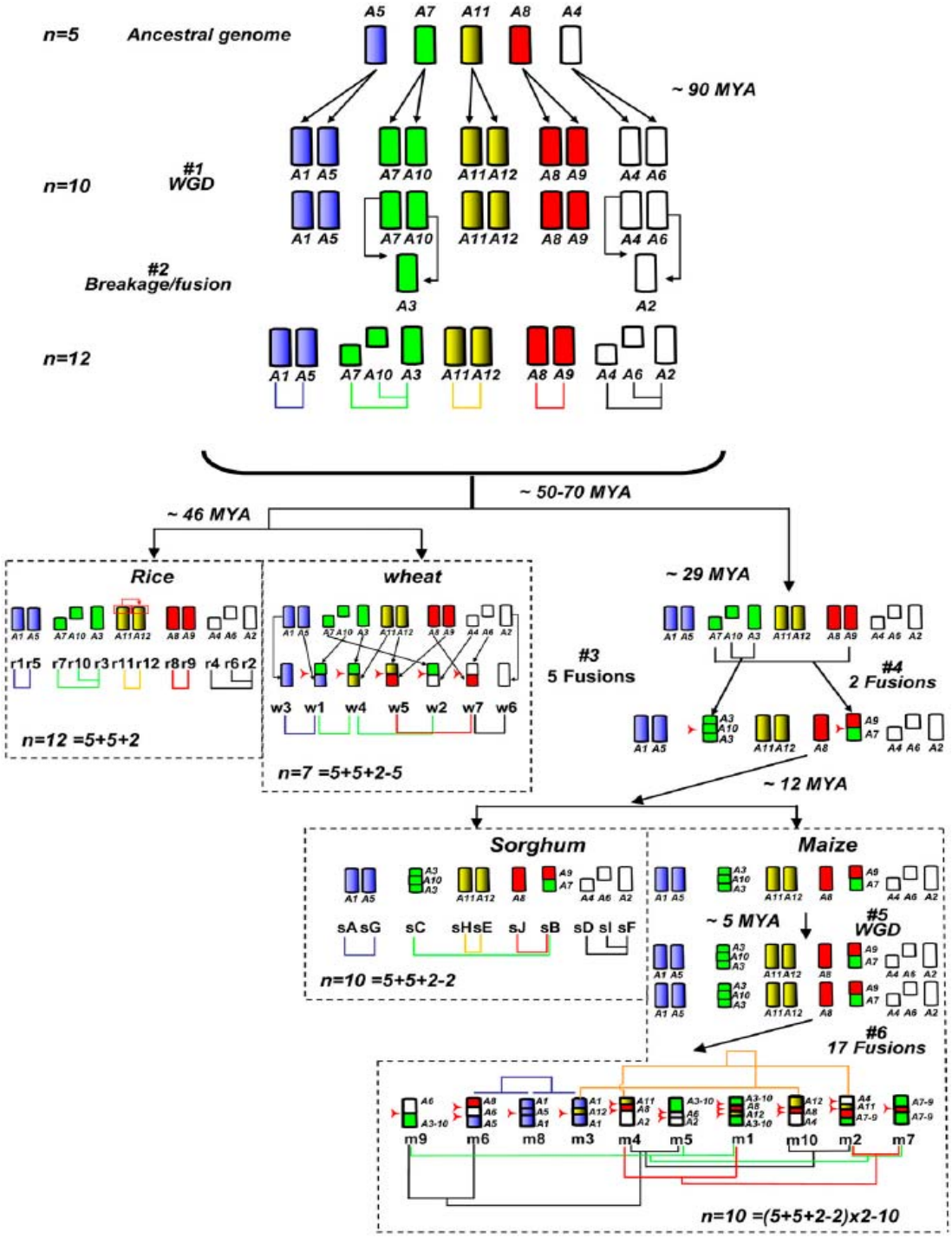


Figure 1.5. Model for structural evolution of rice, wheat, sorghum, and maize genomes from a common ancester with n = 5 chromosomes (Salse et al. 2008).

number of five chromosomes through a series of whole genome and segmental duplications, chromosome fusions and translocations (Salse et al. 2008).

Brachypodium and Triticeae lineage diverged from rice approximately 50 million years ago (Paterson et al. 2004) whereas *Brachypodium* was estimated to have diverged from Triticeae 34-40 million years ago (Bossolini et al. 2007). *Brachypodium* has been reported to be more closely related to barley and wheat than it is to rice, based on combined partial nucleotide sequences of 20 highly expressed genes (Vogel et al. 2006). A 371 kb region sequence of *Brachypodium sylvaticum* was compared with wheat and rice. Macro-collinearity of genetic markers was conserved between *Brachypodium* and wheat whereas rice contained a 220 kb inversion (Bossolini et al. 2007). The sequences of 2185 *Brachypodium* BAC end sequences were compared with the NCBI EST database and there were more hits to wheat sequences than maize supporting the close relation between *Brachypodium* and the Triticeae (Huo et al. 2006). Recently, *Brachypodium* was reported to be a better model for temperate cereals like wheat and barley by comparing *Brachypodium* EST sequences with wheat ESTs and rice genome sequence (Kumar et al. 2009b). However, rice and wheat may be more similar for some genomic regions. For the Q gene region, microcollinearity was more conserved between rice and wheat than between wheat and *Brachypodium* (Faris et al. 2008). In addition, in some regions such as the *Lr34* orthologous region (371 kb), rice contains almost twice the number of genes as *Brachypodium* (66 genes in rice and 43 genes in *Brachypodium*) (Bossolini et al. 2007). Thus the *Brachypodium* genome sequence is largely complementary to rice for genomic studies in a large Triticeae genome in term of gene order and conservation.

REFERENCES

Adkins S, Bellairs S, Loch D (2002) Seed dormancy mechanisms in warm season grass species. Euphytica 126:13-20

Allan RE, Vogel OA (1965) Monosomic Analysis of Red Seed Color in Wheat. Crop Science 5:474-475

Amen R (1968) A model of seed dormancy. Bot Rev 34:1-31

Anderberg RJ, Walker-Simmons MK (1992) Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinases. Proc Natl Acad Sci USA 89:10183-10187

Anderson JA, Sorrells ME, Tanksley SD (1993) RFLP Analysis of Genomic Regions Associated with Resistance to Preharvest Sprouting in Wheat. Crop Sci 33:453-459

Auld AS, Paulsen GM (2003) Effects of Drought and High Temperature During Maturation on Preharvest Sprouting of Hard White Winter Wheat. Cereal Res Comm 31:169-176

Bailey PC, McKibbin RS, Lenton JR, Holdsworth MJ, Flintham JE, Gale MD (1999) Genetic map locations for orthologous Vp1 genes in wheat and rice. Theor Appl Genet 98:281-284

Barnard A, Smith MF (2009) The effect of rainfall and temperature on the preharvest sprouting tolerance of winter wheat in the dryland production areas of the Free State Province. Field Crops Res 112:158-164

Bennetzen JL, Ma J (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 6:128-133

Bentsink L, Koornneef M (2008) Seed Dormancy and Germination. In: The Arabidopsis Book. Rockville, MD, American Society of Plant biologists

Bentsink L, Jowett J, Hanhart CJ, Koornneef M (2006) Cloning of DOG1, a Quantitative Trait Locus Controlling Seed Dormancy in Arabidopsis. Proc Natl Acad Sci USA 103:17042-17047

Bethke PC, Schuurink R, Jones RL (1997) Hormonal signalling in cereal aleurone. J Exp Bot 48:1337-1356

Bewley JD, Black M (1994) Seeds: Physiology of Development and Germination. second edition edn. New York and London, Plenum Press, pp 445

Biddulph TB, Plummer JA, Setter TL, Mares DJ (2008) Seasonal conditions influence dormancy and preharvest sprouting tolerance of wheat (Triticum aestivum L.) in the field. Field Crops Res 107:116

Biddulph TB, Plummer JA, Setter TL, Mares DJ (2007) Influence of high temperature and terminal moisture stress on dormancy in wheat (Triticum aestivum L.). Field Crops Res 103:139-153

Bossolini E, Wicker T, Knobel PA, Keller B (2007) Comparison of orthologous loci from small grass genomes Brachypodium and rice: implications for wheat genomics and grass genome annotation. Plant J 49:704-717

Cai H, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. Theor Appl Genet 104:1217-1228

Cai HW, Morishima H (2000) Genomic regions affecting seed shattering and seed dormancy in rice. Theor Appl Genet 100:840-846

Colucci G, Apone F, Alyeshmerni N, Chalmers D, Chrispeels MJ (2002) GCR1, The Putative Arabidopsis G Protein-Coupled Receptor Gene is Cell Cycle-Regulated, and Its Overexpression Abolishes Seed Dormancy and Shortens Time to Flowering. Proc Natl Acad Sci USA 99:4736-4741

Devos KM, Millan T, Gale MD (1993) Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye and barley. Theor Appl Genet 85:784-792

Dexter JE (1993) End-use quality applications of grading factors in wheat. In: Bass EJ (eds) Grain and oilseeds: handling marketing, processing. Canadian International Grains Institute, Winnipeg, pp 697-722

Dong Y, Tsuzuki E, Kamiunten H, Terao H, Lin D, Matsuo M, Zheng Y (2003) Identification of quantitative trait loci associated with pre-harvest sprouting resistance in rice (Oryza sativa L.). Field Crops Res 81:133-139

Eastmond PJ, Jones RL (2005) Hormonal regulation of gluconeogenesis in cereal aleurone is strongly cultivar-dependent and gibberellin action involves SLENDER1 but not GAMYB. Plant J 44:483-493

Fang J, Chai C, Qian Q, Li C, Tang J, Sun L, Huang Z, Guo X, Sun C, Liu M, Zhang Y, Lu Q, Wang Y, Lu C, Han B, Chen F, Cheng Z, Chu C (2008) Mutations of genes in synthesis of the carotenoid precursors of ABA lead to pre-harvest sprouting and photo-oxidation in rice. Plant J 54:177-189

Fang J, Chu C (2008) Abscisic acid and the pre-harvest sprouting in cereals. Plant Signal Behav 3:1046-1048

Faris JD, Zhang Z, Fellers JP, Gill BS (2008) Micro-colinearity between rice, Brachypodium, and Triticum monococcum at the wheat domestication locus Q. Funct Integr Genomics 8:149-164

Fenner M (1991) The effects of the parent environment on seed germinability. Seed Sci Res 1:75-84

Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. The New phytologist 171:501-23

Finch-Savage WE, Cadman CSC, Toorop PE, Lynn JR, Hilhorst HWM (2007) Seed dormancy release in Arabidopsis Cvi by dry after-ripening, low temperature, nitrate and light shows common quantitative patterns of gene expression directed by environmentally specific sensing. Plant J 51:60-78

Finkelstein R, Reeves W, Ariizumi T, Steber C (2008) Molecular Aspects of Seed Dormancy. Annu Rev Plant Biol 59:387

Flintham JE (2000) Different genetic components control coat-imposed and embryo-imposeddormancy in wheat. Seed Sci Res 10:43-50

Frey A, Audran C, Marin E, Sotta B, Marion-Poll A (1999) Engineering seed dormancy by the modification of zeaxanthin epoxidase gene expression. Plant Mol Biol 39:1267-1274

George DW (1967) High Temperature Seed Dormancy in Wheat (Triticum aestivum L.). Crop Sci 7:249-253

Gill BS, Appels R, Botha-Oberholster A, Buell CR (2004) A Workshop Report on Wheat Genome Sequencing: International Genome Research on Wheat Consortium. Genetics 168:1087-1096

Gomez-Cadenas A, Zentella R, Walker-Simmons MK, Ho TH (2001) Gibberellin/abscisic acid antagonism in barley aleurone cells: site of action of the protein kinase PKABA1 in relation to gibberellin signaling molecules. Plant Cell 13:667-679

Gomez-Cadenas A, Verhey SD, Holappa LD, Shen Q, Ho TH, Walker-Simmons MK (1999) An abscisic acid-induced protein kinase, PKABA1, mediates abscisic acid-suppressed gene expression in barley aleurone layers. Proc Natl Acad Sci USA 96:1767-1772

Gonzalez-Garcia MP, Rodriguez D, Nicolas C, Rodriguez PL, Nicolas G, Lorenzo O (2003) Negative Regulation of Abscisic Acid Signaling by the Fagus sylvatica FsPP2C1 Plays A Role in Seed Dormancy Regulation and Promotion of Seed Germination. Plant Physiol 133:135-144

Gu XY, Kianian SF, Foley ME (2006) Dormancy genes from weedy rice respond divergently to seed development environments. Genetics 172:1199-211

Gu XY, Liu T, Feng J, Suttle JC, Gibbons J (2009) The qSD12 underlying gene promotes abscisic acid accumulation in early developing seeds to induce primary dormancy in rice. Plant Mol Biol doi:10.1007/s11103-009-9555-1

Gu XY, Kianian SF, Foley ME (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (Oryza sativa). Genetics 166:1503-1516

Gualberti G, Papi M, Bellucci L, Ricci I, Bouchez D, Camilleri C, Costantino P, Camilleri C (2002) Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of Arabidopsis seeds. Plant Cell 14:1253-63

Guiltinan MJ, Marcotte WR, Jr, Quatrano RS (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. Science 250:267-271

Hagemann MG, Ciha AJ (1987) Environmental x Genotype Effects on Seed Dormancy and After-Ripening in Wheat. Agron J 79:192-196

Himi E, Mares DJ, 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, Noda K (2004) Isolation and location of three homoeologous dihydroflavonol-4-reductase (DFR) genes of wheat and their tissue-dependent expression. J Exp Bot 55:365-375

Hobo T, Kowyama Y, Hattori T (1999) A bZIP factor, TRAB1, interacts with VP1 and mediates abscisic acid-induced transcription. Proc Natl Acad Sci USA 96:15348-15353

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes Encoding Plastid Acetyl-CoA Carboxylase and 3-Phosphoglycerate Kinase of the Triticum/Aegilops Complex and the Evolutionary History of Polyploid Wheat. Proc Natl Acad Sci USA 99:8133-8138

Huo N, Gu YQ, Lazo GR, Vogel JP, Coleman-Derr D, Luo M-, Thilmony R, Garvin DF, Anderson OD (2006) Construction and characterization of two BAC libraries from Brachypodium distachyon, a new model for grass genomics. Genome 49:1099-1108

Ishimaru K, Yano M, Aoki N, Ono K, Hirose T, Lin SY, Monna L, Sasaki T, Ohsugi R (2001) Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. Theor Appl Genet 102:793-800

Ji HS, Chu SH, Jiang W, Cho YI, Hahn JH, Eun MY, McCouch SR, Koh HJ (2006) Characterization and mapping of a shattering mutant in rice that corresponds to a block of domestication genes. Genetics 173:995-1005

Jing W, Jiang L, Zhang W, Zhai H, Wan J (2008) Mapping QTL for Seed Dormancy in Weedy Rice. Acta Agronomica Sinica 34:737- 742

Johnson RR, Wagner RL, Verhey SD, Walker-Simmons MK (2002) The abscisic acid-responsive kinase PKABA1 interacts with a seed-specific abscisic acid response element-binding factor, TaABF, and phosphorylates TaABF peptide sequences. Plant Physiol 130:837-846

Kawakami N, Miyake Y, Noda K (1997) ABA insensitivity and low ABA levels during seed development of non-dormant wheat mutants. J Exp Bot 48:1415-1421

Kellogg EA (2001) Evolutionary History of the Grasses. Plant Physiol 125:1198-1205

Kermode AR (2005) Role of Abscisic Acid in Seed Dormancy. J Plant Growth Regul 24:319-344

Kilian B, Özkan H, Deusch O, Effgen S, Brandolini A, Kohl J, Martin W, Salamini F (2007) Independent Wheat B and G Genome Origins in Outcrossing Aegilops Progenitor Haplotypes. Mol Biol Evol 24:217-227

King RW, von Wettstein-Knowles P (2000) Epicuticular waxes and regulation of ear wetting and pre-harvest sprouting in barley and wheat. Euphytica 112:157-166

King RW (1993) Manipulation of Grain Dormancy in Wheat. J Exp Bot 44:1059-1066

King RW, Richards RA (1984a) Water uptake in relation to pre-harvest sprouting damage in wheat: ear characteristics. Aust J Agric Res 35:327-336

Kucera B, Cohn MA, Leubner-Metzger G (2005) Plant hormone interactions during seed dormancy release and germination. Seed Sci Res 15:281-308

Kumar A, Kumar J, Singh R, Garg T, Chhuneja P, Balyan HS, Gupta PK (2009a) QTL analysis for grain colour and pre-harvest sprouting in bread wheat. Plant Sci 177:114-122

Kumar S, Mohan A, Balyan HS, Gupta PK (2009b) Orthology between genomes of Brachypodium, wheat and rice. BMC research notes 2:93

La Rota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. Funct Integr Genomics 4:34-46

Lefebvre V, North H, Frey A, Sotta B, Seo M, Okamoto M, Nambara E, Marion-Poll A (2006) Functional analysis of Arabidopsis NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. Plant J 45:309-319

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Lefert PS (1998) Rapid Reorganization of Resistance Gene Homologues in Cereal Genomes. Proc Natl Acad Sci USA 95:370-375

Leubner-Metzger G (2001) Brassinosteroids and gibberellins promote tobacco seed germination by distinct pathways. Planta 213:758

Li AL, Zhu YF, Tan XM, Wang X, Wei B, Guo HZ, Zhang ZL, Chen XB, Zhao GY, Kong XY, Jia JZ, Mao L (2008) Evolutionary and functional study of the CDPK gene family in wheat (Triticum aestivum L.). Plant Mol Biol 66:429-443

Li C, Ni P, Francki M, Hunter A, Zhang Y, Schibeci D, Li H, Tarr A, Wang J, Cakir M, Yu J, Bellgard M, Lance R, Appels R (2004) Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. Funct Integr Genomics 4:84-93

Lin SY, Sasaki T, Yano M (1998) Mapping quantitative trait loci controlling seed dormancy and heading date in rice, Oryza sativa L., using backcross inbred lines. Theor Appl Genet 96:997-1003

Liu S, Cai S, Graybosch R, Chen C, Bai G (2008) Quantitative trait loci for resistance to pre-harvest sprouting in US hard white winter wheat Rio Blanco. Theor Appl Genet 117:691-699

Liu X, Yue Y, Li B, Nie Y, Li W, Wu WH, Ma L (2007) A G Protein-Coupled Receptor Is a Plasma Membrane Receptor for the Plant Hormone Abscisic Acid. Science 315:1712-1716

Mares DJ (1983) Preservation of Dormancy in Freshly Harvested Wheat Grain. Aust J Agric Res 34:33-38

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. Theor Appl Genet 111:1357-1364

Mares D, Mrva K, Tan MK, Sharp P (2002) Dormancy in white-grained wheat: Progress towards identification of genes and molecular markers. Euphytica 126:47-53

Mares DJ (1993) Pre-harvest sprouting in wheat. I. Influence of cultivar, rainfall and temperature during grain ripening. Aust J Agric Res 44:1259-1272

McCrate AJ, Nielsen MT, Paulsen GM, Heyne EG (1982) Relationship between sprouting in wheat and embryo response to endogenous inhibition. Euphytica 31:193-200

McKibbin RS, Wilkinson MD, Bailey PC, Flintham JE, Andrew LM, Lazzeri PA, Gale MD, Lenton JR, Holdsworth MJ (2002) Transcripts of Vp-1 Homeologues are Misspliced in Modern Wheat and Ancestral Species. Proc Natl Acad Sci USA 99:10203-10208

Metzger RJ, Silbaugh BA (1970) Location of Genes for Seed Coat Color in Hexaploid Wheat, Triticum aestivum L. Crop Sci 10:495-496

Miura K, Lin Y, Yano M, Nagamine T (2002) Mapping quantitative trait loci controlling seed longevity in rice (Oryza sativa L.). Theor Appl Genet 104:981-986

Miyamoto T, Everson EH (1958) Biochemical and Physiological Studies of Wheat Seed Pigmentation. Agron J 50:733-734

Munkvold J, Tanaka J, Benscher D, Sorrells M (2009) Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. Theor Appl Genet 119:1223-1235

Nakagawa H, Ohmiya K, Hattori T (1996) A rice bZIP protein, designated OSBZ8, is rapidly induced by abscisic acid. Plant J 9:217-227

Nantel A, Quatrano RS (1996) Characterization of Three Rice Basic/Leucine Zipper Factors, Including Two Inhibitors of EmBP-1 DNA Binding Activity. J Biol Chem 271:31296-31305

Nielsen MT, McCrate AJ, Heyne EG, Paulsen GM (1984) Effect of Weather Variables During Maturation on Preharvest Sprouting of Hard White Wheat. Crop Sci 24:779-782

Noll JS, Czarnecki E (1980) Methods for extending the testing period for harvest-time dormancy in wheat. Cereal Res Commun 8:233-238

Nyachiro JM, Clarke FR, DePauw RM, Knox RE, Armstrong KC (2002) Temperature effects on seed germination and expression of seed dormancy in wheat. Euphytica 126:123-127

Oeda K, Salinas J, Chua NH (1991) A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J 10:1793-1802

Ogbonnaya F, Imtiaz M, Ye G, Hearnden P, Hernandez E, Eastwood R, van Ginkel M, Shorter S, Winchester J (2008) Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. Theor Appl Genet 116:891-902

Ogbonnaya FC, Imtiaz M, DePauw RM (2007) Haplotype diversity of preharvest sprouting QTLs in wheat. Genome 50:107-118

Oh E, Yamaguchi S, Kamiya Y, Bae G, Chung W, Choi G (2006) Light activates the degradation of PIL5 protein to promote seed germination through gibberellin in Arabidopsis. Plant J 47:124-139

Oh E, Kim J, Park E, Kim J, Kang C, Choi G (2004) PIL5, a Phytochrome-Interacting Basic Helix-Loop-Helix Protein, Is a Key Negative Regulator of Seed Germination in Arabidopsis thaliana. Plant Cell 16:3045-3058

Okamoto M, Kuwahara A, Seo M, Kushiro T, Asami T, Hirai N, Kamiya Y, Koshiba T, Nambara E (2006) CYP707A1 and CYP707A2, Which Encode Abscisic Acid 8'-Hydroxylases, Are Indispensable for Proper Control of Seed Dormancy and Germination in Arabidopsis. Plant Physiol 141:97-107

Papi M, Sabatini S, Altamura MM, Hennig L, Schäfer E, Costantino P, Vittorioso P (2002) Inactivation of the Phloem-Specific Dof Zinc Finger Gene DAG1 Affects Response to Light and Integrity of the Testa of Arabidopsis Seeds. Plant Physiol 128:411-417

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903-9908

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20:11-24

Sears ER (1944) Cytogenetic Studies with Polyploid Species of Wheat. II. Additional Chromosomal Aberrations in Triticum Vulgare. Genetics 29:232-246

Simpson GM (1990) Seed Dormancy in Grasses. Cambridge, Cambridge university press

Stand E (1980) A seed dormancy index for selection cereal cultivars resistant to pre-harvest sprouting. Cereal Res Commun 8:233-238

Stoy V, Sundin K (1976) Effects of growth regulation substances on cereal seed germination. Cereal Res Comm 4:157-163

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. Proc Natl Acad Sci USA doi:10.1073/pnas.0911965107

Takeuchi Y, Lin SY, Sasaki T, Yano M (2003) Fine linkage mapping enables dissection of closely linked quantitative trait loci for seed dormancy and heading in rice. Theor Appl Genet 107:1174-1180

Tan MK, Sharp PJ, Lu MQ, Howes N (2006) Genetics of grain dormancy in a white wheat. Aust J Agric Res 57:1157-1166

Tanaka M, Kikuchi A, Kamada H (2008) The Arabidopsis Histone Deacetylases HDA6 and HDA19 Contribute to the Repression of Embryonic Properties after Germination. Plant Physiol 146:149

To A, Valon C, Savino G, Guilleminot J, Devic M, Giraudat J, Parcy F (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. Plant Cell 18:1642-1651

Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Chow TY, Hsing YI, Kitano H, Yamaguchi I, Matsuoka M (2005) GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature 437:693-8

Upadhyay MP, Morris CF, Paulsen GM (1988) Characterization of preharvest sprouting resistance in Clark's Cream white winter wheat. Euphytica 38:85-92

Utsugi S, Nakamura S, Noda K, Maekawa M (2008) Structural and functional properties of Viviparous1 genes in dormant wheat. Genes Genet Syst 83:153-166

Vogel J, Gu Y, Twigg P, Lazo G, Laudencia-Chingcuanco D, Hayden D, Donze T, Vivian L, Stamova B, Coleman-Derr D (2006) EST sequencing and phylogenetic analysis of the model grass Brachypodium distachyon. Theor Appl Genet 113:186-195

Walker-Simmons M, Sesing J (1990) Temperature effects on embryonic abscisic acid levels during development of wheat grain dormancy. J Plant Growth Regul 9:51-56

Walker-Simmons M (1987) ABA Levels and Sensitivity in Developing Wheat Embryos of Sprouting Resistant and Susceptible Cultivars. Plant Physiol 84:61-66

Wan JM, Jiang L, Tang JY, Wang CM, Hou MY, Jing W, Zhang LX (2006) Genetic dissection of the seed dormancy trait in cultivated rice (Oryza sativa L.). Plant Sci 170:786-792

Warner RL, Kudrna DA, Spaeth SC, Jones SS (2000) Dormancy in white-grain mutants of Chinese Spring wheat (Triticum aestivum L.). Seed Sci Res 10:51-60

Wiesner LE, Grabe DF (1972) Effect of Temparature Preconditioning and Cultivar on Ryegrass (Lolium sp.) Seed Dormancy. Crop Sci 12:760-764

Willige BC, Ghosh S, Nill C, Zourelidou M, Dohmann EMN, Maier A, Schwechheimer C (2007) The DELLA Domain of GA INSENSITIVE Mediates the Interaction with the GA INSENSITIVE DWARF1A Gibberellin Receptor of Arabidopsis. Plant Cell doi: 10.1105/tpc.107.051441

Yamauchi Y, Ogawa M, Kuwahara A, Hanada A, Kamiya Y, Yamaguchi S (2004) Activation of Gibberellin Biosynthesis and Response Pathways by Low Temperature during Imbibition of Arabidopsis thaliana Seeds. Plant Cell 16:367-378

Zanetti S, Winzeler M, Keller M, Keller B, Messmer M (2000) Genetic Analysis of Pre-Harvest Sprouting Resistance in a Wheat Spelt Cross. Crop Sci 40:1406-1417

Zhang K, Tian J, Zhao L, Liu B, Chen G (2009) Detection of quantitative trait loci for heading date based on the doubled haploid progeny of two elite Chinese wheat cultivars. Genetica 135:257-265

CHAPTER 2

COMPARATIVE GENETIC ANALYSIS OF A PREHARVEST SPROUTING QTL INTERVAL IN WHITE WINTER WHEAT WITH RICE, *BRACHYPODIUM* AND *AEGILOPS TAUSCHII*

INTRODUCTION

Comparative mapping has become a tool for comparison of gene order across related grass species such as rice, maize, wheat, barley, sorghum and millets. Cross mapping of gene sequences by using molecular markers is the first step to study genome relationships and is useful for developing markers for fine mapping or identifying candidate genes in orthologous regions. Even though the level of collinearity is generally highly conserved among related species, the degree of gene order conservation may vary depending on the region (Devos and Gale 2000). The comparative sequence analysis of a homologous region between rice and wheat revealed numerous chromosomal rearrangements, especially at nonconserved regions (Sorrells et al. 2003). There have been several successful studies that used orthologous regions from divergent grass species for fine mapping and identification of candidate genes. This approach was used to identify a gene controlling vernalization requirement in winter wheat. The vernalization gene, *VRN1*, on wheat (*Triticum monococcum*) chromosome 5A is collinear with rice chromosome 3 and sorghum BACs (Yan et al. 2003). Perfect microcollinearity was reported for a distance of 2.6 cM on wheat chromosome 6B and rice chromosome 2, at the grain protein locus *Gpc-6B1*. New markers based on this collinear region were used to narrow it down to 0.3 cM and identified five candidate genes within 64 kilobasepairs (kb), based on the rice sequence (Distelfeld et al. 2004). In another example, *Rht1*, is a Gibberellin-insensitive gene that controls plant height. It is located on wheat chromosome 4B and

4D, was found to be orthologous with maize *d8* on chromosome 1 (Peng et al. 2004). Collinearity was also reported for wheat chromosome 3BS, rice chromosome 1S and barley chromosome 3HS, at the Fusarium head blight locus (*Fhb1*) (Liu et al. 2006). Some studies have shown that certain wheat regions were not collinear with rice, sorghum and maize, but were collinear with barley and rye. The region at shrunken 2 (*Sh2*) and the anthocyaninless1 (*A1*) regions were collinear among rice, sorghum and maize but not between wheat and barley. Twenty kb separates *Sh2* and *A1* in rice and sorghum but 140 kb separates them in maize (Li and Gill 2002). Collinearity between barley and wheat was more conserved near the wheat leaf rust disease locus *Lr1* but was organized differently in rice (Gallego et al. 1998). A translocation in rye, in the same region on 2R, is comparative to that of wheat on 2B (Devos et al. 1993). Even though the collinearity within grass species is generally conserved, Liu et al. (2006) reported complexity in microcollinearity due to inversion and insertions/deletions. In wheat, Devos et al (1999) reported that the polymorphisms, deletions and translocations have been found in Chinese Spring nullisomic-tetrasomic and ditelosomic lines. Disease resistance NBS-LRR (nucleotide binding site-leucine rich repeat)-containing genes may evolve rapidly, thus complicating comparative analysis (Leister et al. 1998).

Wheat preharvest sprouting (PHS) is the germination of grain on the mother spike before harvesting, resulting in reduced grain quality and harvest yield. PHS and grain dormancy in wheat are expressed as a quantitatively inherited trait that is strongly influenced by environment. Quantitative Trait Loci (QTL) studies are useful to identify the genome regions that control PHS and their relative importance for further studies such as fine mapping. Fine mapping can narrow down the specific QTL regions, perhaps to the gene level. In previous work done by Munkvold (2009), PHS QTLs were mapped in a doubled haploid population of soft white winter wheat from a

cross between Cayuga (PHS resistant wheat) and Caledonia (PHS susceptible) lines. The major PHS QTL located on chromosome 2B was significant in all 16 environments and explained from 5-31% of phenotypic variation. However the location of the QTL mapping interval varied somewhat in different environments. In all the environments, those intervals were overlapping. Based on PHS mean overall environments, the QTL was close to Barc55 and flanked by markers Gwm 429 and Wmc474. Major QTLs were also detected on chromosomes 2D, 3D and 6D. They were significant in 7, 4 and 10 out of 16 environments, respectively. The PHS QTL on 2B and 2D may be homoeologous but more mapping is required. In addition, seed dormancy (SD) QTL was located in the same interval as the 2B PHS QTL (Munkvold et al. 2009). Fine mapping would help narrow down the region on this QTL and define the gene(s) contributing to both PHS and SD.

For developing markers and defining the orthologous regions for wheat, rice and *Brachypodium*, knowledge about the location of flanking markers on wheat chromosomes was important. Barc55 is a flanking marker, located at the wheat deletion interval 2BS1-0.53-0.75 (Somers et al. 2004). The comparative map between rice and *Brachypodium*, and wheat Expressed Sequence Tags (ESTs) located in this bin was used for developing new markers for fine mapping and identifying the candidate genes that might influence the PHS and SD phenotypes. This information will help to identify gene order on the microcollinear level. The objectives of this study were to

1) develop new markers designed from wheat ESTs, EST contigs and transcripts in a specific major PHS QTL region on 2B by using available bioinformatics .

2) use the comparative map as a tool to identify the useful ESTs, transcripts and candidate known-function genes located in the orthologous region that would be used for marker development and candidate gene identification.

MATERIAL AND METHODS

Bioinformatics

The location and primer information for marker Barc55 was obtained from GrainGenes, a database for Triticeae and *Avena*, http://wheat.pw.usda.gov/GG2/index.shtml. The ESTs (Expressed Sequence Tags) were located in the deletion interval 2BS1-0.53-0.75. The details of these ESTs were obtained from GrainGenes–wEST-SQL resources http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi). The comparisons of the homologous rice genes and wheat ESTs were obtained from Rice genome browser-TIGR Release 5.0 http://www.modelcrop.org/cgi-bin/gbrowse/rice/. The comparisons between wheat and *Brachypodium* were obtained from the Brachy 4x preliminary assembly http://www.modelcrop.org/cgi-bin/gbrowse/brachy4x/. The rice QTLs that were homologous with the region of deletion interval 2BS1-0.53-0.75 can be found in Gramene; a resource for Comparative Grass Genomics http://www.gramene.org/. The *Aegilops tauschii* contigs containing BACs were found using the wheat genome D-FPC map: http://phymapgb.bioinformatics.ucdavis.edu/cgi-bin/gbrowse/wheatfpc11/#search

Marker development

1) *Designing polymorphic markers based on intron variation.*

Currently, there is no complete genome sequence for wheat, but there is information available about ESTs, EST contigs and wheat transcripts. Wheat ESTs and EST contigs have both been used for marker design. EST contigs were provided by the gene index project, computational biology and functional genomics laboratories http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/Blast/index.cgi. Exon sequences were used to designed primers for amplifying PCR products across introns. So each EST

and EST contig was compared with rice homologous genes whose sequences were provided by the TIGR rice genomic annotation project http://blast.jcvi.org/euk-blast/index.cgi?project=osa1 and the Gramene database. The alignment between sequences was compared by NCBI blast http://blast.ncbi.nlm.nih.gov/Blast.cgi. Sequences were also compared directly by using a Vector NTI (Lu and Moriyama 2004) program, to avoid the splicing junction. The primers were designed by Primer3 (http://frodo.wi.mit.edu/primer3/) so that targets and excluded regions could be specified.

2) *Designing primers based on variation in coding sequences.*

The variation between the two parents, Cayuga and Caledonia, might be located not only in introns but also in coding regions. The type of variation might include indels (insertion and deletions), and/or SNPs (single nucleotide polymorphisms). The long nucleotide sequences were useful for designing new polymorphic markers, such as CAPS (cleaved amplified polymorphic sequence), EST-SSR (simple sequence repeat), SNP and indel markers. CAPS were designed using the Solanaceae Genome network tool http://sgn.cornell.edu/tools/caps_designer/caps_input.pl and SSRs were identified using the Gramene Simple Sequence Repeat Identification Tool http://www.gramene.org/db/markers/ssrtool.

The polymorphic primers were first identified in winter white hexaploid wheat (*Triticum aestivum*) cultivars, Cayuga (PHS resistant) and Caledonia (PHS susceptible). This bi-parental cross population was used for coarse mapping. PCR products from both parents were cloned into a pGEM-T Easy vector for sequencing. Because multiple copies occur in wheat, several copies were sequenced and compared by using a Vector NTI program. The 500-1000 bp sequences were also used to design primers for amplifying short 150-300 bp sequences. The shorter sequences could be

compared by polyacrylamide gel electrophoresis (PAGE), based on size differences. If PCR products were still monomorphic on PAGE gels, then single-strand conformation polymorphism (SSCP) gels were used to compare products, based on differences in mobility resulting from differences in sequences (often a single base pair change) that alter secondary structure (Sunnucks et al. 2000; Liu et al. 1999) or tertiary structure of the DNA molecules (Liu et al. 1999).

3) *Designing primers based on Simple sequence repeats (SSRs) in the genomic sequence.*

Some genes, like *Myb*-related genes, contain SSR sequences. The primer sequences for the *Myb*-related genes were designed from sequences in the TIGR plant transcript assemblies http://plantta.jcvi.org/index.shtml. For some genes, such as the Calmodulin/calcium dependent protein kinase, we used the SSR identification tool (http://www.gramene.org/db/markers/ssrtool) and primer3 http://frodo.wi.mit.edu/primer3/ to design primers.

Testing for polymorphic markers

Parental DNA was used for testing polymorphism. Subsequently, the location of polymorphic markers was determined in a doubled haploid (DH) population of 149 individuals from a cross between Cayuga and Caledonia. The DH population was used in previous studies on PHS QTL mapping (Munkvold 2007). The PCR annealing temperature was different for each primer. The PCR products were visualized on a 4% PAGE, based on size difference and a 10% SSCP gel, based on differences in mobility. The locations of new markers were determined by using the previous map (Munkvold et al. 2009) and MapManager QTXb20 software (Manly et al. 2001).

RESULTS

Comparative analysis between wheat bin 2BS1-0.53-0.75 (2BS1.1 bin) and rice

From the GrainGenes database, Barc55, a marker centered in the PHS 2B.1 QTL interval (Munkvold et al. 2009), was mapped in the deletion interval 2BS1-0.53-0.75 (2BS1.1). In order to saturate the PHS QTL interval in chromosome 2B, new markers were designed based on ESTs, EST contigs and the known functional genes that might be involved in seed dormancy. GrainGenes-wEST-SQL-resources reported 104 ESTs in this bin. There were 93 (89%) ESTs that showed high homology with rice genes located on almost all rice chromosomes except chromosome 10. Some ESTs were homologous with genes that were located on multiple rice chromosomes, but only the ESTs that gave the higher E-value were included in Figure 2.1. Eleven ESTs had no match with any rice gene. Out of 93 ESTs in the 2BS1.1 bin, 78 (84%) were homologous with rice genes whose functions were known. Rice chromosome 7 had the most genes (n = 49) homologous to the ESTs (Figure 2.1 & Table 2.1). Chromosome 3 was the second, with 14 genes homologous to ESTs (Figure 2.1 & Table 2.2). Some ESTs were homologous with the same rice gene. For example, there were 4 pairs of ESTs showing homology with the same genes. BE 446480 and BG314234 were homologous with histone acetyltransferese (LOC_Os07g43360) whereas the other 3 pairs were homologous with UBP24 (LOC_Os07g46660), ATP binding protein (LOC_Os07g47530) and oxidoreductase (LOC_Os07g48640), respectively, on rice chromosome 7. Comparative mapping of these ESTs was used to delimit the orthologous region to the major PHS QTL interval on 2B.1 and rice chromosome 7 and 3.

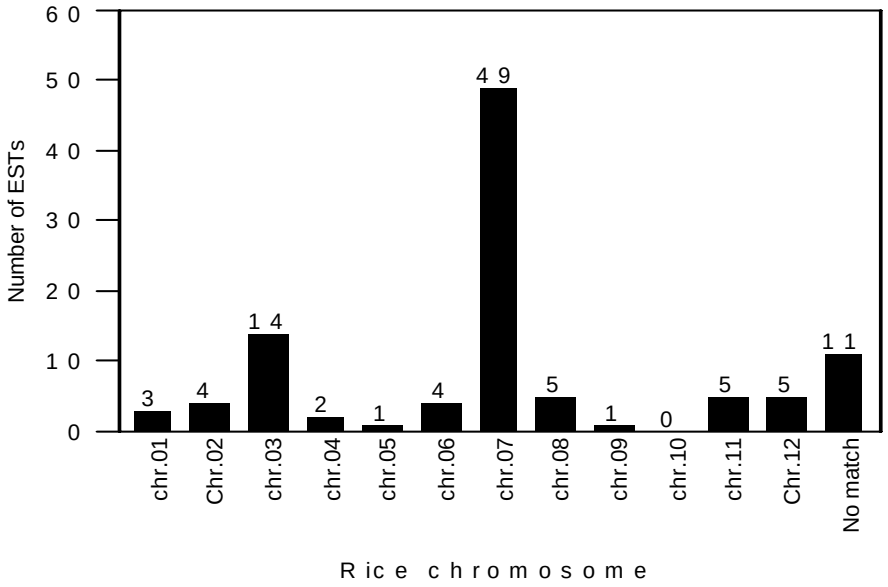


Figure 2.1. The number of wheat ESTs (wESTs) in deletion line interval 2BS1-0.53-0.75 homologous with rice genes in each chromosome.

Rice chromosome 7 has the most genes homologous with wESTs. Rice chromosome 3 has the second most homologous genes. Some ESTs were homologous with genes that were located on multiple rice chromosomes. But only the EST that gave the highest E-value was accounted for here.

Table 2.1. The homology between wheat ESTs on deletion line interval 2BS1-0.53-0.75 and rice chromosome 7, with the approximate location of the rice seed dormancy QTLs.

Dark black letters represent the comparative map region on the PHS QTL of wheat chromosome 2B.1. Arrows represent some rice QTL position on chromosome 7.

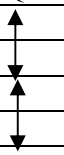
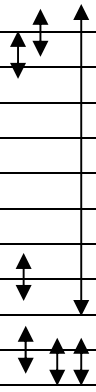
wheat 5'EST	TIGR rice loci	Annotation gene	E-valve	The location of rice QTL
BF478850	LOC_Os07g05010	PIL6	9.70E-02	
BE426431	LOC_Os07g23200	F-actin capping protein alpha subunit	2.60E-23	
BF474758	LOC_Os07g26440	CIP7	8.50E-09	
BF293744	LOC_Os07g34190	chalcone synthase DII	2.30E-49	
BF293463	LOC_Os07g36180	EMB1417	2.40E-27	
BE443822	LOC_Os07g36880	expressed protein	8.90E-31	
BM138155	LOC_Os07g38240	AN1-type zince finger protein 2B	4.40E-43	
BE500443	LOC_Os07g38580	zinc ion binding protein	1.60E-49	
BE446480	LOC_Os07g43360	histone acetyltransferase MYST1	7.10E-65	
BG314234	LOC_Os07g43360	histone acetyltransferase MYST1	2.70E-91	
BE442746	LOC_Os07g43430	expressed protein	2.40E-22	
BE494262	**LOC_Os07g43470**	**developmentally-regulated GTP-binding protein1**	**1.50E-44**	
BE399688	**LOC_Os07g43540**	**Origin recognition complex**	**3.00E-22**	
BE517987	**LOC_Os07g43670**	**ribonuclease 1 precusor**	**4.40E-46**	
BQ294702	**LOC_Os07g43970**	**calmodulin binding protein**	**5.60E-55**	
BF201533	**LOC_Os07g44060**	**catalytic/hydrolase**	**1.70E-54**	
BE498254	**LOC_Os07g44070**	**carboxylic ester hydrolase**	**4.00E-39**	
BE424118	**LOC_Os07g44190**	**h/ACA ribonucleaoprotein complex**	**1.10E-77**	
BG314582	**LOC_Os07g44400**	**kinesin POK1**	**3.20E-11**	
BF202681	**LOC_Os07g44660**	**xylulose kinase**	**3.60E-51**	
BE442903	**LOC_Os07g44744**	**expressed protein**	**2.90E-32**	
BE444297	**LOC_Os07g44950**	**DNA binding protein**	**3.30E-25**	
BE406677	**LOC_Os07g44960**	**expressed protein**	**6.10E-15**	
BE499478	**LOC_Os07g45064**	**ATM-like protein**	**2.70E-90**	
BE499671	**LOC_Os07g45160**	**SAC3/GANP family protein**	**1.20E-63**	
BE422835	**LOC_Os07g46190**	**pectinasterase VGDH2 precursor**	**6.90E-21**	
BG607608	**LOC_Os07g46300**	**expressed protein**	**4.10E-75**	
BE443026	**LOC_Os07g46340**	**metylase**	**2.00E-59**	
BE445628	**LOC_Os07g46450**	**rho GTPase activator**	**2.30E-47**	
BE497494	**LOC_Os07g46460**	**ferredoxin-dependent glutamate synthase**	**2.50E-53**	
BG605089	**LOC_Os07g46540**	**condensin complex subunit 1**	**2.50E-92**	
BG314532	**LOC_Os07g46660**	**UBP24**	**2.60E-66**	
BM140364	**LOC_Os07g46660**	**UBP24**	**9.60E-59**	

Table 2.1. (Continued)

BE438266	**LOC_Os07g46750**	**elongation factor 1 beta**	**1.10E-23**	
BE405597	**LOC_Os07g47530**	**ATP-binding protein**	**1.30E-51**	
BF478837	**LOC_Os07g47530**	**ATP-binding protein**	**2.10E-42**	
BE488865	**LOC_Os07g47670**	**hypoxia induced protein**	**1.20E-15**	
BE636824	**LOC_Os07g48090**	**CBL-interacting serine/threonine-protein kinase 11**	**3.20E-68**	
BE606438	**LOC_Os07g48100**	**CBL-interacting serine/threonine-protein kinase 15**	**2.00E-63**	
BG262864	**LOC_Os07g48310**	**ATP binding protein**	**7.00E-47**	
BM134309	**LOC_Os07g48350**	**expressed protein**	**1.10E-33**	
BF428792	**LOC_Os07g48430**	**nucleoside-triphosphatase**	**2.40E-58**	
BE490666	**LOC_Os07g48640**	**oxidoreductase**	**3.80E-41**	
BE425990	**LOC_Os07g48640**	**oxidoreductase**	**2.50E-28**	
BE500206	**LOC_Os07g48770**	**expressed protein**	**1.70E-46**	
BM138067	LOC_Os07g48920	betaine-aldehyde dehydrogenase	1.40E-37	
BF484399	LOC_Os07g48920	pre-mRNA processing factor	7.30E-75	
BE518306	LOC_Os07g48980	nicotianamine syntase 3	4.40E-21	
BF485144	LOC_Os07g49280	expressed protein	4.10E-79	

Table 2.2. The homology between wheat ESTs on deletion line interval 2BS1-0.53-0.75 and rice chromosome 3, with the approximate location of the rice seed dormancy QTLs.

Dark black letters represents the comparative map regions on the PHS QTL of wheat chromosome 2B.1

wheat 5'EST	TIGR rice loci	Annotation rice genes	E-value	The location of rice QTL
BE403863	LOC_Os03g02530	membrane protein	5.60E-28	
BM138089	LOC_Os03g03150	fizzy-related protein	4.30E-53	
BE498320	LOC_Os03g04760	expressed protein	6.10E-25	
BG262560	LOC_Os03g07290	ribosomal protein L32 containing protein	1.20E-34	
BF202468	**LOC_Os03g20380**	**CIPK-like protein1**	**2.40E-25**	
BE422913	**LOC_Os03g20700**	**Mg-chelatase subunit XANTRA-F**	**7.10E-84**	
BE406474	**LOC_Os03g22460**	**expressed protein**	**1.10E-24**	
BE403387	**LOC_Os03g22730**	**necleolar NOP5**	**1.30E-55**	
BF483620	**LOC_Os03g24920**	**ubiquitin-like protein**	**1.60E-23**	
BF292706	**LOC_Os03g25110**	**expressed protein**	**9.40E-35**	
BE636923	**LOC_Os03g25760**	**calmodulin binding protein**	**4.70E-21**	
BE495275	LOC_Os03g41100	cyclin-A2	6.30E-41	
BE426646	LOC_Os03g49400	ethylene insensitive 2	3.60E-50	
BE490286	LOC_Os03g56280	malate dehydrogenase	3.50E-75	

Comparative analysis between wheat ESTs in the PHS interval on 2B.1 and rice chromosomes 7 and 3

To construct the comparative map between the major PHS QTL interval on 2B.1 with the orthologous region in rice and *Brachypodium*, new markers close to flanking markers Wmc 474 and Gwm 429 were useful for locating the distal and proximal regions. Designing polymorphic markers based on intronic variation is described later in the marker development section.

A comparative map was constructed based on the ESTs located closest to flanking markers Wmc 474 and Gwm 429 within the PHS QTL interval on wheat chromosome 2B in homologous regions on rice chromosomes 7 and 3 (Figure 2.2). Eleven wheat ESTs were homologous with genes that occurred on both rice chromosomes 7 and 3. However, different E-values suggested varying amounts of homology between the genes on the two chromosomes (Table 2.3). These genes were useful for constructing the comparative map and localized ten of the eleven ESTs in chromosome 2B (interval 2BS1-0.53-0.75). All ESTs were also located on one or both of the other two, group 2 wheat chromosomes. However, one EST, BF428609 that was not located on 2B was also included in the map, to assess collinearity. Three of those ESTs were homologous with genes previously reported to be involved in seed dormancy including Calmodulin binding protein, CBL-interacting serine/threonine-protein kinase 15 (CIPK15) and CIPK like protein 1 (Reddy et al. 2002; Guo et al. 2002; Ok et al. 2005).

The PHS QTL interval on chromosome 2B showed homology with the long arm end of rice chromosome 7 (around 3 Mb, from the distal position 26,017,844 bp to position 29,195,770 bp) and with rice chromosome 3 (around 3 Mb, from a proximal position 11,509,139 bp to position 14,720,405 bp).

Figure 2.2. A comparative map representing the collinear regions between wheat ESTs (wESTs) on deletion interval 2BS1-0.53-0.75 and rice chromosomes 7 and 3.

The comparative map covered the PHS QTL on wheat chromosome 2B.1, between EST BE494262 and BE 500206. The orthologous regions are the distal part of rice chromosome 7 long arm (around 3 Mb) from position 26,017,844 bp to position 29,195,770 bp and the middle part of rice chromosome 3 (around 3 Mb), from position 11,509,139 bp to position 14,720,405 bp. A marker designed from BE494262, localized closet to flanking SSR marker Wmc474. A marker designed from BE 500206, localized closest to flanking SSR marker Gwm 429. The rice gene loci on chromosome 7 and 3 were placed in order. The homologous wESTs were placed according to the order of TIGR rice gene loci. This figure represents only collinearity between rice genes and ESTs on wheat group 2 chromosomes (2A, 2B and 2D). These genes were useful for saturating the PHS QTL region. (The red letters represent the collinear genes on both rice chromosomes 7 and 3, and wheat ESTs. The blue letters in parentheses represent the wheat ESTs located on 2BS1-0.53-0.75.

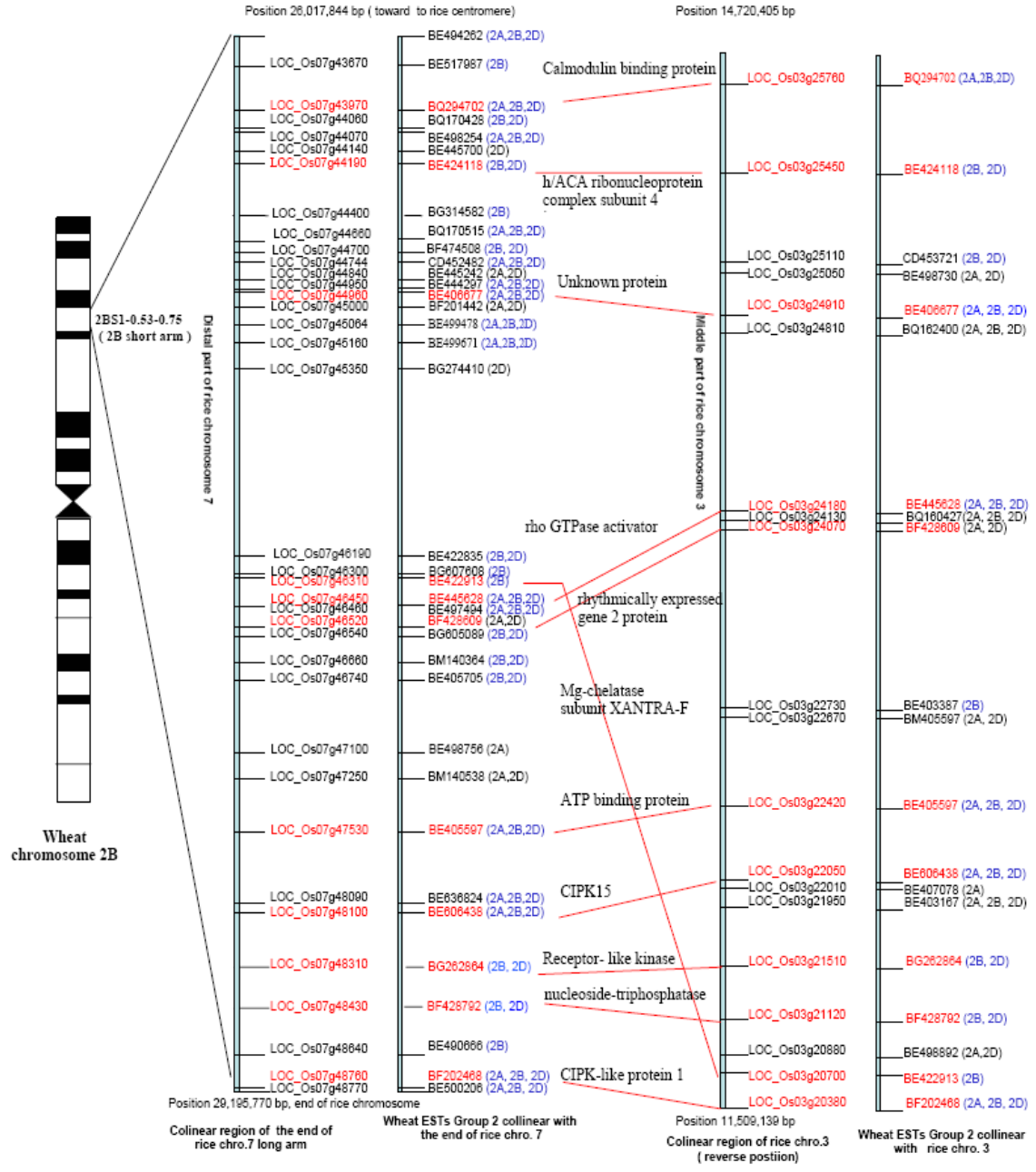

Position 26,017,844 bp (toward to rice centromere)
Position 14,720,405 bp
Wheat chromosome 2B
2BS1-0.53-0.75 (2B short arm)
Distal part of rice chromosome 7
Middle part of rice chromosome 3
Position 29,195,770 bp, end of rice chromosome
Position 11,509,139 bp
Colinear region of the end of rice chro.7 long arm
Wheat ESTs Group 2 collinear with the end of rice chro. 7
Colinear region of rice chro.3 (reverse postiion)
Wheat ESTs Group 2 collinear with rice chro. 3
Calmodulin binding protein
h/ACA ribonucleoprotein complex subunit 4
Unknown protein
rho GTPase activator
rhythmically expressed gene 2 protein
Mg-chelatase subunit XANTRA-F
ATP binding protein
CIPK15
Receptor- like kinase
nucleoside-triphosphatase
CIPK-like protein 1
LOC_Os07g43670
LOC_Os07g43970
LOC_Os07g44060
LOC_Os07g44070
LOC_Os07g44140
LOC_Os07g44190
LOC_Os07g44400
LOC_Os07g44660
LOC_Os07g44700
LOC_Os07g44744
LOC_Os07g44840
LOC_Os07g44950
LOC_Os07g44960
LOC_Os07g45000
LOC_Os07g45064
LOC_Os07g46160
LOC_Os07g46350
LOC_Os07g46190
LOC_Os07g46300
LOC_Os07g46310
LOC_Os07g46450
LOC_Os07g46460
LOC_Os07g46520
LOC_Os07g46540
LOC_Os07g46660
LOC_Os07g46740
LOC_Os07g47100
LOC_Os07g47250
LOC_Os07g47530
LOC_Os07g48090
LOC_Os07g48100
LOC_Os07g48310
LOC_Os07g48430
LOC_Os07g48640
LOC_Os07g48760
LOC_Os07g48770
BE494262 (2A,2B,2D)
BE517987 (2B)
BQ294702 (2A,2B,2D)
BQ170428 (2B,2D)
BE498254 (2A,2B,2D)
BE446700 (2D)
BE424118 (2B,2D)
BG314582 (2B)
BQ170515 (2A,2B,2D)
BF474508 (2B, 2D)
CD452482 (2A,2B,2D)
BE445242 (2A,2D)
BE444297 (2A,2B,2D)
BE406677 (2A,2B,2D)
BF201442 (2A,2D)
BE499478 (2A,2B,2D)
BE499671 (2A,2B,2D)
BG274410 (2D)
BE422835 (2B,2D)
BG607608 (2B)
BE422913 (2B)
BE445628 (2A,2B,2D)
BE497494 (2A,2B,2D)
BF428609 (2A,2D)
BG605089 (2B,2D)
BM140364 (2B,2D)
BE405705 (2B,2D)
BE498756 (2A)
BM140538 (2A,2D)
BE405597 (2A,2B,2D)
BE636824 (2A,2B,2D)
BE606438 (2A,2B,2D)
BG262864 (2B, 2D)
BF428792 (2B, 2D)
BE490666 (2B)
BF202468 (2A, 2B, 2D)
BE500206 (2A,2B, 2D)
LOC_Os03g25760
LOC_Os03g25460
LOC_Os03g25110
LOC_Os03g25050
LOC_Os03g24910
LOC_Os03g24810
LOC_Os03g24180
LOC_Os03g24130
LOC_Os03g24070
LOC_Os03g22730
LOC_Os03g22670
LOC_Os03g22420
LOC_Os03g22050
LOC_Os03g22010
LOC_Os03g21950
LOC_Os03g21510
LOC_Os03g21120
LOC_Os03g20880
LOC_Os03g20700
LOC_Os03g20380
BQ294702 (2A,2B,2D)
BE424118 (2B, 2D)
CD453721 (2B, 2D)
BE498730 (2A, 2D)
BE406677 (2A, 2B, 2D)
BQ162400 (2A, 2B, 2D)
BE445628 (2A, 2B, 2D)
BQ160427(2A, 2B, 2D)
BF428609 (2A, 2D)
BE403387 (2B)
BM405597 (2A, 2D)
BE405597 (2A, 2B, 2D)
BE606438 (2A, 2B, 2D)
BE407078 (2A)
BE403167 (2A, 2B, 2D)
BG262864 (2B, 2D)
BF428792 (2B, 2D)
BE498892 (2A,2D)
BE422913 (2B)
BF202468 (2A, 2B, 2D)

Table 2.3 The wheat ESTs that are homologous with both rice chromosomes 7 and 3

wheat EST	Annotataed rice gene	TIGR gene loci	E-valve
BQ294702	calmodulin binding protein	LOC_Os07g43970	5.60E-55
		LOC_Os03g25760	7.50E-41
BE424118	h/ACA ribonucleoprotein complex subunit 4	LOC_Os07g44190	1.10E-77
		LOC_Os03g25450	3.30E-73
BE422913	Mg-chelatase subunit XANTRA-F	LOC_Os07g46310	1.50E-53
		LOC_Os03g20700	7.10E-84
BE445628	rho GTPase activator	LOC_Os07g46450	2.30E-47
		LOC_Os03g24180	2.40E-37
BE405597	ATP biding protein	LOC_Os07g47530	1.30E-51
		LOC_Os03g22420	2.00E-41
BE606438	CBL-interacting serine/threonine-protein kinase 15	LOC_Os07g48100	2.00E-63
	(CIPK15)	LOC_Os03g22050	6.30E-51
BF202468	CIPK-like protein 1	LOC_Os07g48760	7.60E-22
		LOC_Os03g20380	2.40E-25
BG262864	ATP binding protein/ receptor-like kinase	LOC_Os07g48310	7.00E-47
		LOC_Os03g21510	1.20E-44
BF428792	nucleoside-triphosphatase	LOC_Os07g48430	2.40E-58
		LOC_Os03g21120	4.80E-48

The region on rice chromosome 7 that is orthologous to the PHS QTL interval contained approximately sixty-one rice genes that matched wheat ESTs. Thirty-seven of these matching genes were on wheat chromosomes 2A, 2B and 2D. Thirty out of 37 matching with the ESTs on group 2B chromosomes were mapped in 2BS1-0.53-0.75 interval. For the orthologous region on rice chromosome 3, approximately 58 rice genes matched wheat ESTs. Twenty of these ESTs mapped to wheat group 2 chromosomes. Of the 20 group 2 chromosome ESTs matching rice genes, 12 were mapped in the 2BS1-0.53-0.75 interval. Most of the wheat ESTs in the comparative map on chromosome 2B (2BS1-0.53-0.75) are also located on homoeologous regions of 2A (2AS5-0.78-1.00) and 2D (2DS5-0.47-1.00). Forty-seven percent of the ESTs in the comparative map were located on all three wheat group 2 chromosomes. Thirty-two percent of the ESTs in the comparative map are located on 2 of the 3 group 2 wheat chromosomes. Twenty-one percent of the ESTs are located on only one group 2 chromosome.

A new polymorphic marker, ss26.3, was designed from EST BE494262. It is located near Wmc474 on the short arm. This EST was homologous with rice gene LOC_Os07g43470, located on rice chromosome 7 but no orthologous gene was on rice chromosome 3. Wheat EST BQ294702 was homologous with the nearby calmodulin-binding protein gene in both rice chromosome 7 (LOC_Os07g43970), and rice chromosome 3 (LOC_Os03g25760) at the distal end of the comparative map. On the proximal region of the map, the new polymorphic marker, ss57.2, was designed from EST BE405569. But this marker, located near Gwm429, did not show homology with rice chromosome 7 nor 3. The nearby expressed protein gene, LOC_Os07g48770 on rice chromosome 7, was homologous with BE500206 but this gene was not found on rice chromosome 3. The nearby wheat EST BF202468 is homologous with CIPK-like protein 1 gene, was located on both rice chromosomes 7 (LOC_Os07g48760) and

3 (LOC_Os03g20380) at the proximal end of the comparative map

In summary, there was macrocollinearity between wheat ESTs and rice chromosomes 7 and 3 in terms of gene content and order. The homologous wheat ESTs in the PHS QTL interval were useful to delimit the ESTs that could be used to develop markers.

Comparative analysis of rice seed dormancy QTL with the 2BS1.1 bin and the PHS QTL interval

A comparative analysis of the rice SD QTL locations in an orthologous region was used to identify the possible homologous wheat ESTs that were useful for marker development. Previous QTL studies reported 12 SD QTLs on rice chromosome 7 (Lin et al. 1998; Ishimaru et al. 2001; Cai and Morishima 2002; Miura et al. 2002; Jiang et al. 2003; Gu et al. 2004; Gu et al. 2005; Wan et al. 2005) and 7 SD QTLs on rice chromosome 3 (Ishimaru et al. 2001; Cai and Morishima 2002; Jiang et al. 2003; Takeuchi et al. 2003). Based on Gramene data, the position of rice SD QTLs was compared with Gramene annotated Nipponbare sequence 2006. The location of genes underlying those SD QTLs was compared with rice TIGR gene loci (Table 2.4). QTL intervals on rice chromosome 7 covered from 125 bp to 7 Mb, from position 2,080,474 to 26, 528,484 bp covering nearly the entire rice chromosome (See Table 2.4a). QTL intervals on rice chromosome 3 started at 136 kb and covered the interval to 12 Mb, from position 8,389,519 to 33, 787, 780 bp, again nearly covering the entire rice chromosome (See Table 2.4b). Some QTLs are located in the same region but in different QTL intervals. Some QTL regions were small enough to allow identification of the nearest genes within that rice SD QTL interval. For example, SD QTL, Gramene QTL Accession ID AQC2009 is located at the kinesin POK1 locus.

Table 2.4. Rice seed dormancy (SD) QTLs on chromosomes 7 (A) and 3 (B), based on Gramene information (http://www.gramene.org/)

Gramene QTL assession ID	QTL position	QTL Length	TIGR rice loci	Gene, nearby gene or SD-known functional gene at QTL location	References
CQAH33	2,080,474-2,316,508	236kb			(Cai et al. 2002)
AQGF013	5,734,495-12,782,939	7.04Mb			(Gu et al. 2005)
AQGD004	5,734,495-5,734,684	189bp	LOC_Os07g10570	prolamin precursor	(Gu et al. 2005)
CQE56	9,103,038-18,335,769	9.23Mb			(Ishimaru et al. 2001)
AQCK003	15,447,202-18,335,769	2.9Mb			(Lin et al. 1998)
EQA0003	19,256,213-19,256,338	125bp	LOC_Os07g32390	targeting protein for Xk1p2 containing protein	(Wan et al. 2005)
CQE59	21,439,354-25,302,098	3.8Mb			(Ishimaru et al. 2001)
AQN004	22,247,245-22,251,048	3803bp	LOC_Os07g37140	expressed protein	(Miura et al. 2002)
AQCZ006	22,857,016-22,859,203	2187bp	LOC_Os07g38120	calcium dependent protein kinase (CDPK) isoform AK1	(Jiang et al. 2003)
AQGD008	25,471,987-25,472,119	132bp	LOC_Os07g42570	dirigent- like protein pDIR17	(Gu et al. 2005)
AQFB008	25,471,987-25,651,967	180kb	LOC_Os07g42740, LOC_Os07g42770	Calmodulin binding protein, CDPK related protein kinase	(Gu et al. 2004)
AQCZ009	26,527,687-26,528,484	797bp	LOC_Os07g44400	Kinesin POK1	(Jiang et al. 2003)

A.

Gramene QTL assession ID	QTL position	QTL Length	TIGR rice locci	Gene and nearby gene at QTL location	References
AQDW002	10,166,858-10,303,012	136 kb			(Takeuchi et al. 2003)
AQCZ001	10,349,420-11,509,132	1.15 Mb			(Jiang et al. 2003)
AQF008	15,315,455-27,084,712	11 Mb			(Cai et al. 2002)
AQCZ005	31,571,369-34,177,546	2.6 Mb			(Jiang et al. 2003)
CQAH13	32,255,077-32,256,761	1684 bp	LOC_Os03g56710	hypotical protein	(Cai and Morishima 2000)
CQAH16	33,786,158-33,787,780	1622 bp	LOC_Os03g59470	fos intronic gene CG7615 PA	(Cai and Morishima 2000)
CQE17	8,389,519-21,194,221	12 Mb			(Ishimaru et al. 2001)

B.

AQC2006 is a calcium dependent protein kinase isoform AK1 and AQN004 is an unknown gene. Based on Gramene data, other agronomic QTLs such as heading date, grain weight, abiotic stress, and biotic stress tolerances are also located on rice chromosome 3 while heading date (HD), seed set percent, amylase content, and leaf nitrogen content are located on rice chromosome 7. SD and HD QTLs are located on both rice chromosomes, 7 and 3. Some QTLs for those agronomic traits are also located in the orthologous regions in wheat.

Comparative analysis of the location of rice SD QTLs indicated that there was only one rice SD QTL (Gramene QTL Accession ID AQCZ009) on chromosome 7 (Jiang et al. 2003), located in the region orthologous to the PHS QTL interval in the Cayuga x Caledonia population. That rice location was between 26,527,687-26,528,484 bp. The other eleven rice SD QTLs on chromosome 7 were located outside the PHS QTL interval. They were all located proximal to the rice centromere from 22,247,245 to 25,651,967bp but in the distal region of wheat chromosome 2B. The QTL, AQCZ009, was predicted to lie within the wheat PHS QTL interval. AQCZ009 is reported to be a Kinesin POK1 gene that has 67% homology to wheat EST BG314582. Interestingly, this EST also showed 62% homology with the FAD binding protein whose expression levels co-segregated with the PHS QTL on wheat chromosome 2B (Munkvold 2007).

In addition, there were two rice SD QTLs on chromosome 3 that were predicted to lie within the orthologous PHS QTL interval on 2B.1. The chromosome 3 regions were Gramene ID CQE17, from position 8,389,519-21,194,221 (Ishimaru et al. 2001) and AQCZ001, from position 10,349,420-11,509,132 (Jiang et al. 2003). Five SD QTLs were reported to be outside of the comparative mapping region. The QTL interval is very large, indicating that many closely linked regions contribute to SD on rice chromosome 3.

In summary, there were 3 rice SD QTLs predicted to lie within the region orthologous to the PHS QTL interval and new markers based on the genes Kinesin POK1 and FAD-binding protein were used for the comparative mapping.

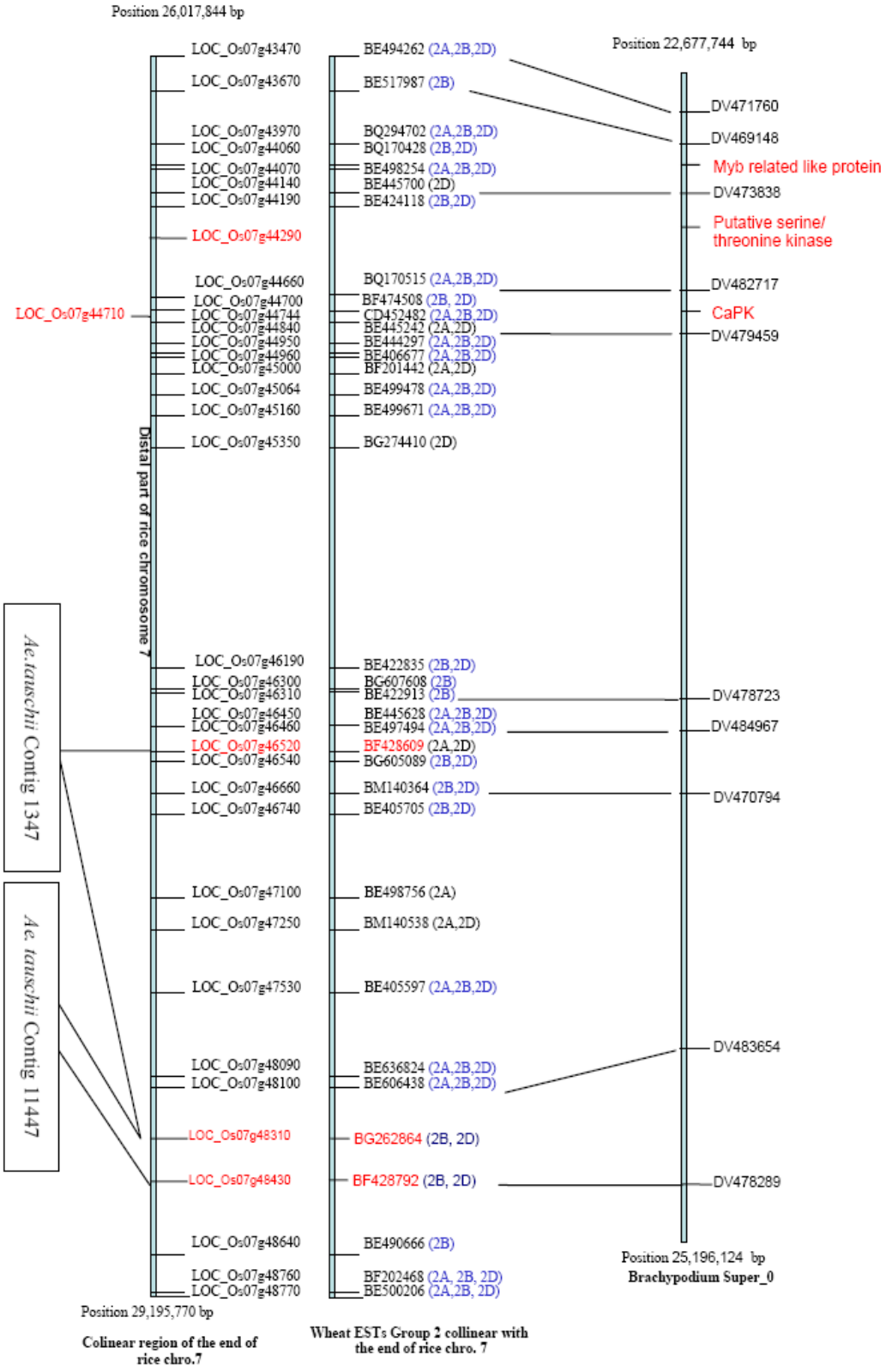
Comparative analysis among wheat ESTs, rice, Brachypodium and Ae. tauschii BACs

The region on the comparative map was also homologous with *Brachypodium* Super_0. There were 10 wheat ESTs that are homologous with *Brachypodium* ESTs (Figure 2.3). The gene order of *Brachypodium* was the same as that of rice chromosome 7 and the wheat PHS QTL interval in the 2B.1 bin at the macro level. All three wheat transcripts, BQ743864, BT009309.1 and TA32305, homologous to myb DNA binding protein, Calmodulin-dependent protein kinase1 and CBL-interacting serine/threonine-protein kinase 1, respectively in rice chromosome 7, also occurred in *Brachypodium* in the same order. Those known function genes have been reported to be involved in seed dormancy (Ok et al. 2005; Woodger et al. 2003; Sheen 1996). Wheat transcripts of those genes were used to develop markers.

Aegilops tauschii is a diploid wheat relative that the D-genome. There is information available for *Ae. tauschii* Bacterial artificial chromosome (BAC) libraries and a partial physical map (Akhunov et al. 2005). If a BAC sequence were available it may reveal the gene order in hexaploid wheat, which also contains a D-genome. There are 3 wheat ESTs, BF428609, BG262864 and BF428792 in the PHS QTL interval that have been located on *Ae. tauschii* BACs (Figure 2.3). These BACs are located in two contigs, 1347 and 11447. Interestingly, BF428609 (located on 2A and 2D) and BG262864 (located on 2B and 2D) are both located in the same BACs (bold letters in following sentence) on contig 1347.

Figure 2.3 A comparative map representing the collinear regions between wheat ESTs on deletion interval 2BS1-0.53-0.75, rice chromosome 7, *Brachypodium* and *Ae. tauschii* BACs.

The comparative map covered the PHS QTL on wheat chromosome 2B.1 between EST BE494262 and BE 500206. The orthologous regions are the distal part of rice chromosome 7 (around 3 Mb) from position 26,017,844 bp to position 29,195,770 bp and *Brachypodium* (around 2.5Mb), from position 22,677,744 bp to position 25,196,124 bp. A marker designed from BE494262, localized closest to flanking SSR marker Wmc474. A marker designed from BE 500206, localized closest to flanking SSR marker Gwm 429. The rice gene loci on chromosome 7 and ESTs of *Brachypodium* were placed in order. The homologous wESTs were placed according to the order of the TIGR rice gene loci. This figure represents only collinearity between rice genes and wESTs on wheat group 2 chromosomes (2A, 2B and 2D). Wheat ESTs BF428609 (2A and 2D), BG262864 and BF428792, in red color, were located in *Ae.tauschii* BACs. These BACs might be useful for revealing wheat gene content and order. The known-function genes, in red, were used to design markers. The blue letters in parentheses represent the wheat ESTs located on 2BS1-0.53-0.75.


Position 26,017,844 bp
LOC_Os07g43470
LOC_Os07g43670
LOC_Os07g43970
LOC_Os07g44060
LOC_Os07g44070
LOC_Os07g44140
LOC_Os07g44190
LOC_Os07g44290
LOC_Os07g44660
LOC_Os07g44700
LOC_Os07g44744
LOC_Os07g44840
LOC_Os07g44950
LOC_Os07g44960
LOC_Os07g45000
LOC_Os07g45064
LOC_Os07g45160
LOC_Os07g45350
LOC_Os07g44710
Distal part of rice chromosome 7
LOC_Os07g46190
LOC_Os07g46300
LOC_Os07g46310
LOC_Os07g46450
LOC_Os07g46460
LOC_Os07g46520
LOC_Os07g46540
LOC_Os07g46660
LOC_Os07g46740
LOC_Os07g47100
LOC_Os07g47250
LOC_Os07g47530
LOC_Os07g48090
LOC_Os07g48100
LOC_Os07g48310
LOC_Os07g48430
LOC_Os07g48640
LOC_Os07g48760
LOC_Os07g48770
Position 29,195,770 bp
Colinear region of the end of rice chro.7
Ae.tauschii Contig 1347
Ae. tauschii Contig 11447
BE494262 (2A,2B,2D)
BE517987 (2B)
BQ294702 (2A,2B,2D)
BQ170428 (2B,2D)
BE498254 (2A,2B,2D)
BE445700 (2D)
BE424118 (2B,2D)
BQ170515 (2A,2B,2D)
BF474508 (2B, 2D)
CD452482 (2A,2B,2D)
BE445242 (2A,2D)
BE444297 (2A,2B,2D)
BE406677 (2A,2B,2D)
BF201442 (2A,2D)
BE499478 (2A,2B,2D)
BE499671 (2A,2B,2D)
BG274410 (2D)
BE422835 (2B,2D)
BG607608 (2B)
BE422913 (2B)
BE445628 (2A,2B,2D)
BE497494 (2A,2B,2D)
BF428609 (2A,2D)
BG605089 (2B,2D)
BM140364 (2B,2D)
BE405705 (2B,2D)
BE498756 (2A)
BM140538 (2A,2D)
BE405597 (2A,2B,2D)
BE636824 (2A,2B,2D)
BE606438 (2A,2B,2D)
BG262864 (2B, 2D)
BF428792 (2B, 2D)
BE490666 (2B)
BF202468 (2A, 2B, 2D)
BE500206 (2A,2B, 2D)
Wheat ESTs Group 2 collinear with the end of rice chro. 7
Position 22,677,744 bp
DV471760
DV469148
Myb related like protein
DV473838
Putative serine/ threonine kinase
DV482717
CaPK
DV479459
DV478723
DV484967
DV470794
DV483654
DV478289
Position 25,196,124 bp
Brachypodium Super_0

Those BACs in contig1347 (467 kb) were HB096E14, HI143A9, HD122I6, HB059D13, HI078M16, HD062D23, **HD079P6**, **RI095G8**, **HI119H7**, **HB132M21**, **HD080C17**, **HD064P22**, **HI112P17** and **HB022E10**. Some of those smaller BACs were imbedded on larger BACs (see more details on http://phymapgb.bioinformatics.ucdavis.edu/cgi-bin/gbrowse/wheatfpc11/#search). The distance between BF428609 and BG262864 was around 1.07 mb based on a rice physical map. The BF428609 is located in the middle of the comparative map and may be close to that of the target marker, Barc55. A polymorphic marker close to this EST would prove that this BAC sequence is needed. BG262864 is located on contig 1347, but also occurs on the same BAC such as HBO22G2 and BF428792 (located on 2B and 2D) on contig 11447. The distance between BG262864 and BF428792 was around 99 kb, based on the rice physical map. BG262864 was not only located on contig 1347 and 11447 but on the other 5 contigs as well. It is possible that a duplication of this gene occurred within chromosome 2B and 2D. BF428609 was located on contig 1347 and contig 5778, whereas BF428792 was located on contig11447 only. There are three ESTs mapped in contig 5778, BF428609, BE494076 and BG262340. BE494076 and BG262340 did not locate in the PHS interval. BE494076 was mapped on 1B and 1D whereas BG262340 was mapped on 3B, 4B and 6B. This contig might not be located on a PHS interval.

In summary, macrocollinearity not only occurred between wheat ESTs and rice chromosomes but also with *Brachypodium*, in terms of gene content and order. In addition, *Ae. tauschii* contig 1347 was the most likely to be a homologous region containing Barc55, based on comparative mapping, because two ESTs in the PHS interval are located in this contig.

The known function genes involved in seed dormancy in the orthologous region

Based on the comparative map, there were 7 ESTs homologous to known-function genes involved in seed dormancy (Table 2.5). Those genes code for chalcone synthase (Debeaujon et al. 2000), CBL-interacting serine/threonine-protein kinase (CIPK)-like protein 1(Ok et al. 2005), calmodulin binding protein (Reddy et al. 2002), developmentally-regulated GTP-binding protein 1, CBL-interacting serine/threonine-protein kinase 11 (Mahajan et al. 2006; Gong et al. 2004) and CBL-interacting serine/threonine-protein kinase 15 (Guo et al. 2002). New markers based on wheat transcripts were developed for the myb DNA binding protein (Woodger et al. 2003), Calmodulin-dependent protein kinase1 (Sheen 1996) and CBL-interacting serine/threonine-protein kinase 1 (Ok et al. 2005). BF293744 matched chalcone synthase DII, but mapped outside of the PHS interval. However it may be a candidate gene for PHS in red-grained wheat because of its role in the pigment pathway

Marker development

1) ***Designing polymorphic markers based on intron variation.***

Only 11 primers were polymorphic between the parents and could be mapped in the DH population (Figure 2.4 and Table 2.7). The aim of this step was to develop all possible markers from ESTs located in the wheat deletion interval 2BS1-0.53-0.75 (2BS1.1) containing the PHS QTL interval and flanked by markers Wmc474 and Gwm 429. There were 153 primers designed from sequences of 71 of the 104 ESTs. The other ESTs and their contigs were either too short or lacked introns, when compared to homologous rice genes. Ninety-four of the primers produced monomorphic amplicons between the two white wheat parents, Cayuga and Caledonia. Fifteen primers amplified non-specific faint bands and 11 primer pairs produced no product. Twelve primers amplified products for only one parent.

Table 2.5 The known-function genes located at collinear regions that have been reported to be involved in seed dormancy or preharvest sprouting.

Most of these wheat ESTs and wheat transcripts were used to design new markers for fine mapping. Since the markers were designed from known-function genes, they usually led to candidate genes that showed seed dormancy and preharvest sprouting phenotype and genotype co-segregation.

wheat 5'EST	wheat transcipt	TIGR rice loci	Annotation gene	Reference
BF293744		LOC_Os07g34190	chalcone synthase DII	(Debeaujon et al. 2000)
BF202468		LOC_Os03g20380	CIPK-like protein1	(Ok et al. 2005)
BE636923		LOC_Os03g25760	calmodulin binding protein	(Reddy et al. 2002)
BE494262		LOC_Os07g23200	developmentally-regulated GTP-binding protein1	?
BQ294702		LOC_Os07g43970	calmodulin binding protein	(Reddy et al. 2002)
BE636824		LOC_Os07g48090	CBL-interacting serine/threonine-protein kinase 11 (CIPK11)	(Mahajan et al. 2006;Gong et al. 2004)
BE606438		LOC_Os07g48100	CBL-interacting serine/threonine-protein kinase 15 (CIPK15)	(Guo et al. 2002)
	BQ743864	LOC_Os07g43420	myb related protein	(Woodger et al. 2003)
	BT009309.1	LOC_Os07g44710	Calmodulin-dependent protein kinase1	(Sheen 1996)
	TA32305	LOC_Os07g44290	CBL-interacting serine/threonine-protein kinase 1 (CIPK1)	(Ok et al. 2005)

The 21 remaining primer pairs were polymorphic between the parents and were tested for mapping on the Cayuga x Caledonia DH population. Some polymorphic primers were designed from the same EST and some produced non-specific banding patterns. The polymorphic markers generated from this step were ss57.2-BE405569, ss3.2-BF202468, ss66.2-BE500206, ss4.4-BE606438, ss47.5-BE498254, ss26.3-BE494262, ss11.4-BE424562, ss11.1-BE424562, ss61.2-BG314234, ss64.2-BG274905 and ss38.1-BE488865 (Table 2.7). The product length varied from 200 to 600 bp, depending on the intron length. The three markers ss61.2-BG314234, ss64.2-BG274905 and ss38.1-BE488865 amplified fragments located on chromosomes either 2A or 2D. Markers ss3.2- BF202468, ss4.4- BE606438 and ss26.3- BE494262 were among the markers designed from known-function genes that might affect seed dormancy (Table 2.5).

The next step of marker development was conducted after constructing the comparative map. The comparative map was then used as the tool to identify the useful ESTs, transcripts and candidate known-function genes located in the orthologous region

2) ***Designing primers based on variation in coding region.***

As a result, three out of 22 primer sets, ss44-BE636824, ss31-BF201533 and ss20-BQ294702, showed polymorphism (Figure 2.4 and marker details in Table 2.7). The aim of this step was to develop CAPS (cleaved amplified polymorphic sequence) by using long sequences and also to redesign primers that amplify smaller fragments, in case CAPS were not successful. Only the contigs of wheat ESTs in the orthologous region were used to design primers in this step. The contigs of 22 wheat ESTs were used to design 44 primer pairs. The amplified products were around 500-1000 bp, depending on the contig length. Sequencing directly from the PCR products from both parents Cayuga and Caledonia failed because of the large number of duplicated loci,

so the PCR products from each parent were cloned into vectors to separate each copy. A CAPS design was unsuccessful because the hexaploid wheat genome made it difficult to compare complementary sequences between two parents. For example, a polymorphic CAPS primer was found by a comparison between sequences from non-matching loci from both parents. However, after using the CAPS on Cayuga and Caledonia, it was still monomorphic. Unknown matching loci make CAPS difficult to design in hexaploid wheat.

A second approach was to use long sequences to design primers that amplified many small fragments. The primers were designed from conserved regions that covered the variable region from both parents. After knowing the sequence, the new primers were designed to be short fragments between 200-400 bp. These small fragments were visualized on PAGE and SSCP gels. Markers ss44-BE636824 and ss20-BQ294702 were designed from known-function genes that might contribute to seed dormancy.

3) ***Designing primers based on candidate genes that might affect seed dormancy.***

Eleven genes were selected for mapping based on previous reports that they may play a role in seed dormancy or PHS (Table 2.5). Primers for five genes had already been developed as described above; ss3.2-BF202468 was homologous with CBL-interacting serine/threonine kinase-like protein 1, ss4.4-BE606438, with CBL-interacting serine/threonine kinase 15, ss26.3-BE494262 with developmentally-regulated GTP binding protein, ss44-BE636824 with CBL-interacting serine/threonine kinase 11 and ss20-BQ294702 with Calmodulin binding protein. Additional primers were designed based on Calmodulin-dependent protein kinase 1, CBL-interacting serine/threonine-protein kinase 1, Myb binding protein and FAD-binding protein. FAD-binding protein was one of the expression QTL genes on 2B.1 (Munkvold, personal communication). Twenty primer pairs were designed from Calmodulin-

dependent protein kinase1, but only two markers were polymorphic; one located on 2A (CaPK1-3) and one on 2B (rCaPK1-2&1-1-1 or rCaPK). Thirteen primer pairs were designed for CBL-interacting serine/threonine-protein kinase 1. All 13 primers were monomorphic between the parents after surveying them on both PAGE and SSCP. Six primer pairs were designed from Myb-related protein. Two polymorphic markers for Myb-related protein were developed. Marker ssMyb1&12 was on 2B.1 whereas Myb1&13 was an unlinked marker. Twelve primer pairs were designed for the FAD-binding protein and one was polymorphic on 2B.1. (See primer details on Table 2.7)

4) *Designing primers based on wheat transcripts surrounding EST BF428609, a potential for Ae. tauschii BAC sequencing .*

To prove that *Ae. tauschii* BACs in a contig1347 that contain ESTs BF428609 and BG262864 are needed, new markers were designed based on ESTs surrounding EST BF428609. The information of EST and EST contigs was used in the previous steps in marker development so only the homologous wheat transcripts of these ESTs, BG607608, BE422913, BE445628, BG605089 and BM140364, were used to designed new primers in both intronic and coding variations. Twenty-eight primers pairs were designed from 5 wheat transcripts. Twenty-four primers were monomorphic in both PAGE and SSCP, 2 primers were non-specific and 2 primers were polymorphic. The polymorphic primers produced in this step were ss56.3-BG605089 located on 2A.2 and ss56.2-BG605089 located on 2D. These markers were both designed from the same wheat transcript (CA497948) that was homologous with EST BG605089 but in a different region. The distance between the BG605089 and BF428609, that was mapped on *Ae. tauschii* contig 1347, was around 7.4 kb. Marker ss56.3-BG605089 on 2A.2 was located between marker ss4.4a and ss20a, closer to marker ss4.4a. The distance between marker ss56.3-BG605089 and ss4.4a was 5.4 cM, the same distance

between markers ss4.4b and Barc55 at linkage 2B.1. Marker ss53.2-BG605089 on 2D was 5.8cM away from the other two new markers and 10.8cM away from Wmc453a. The distance was almost the same as the distance between ss56.3-BG605089 and Wmc453b (Figure 2.4 and primer detail on Table 2.7). No markers developed in this step located on 2B. However, a comparison with 2A and 2D suggested that *Ae. Tauschii* contig 1347 would reveal the genes. There were 14 *Ae. tauschii* BAC clones located in contig 1347, covering 467 kb. *Ae. Tauschii* BAC clones containing ESTs BF428609 and BG262864 cover about half of contig 1347. The largest BAC clone, HI119H7, covered around 242 kb.

In summary, *Ae. Tauschii* BAC sequence around 467 kb could reveal the numerous genes that are located close to marker Barc55, based on the comparison of the location of new markers, ss56.3- BG605089 and ss56.2- BG605089. One of those genes may be a candidate gene contributing to PHS.

The updated linkage maps for chromosome bins 2A.2, 2B.1 and 2D.1

Fourteen new marker loci were mapped to 2B.1, five were in 2A.2 and three loci were on 2D. In bin 2B.1, nine out of 14 markers were located within the PHS QTL interval. Five out of 14 markers, including a marker ssFAD-5 designed from FAD binding protein, were located distal to the PHS QTL marker, Wmc474.

All of the new markers detected on 4% PAGE were dominant whereas markers detected on 10% SSCP were either dominant (60%) or co-dominant (40%). Example pictures of the polymorphic marker products are shown in Appendix Figure A1.1 (4% PAGE) and Appendix Figure A1.2 (10%SSCP). Ten markers were used for further fine mapping. The 8 markers located on 2A and 2D were useful to localize the homoeologous regions between group 2 chromosomes on the comparative map. In a previous linkage map (Munkvold, 2007), Wmc453 amplified three loci on 2A, 2B and

2D. The linkage maps for wheat chromosomes, 2A, 2B and 2D and homoeologous regions between the genomes were updated after adding new markers (Figure 2.4). Marker Xe40m49-83 on chromosome 2A was linked to a minor PHS QTL. The PHS QTLs linked to Barc55 on 2B and to Xgwm71b on chromosome 2D had larger effects.

Table 2.6 The homology between wheat ESTs, *Brachypodium* ESTs and rice chromosome 7 on a comparative map.

wheat EST	BrachyEST	Annotataed rice gene	TIGR gene loci rice chro.7
BE494262	DV471760	developmentally-regulated GTP-binding protein1	LOC_Os07g43430
BE517987	DV469148	ribonuclease 1 precusor	LOC_Os07g43670
BE445700 (2D)	DV473838	cytochrome P450 42A1	LOC_Os07g44140
BF202681, BQ170515	DV482717	xylulose kinase	LOC_Os07g44660
BE445242 (2A,2D)	DV479459	protein yrdA	LOC_Os07g44840
BE422913	DV478723	Mg-chelatase subunit XANTRA-F	LOC_Os07g46310
BE497494	DV484967	ferrodoxin-dependent glutamate synthase	LOC_Os07g46460
BM140364	DV470794	UBP 24	LOC_Os07g46660
BE606438	DV483654	CBL-interacting serine/threonine-protein kinase 15	LOC_Os07g48100
BF428792	DV478289	nucleoside-triphosphatase	LOC_Os07g48430

Figure 2.4 The updated linkage maps among wheat 2A.2, 2B.1 and 2D.1, and homologous regions between the genomes, after adding new markers.

New polymorphic markers were added to the previous linkage groups, 2A.2, 2B.1 and 2D.1, of a doubled haploid population. The population is a cross between winter white Cayuga wheat and Caledonia wheat (population size =149). SSR marker Wmc453, from previous linkage map, was located at all three group 2 chromosomes and represented the collinear region between the genomes. Wmc474 and Gwm429 were the flanking markers and Barc55 are the middle marker within the major PHS QTL on 2B.1. The black bars represent the PHS QTL location on the DH population. The minor PHS QTL on 2A.2 was significant in one environment ($p < 0.05$, Population size =149) (Munkvold 2007). The major PHS QTL on 2B.1 was highly significant ($p < 0.01$) in all 16 environments (Population size = 209) (Munkvold et al. 2009). The major PHS QTL on 2D was highly significant ($p < 0.01$) in 7 environments (Population size =209) (Munkvold et al. 2009). The comparative map represents the major PHS QTL on 2B.1, based on the PHS mean of 16 environments (in black bar on 2B.1). The broken bar represents multiple overlapping PHS QTL intervals on 2B.1, whose flanking markers sometimes change in different environments. New markers (in bold script) are designed based on ESTs, EST contigs and gene transcripts. Those markers are located in all three homoeologous regions in the wheat genomes 2A, 2B and 2D.

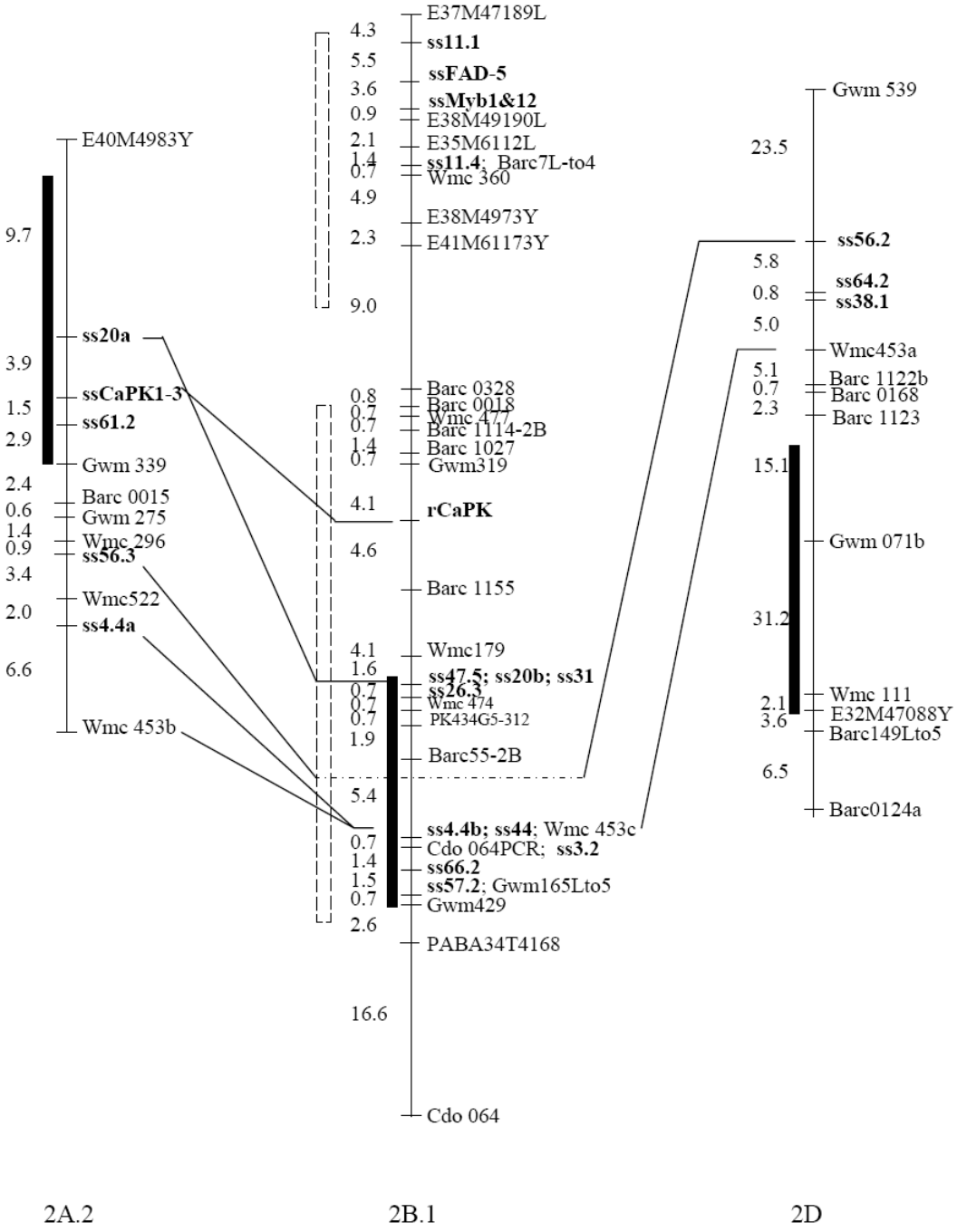

E40M4983Y
9.7
ss20a
3.9
ssCaPK1-3'
1.5
ss61.2
2.9
Gwm 339
2.4
Barc 0015
0.6
Gwm 275
1.4
Wmc 296
0.9
ss56.3
3.4
Wmc522
2.0
ss4.4a
6.6
Wmc 453b
E37M47189L
4.3
ss11.1
5.5
ssFAD-5
3.6
ssMyb1&12
0.9
E38M49190L
2.1
E35M6112L
1.4
ss11.4; Barc7L-to4
0.7
Wmc 360
4.9
E38M4973Y
2.3
E41M61173Y
9.0
0.8
Barc 0328
0.7
Barc 0018
0.7
Wmc 477
Barc 1114-2B
1.4
Barc 1027
0.7
Gwm319
4.1
rCaPK
4.6
Barc 1155
4.1
Wmc179
1.6
ss47.5; ss20b; ss31
0.7
ss26.3
0.7
Wmc 474
0.7
PK434G5-312
1.9
Barc55-2B
5.4
ss4.4b; ss44; Wmc 453c
0.7
Cdo 064PCR; ss3.2
1.4
ss66.2
1.5
ss57.2; Gwm165Lto5
0.7
Gwm429
2.6
PABA34T4168
16.6
Cdo 064
Gwm 539
23.5
ss56.2
5.8
ss64.2
0.8
ss38.1
5.0
Wmc453a
5.1
Barc 1122b
0.7
Barc 0168
2.3
Barc 1123
15.1
Gwm 071b
31.2
Wmc 111
2.1
E32M47088Y
3.6
Barc149Lto5
6.5
Barc0124a
2A.2
2B.1
2D

Table 2.7. New markers designed from wheat ESTs, their contigs, and wheat transcripts.

Some of these markers were used for fine mapping for the PHS QTL on wheat chromosome 2B.1 of a backcross white winter Cayuga x Caledonia wheat population. The PCR annealing temperatures were at 56°C except where indicating otherwise. The estimated DNA fragment size is based on a 4% PAGE gel. F = Forward primer, R = Reverse primer and chr. = Chromosome.

Primer name	primer sequences	EST or contig or gene	Rice annotation gene	Gel detection	Estimated DNA size-chr. location
ss57.2	F--TGGGACACTTTCACTCATGC R--GCCTCTAGGCCTGTCTCAGAT	BE405569	Non-cyanogenic beta-glucosidase precursor	SSCP	500bp-2B
ss3.2	F--CCCACCATACTTTGAGTGGTTT R--GCCCGTTTTTCACCATGTAA	BF202468	CBL-interacting serine/threonine kinase -like protein 1	SSCP	200bp-2B
ss66.2	F--CCCACCATACTTTGAGTGGTTT R--GCCCGTTTTTCACCATGTAA	BE500206	expressed protein	PAGE	200bp-2B
ss4.4	F--GGAACATCCTTGGTTCAGGA R--GGTCGAGTCCAGTTGAGAGG	BE606438	CBL-interacting serine/threonine kinase 15	SSCP	210bp-2A,2B
ss44c2c4 (ss44)	F--GACGGAGTTCAAGGAGGACA R--GGCCTCAGCCTCCACGTA	BE636824	CBL-interacting serine/threonine kinase 11	SSCP	240bp-2B
ss47.5	F--GAGTCAAGTGCCTCGCTGAT R--CAGAACCCTGCAATCTCACA	BE498254	Carboxylic ester hydrolase	SSCP	250bp-2B
ss31c2c3 (ss31)	F--ACTGCCCTACGAGAAGCTGA R--GCATATGATTCCCTCAAAGCA	BF201533 BQ170428	Catalytic hydrolase	SSCP	280bp-2B
ss20c3c1 (ss20)	F--GAAGAAGAAATGGTTCCTGCTG R--TCCCAATGACTTCTCCCTTG	BQ294702	Calmodulin binding protein	SSCP	250bp-2A, 2B
ss26.3	F--TTCCTTCTGTTGGGAAATCAA R--GCCCCTTTGTACATAATAACTCCA	BE494262	developmentally-regulated GTP binding protein	PAGE	220bp-2B

Table 2.7 (Continued)

rCaPK1-2&1-1-1 (rCaPK)	F--GCCGAGCCAAATCCCGGAGC R--CCTTGAAGATGGCAAAATGA	BT009309.1 (transcript)	Calmodulin-dependent protein kinase1	PAGE (50 °C)	360bp-2B
ss11.4	F--GTGTTTTGGCGTGGTGTTCT R--CGCATTGCCATCCCATATAC	BE424562	Dihydrofolate reductase	SSCP	650bp-2B
ssMyb1&12	F--GGAACATGCGGGATCTGG R--ACGAAGCACGTCATCCTCCT	TA52577 (transcript)	myb related like protein	PAGE (50°C)	230bp-2B
ssFAD-5	F--CAGGAACCGTGCTGATTCTT R--CACCATTTCCAATGACATGC	TA12248_4513 (transcript)	FAD-binding protein	SSCP (50°C)	175bp-2B
ss11.1	F--CTAGCATCCGCTCCCTACAG R--CAGTAAGATGGATGGCCTCA	BE424562	Dihydrofolate reductase	SSCP	215bp-2B
ssCaPK1-3	F--GCTCAACGACATTGTTGGAA R--TCAGAGCTTGTGAGGCAGTC	BT009309.1 (transcript)	Calmodulin/Ca^{2+}-dependent protein kinase1	SSCP	290bp-2A
ss61.2	F--ACAGGAGGTTGGATGAGTGG R--TCTGGTGTCGTGTCATCTTCA	BG314234	putative histone acetyltransferase	SSCP	320bp-2A
ss56.3	F--GGATCTCAAGCCTTGCAAAA R--CCTTGAGGTGTTGATTGCAC	BG605089 CA497948 (transcript)	Condensin complex subunit 1	PAGE	275bp-2A
ss64.2	F--TTGGTGAAAATTTATTGAAGAACA R--TCCTAGGGTCCTTGGTCTTGT	BG274905	Protein AIG1	PAGE	225bp-2D
ss38.1	F--GAATGGCGGAGGAGAAGG R--ACCAGTTGTAGGCGATGGAT	BE488865	hypoxia induced protein	PAGE	250bp-2D
ss56.2	F--GAGGCAAATCTCCAGATCCT R--GCAGCATTCTTCCTCACAGA	BG605089 CA497948 (transcript)	Condensin complex subunit 1	SSCP	180bp-2D
ssMyb1&13	F--GAATGGCGGAGGAGAAGG R--ATGTGGCGGTCCTTCTTG	TA52577 (transcript)	myb related like protein	PAGE (50°C)	125bp-unknown

DISCUSSION

Comparative analysis of a region orthologous to the PHS QTL interval in the 2B.1 bin

Useful ESTs for marker development in wheat were revealed in the PHS interval. Also, the location of rice seed dormancy QTL and known-function genes on the orthologous region within the interval were revealed. In this study, comparative mapping revealed macrocollinearity between the wheat interval on chromosome 2B containing a PHS QTL (bin 2BS1-0.53-0.75), rice chromosome 7, rice chromosome 3 and *Brachypodium* Super_0 (Super represents the super contigs). This 2BS1-0.53-0.75 bin (2BS1 bin) was previously reported to be homologous to rice chromosome 7 (La Rota and Sorrells 2004), but these results suggest that it is also partially homologous to rice chromosome 3. Not all wheat ESTs located in the 2BS1-0.53-0.75 bin were located within the PHS QTL interval flanked by markers Wmc 474 and Gwm 429. For example, 70% of all wheat ESTs in the 2BS.1 bin that were homologous to rice chromosome 7 sequences were located in the PHS QTL interval whereas 30% were located distal to Wmc474 and proximal to Gwm429. Similarly, only 50% of all wheat ESTs in the 2BS.1 bin that were homologous to rice chromosome 3 were located in the PHS QTL interval. Consequently, the comparative map was used to narrow down the possible wheat ESTs and candidate genes that were located the PHS QTL interval.

The comparative map showed that the PHS QTL interval on wheat 2B.1 was homologous to a 3 Mb segment on rice chromosomes 7 and 3. The homologous segment in *Brachypodium* is about 2.5 Mb. Fifty-one percent of the wheat ESTs that were homologous with genes on rice chromosome 7 and located in the 2B.1 interval, mapped to group 2 chromosomes. Thirty-four percent of the wheat ESTs that were homologous with genes on rice chromosome 3 and located in the 2B.1 interval, mapped to group 2 chromosomes.

Seed dormancy QTL and known-function genes involved in seed dormancy

There were several rice seed dormancy (SD) QTLs located in the homologous rice region, suggesting that there may also be genes that are involved in wheat SD. The regulation may include ABA regulation or GA regulation or recognition. CBL-interacting serine/threonine kinase-like protein 1 (CIPK1 like protein), CBL-interacting serine/threonine kinase 15 (CIPK15), Calmodulin/Ca^{2+}-dependent protein kinase 1 (CDPK), and CBL-interacting serine/threonine-protein kinase 1 (CIPK1) were all genes of the threonine/serine protein kinase family (Ok et al. 2005; Anderberg and Walker-Simmons 1992). CIPK1, CIPK15 were reported to be involved in ABA responsiveness (Guo et al. 2002; D'Angelo et al. 2006). Wheat Ca^{2+}-dependent protein kinases were found to respond to both ABA and GA (Li et al. 2008). Calmodulin binding protein and a Myb related protein were part of GA-down regulation (Bethke et al. 1997). These genes could be involved in embryo dormancy in white grain wheat. Chalcone synthase DII is known to be an intermediate in the flavonoid biosynthetic pathway (Debeaujon et al. 2000) and could influence seed dormancy in red kernel wheat.

Not only rice SD QTLs but also QTLs for other agronomic traits such as heading date, grain weight, abiotic stress, and biotic stress tolerances are located in the homologous regions in rice. Both QTLs for seed dormancy and heading date were closely linked on the wheat 2B chromosome (Munkvold 2007), and were reported to be linked on rice chromosomes 7 and 3 (Takeuchi et al. 2003). Several traits, including resistance to diseases like stem rust (Wu et al. 2009), leaf rust (Leonova et al. 2007), and yellow rust (Mallard et al. 2005) were also controlled by genes on wheat chromosome 2B. Other genes controlling agronomic traits include semi-dwarfing genes *Rht7* (Worland et al. 1980) on chromosome 2A; *Rht8* on 2D (Worland et al. 1998) and photoperiod response genes *Ppd1*, *Ppd2* and *Ppd3*, controlling ear-

emergence time, on 2A, 2B, and 2D (Scarth and Law 1983). *Ppd2* is the major photoperiod gene on the short arm of chromosome 2B.

The markers for fine mapping that have been developed in this project will help to reveal the genes contributing to PHS and be advantageous for other traits that are located in this region. For example, the stem rust genes *sr36* and *sr40* are located between wmc 474 and gwm429 that are flanking markers for the PHS QTL (Wu et al. 2009).

Sixty percent of the wheat ESTs in the 2B.1 bin were homologous with genes on rice chromosomes 7 and 3 whereas 30% were homologous with genes on other rice chromosomes and 10% were not homologous with any rice genes. Consequently, not all possible candidate genes were identified on the comparative map. For example, wheat EST BG314582 was annotated as a Kinesin POK1 gene at 67% homology and matched one of the rice SD QTLs (Jiang et al. 2003). Interestingly, this EST also showed 62% homology with FAD-binding protein, whose expression QTL (eQTL) (r^2 = 0.45) co-segregated strongly with the PHS QTL phenotype (Munkvold, personal communication). This FAD-binding protein was homologous to a rice gene on chromosome 4 (LOC_Os04g02730). Wheat ESTs BG605273 and CD454629 are homologous with this gene and have been mapped on wheat group 2 chromosomes (Rice Genome Annotation TIGR release 5, http://www.modelcrop.org/cgi-bin/gbrowse/rice/). Marker ssFAD-5, designed from FAD-binding protein (LOC_Os04g02730), was located outside of the 2B.1 QTL interval, close to ssMyb1&12. How this gene might interact with other genes to affect PHS is still unknown. There were 14 eQTL genes reported for 2B.1 (Munkvold 2007). The most significant eQTL (r^2 = 0.52) did not match any genes in a BLAST search. The r^2 values were often quite high because the heritability of individual gene expression is often quite high. The next significant eQTL matched FAD-binding protein (r^2 = 0.45).

Nine eQTLs matched known function genes (r^2 = 0.14-0.40) and three were unknown proteins (r^2 = 0.15, 0.24 and 0.29). It was difficult to develop a marker from an eQTL in the PHS interval because eQTLs may be either cis- or trans-regulated.

Homoeologous relationship among group two on linkage 2A.2, 2B.1 and 2D

Aegilops tauschii BAC sequence is useful for studying gene order. However, the D-genome may not be a good representative for the QTL region on chromosome 2B. Our results suggest that chromosome 2A and 2B are more closely related to each other than they are to 2D in gene order and content because some markers that occur on 2A and 2B do not occur on 2D. However, it might be the low polymorphism contributing fewer markers on 2D. Comparative mapping indicated that all three genomes were highly similar in gene content and order (Chao et al. 1989). The collinearity within homoeologous chromosome 2A, 2B and 2D was conserved along the whole chromosome, except the distal region of the short arm of chromosome 2B, which was likely involved in interchromosomal translocation (Devos et al. 1993). However, our results seem to suggest that a translocation has not occurred within the PHS QTL interval because most of the ESTs located within the QTL interval were also located on all homoeologous group 2 chromosomes.

New markers were used to compare three linkage groups, 2A.2, 2B.1 and 2D. All three linkage groups contained a region homoeologous to bin 2BS1-0.53-0.75. There was a minor PHS QTL mapped on 2A.2 by Munkvold et al. (2009) suggesting that it might be homologous to the PHS QTLs on 2B.1 or 2D. However, our results indicate that the PHS QTL on 2A.2 is not homologous with the PHS QTL interval on 2B.1 but is homologous to a minor PHS QTL on 2B.1, which was located outside Wmc 474 in a distal region. Also, the PHS QTL on 2A.2 is not homologous with the PHS QTL interval on 2D. Similarly, the major PHS QTL on 2B.1 is not homologous

with the major QTL on 2D because all three new markers locate distal to the QTL on 2D. There were two markers designed from known-function genes, calmodulin binding protein and calmodulin/Ca^{2+} dependent protein kinase, located on both 2A.2 and 2B.1. The function of these two genes was reported to affect seed dormancy (Reddy et al. 2002; Sheen 1996). If these two genes are involved in PHS, then their effect would be small because the PHS QTL on 2A.2 was barely significant in only one environment at $p < 0.05$ (Munkvold 2007) but not significant at $p < 0.01$ (Munkvold et al. 2009). The eQTL mapping done by Munkvold (2007) showed that Ca^{2+} dependent protein kinase was one of the eQTL genes in the 2A.2 bin and Calmodulin was one of the eQTL genes in 3D.1. In addition, the location of marker rCaPK (designed from Calmodulin/Ca^{2+} dependent protein kinase) distal to Wmc474 matched with two rice SD QTLs, Gramene ID AQCZ006 (Jiang et al. 2003) and AQFB008 (Gu et al. 2004). These QTLs covered Calmodulin and/ or Ca^{2+} dependent protein kinase. This may suggest that calmodulin/Ca^{2+} dependent protein kinase and Calmodulin binding protein may be responsible for the PHS phenotype. Fine mapping would help to prove this suggestion.

There were three markers designed from ESTs in bin 2BS1-0.53-0.75, that mapped on chromosome 2D. These three markers revealed that the region homologous to the PHS QTL interval on 2B.1 was distal from the Wmc 453a locus, but the PHS QTL on 2D was proximal to the Wmc453a locus. It is possible that this major QTL on 2D was located in the bin close to the centromere or on the long arm of 2D, but the lack of markers in that region prevented determination of the exact location. Munkvold (2007) did not report any matching eQTLs between chromosome 2B.1 and 2D, suggesting that the PHS QTL on 2B.1 is not homologous with the PHS QTL on 2D. Interestingly, the genetic distance between new markers and wmc453a was 5-5.8 cM on 2D. This distance matches the distance from barc55 to wmc453c on 2B.1. If

the genetic distance between the two regions is similar to the physical distance in both regions on 2B.1 and 2D, then these two new markers will be useful for probing with *Ae. tauschii* BACs to reveal the gene content and gene order of the region close to Barc55.

These results indicate that the PHS interval on 2B.1 was not homologous to the PHS QTL on 2D. Interestingly, chalcone synthase and FAD-binding protein are 2 of 8 eQTL genes that were located on chromosome 6D (the location of a second major PHS QTL). Two ESTs, located in the deletion 2BS1 bin, match with these two genes. Some chromosome regions on 2BS, 6AS and 6DS were reported to be homologous. They used RFLP to study the relationship between homoeologous chromosomes. Generally, the gene order was highly conserved in homoelogous groups. For example, some probes hybridized in all three 2A, 2B and 2D but some probes, designed from ABA- induced cDNA, hybridized on 6AS, 2BS and 6DS but not on 2A and 2D (Devos 1999). Also, Devos (1999) reported again that the gene duplication between chromosome 2BS and 6BS occurred on Chinese spring nullisomic-tetrasomic line and ditelosomic line (Dt). The terminal deletion was observed for chromosome arm 2BS in Dt6BS. There may be some interaction between genes on 2B and 6D, contributing to PHS. Also it could suggest that there might be some gene duplication located in both PHS intervals that would be revealed at the DNA sequence level.

Marker development for polyploid species

The difficulty of working in a polyploid has limited progress of marker development in this project. Polyploidy, autogamy and the similarity between the parents are the limiting factors to finding the necessary polymorphism for mapping in a population (Tanksley and Nelson 1996). During this study, 22 polymorphic loci (8%) were found between the two parents, from the total 278 designed primers. Not

only parental relationship but also the types of markers that are developed limit polymorphism. The markers in this project were developed from ESTs, EST contigs and gene transcripts which exclude the 5' untranslated region (UTR) of genes. It is possible that more polymorphism will be in a 5' UTR, such as a promoter region. The information for the 5'UTR sequence was not available in wheat but already available in rice. However, it is difficult to use 5'UTR from distantly related species because this region is variable even in closely related species. However, markers developed from coding regions were more advantageous than SSR markers because the sequences in coding regions are conserved among species. Gene prediction and map-based cloning was easier for markers designed from coding regions.

Comparative mapping was useful for developing markers and predicting the candidate genes in orthologous regions of related species. This would work best at the level of macrocollinearity but smaller scale comparisons may also be possible, depending on the region. The orthologous regions between sequenced grasses, such as rice and *Brachypodium* were useful to determine the candidate genes at orthologous regions that might contribute to the PHS phenotype in wheat. The candidate genes were used to develop markers in wheat. We focused on saturating the major PHS QTL on chromosome 2B.1 in white winter wheat, by using a comparative map as the tool to develop new markers. *Aegilops tauschii* may also be useful for marker development because it already has BAC libraries. For example, SSR markers were developed for the wheat *Lr34* rust resistance region (Bossolini et al. 2006) .

Potential for Ae. tauschii BAC sequencing

There was some success using *Ae. tauschii BAC* clones to identify the genes that affect some agronomic traits. Libraries of BAC from both *Ae. tauschii* and hexaploid wheat (approximately 400 kb) at leaf rust resistance gene *Lr1* revealed that

the *LR1RGA1* gene co-segregated with the disease resistance (Qiu et al. 2007). *Ae. tauschii BAC* clone HI6P23 (approximately 145 kb), used at the gap of a *Brachypodium* BAC clone, revealed 3 genes shared by 3 species, wheat, *Ae. tauschii* and *Brachypodium*. Two of these three genes, *Mot1* and and *FtsH4* completely linked to the earliness *per se* phenotype in *T. monococcum* (Faricelli et al. 2009). In this study, sequences of an *Ae. tauschii* BAC clone (approximately 242 kb) in contig1347 may reveal numerous genes and possibly identify a candidate gene (s) that contribute to the PHS phenotype in winter white wheat.

In conclusion, the regions in rice and *Brachypodium* orthologous to the PHS QTL interval on 2B.1 were generally collinear. In addition, the new markers were developed to compare the relationships among the PHS QTLs on homoeologs of wheat group 2. Only the new markers on linkage 2B.1 were used for fine mapping in the selected recombinant population.

REFERENCES

Akhunov E, Akhunova A, Dvořák J (2005) BAC libraries of Triticum urartu, Aegilops speltoides and Ae. tauschii, the diploid ancestors of polyploid wheat. Theor Appl Genet 111:1617-1622

Anderberg RJ, Walker-Simmons MK (1992) Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinases. Proc Natl Acad Sci USA 89:10183-10187

Bethke PC, Schuurink R, Jones RL (1997) Hormonal signalling in cereal aleurone. J Exp Bot 48:1337-1356

Bossolini E, Krattinger S, Keller B (2006) Development of simple sequence repeat markers specific for the Lr34 resistance region of wheat using sequence information from rice and Aegilops tauschii. Theor Appl Genet 113:1049-1062

Cai H, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. Theor Appl Genet 104:1217-1228

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor Appl Genet 78:495-504

D'Angelo C, Weinl S, Batistic O, Pandey GK, Cheong YH, Schultke S, Albrecht V, Ehlert B, Schulz B, Harter K, Luan S, Bock R, Kudla J (2006) Alternative complex formation of the Ca-regulated protein kinase CIPK1 controls abscisic acid-dependent and independent stress responses in Arabidopsis. Plant J 48:857-872

Debeaujon I, Leon-Kloosterziel K, Koornneef M (2000) Influence of the Testa on Seed Dormancy, Germination, and Longevity in Arabidopsis. Plant Physiol 122:403-414

Devos KM, Millan T, Gale MD (1993) Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye and barley. Theor Appl Genet 85:784-792

Devos KM, Sorrells ME, Anderson JA, Miller TE, Reader SM, Lukaszewski AJ, Dubcovsky J, Sharp PJ, Faris J, Gale MD (1999) Chromosome aberrations in wheat nullisomic-telosomic and ditelosomic lines. Cereal Res Commun 27:231-239

Devos KM, Gale MD (2000) Genome Relationships: The Grass Model in Current Research. Plant Cell 12:637-646

Distelfeld A, Uauy C, Olmos S, Schlatter A, Dubcovsky J, Fahima T (2004) Microcolinearity between a 2-cM region encompassing the grain protein content locus

Gpc-6B1 on wheat chromosome 6B and a 350-kb region on rice chromosome 2. Funct Integr Genomics 4:59-66

Faricelli M, Valárik M, Dubcovsky J (2009) Control of flowering time and spike development in cereals: the earliness per se Eps-1 region in wheat, rice, and Brachypodium. Funct Integr Genomics doi:10.1007/s10142-009-0146-7

Gallego F, Feuillet C, Messmer M, Penger A, Graner A, Yano M, Sasaki T, Keller B (1998) Comparative mapping of the two wheat leaf rust resistance loci Lr1 and Lr10 in rice and barley. Genome 41:328-336

Gong D, Guo Y, Schumaker KS, Zhu JK (2004) The SOS3 family of calcium sensors and SOS2 family of protein kinases in Arabidopsis. Plant Physiol 134:919-926

Gu XY, Kianian SF, Foley ME (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (Oryza sativa). Genetics 166:1503-1516

Gu X, Kianian SF, Hareland GA, Hoffer BL, Foley ME (2005) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (Oryza sativa). Theor Appl Genet 110:1108-1118

Guo Y, Xiong L, Song C, Gong D, Halfter U, Zhu J (2002) A Calcium Sensor and Its Interacting Protein Kinase Are Global Regulators of Abscisic Acid Signaling in Arabidopsis. Dev Cell 3:233-244

Ishimaru K, Yano M, Aoki N, Ono K, Hirose T, Lin SY, Monna L, Sasaki T, Ohsugi R (2001) Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. Theor Appl Genet 102:793-800

Jiang L, Cao YJ, Wang CM, Zhai HQ, Wan JM, Atsushi Y (2003) Detection and analysis of QTL for seed dormancy in rice (Oryza sativa L.) using RIL and CSSL population. Acta Genetica Sinica 30:453-458

La Rota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. Funct Integr Genomics 4:34-46

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Lefert PS (1998) Rapid Reorganization of Resistance Gene Homologues in Cereal Genomes. Proc Natl Acad Sci USA 95:370-375

Leonova I, Laikova L, Popova O, Unger O, Börner A, Röder M (2007) Detection of quantitative trait loci for leaf rust resistance in wheat-T. timopheevii / T. tauschii introgression lines. Euphytica 155:79-86

Li AL, Zhu YF, Tan XM, Wang X, Wei B, Guo HZ, Zhang ZL, Chen XB, Zhao GY, Kong XY, Jia JZ, Mao L (2008) Evolutionary and functional study of the CDPK gene family in wheat (Triticum aestivum L.). Plant Mol Biol 66:429-443

Li W, Gill BS (2002) The Colinearity of the Sh2/A1 Orthologous Region in Rice, Sorghum and Maize Is Interrupted and Accompanied by Genome Expansion in the Triticeae. Genetics 160:1153-1162

Lin SY, Sasaki T, Yano M (1998) Mapping quantitative trait loci controlling seed dormancy and heading date in rice, Oryza sativa L., using backcross inbred lines. Theor Appl Genet 96:997-1003

Liu Q, Feng J, Buzin C, Wen C, Nozari G, Mengos A, Nguyen V, Liu J, Crawford L, Fujimura FK, Sommer SS (1999) Detection of virtually all mutations-SSCP (DOVAM-S): a rapid method for mutation scanning with virtually 100% sensitivity. BioTechniques 26:932-942

Liu S, Zhang X, Pumphrey M, Stack R, Gill B, Anderson J (2006) Complex microcolinearity among wheat, rice, and barley revealed by fine mapping of the genomic region harboring a major QTL for resistance to Fusarium head blight in wheat. Funct Integr Genomics 6:83-89

Lu G, Moriyama EN (2004) Vector NTI, a balanced all-in-one sequence analysis suite. Briefings in Bioinformatics 5:378-388

Mahajan S, Sopory SK, Tuteja N (2006) Cloning and characterization of CBL-CIPK signalling components from a legume (Pisum sativum). FEBS J 273:907-925

Mallard S, Gaudet D, Aldeia A, Abelard C, Besnard AL, Sourdille P, Dedryver F (2005) Genetic analysis of durable resistance to yellow rust in bread wheat. Theor Appl Genet 110:1401-1409

Manly KF, Cudmore, Jr.,Robert, H., Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. Mammalian Genome 12:930-932

Miura K, Lin Y, Yano M, Nagamine T (2002) Mapping quantitative trait loci controlling seed longevity in rice (Oryza sativa L.). Theor Appl Genet 104:981-986

Munkvold J, Tanaka J, Benscher D, Sorrells M (2009) Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. Theor Appl Genet 119:1223-1235

Munkvold JD (2007) Mapping of Phenotypic and Gene Expression QTL related to preharvest sprouting and seed dormancy in white winter wheat. Thesis (Ph.D.) edn. Cornell University

Ok SH, Jeong HJ, Bae JM, Shin JS, Luan S, Kim KN (2005) Novel CIPK1-associated proteins in Arabidopsis contain an evolutionarily conserved C-terminal region that mediates nuclear localization. Plant Physiol 139:138-150

Peng JH, Zadeh H, Lazo GR, Gustafson JP, Chao S, Anderson OD, Qi LL, Echalier B, Gill BS, Dilbirligi M, Sandhu D, Gill KS, Greene RA, Sorrells ME, Akhunov ED, Dvorak J, Linkiewicz AM, Dubcovsky J, Hossain KG, Kalavacharla V, Kianian SF, Mahmoud AA, Miftahudin, Conley EJ, Anderson JA, Pathan MS, Nguyen HT, McGuire PE, Qualset CO, Lapitan NL (2004) Chromosome bin map of expressed sequence tags in homoeologous group 1 of hexaploid wheat and homoeology with rice and Arabidopsis. Genetics 168:609-623

Qiu J, Schürch A, Yahiaoui N, Dong L, Fan H, Zhang Z, Keller B, Ling H (2007) Physical mapping and identification of a candidate for the leaf rust resistance gene Lr1 of wheat. Theor Appl Genet 115:159-168

Reddy VS, Ali GS, Reddy ASN (2002) Genes Encoding Calmodulin-binding Proteins in the Arabidopsis Genome. J Biol Chem 277:9840-9852

Scarth R, Law CN (1983) The location of the photoperiod gene, Ppd2 and an additional genetic factor for ear-emergence time on chromosome 2B of wheat. Heredity 51:607-619

Sheen J (1996) Ca2+-dependent protein kinases and stress signal transduction in plants. Science 274:1900-1902

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

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen HT, Peng J, Lapitan NL, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi DW, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818-1827

Sunnucks P, Wilson AC, Beheregaray LB, Zenger K, French J, Taylor AC (2000) SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. Mol Ecol 9:1699-1710

Takeuchi Y, Lin SY, Sasaki T, Yano M (2003) Fine linkage mapping enables dissection of closely linked quantitative trait loci for seed dormancy and heading in rice. Theor Appl Genet 107:1174-1180

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191-203

Wan JM, Cao YJ, Wang CM, Ikehashi H (2005) Quantitative Trait Loci Associated with Seed Dormancy in Rice. Crop Sci 45:712-716

Woodger F, Millar A, Murray F, Jacobsen J, Gubler F (2003) The Role of GAMYB Transcription Factors in GA-Regulated Gene Expression. J Plant Growth Regul 22:176-184

Worland AJ, Korzun V, Röder MS, Ganal MW, Law CN (1998) Genetic analysis of the dwarfing gene Rht8 in wheat. Part II. The distribution and adaptive significance of allelic variants at the Rht8 locus of wheat as revealed by microsatellite screening. Theor Appl Genet 96:1110-1120

Worland AJ, Law CN, Shakoor A (1980) The genetical analysis of an induced height mutant in wheat. Heredity 45:61-71

Wu S, Pumphrey M, Bai G (2009) Molecular Mapping of Stem-Rust-Resistance Gene Sr40 in Wheat. Crop Sci 49:1681-1686

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci USA 100:6263-6268

CHAPTER 3

FINE MAPPING OF A PREHARVEST SPROUTING QTL INTERVAL ON CHROMOSOME 2B IN WHITE WINTER WHEAT

INTRODUCTION

Seed dormancy (SD) in wheat (*Triticum aestivum*) is an important agronomic trait. Lack of dormancy results in preharvest sprouting (PHS). PHS in wheat is the condition where germination of grains occurs on the spike before harvest. Prolonged rainfall and high humidity contribute to premature germination of grains. The main effects of PHS are a lower yield due to harvest losses and, more importantly, a reduction in end-product quality. Resistance to PHS is a quantitative trait that is affected by genotype, environment and genotype by environment interaction (Anderson et al. 1993; Imtiaz et al. 2008; Zanetti et al. 2000). Consequently, PHS resistance improvement is difficult to achieve because of the influence of environmental conditions during the harvest season, such as rainfall, humidity and temperature.

Fine mapping has been used to narrow the Quantitative Trait Loci (QTL) intervals sufficiently for physical mapping and map-based cloning procedures and to determine if a QTL for some traits have pleiotropic effects. It also helps identify precise markers for marker-assisted selection (MAS). In tomato, fine mapping in selected overlapping recombinants narrowed down the QTL region contributing to agronomic traits to as little as 3 cM and separated the phenotypic effect of the close linked QTL from pleiotropic effects (Paterson et al. 1990). In rice, studies have reported that the significant correlation between heading date and SD might be due either to a pleiotropic effect of one gene or two tightly linked genes (Lin et al. 1998). However, in a later study fine mapping indicated that the correlation between heading

date and seed dormancy QTLs on rice chromosome 3 is due to linkage (Takeuchi et al. 2003). In durum wheat, fine mapping provided a more precise location of grain protein content (GPC) within the QTL interval on chromosome 6B. Fine mapping was accomplished by adding new markers and mapping on recombinant substitution lines (RSLs) that contained additional crossover events in that region. The QTL interval was narrowed to 2.7 cM (Olmos et al. 2003) and then narrowed to 0.3 cM. Five candidate genes were identified within a collinear 64 kb region of rice chromosome 2 (Distelfeld et al. 2004). Later in this interval a single gene was found to control wheat grain protein, zinc and iron content pleiotropically (Uauy et al. 2006). Some traits such as yield and protein content have been shown to be inversely related. The negative relationship might be due to very tight linkage or to pleiotropy (Chung et al. 2003). Fine mapping of additional recombinants could show a reduction of the inverse phenotypic correlation as linkage is broken. In soybean protein content is inversely related to yield and oil content. A seed protein QTL controlling protein content, grain yield and oil content was fine mapping in BC4 and BC5 down to a 3 cM interval. However, this study was unable to determine if the protein QTL controls yield and oil content through pleiotropy (Nichols et al. 2006).

For fine mapping to be successful, markers must saturate the QTL region. Comparative mapping among the grass species can be used as a tool to develop new markers for fine mapping. Collinearity between wheat, barley and rice over a 52 kb region was used to increase density of markers and reduce the genetic interval to around 0.3 cM at the *SKr* gene region controlling crossability between wheat and rye (Alfares et al. 2009). Five markers were located in a 0.3 cM region. A physical map was later constructed based on a 400 kb wheat BAC contig on chromosome 5BS. The markers that were developed indicated that the *SKr* gene region was close to *GSP-1* (Grain softness protein) and 5 other genes, excluding *GSP-1*, were found in that

region. In another fine mapping study of wheat (Röder et al. 2008) the grain weight QTL (*gw1*) on chromosome 7D, explained 84.7% of phenotypic variation. The interval was narrowed to 7.6 cM and cosegregated with a plant height reducing locus (*Rht),* which explained 70.7% of the phenotypic variation. However, it was not possible to determine if it was a case of pleiotropy or closely linked genes (Röder et al. 2008). The SD QTL interval on rice chromosome 12 was narrowed to less than 75 kb, containing three genes, PIL5 (phytochrome-interacting factor3-like 5), hypothetic protein and bHLH (basic helix loop helix) DNA binding domain containing protein (Gu et al. 2009).

In addition to dense markers, appropriate populations for fine mapping must be used for the trait to be mapped as a single Mendelian factor if cloning is to be achieved. The population should contain enough break points in recombinant lines and recombinant population size should be large enough for sufficient statistical power to detect associations between the trait and the markers. The population can be recombinant inbred lines (RILs), nearly isogenic lines (NILs), advanced backcrossed lines or chromosome segment substitution lines (CSSLs). Nearly isogenic lines have been used to narrow down a QTL, such as a SD QTL (SD1) in barley (Han et al. 1999), a major fruit weight QTL in tomato (Alpert and Tanksley 1996) and a heading date QTL (Hd6) in rice (Yamamoto et al. 2000). NILs have the advantage of uniform genetic background except in the targeted region. However, population development is very time-consuming and many individuals are needed. For example, to have a 95% recovery rate of the target region, a 0.1 cM interval would require approximately 3000 individuals (Tanksley 1993). Alternatively, advanced backcrossed lines could decrease the time for development. The populations such as BC3 and BC4 have been used successfully in many fine mapping studies of rice (Takeuchi et al. 2003; Yamamoto et al. 1998), maize (Vladutu et al. 1999) soybean (Nichols et al. 2006), wheat (Röder et

al. 2008), and tomato (Paterson et al. 1990). However, the disadvantage of an advanced backcrossed backcross line is that it contains less uniform genetic background than NILs.

Once a given QTL can be mapped as a single Mendelian factor, chromosome walking and landing methods can be used to identify the genes underlying QTLs. After the candidate genomic region has been defined, several molecular approaches, such as sequencing of the candidate region, gene prediction, expression profiling, and genetic complementation of the candidate gene, can be applied for identification of genes (Yano 2001). Map-based cloning in rice has been done for genes of major agronomic traits, such as heading date, submergence tolerance, seed number, salt tolerance and seed shattering (Yamamoto and Yano 2008) as well as seed dormancy (Sugimoto et al. 2010) .

QTL studies of PHS and SD phenotypes revealed genetic regions contributing to these traits. The PHS and SD QTLs coincided in many studies, especially the major SD and PHS QTLs for both traits (Tan et al. 2006; Mares et al. 2005; Ogbonnaya et al. 2008; Munkvold et al. 2009). In previous work, PHS QTL mapping utilized a doubled haploid (DH) white winter wheat population from a cross between Cayuga and Caledonia. Approximately 15 different PHS QTLs, located across the whole wheat genome, were detected at $p < 0.05$ in at least one environment (Munkvold et al. 2009). The major QTLs were located on chromosomes 2B.1, 2D, 3D.1 and 6D and were significant ($p < 0.01$) in four or more environments. All PHS QTL resistance alleles were contributed by Cayuga. The PHS QTL on 2B.1 was significant ($p < 0.01$) in all 16 environments and explained approximately 24% of the phenotypic variation. The QTL interval was broad, ranging from 2 to 31 cM across environments. The QTL peak for the PHS QTL mean from all environments was at 14 cM between, the SSR markers Barc55 and Wmc474 on chromosome 2B (Munkvold et al. 2009). The

comparative map from the previous chapter represented the major PHS QTL interval from the 2 flanking markers Gwm429 and Wmc474, whereas Barc55 was located between these 2 markers. Based on PHS QTL data for 2B.1 on 209 DH lines, the QTL interval covered nearly the entire 2B.1 deletion. The broad QTL interval could result from either a major QTL or multiple closely linked smaller QTLs. The QTL mapping resolution was limited to approximately 10-20 cM, which is inadequate for determine if there is one major QTL or multiple smaller, closely linked QTL in the interval. If the QTL is the result of two or more linked genes, fine mapping can be resolved by allowing sufficient recombination to occur (Tanksley 1993). Fine mapping of PHS QTL requires comparing PHS means of individual recombinants with those of non-recombinants. A high proportion of recombinant individuals can provide the statistical test with sufficient power to resolve multiple linked QTLs. In this study, fine mapping in recombinant populations for the 2B.1 was used to determine if there is one or more linked QTLs.

In addition to fine mapping, the origin of Cayuga PHS allele on 2B.1 interval was identified in this study. The Cayuga PHS resistance allele on 2B was assumed to come from Clark's Cream. However, a previous experiment in our lab showed that Clark's Cream has three different alleles at Barc55, from different sources. This could suggest that Clark's Cream was a heterogeneous variety. Clark's Cream, hard white winter wheat, was developed by Mr. Earl G. Clark before 1952 and it exhibits a high level of preharvest sprouting resistance (Heyne 1956). The pedigree of Clark's Cream was KanKing/Golden 50 or Golden Chief, all hard red winter wheat varieties (Upadhyay et al. 1988). A study of the relationship between Cayuga and 14 Clark's Cream-related lines would help to determine the origin of the Cayuga PHS allele on chromosome 2B. Blackhull was the progenitor of most of the Clark's Cream-related lines (Heyne 1956).

The objectives of this study were to

1) develop fine mapping populations that contain additional recombinants at the major PHS QTL interval on chromosome 2B.1, in a reasonable amount of time by using flanking marker-assisted selection.

2) identify the precise location of the flanking markers that are closely linked to the major PHS QTL on chromosome 2B.1 by fine mapping.

3) confirm that the major PHS QTL is collocated with the SD QTL by fine mapping.

4) determine which Clark's Cream-related lines contributed the PHS allele of Cayuga at the PHS interval on 2B.1.

MATERIAL AND METHODS

Development of populations for fine mapping

In this study, the mapping populations used for fine mapping were selected homozygous recombinant BC1F4 and BC1F5 lines. They were developed from a cross between Cayuga and Caledonia, followed by a backcross to Caledonia. Cayuga is a PHS resistant variety derived from a cross between Geneva (PHS susceptible) and Clark's Cream (PHS resistant), which was then backcrossed to Geneva. Caledonia is a PHS susceptible variety selected as an off-type from Geneva. Cayuga was used as the female plant and Caledonia was used as the male plant. Population development was started by crossing Cayuga and Caledonia in a greenhouse. The F1 plants were grown in the field and used for doubled haploid (DH) development using the maize pollinator method. The DH lines were developed by Thompson Limited, Ontario, Canada. The selected DH individuals that contained high percentage of Caledonia background (63.5-72.6%), but with the Barc55 region within the PHS QTL on 2B.1 coming from Cayuga, were chosen to backcross with Caledonia. The DH # 86, that was used to produce backcross family 04173, contained about 63.5 % Caledonia background. The

DH # 95, which was used to produce backcross family 04175, contained about 64.5% Caledonia background. The DH # 87, which was used to produce backcross family 04174, contained about 72.6% Caledonia background. One selected DH individual (DH # 129) that contained a high percentage of Cayuga background, but with Barc55 within the PHS QTL on 2B.1 from Caledonia, was chosen to cross with Cayuga. DH #129 was used to produce backcross family 04176, containing around 60.8% Cayuga background. BC1F1 of the backcrossed DH lines with Caledonia and Cayuga generated 13 families of BC1F2. All 13 families were grown in the field in separate rows designated 04173-1, 04173-2, 04175-0, 04176-1, 04176-2, 04174-1, 04174-2, 04174-3, 04174-4, 04174-5, 04174-6, 04174-7 and 04174-8. Almost all families contained Caledonia background, except 04176-1 and 04176-2, which contained Cayuga background. Two spikes of all individual plants in each family were collected to produce 1087 BC1F3. The number of individual plants varied in each family. There were 129 individual plants from family 04173-1, 108 plants from family 04173-2, 95 plants from 04175-0, 110 plants from 04176-1, 20 plants from 04176-2, 100 plants from 04174-1, 78 plants from 04174-2, 83 plants from 04174-3, 70 plants from 04174-4, 83 plants from 04174-5, 70 plants from 04174-6, 50 plants from 04174-7, and 91 plants from 04174-8. All BC1F3s were grown in the field location Snyder, Ithaca, NY during the fall 2006. All BC1F3s were screened for heterozygous and homozygous recombinants containing a break point at the PHS QTL on 2B.1, by using 3 flanking markers, Wmc474, Barc55 and Gwm429. All recombinant plants were used to produce 1847 BC1F4. All BC1F4s were grown at two locations; Synder and Ketola, Ithaca, NY, during the fall of 2007.

For fine mapping of the BC1F4 population, homozygous recombinants containing the break point at the PHS QTL on 2B.1 were selected by screening all 1847 BC1F4 individuals, using the three flanking markers. More additional

homozygous recombinants were needed in BC1F5, so both selected heterozygous and homozygous recombinants were grown at two locations, Synder and Ketola, Ithaca, NY, during the fall of 2008. For fine mapping of the BC1F5 population, homozygous recombinants containing recombinations in the PHS QTL interval on 2B.1 were selected by screening all 925 BC1F5 individuals with the same three flanking markers. BC1F4 and BC1F5 recombinants that we used for fine mapping contained Caledonia background. The number of recombinants that contained Cayuga background was not sufficient for fine mapping.

DNA isolation

All markers used in this project were PCR-based primers. All populations, BC1F3, BC1F4 and BC1F5 grown in the field were used for DNA extraction. After planting in the field for 3 weeks, 5 leaves of different plants in the same rows were collected for each family number. Approximately 0.8-1.0 cm of each leaf was collected. Leaf samples from one row were added to each well of a 96 well plate containing 0.25 grams of silica gel in each well. The leaf samples were dried for 5-10 days at room temperature until dry. The dried leaf samples were ground by a homogenizer (TALBOYS). Several rounds of shaking were used to make sure that the samples were completely ground. All 96 well plates containing ground tissues and silica gel were further used in small-scale DNA extraction. The method of DNA extraction used in this project was modified from other small-scale extractions. A volume of 500 µl of the extraction buffer (0.1 M Tris-HCl pH 7.5, 0.05 M EDTA pH 8, 1.25% (w/v) SDS) was added to the ground samples, mixed by vortexing before incubation at 65 °C for 30 min with occasional shaking every 15 min. The solution was kept refrigerated for 15 min until cool. Samples were centrifuged at 4,000 rpm by a table centrifuge (Eppendorf Centrifuge 5810) and then the supernatant was

transferred to the new 96 well plates. The supernatant was combined with 250 µl of cold 5M potassium acetate containing acetic acid solution (60% 5M potassium acetate, 11.5% acetic acid and 28.5% distilled water). The solution was refrigerated for 15 min after vigorous shaking. The solution was centrifuged for 25 min and the supernatant was transferred to new 96 well plates. The supernatant was added with 1 volume of chloroform: isoamyl alcohol (24:1 ratio) to precipitate protein and then it was shaken vigorously. The solution was centrifuged for 20 min at 4,000 rpm. The supernatant was transferred to new 96 well plates and then 1 volume of isopropanol was added to precipitate DNA and left at -20°C overnight. The solution was centrifuged at 4,000 rpm for 25 min. The DNA pellet was dissolved in 200 µl TE. The dissolved solution was added with 3 M Sodium acetate pH 5.2 and 2 volumes of 95 % ethanol, to precipitate DNA again. After mixing the solution and leaving at room temperature at 15 min, the solution was centrifuged at 4,000 rpm for 20 min. The pellet was washed with 2 volumes of 70% ethanol and centrifuged for 10 min. The dried DNA pellet was re-suspended in 200 µl of TE buffer.

DNA genotyping

Sixteen markers were used for fine mapping, including six SSR markers (Barc55, Wmc474, Gwm429, Cdo64PCR, Wmc453c and Gwm319) and 10 newly designed primers based on ESTs on the 2B.1 PHS interval (ss57.2, ss66.2, ss3.2, ss4.4b, ss44, ss26.3, ss47.5, ss20b, ss31 and rCaPK) that were described in the previous chapter.

1) Screening for homozygous recombinants using flanking markers

Three co-dominant markers, Barc55, Wmc474 and Gwm429, were used to screen heterozygous and homozygous recombinants. There were 1,087 individuals screened from the BC1F3 population, by selecting heterozygous and homozygous

recombinants, using Barc55 primer with an additional 18 base pair M13 sequence (5'TGTAAAACGACGGCCAGT3') at the 5' end of the Barc55 forward primer (Barc55F-M13). The PCR reaction was similar to general PCR, except for the addition of 10 pmole of M13 dye. The amount of Barc55F-M13 was 4 times less than the reverse primer, allowing for the forward primer to be expended before binding with the fluorescence dye during the dye binding cycle. The PCR cycles were 30 cycles of normal PCR and 7 additional cycles at 53 °C annealing temperature for dye binding. For BC1F3 genotype screening, the PCR product amplified by Wmc474 and Gwm429 was run on a 4% PAGE gel. For the 1847 BC1F4 and 925 BC1F5 individual screening, all three flanking markers were labeled directly, each with a different fluorescence color. The PCR cycles and reaction for three fluorescence-labeled markers was the same as in a general PCR reaction. The DNA fragments that were amplified from both markers labeled with M13 and directly labeled fluorescence markers were analyzed by using a fluorescence-detection system at Cornell University Life Science Core Laboratories Center. The results were visualized on Peak Scanner ™ Software v.10 (http://marketing.appliedbiosystems.com/mk/get/PS1_login). After screening by 3 flanking markers, the homozygous recombinants were used in genotyping for fine mapping.

The initial population number and resulting screening population were described in Table 3.1. The three flanking markers used, Wmc474, Barc55 and Gwm429, were co-dominant SSR markers. The BC1F3 population, containing 1087 individuals, was the first population to be screened resulting in 119 families (11%) out of 1087 BC1F3 families that were heterozygous and homozygous recombinants. For

Table 3.1 Summary of population sizes of BC1F3, BC1F4 and BC1F5 used in fine mapping population development.

The population was developed in both populations that containing Cayuga and Caledonia background. Only Caledonia background was used for fine mapping.

	Initial Population Size			non-recombinants			Heterozygous recombinants		Homozygous recombinants	
	Cayuga Background (Cay-BG)	**Caledonia Background (Cal-BG)**	Total	Cay-BG	**Cal-BG**	**Missing**	Cay-BG	**Cal-BG**	Cay-BG	**Cal-BG**
BC1F3	130	**957**	1087	121	**841**	**6**	5[a]	**72[a]**	4[a]	**38[a]**
BC1F4	131	**1716**	1847	91	**633**	**18**	32	**706**	8	**359[b]**
BC1F5	10	**915**	925	0	**391**	**12**	3	**42**	7	**470[c]**

a = The heterozygous and homozygous recombinant BC1F3 that were used to produce the 1847 BC1F4
b = The homozygous recombinant BC1F4 that were used for fine mapping
c = The homozygous recombinant BC1F5 that were used for fine mapping

Caledonia background, 4% were homozygous recombinants and 7.5% were heterozygous recombinants. For the next population development, in order to increase recombinants without increasing the size of the population that was used for fine mapping, only heterozygous and homozygous recombinant BC1F3 families were used to produce 1847 BC1F4 individuals. After screening by flanking markers, 1065 of 1847 individuals were heterozygous and homozygous recombinants for Caledonia background. Seven hundred and six of 1847 individuals were heterozygous BC1F4 recombinants and 359 out of 1847 were homozygous BC1F4 recombinants that were used for fine mapping. For BC1F5 development, not all heterozygous and homozygous BC1F4 recombinant families were used. The selection was based on both genotyping and families of BC1F4. Two BC1F4 families that were developed from the same BC1F3 family, with the same genotype, were selected from heterozygous and homozygous recombinants. For BC1F5 population development, five individual plants of each selected BC1F4 family were used. One hundred families out of 706 heterozygous BC1F4 recombinants containing Caledonia background were selected. Also, the selection of homozygous recombinants was based on the fine mapping results of 359 BC1F4 recombinants. Eighty four families out of 359 homozygous BC1F4 recombinants, with recombination between Barc55 and Wmc453c as well as between Barc55 and Wmc474, were selected for the BC1F5 population. In summary, 184 families of both heterozygous and homozygous BC1F4 recombinants were used to produce 915 BC1F5 individuals. As a result, 470 homozygous BC1F5 recombinants containing Caledonia background were used further for fine mapping at the PHS QTL interval on 2B.1.

Homozygous BC1F5 recombinants containing Cayuga background were not used for fine mapping analysis because of a lack of recombinants. Fine mapping analysis was conducted on the population containing Caledonia background only.

2) Selected recombinant population genotyping

Fifteen markers were used for fine mapping in 359 homozygous BC1F4 recombinants containing Caledonia background. Additional markers used in this step, from flanking markers, were Cdo64PCR, Wmc453c, Gwm319, ss57.2, ss66.2, ss3.2, ss4.4b, ss44, ss26.3, ss47.5, ss20b, ss31 and rCaPK. Fine map genotyping of homozygous BC1F5 recombinants was based on the result of BC1F4 fine mapping. So, six markers; Gwm429, Wmc453c, Barc55, ss26.3, Wmc474 and rCaPK, were used in BC1F5 fine mapping.

3) PCR reaction and PCR product detection

Approximately 5-10 pmole of genomic DNA was used for PCR reaction. The annealing temperatures of some markers were different. Barc55 annealed at 55 °C, Wmc474 at 56 °C, Gwm429 at 50 °C, Cdo64PCR at 56 °C, Wmc453c at 61 °C and Gwm319 at 55 °C. The PCR products were amplified by SSR markers and detected on a 4% PAGE gel, except the three flanking markers. The annealing temperature of newly designed markers and the types of detection gels were described in Chapter two. There were three types of PCR machines used. MJ Research PTC-225 Peltier Thermal Cycler and BIO-RAD C1000^{TM} Thermal Cycler for 384 well plates and BIO-RAD MyCycler^{TM} Thermal cycler for 96 well plates. Running on SSCP gels was conducted overnight, from 16-19 hrs, depending on the PCR product size. The general time was 18 hrs at 3 voltages for one gel.

PHS phenotyping

The 13 families in the BC1F2 were planted at Caldwell field, Ithaca, NY during the fall of 2005. Each family was grown in a long row. PHS phenotyping was done during the summer of 2006. 1087 BC1F3 individuals were planted at Synder, Ithaca, NY during the fall of 2006, in 2 meter rows with a family-based planting. A few non-

recombinant families were phenotyped for PHS during the summer of 2007 as controls. During fall of 2007, 1847 BC1F4 individuals were planted at Snyder and Ketola fields, Ithaca, NY, in 2 meter rows with a family-based planting. Only 367 BC1F4 recombinants and some BC1F4 non-recombinants were phenotyped during summer of 2008. During the fall of 2008, 925 BC1F5 individuals were planted at two locations, Snyder and Ketola, Ithaca, NY, in one meter rows with a randomized block design. Only 477 BC1F5 recombinants and some non-recombinants were phenotyped during the summer of 2009.

For phenotyping, five spikes were sampled on the same day for BC1F2, BC1F3 and BC1F4, at each location. Because of climatic conditions, the BC1F5 spikes had a wider range in maturity, thus requiring two harvests on different days in each location. The spikes were determined to be at physiological maturity by loss of green color from the glumes. The spikes were dried indoors for five days at ambient humidity and temperature. Three additional spikes were collected for a BC1F4 germination test. After drying for five days, the phenotyping was conducted as described by Anderson (1993). For germination testing, three additional spikes were dried for five days then stored at -20 °C to preserve dormancy until the germination test was conducted.

SD phenotyping

The germination test was used to measure SD in 359 BC1F4 recombinants at one location, Ketola 2008. The spikes were stored at -20 °C for 8 months. The primary and secondary seeds from the middle two-thirds of the spike were used for the germination test. Seeds having blackpoint, disease, or shriveling were discarded. Twenty seeds from each spike were sterilized by soaking in 10% Clorox® bleach and rinsing with sterile water 5 times. The sterilized seeds were placed crease-down on a 9

cm diameter germination paper soaked with 3 ml of sterilized water in Petri plates. The plates were placed in plastic bags to retain humidity and prevent drying and stored at 22 °C in the dark. Germinated seeds were evaluated every 24 hours for 10 days by looking at radicle protrusion. All germinated seeds were discarded. The seeds contaminated with fungus during germination were discarded and excluded from calculations. The germination test consisted of 3 replicates, each with seed from a different spike. The germination index (GI) was calculated by the following method: GI = (10 x g1 + 9 x g2 +…….+1 x g10)/ (# of viable seeds), where g is the number of seeds that germinated on the given day (Munkvold et al. 2009).

Fine mapping analysis

1) Statistical data analysis

PHS scores were compared for non-recombinant families, containing Barc55 and flanking markers for Cayuga and Caledonia alleles at the 2B.1 PHS interval. Also, homozygous recombinant families containing different recombination break points at the 2B.1 PHS interval were evaluated for PHS. The t-test of the JMP analysis from JMP statistical software package (http://www.jmp.com/) was used to determine the level of significance and to determine phenotype distribution of recombinants that were used for fine mapping.

2) Linkage map construction

All markers used in this project were located in the same PHS QTL interval on 2B.1 based on linkage analysis in the Cayuga x Caledonia population. Fine mapping of BC1F4 recombinants increased the distance between some markers due to an increase in recombination. The recombinants contained break points between Gwm429 and Barc55 and between Barc55 and Wmc474. MapManager QTXb20 (Manly et al. 2001) was used to determine the marker order and a new linkage map

was created because only selected homozygous recombinants were used in this analysis. The MapManager setting was Kosambi mapping function, with linkage threshold significance of $p < 0.001$. Linkage evaluation was set at self RI because only homozygous recombinants were used.

3) QTL analysis

The QTL analyses used the software package Windows QTL Cartographer (Version 2.5 http://statgen.ncsu.edu/qtlcart/WQTLCart.htm) to evaluate the new PHS QTL data from BC1F4 recombinants. Results from the t-test of the JMP analysis were used to confirm the results from this analysis. Only the t-test was used to analyze homozygous BC1F5 recombinants. For QTL Cartographer, composite interval mapping (CIM) with model 6 was used. Precise selection walking speed (cM) was 0.5. The significance threshold was calculated at $p < 0.01$ by 1,000 permutations.

Genotyping Clark's Cream- related lines

There were 18 wheat lines used for genotyping; Caledonia, Geneva, Cayuga, Clark's Cream and Clark's Cream-related lines; KanKing, Golden 50, Golden Chief, PI 520756, PI520757, PI520758, PI 520759, PI 520760, ChiefKan, KanQueen, Clarkan, Redchief, Comanche and Blackhull. After growing seeds for one week, five leaves in each line were collected for DNA extraction. Mini scale DNA extraction was similar with previous step, except using liquid nitrogen instead of silica gel in this step. Six markers at the PHS interval on 2B.1 were used for genotyping. Those markers were Gwm429, Wmc453c, Barc55, ss26.3, Wmc474 and rCaPK. The phylogenetic tree relationship was constructed by using TASSEL program (http://www.maizegenetics.net/). The PHS phenotype of those lines was evaluated in year 2006 in one location for Clark's Cream related lines but 4 locations for the rest of the lines. The PHS mean score across locations were evaluated by using BLUP value

calculated by Fit Model in JMP statistical software package (http://www.jmp.com/). T-test in JMP analysis was used to compare the significant BLUP-PHS score value in each allele at markers Wmc453c, Barc55, Wmc474 and rCaPK associated with PHS resistance.

Barc55 allele relationship in white winter wheat

PHS haplotype was evaluated only at Barc55 by using 208 white wheat lines. The genotyping at Barc55 was provided by Zongyun Feng (Personal communication). The PHS phenotype of those lines was evaluated for 7 years, 2003-2009 but not all lines were scored the PHS phenotype from 2003-2007. The PHS mean score across years were evaluated by using BLUP value calculated by Fit Model in JMP statistical software package (http://www.jmp.com/). However, only BLUP value for 166 lines was used for Barc55 allele evaluation in this study because some lines were heterozygous and some lines contained missing data. T-test in JMP analysis was used to compare the significant BLUP-PHS score value in each Barc55 alleles.

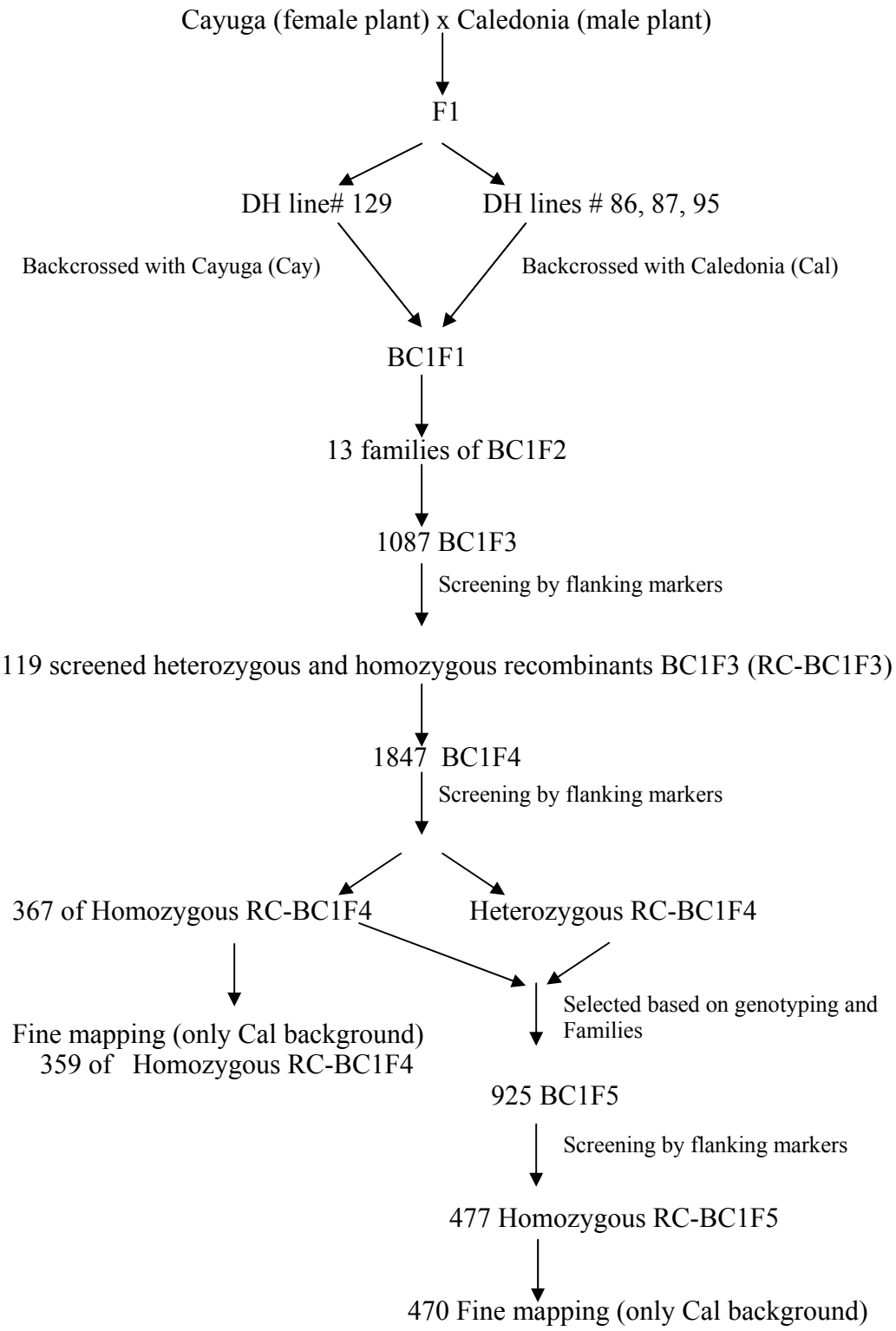
RESULTS

Development of populations for fine mapping

The development of populations for fine mapping in this project is diagramed in figure 3.1. After first backcrossing selected DH individuals with Caledonia and Cayuga (BC1), the background of BC1 families in this project contained more than a normal BC1 ranging from 81.75 %to 86.3% Caledonia background. This background was almost comparable to a BC2, which would average 87.5% of the recurrent parent background. Almost all BC1F2 families contained Caledonia background (88% of BC1F2). PHS score means of families containing Caledonia background ranged from 3.52 in family 04174-6 to 5.53 in family 04174-8 with a mean of 4.53. The PHS score mean was low in two families that contain Cayuga background, ranging from 1.94, in

Figure 3.1. Population development for fine mapping.

The populations used for fine mapping were screened for homozygous BC1F4 and BC1F5 recombinants**.** After developing the double haploid (DH) that were used in QTL mapping (Munkvold et al 2009), three selected DH lines that contained high background of Caledonia were backcrossed with Caledonia and one selected DH that contained high background of Cayuga were backcrossed with Cayuga. The recombinant screening was started at BC1F3 generation by selecting both heterozygous and homozygous recombinants resulting from genotyping with three flanking markers; Gwm429, Barc55 and Wmc474. These recombinants were used to produce 1087 BC1F4 individuals. For fine mapping in this generation, only homozygous recombinants were used. After screening by three flanking markers, 359 BC1F4 contained Caledonia background were homozygous recombinants. After narrowing the QTL region in homozygous BC1F4 recombinants, more recombinants in BC1F5 were needed at the new interval region. Both selected homozygous and heterozygous BC1F4 recombinants were used to produce 925 BC1F5 individuals. After screening with three flanking markers, 470 BC1F5 were found to be homozygous recombinants that were used for fine mapping analysis.


Cayuga (female plant) x Caledonia (male plant)
F1
DH line# 129
DH lines # 86, 87, 95
Backcrossed with Cayuga (Cay)
Backcrossed with Caledonia (Cal)
BC1F1
13 families of BC1F2
1087 BC1F3
Screening by flanking markers
119 screened heterozygous and homozygous recombinants BC1F3 (RC-BC1F3)
1847 BC1F4
Screening by flanking markers
367 of Homozygous RC-BC1F4
Heterozygous RC-BC1F4
Fine mapping (only Cal background)
359 of Homozygous RC-BC1F4
Selected based on genotyping and Families
925 BC1F5
Screening by flanking markers
477 Homozygous RC-BC1F5
470 Fine mapping (only Cal background)

family 04176-2, to 2.54, in family 04176-1 The PHS mean score of all families containing Cayuga background was 2.24 (Figure 3.2).

Individual plants in BC1F2 families were used to produce 1087 BC1F3 families. Three flanking markers, Wmc474, Barc55 and Gwm429 were used to screen for heterozygous and homozygous recombinants containing a break point at PHS QTL on 2B.1. There were 46 BC1F3 families shown to be recombinant between Barc55 and Wmc474. Forty-two BC1F3 families contained Caledonia background and 4 families contained Cayuga background. There were 75 BC1F3 families shown to be recombinant between Barc55 and Gwm429. Seventy BC1F3 families contained Caledonia background and 5 families contained Cayuga background. Of those recombinants, two families were recombinants between both Barc55 and Wmc474 and Barc55 and Gwm 429 resulting from double crossover. As a result, 119 BC1F3 families were either heterozygous or homozygous recombinants.

Approximately 15 individual plants of these recombinant BC1F3 families were selected to produce 1847 BC1F4 individuals. For fine mapping, 367 homozygous recombinants of BC1F4 were identified from all BC1F4 individuals. Almost all were Caledonia background except 8 homozygous recombinants in Cayuga background. Because the few recombinants contained Cayuga background, only 359 BC1F4 Caledonia families were used for QTL analysis and fine mapping. Some additional recombinants that contained different break points were identified in BC1F5. Both homozygous and heterozygous recombinants containing different recombinant break points were selected to produce 925 BC1F5 individuals. For fine mapping, 477 homozygous recombinants of BC1F5 were selected from screening all BC1F4 individuals. 470 BC1F5 individuals containing Caledonia background were used further for fine mapping.

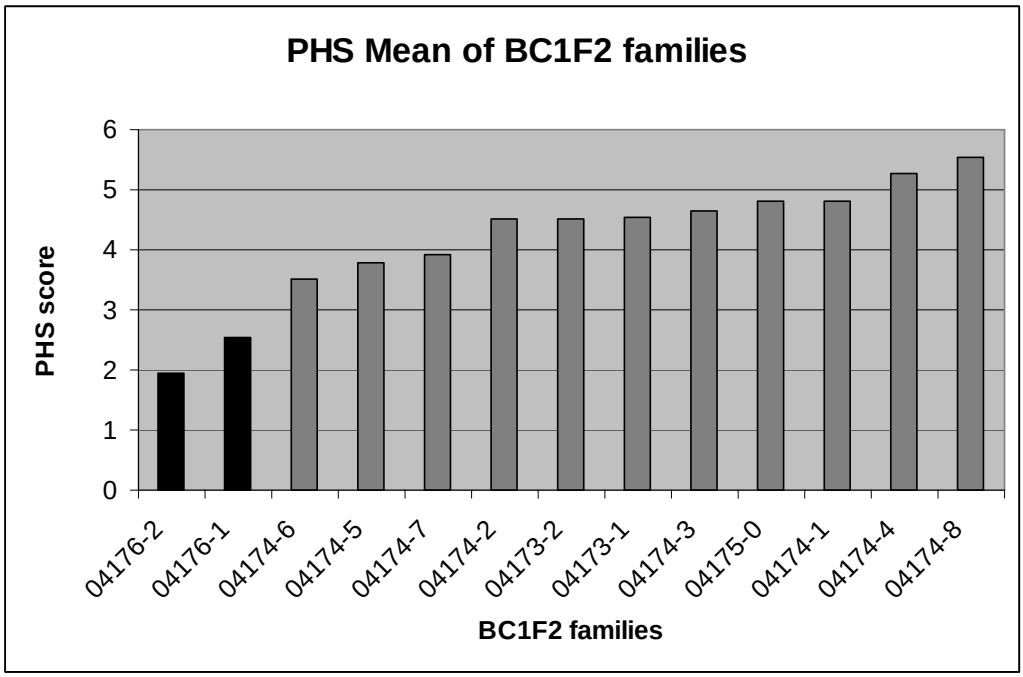


Figure 3.2. PHS mean of BC1F2 families that were used further to develop BC1F3. The black bars represent the families containing Cayuga background and the gray bar represents the families containing Caledonia background. The PHS mean score of all families containing Caledonia background is 4.53. The PHS mean score of all families containing Cayuga background is 2.24

Fine mapping analysis

QTL effect on preharvest sprouting and seed dormancy in non-recombinant groups

The PHS or SD QTL effect from the 2B interval was determined by comparing the phenotype mean between non-recombinant families containing Cayuga alleles (NR-cay) in the PHS interval with non-recombinant families containing Caledonia alleles (NR-cal) in the PHS interval (Table 3.2, Figure 3.3). In 2007 (BC1F3), the PHS QTL effect on 2B was significant ($p < 0.0001$). The NR-cay mean score was 3.95 compared to the NR-cal mean score of 5.39. The NR-cay mean was higher than the Cayuga parent, which was 3.17 and the NR-cal mean score was lower than the Caledonia parent, which was 6.17.

In 2008 (BC1F4), the PHS mean scores for 20 non-recombinant BC1F4 families in each group were compared at two locations, Ketola and Snyder. The PHS QTL effect on 2B was significant at Ketola ($p < 0.05$) but not at Snyder (ns). At the Ketola location, the NR-cay mean score was 4.0, whereas the NR-cal mean score was 4.55. At Snyder, the NR-cay mean score was 4.17 whereas the NR-cal mean score was 4.73. Even though the mean difference between NR-cay and NR-cal (mean different =0.5) was the same in the two locations, the standard deviation at Synder was double than at Ketola resulting in non-significance.

In 2009 (BC1F5), the PHS mean score of 19 non-recombinant BC1F5 families in each group was compared at two locations and the PHS QTL effect on 2B was significant at both Ketola ($p < 0.01$) and Snyder ($p < 0.0001$). At Ketola, the NR-cay mean score was 1.90 whereas the NR-cal mean score was 3.21. At Snyder, the NR-cay mean score was 2.07 whereas the NR-cal mean score was 3.89.

The SD phenotype was evaluated at one location, Ketola in 2008. The GI mean score of both parents was significantly different ($p < 0.0001$). The germination percentage between two parents was determined within 10 days indicating that 10

Table 3.2. PHS mean score and Germination index (GI) mean score for seed dormancy in non-recombinant groups at different populations and years.

	Parent score			Non-recombinant (NR)		
	Cayuga (cay)	Caledonia (Cal)	Significant p-value	NR-Cay	NR-Cal	Significant p-value
PHS-BC1F3 -Synder 2007	3.17	6.17	<0.0001	3.95±0.67	5.39±0.37	<0.0001
PHS-BC1F4 -Ketola 2008	2.48±0.92	6.16±0.35	<0.0001	4.0±0.71	4.55±0.86	<0.05
PHS-BC1F4 -Snyder 2008	1.64±0.58	5.42±1.06	<0.0001	4.17±1.53	4.73±1.52	NS (0.18)
PHS-BC1F5 -Ketola 2009	0.93±0.78	3.27±0.85	<0.01	1.90±0.95	3.21±1.88	<0.01
PHS-BC1F5 -Snyder 2009	1.03±0.83	5.67±1.26	<0.0001	2.07±1.13	3.89±1.48	< 0.0001
SD-BC1F4 -Ketola 2008	5.86±0.61	8.14±0.70	<0.0001	6.99±0.62	7.49±0.89	<0.05

PHS = Preharvest sprouting, scoring from 0-9, 0=no sprouting 9= the most sprouting
SD = Seed dormancy, evaluated by germination index from day 1-day 10. score 0 = no seed germination at all at day 10, score 10 mean all seed germination at day 1
NS = Non-significance

days is the optimal number of days for germination in non-recombinant and recombinant groups (Munkvold et al. 2009). Seed germination of Caledonia began on Day 1 with an average of 15 % germination whereas Cayuga began on Day 3 with an average of 12% germination (Figure 3.3B,C). The highest percent germination of Caledonia was on Day 2 (30% germination), whereas the highest percent germination for Cayuga was on Day 5 (30% germination). After Day 5, more than 95% of seeds from Caledonia had germinated whereas Cayuga exceeded 95% after Day 8. At Day 10, almost 100 % of seeds had germinated in both parents. Consequently, the number of days to evaluate seed germination for SD was 10 days. Mean GI score for both parents was significantly different ($p < 0.0001$) and the score for Cayuga was 5.86 versus 8.14 for Caledonia. Also, the SD QTL effect was significant for both non-recombinant families ($p < 0.05$). Mean GI scores for NR-cay and NR-cal were 6.99 and 7.49, respectively. The level of significance for both PHS and SD effects was similar at the Ketola location. It is possible that PHS and SD effects in this PHS interval were contributed by the same gene.

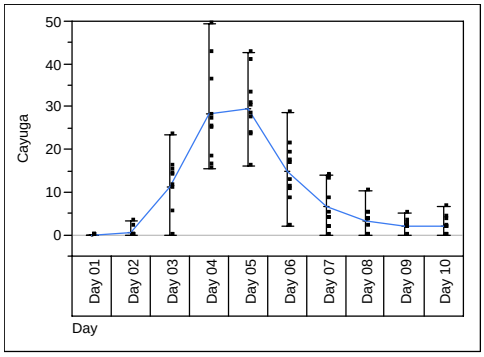
QTL effect on preharvest sprouting and seed dormancy in BC1F4 recombinants

For QTL analysis, 359 BC1F4 recombinants in the Caledonia background were genotyped and phenotyped. The distribution of phenotypes of this recombinant population appeared normal however the Goodness-of-Fit Test was significant indicating that it was not normally distributed (figure 3.4). At Ketola, the PHS phenotype peaks were at PHS scores 4-4.5 and 5-5.5 suggesting that there may be two QTLs contributing to PHS at the 2B PHS interval. Because the recombinants contained 81.75-86.30 % Caledonia background, the segregating background could contribute phenotypic variation. The distribution of PHS phenotype at Snyder 2008 was broader than at Ketola 2008.

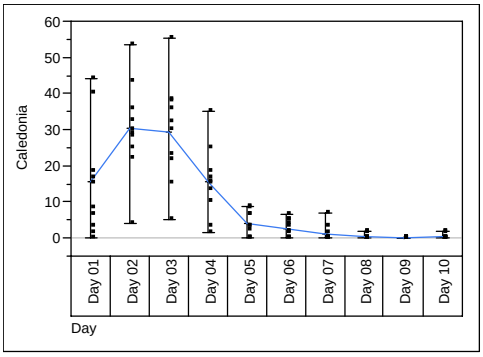
Cayuga Caledonia

0 1 2 3 4 5 6 7 8 9

A



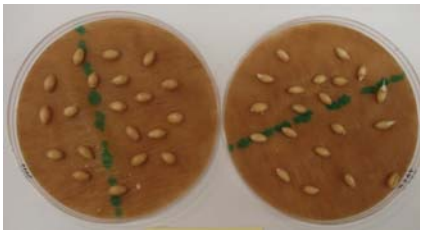
Cayuga



Caledonia

B

Cayuga Caledonia



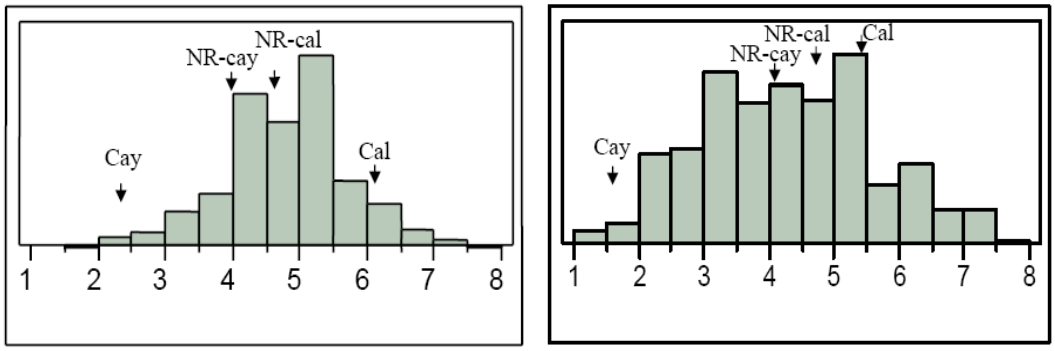
Germination at Day 1

Germinated (Top) and non-germinated seeds (below)

C

Figure 3.3. PHS and SD phenotyping.

PHS score is evaluated from 0-9. 0= no any sprouting on a spike at all. Score 9 is the most sprouting for the whole spike (A). SD is evaluated from germination index and percent germination between two parents was evaluated from Day 1 to Day 10 (B). Score 1= mean no any seed germination at all on day 10. Score 10 = all 20 seeds germinated at day 1 (C).

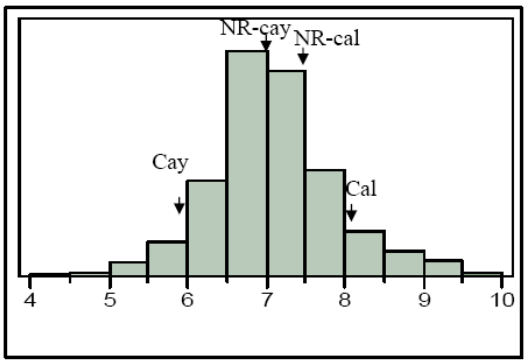


PHS at Ketola 2008

A

PHS at Snyder 2008

B



Germination Index at Ketola 2008

C

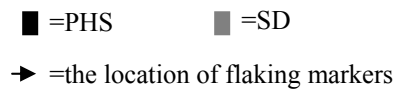
Figure 3.4. PHS phenotype distribution and germination index of 359 recombinant BC1F4 at Ketola 2008 (A,C) and Snyder 2008 (B).

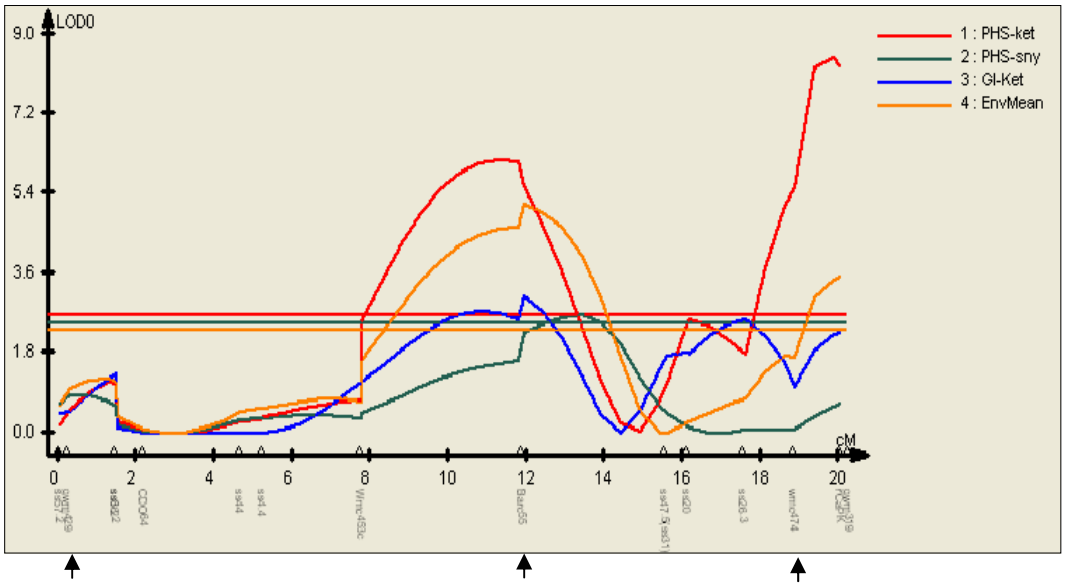
Cay= Cayuga, Cal= Caledonia, NR-cay = non-recombinant containing Cayuga at PHS QTL interval. NR-cal = non-recombinant containing Caledonia at PHS QTL interval.

After linkage map analysis by MapManager QTXb20, the increase in recombinants resulted in almost all of the markers being separated. The results of QTL analysis of the BC1F4 recombinants and genetic distances was shown in Figure 3.5. Because only recombinants were used for QTL analysis, the genetic distance was normalized with 633 BC1F4 non-recombinants (Table 3.1). The 2B.1 linkage distance in BC1F4 population was around 20 cM that higher than previous 2B.1 linkage distance in DH population that was 10 cM because of an increase of new markers and recombinants. The average marker distance was one marker per 0.91 to 1.32 cM. Based on comparisons between markers ss3.2 and ss4.4 and markers ss47.5 and ss26.3 and their location on the rice physical map, there was about 120-128 kb of rice physical map distance per one wheat cM. Results of the QTL analysis were summarized on Table 3.3. The major PHS QTL on 2B.1, contributed from Cayuga, was between marker wmc453c and Barc 55. This interval distance was 4.2 cM. There were 3 co-localized significant QTLs on this interval; PHS QTL at Ketola 2008, Mean PHS QTL at Ketola and Snyder2008 and SD QTL at Ketola 2008. All three significant QTL LOD scores were from 3.10 – 6.19 ($p < 0.01$). No new markers designed from ESTs were located in this interval so it was not possible to identify the candidate gene for this QTL. Another minor PHS QTL at Wmc474, also significant at Ketola 2008 and Mean PHS QTL at Ketola and Snyder2008 was detected at rCaPK marker. Linked marker was designed for known function gene that might involve SD. Marker rCaPK was designed from the transcript for gene Calmodulin/Ca^{2+} dependent protein kinase. The QTL LOD scores were 3.26 and 8.07 between markers Wmc 474 and rCaPK ($p < 0.01$). Based on QTL analysis, initially it was not clear which parent contributed to this minor QTL because of the high frequency of Caledonia background at marker Wmc474 and surrounding markers. Only 29 out of 359 BC1F4 recombinants contained a Cayuga allele at marker Wmc474 whereas 251 out 359 BC1F4

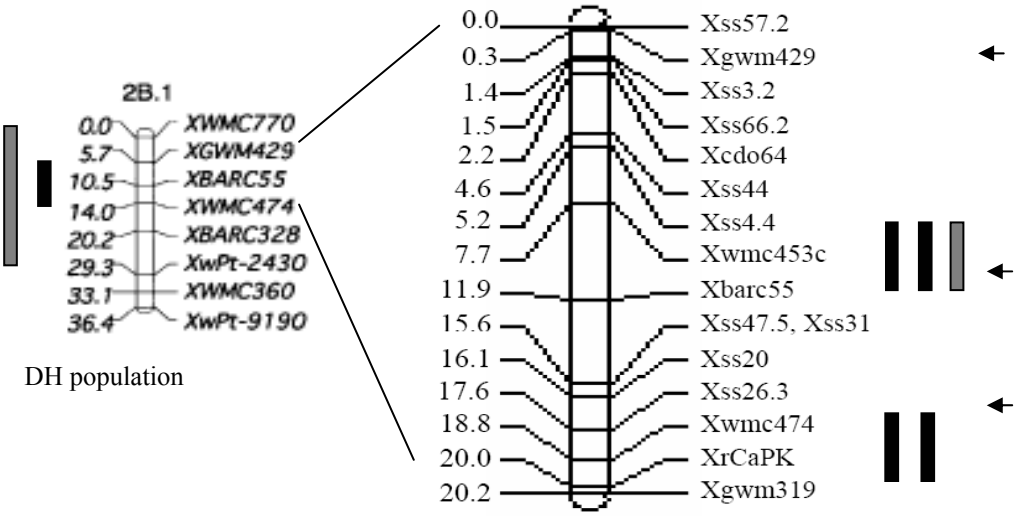
Figure 3.5. QTL analysis in recombinant BC1F4 results (A) and the genetic distance after normalization with non-recombinant BC1F4 (B).

The major PHS QTL on 2B.1 was narrowed down to be between marker Wmc453c and Barc55 that contributed from Cayuga parent. The significant include 2 environments of PHS at ketola 2008 and Mean for both location of Ketola and Snyder 2008. SD effect contributed from Cayuga parent also significant between marker Wmc453c and Barc55. Another PHS QTL also detected located between marker Wmc474 and rCaPK that designed from Calmodulin/Ca^{2+} dependent protein kinase and significant in 2 environments of PHS at ketola 2008 and Mean for both location of Ketola and Snyder 2008




LOD0
9.0
7.2
5.4
3.6
1.8
0.0
cM
0
2
4
6
8
10
12
14
16
18
20
1 : PHS-ket
2 : PHS-sny
3 : GI-Ket
4 : EnvMean

A.


2B.1
0.0 XWMC770
5.7 XGWM429
10.5 XBARC55
14.0 XWMC474
20.2 XBARC328
29.3 XwPt-2430
33.1 XWMC360
36.4 XwPt-9190
DH population
0.0 Xss57.2
0.3 Xgwm429
1.4 Xss3.2
1.5 Xss66.2
2.2 Xcdo64
4.6 Xss44
5.2 Xss4.4
7.7 Xwmc453c
11.9 Xbarc55
15.6 Xss47.5, Xss31
16.1 Xss20
17.6 Xss26.3
18.8 Xwmc474
20.0 XrCaPK
20.2 Xgwm319
Recombinant BC1F4 population

B.

Table 3.3. Summary of a PHS and SD QTL analysis of 359 BC1F4 recombinants.

	QTL Test at 99% Confidence interval	QTL LOD score	QTL interval peak	QTL interval after normalization with non-recombinant population	Parent contributed to the phenotype	The Annotation of gene that been used to design primers that co-linked with QTL
PHS-BC1F4 -Ketola 2008	Significant	6.19	Wmc453c -Barc55	4.20	Cayuga	
		2.59	ss20		?	Calmodulin binding protein
	Significant	8.07	Wmc474 -rCaPK2	1.20	?	Calmodulin/Ca^{2+} dependent protein kinase
PHS-BC1F4 -Snyder 2008		2.70	Barc55-ss47.5	2.20	Cayuga	
PHS-BC1F4 -Mean at 2 envirs	Significant	5.19	Wmc453c -Barc55	4.20	Cayuga	
	Significant	3.26	Wmc474-rCaPK	1.20	?	
SD-BC1F4 -Ketola 2008	Significant	3.10	Wmc453c -Barc55	4.20	Cayuga	
		2.57	ss26.3		?	Developmentally -regulated binding protein 1

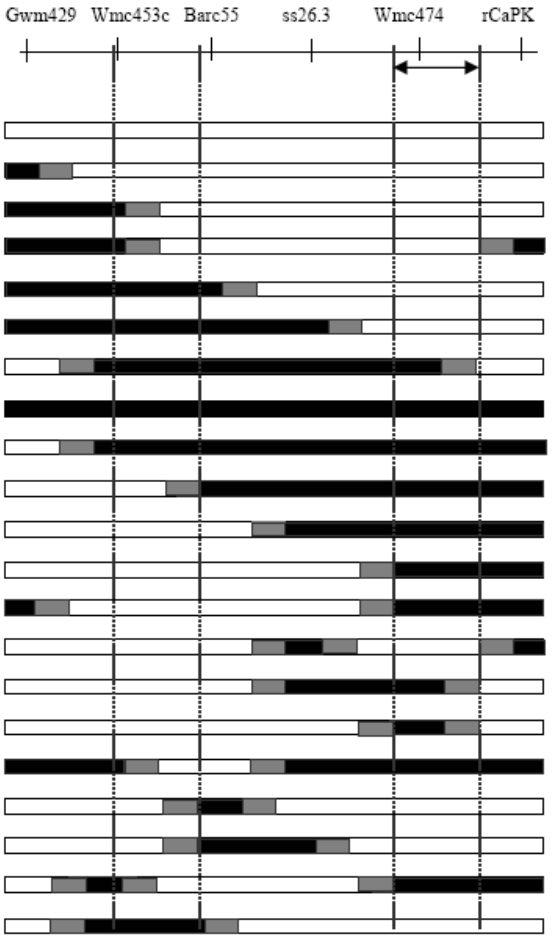
recombinants contained a Cayuga allele at marker Barc55. After comparing the PHS mean scores with NR-cal, the QTL between wmc474 and rCaPK was contributed from Cayuga ($p < 0.05$). This result was confirmed in fine mapping of selected BC1F5 recombinants because the frequency of the Cayuga allele was increased for Wmc474 to evaluate which parent contributed the effect for the interval between Wmc474 and rCaPK. Due to the QTL LOD score of 8.07 between markers Wmc474 and rCaPK, being higher than the PHS effect between Wmc453c and Barc55, it is possible that this region is also contributing to the PHS QTL in the 2B.1 interval.

Fine mapping of preharvest sprouting QTL in BC1F5 recombinants

The BC1F5 population was used for fine mapping to increase recombinants as well as the number of lines with Cayuga alleles at Wmc474 and surrounding markers. This would help to better estimate the Cayuga-derived PHS QTL effect at Wmc474 compared to the PHS QTL effect between Wmc453c and Barc55. For fine mapping analysis, 6 markers; Gwm429, Wmc453c, Barc55, ss26.3, Wmc474 and rCaPK were used. These markers were all flanking markers based on the QTL result from BC1F4 recombinants. The order of these 6 markers was based on the marker order determined by mapping in the Cayuga x Caledonia population and selected BC1F4 recombinant genotypes. After screening with three flanking markers, 470 BC1F5 individuals were homozygous BC1F5 recombinants but only 456 BC1F5 families were used in fine mapping. Fourteen out of 470 BC1F5 recombinants were not included in the analysis because there were missing data for Wmc453c or rCaPK. Also, some of these recombinants could not be assigned to any recombinants group. The 456 recombinant BC1F5 families were grouped according their recombination break points at six flanking markers at PHS interval on 2B. Nineteen recombinant groups and two non-recombinant families were assigned group 01 to group 21 (gr. 01 to gr. 21) according their breakpoint and each group contained at least 3 recombinant families to 120

families (Figure 3.6). Non-recombinant groups for Caledonia (NR-cal) and Cayuga (NR-cay) were also analyzed. The PHS mean score of 21 groups including 19 recombinant and 2 non-recombinant groups was used to estimate the PHS effect in the PHS interval on 2B.1. PHS scores were evaluated for two environments, Ketola in 2009 (Ketola09), Snyder in 2009 (Snyder09) and the Mean (Mean09). The QTL analysis resolved two Cayuga-derived QTL alleles contributing to PHS in the 2B.1 PHS interval (Figure 3.6). The first was between markers Wmc453c and Barc55 and was significant for both environments and the mean. The second PHS QTL was between markers Wmc474 and rCaPK, also significant for both environments and the mean. Each recombinant group contained from zero to two PHS QTLs. Additionally, nineteen groups of recombinant families containing different break point were assigned into 4 categories according whether having no PHS QTL, having one PHS QTL between Wmc453c and Barc55, having one PHS QTL between Wmc474 and rCaPK and having two the PHS QTLs. The recombinant groups of first category included recombinant group 02, group 14, group 18 and group 19. These groups had no PHS QTL effect because they were not significantly different from NR-cal (ns). The PHS mean scores ranged from 3.31-3.75 based on the Mean09 whereas mean score for NR-cal was 3.55. The recombinant groups of second category included recombinant group 03, group 04, group 05 and group 06. These recombinant groups had one PHS QTL between Wmc453c and Barc55. The PHS mean scores ranged from 3.06- 3.22 and were significantly different from recombinant groups that had no PHS QTL (p-value ranged from 0.0023-0.0383) based on Mean09, The recombinant groups of third category include recombinant group 10, group 11, group 12, group 15, group 16 and group 20. These recombinant groups had one PHS QTL between Wmc474 and rCaPK with mean PHS scores ranging from 1.70-2.85 and were significantly different from recombinant groups that had no PHS QTL ($p < 0.001$) based on Mean09.

Figure 3.6. Fine mapping at PHS interval at 2B in selected homozygous recombinant BC1F5.
Six markers were used in fine mapping genotyping. The order of markers was based on the marker order of the double haploid and selected recombinant BC1F4 genotypes from Gwm429 to rCaPK. There were 19 recombinant groups (456 individuals) and two non-recombinant groups (35 individuals) for Caledonia (NR-cal, gr.01) and Cayuga (NR-cay, gr.08). PHS scores were evaluated at two environments, Ketola09 and Snyder09, and the mean between these two environments (Mean09). As the result, there were two major QTL regions coming from Cayuga contributing to PHS on 2B.1. The first QTL region was located between marker Wmc453c and Barc55, significant in all two environments and also mean09. The second QTL region was between Wmc474 and rCaPK, also significant in all two environments and also mean09. Each recombinant group contained either no QTL region, one QTL region or two QTL regions. Recombinant groups with a p-value less than 0.05 contained PHS QTL coming from Cayuga. The recombinant (gr.07, gr.09 and gr.17) and NR-cay (gr.08) groups containing two QTL regions were significantly different from recombinant groups that have no QTL region and recombinant groups that have one QTL region. The PHS mean score suggests that the contribution from the second PHSQTL region, between Wmc474 and rCaPK, was more than the first PHS QTL region, between marker Wmc453c and Barc55.
RC-group = Recombinant group, each group also was assigned as genotype such as 000111; 0 = for Cayuga (PHS resistant), 1= for Caledonia (PHS susceptible) parents
RC # = The number of recombinants
ME±SD = PHS score mean ± Standard deviation
# of QTL regions = The number of PHS-affecting region based on Mean09.
P-value = The significance value of each recombinant group, compared with a recombinant group that has no PHS-affecting region (gr.02)
a = The comparison between recombinant groups containing two PHS-affecting regions with the recombinant groups containing one PHS-affecting regions with gr.06 and with gr.10, respectively (* = p-value <0.05, ** =p-value < 0.01) .
NS = Non-significant p-value
gr. = recombinant group
The 2 regions between the vertical dotted lines on the figure represent PHS-affecting regions.
Empty bar = Caledonia (PHS susceptible parent) region
Black bar = Cayuga (PHS resistant parent) region
Grey bar = Cayuga (PHS resistant parent) region without knowing the boundary region.



		Ketola09		Snyder09		Mean09		
RC-group	RC # (491)	ME±SD	P-value	ME±SD	P-value	ME±SD	P-value	# of QTL regions
gr.01-111111	16	3.21±1.88	NS	3.89±1.48	NS	3.55±1.38	NS	No
gr.02-011111	33	3.41±1.06	NS	4.08±0.91	NS	3.75±0.74	NS	No
gr.03-001111	53	2.85±1.18	0.0486	3.43±1.27	0.0113	3.15±0.93	0.0059	1
gr.04-001110	12	2.71±1.27	NS	3.42±1.16	NS	3.06±1.15	0.0383	1
gr.05-000111	120	2.89±1.22	0.0394	3.42±1.18	0.0035	3.16±1.00	0.0023	1
gr.06-000011	35	3.03±1.23	NS	3.41±0.90	0.018	3.22±0.82	0.0264	1
gr.07-100001	4	1.75±0.91	0.0138	2.55±0.64	0.0123	2.15±0.49	0.0021	$2^{*,NS}$ a
gr.08-000000	19	1.90±0.95	<.0001	2.07±1.13	<.0001	1.98±0.65	<.0001	$2^{**,**}$ a
gr.09-100000	15	1.69±1.41	<.0001	2.73±1.34	0.0002	2.21±1.17	<.0001	$2^{**,*}$ a
gr.10-110000	22	2.88±1.63	NS	2.84±1.33	<.0001	2.85±1.25	0.001	1
gr.11-111000	72	2.23±1.20	<.0001	3.28±1.09	0.0009	2.75±0.91	<.0001	1
gr.12-111100	22	2.29±1.36	0.0014	3.37±0.90	0.0255	2.83±0.93	0.0007	1
gr.13-011100	3	2.67±0.23	NS	3.12±1.14	NS	2.89±0.66	NS	?
gr.14-111010	3	2.40±0.92	NS	4.27±1.53	NS	3.33±1.07	NS	No
gr.15-111001	12	1.60±1.75	<.0001	1.82±1.60	<.0001	1.70±1.50	<.0001	1
gr.16-111101	6	1.96±1.17	0.0177	2.27±1.01	0.0004	2.15±1.03	0.0002	1
gr.17-001000	4	1.50±0.50	0.0047	1.35±1.02	<.0001	1.43±0.69	<.0001	$2^{**,**}$ a
gr.18-110111	16	3.26±1.49	NS	3.78±0.88	NS	3.52±0.89	NS	No
gr.19-110011	10	2.91±1.49	NS	3.78±0.92	NS	3.31±0.96	NS	No
gr.20-101100	4	2.75±0.34	NS	2.60±1.34	0.0154	2.68±0.80	0.0382	1
gr.21-100111	10	3.22±0.92	NS	4.16±1.36	NS	3.69±0.76	NS	?

The recombinant groups of fourth category included group 07 and group 09. These recombinant groups as well as all NR-cay groups had both PHS QTLs. The PHS mean scores of these groups ranged from 1.43-2.21 based on Mean09. These groups were compared with group 06, which had one PHS QTL region between Wmc453c and Barc55 and with group 10, which had one PHS QTL region between Wmc474 and rCaPK. These groups with two PHS QTL regions were significantly different ($p < 0.01$) from the groups that had only one PHS QTL region. The PHS mean scores suggest that the contribution from the second PHS QTL region, between Wmc474 and rCaPK (mean score = 2.67), had a larger effect than the first PHS QTL between markers Wmc453c and Barc55 (mean score = 3.16) and there were significantly different ($p<0.0001$). The second PHS QTL between Wmc474 and rCaPK contributed two-thirds of the PHS effect whereas the first PHS QTL between markers Wmc453c and Barc55 contributed one-third of the PHS effect for the 2B.1 interval. Some recombinant groups that had only one PHS QTL region (group 15 and group 16) between Wmc474 and rCaPK had a larger PHS effect ($p<0.0001$). Moreover, there were some exceptions in some groups. For example, two recombinant groups (group13 and group 21) should contain one PHS QTL but their mean scores were not significantly different (ns) from the recombinant groups that had no PHS QTL region. Group 13 was non-significant because it contained only 3 recombinants, even though the PHS score was 2.89 based on Mean09. Group 21 mean score was 3.69, based on Mean09 suggesting that it might contain the break point between markers Wmc453c and Barc55, resulting in no-PHS effect.

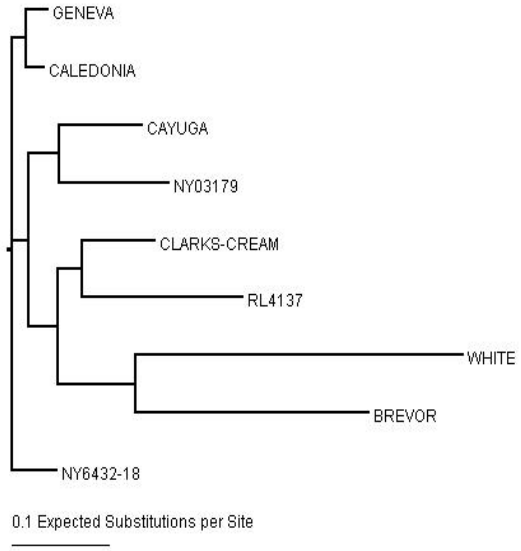
In summary, fine mapping analysis was conducted using non-recombinant groups (NR-cay and NR-cal) and recombinant groups in BC1F4 (359 individuals) and in BC1F5 (456 individuals). The PHS effect coming from Cayuga was evaluated by comparing NR-cay and NR-cal families. The PHS effect of NR-cay was significantly

different from that of NR-cal in all three generations, BC1F3, BC1F4 and BC1F5. Also SD effect of NR-cay was significantly different from that of NR-cal in the only environment it was tested, Ketola 2008. Fine mapping in BC1F4 and BC1F5 recombinants narrowed the PHS QTL (14 cM) and resolved two closely linked PHS QTLs. Two Cayuga-derived PHS QTL regions contributed to the 2B.1 PHS QTL interval. The first PHS QTL region was between Wmc453c and Barc55 (around 4.2 cM based on BC1F4 linkage map). The effect of this region was significant for one environment and the mean for the BC1F4 recombinants, and significant in both environments and the mean for the BC1F5 recombinants. The second PHS QTL region was between Wmc474 and rCaPK (1.2 cM based on the BC1F4 linkage map). The effect of this region was significant in one environment for the BC1F4 recombinants the mean for the BC1F4 recombinants, and significant in both environments and the mean for the BC1F5 recombinants. Based on BC1F5 fine mapping, the PHS QTL region between Wmc474 and rCaPK contributed two-thirds of the PHS QTL effect (around 16 % of phenotypic variation), whereas the PHS QTL region, between markers Wmc453c and Barc55 contributed one-third of PHS QTL effect (around 8 % of phenotypic variation) for the 2B.1 interval. These results indicate that multiple genes contributed to the PHS effect of the 2B.1 interval. One of those candidate genes was Calmodulin/Ca^{2+} dependent protein kinase that was used to design marker rCaPK. To identify the other candidate genes, genome or positional walking will be needed.

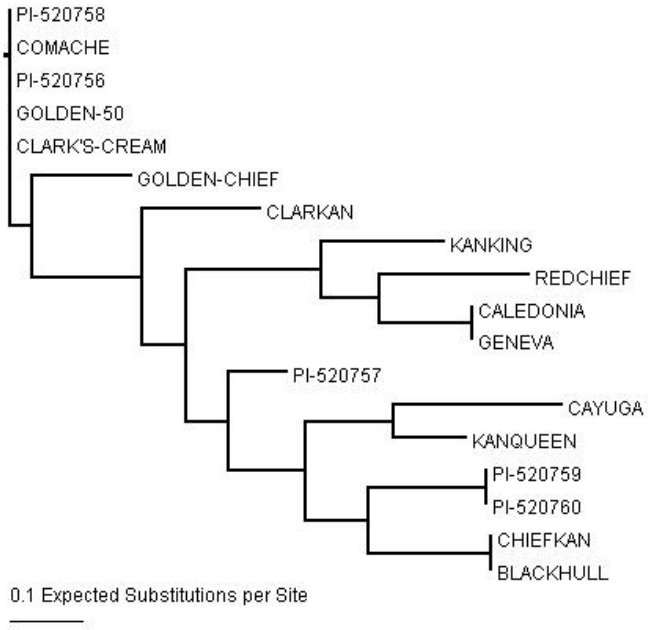
The relationship between PHS Cayuga allele with Clark's Cream-related lines

The phylogenetic relationship between Cayuga, Clark's Cream, Caledonia and Geneva was determined by using 50 DArT markers located on 2B (Figure 3.7A). Those parents were genotyped by Long-Xi Yu (unpublished data). Caledonia shared almost

100 % of its alleles with Geneva. This suggests that almost all PHS-susceptible alleles on Caledonia chromosome 2B were inherited from Geneva. Also, a very close relationship between Caledonia and Geneva was found by genotyping with only 6 flanking markers at the PHS interval on 2B (Figure 3.7B). This supports the possibility that Caledonia alleles at the PHS interval were inherited from Geneva. By genotyping 50 DArT markers, it was determined that Cayuga chromosome 2B was closer to that of Clark's Cream than it was to Geneva. However, after genotyping by using 6 flanking markers on 2B.1 PHS interval, we found that the Cayuga alleles were different from Clark's Cream at markers Gwm429, Wmc453c, Barc55, ss26.3 and Wmc474. Only one Clark's Cream allele at marker rCaPK matched a Cayuga allele. There are 3 different alleles at Barc55 for Cayuga, Caledonia/Geneva and Clark's Cream. Furthermore, 14 Clark's Cream related lines were genotyped to identify relationship with Cayuga. The phylogenic relationship was shown on Figure 4.2B. There were four groups of those lines whose members shared the same allele at all six flanking markers. The first group contained PI 520758, Comache, PI520756, Golden 50 and Clark's Cream. The second group contained Caledonia and Geneva. The third group contained PI 520757 and PI 520760. The fourth group contained ChiefKan and Blackhull. KanQueen shared the most alleles with Cayuga at Wmc474, Barc55, Wmc474 and rCaPK (Table 3.4) However, the relationship between KanQueen and Cayuga still isn't clear. Golden Chief was shown to share the same allele as Cayuga at Barc55 and rCaPK. This suggests that one of the PHS resistance alleles at Barc55 and rCaPK may come from Golden Chief, one of the parents used to develop Clark's Cream. There are 3 different alleles at Barc55 shown on Figure 3.8. The size was around 132 bp for Cayuga, around 128 bp for Caledonia and around 124 bp for Clark's Cream.



A



B

Figure 3.7. Phylogenetic relationships between wheat varieties.

The relationship was found by using 50 DArT markers located on chromosome 2B (provided by Long-Xi Yu) (A) Cayuga and Clark's Cream-related lines at the PHS interval on 2B.1, found by using 6 flanking markers. Caledonia and Geneva were used as controls (B).

Table 3.4 Relationship of PHS resistance alleles for Cayuga with Clark's Cream related lines.

			Same allele with Cayuga at marker						
Wheat lines	PHS Mean - 2006	wheat type	Gwm429	**Wmc453c**	**Barc55**	ss26.3	**Wmc474**	**rCaPK**	pedigree
Cayuga	0.1644	SWW	Y	Y	Y	Y	Y	Y	Clark's Cream/Geneva //Geneva
Caledonia	3.1465	SWW							Off-type of Geneva
Geneva	2.1955	SWW							
Clark's Cream	0.5657	HRW						Y	composite bulk of KanKing / Golden 50 or KanKing / Golden Chief
KanKing	0.9147	HRW		Y				Y	Kansas bulk hybrid 43A31 obtained from E.G. Heyne
Golden 50	0.9147	HRW						Y	selection from Rodco (MIX)CI-12406+CONCHO)
Golden Chief	0.9147	HRW			Y			Y	Golden 50 / Red Chief
PI 520756	1.1462	HWW						Y	KS75216 / Clark's Cream = Pitic 62 / II53-526 (Chris sib)//2* Sonora 64 /3/ Klein Rendidor /4/ Scout /5/ Clark's Cream
PI 520757	1.1462	HWW		Y				Y	KS82W445 / Clark's Cream = Plainsman V // Caprock / Purdue HP /3/Clark's Cream
PI 520758	1.1462	HWW		N				Y	KS82W404 / Clark's Cream = KS75216 / Plainsman V // Clark's Cream
PI 520759	1.4934	HWW		Y		Y		Y	KS82W439 / Clark's Cream = Timwin / Bezostaja 1 // Clark's Cream
PI 520760	2.0720	HWW		Y		Y		Y	KS82W450 / Clark's Cream = Plainsman V / Newton // Clark's Cream
ChiefKan	0.9147	HRW		Y		Y		Y	Blackhull / soft wheat // Superhard Blackhull
KanQueen	0.9147	SRW		Y	Y		Y	Y	Clarkan / Blackhull
Clarkan	0.9147	SRW						Y	Black Hull / Harvest Queen
Red Chief	1.4934	HRW	Y	Y					Blackhull selection [or] Redhull / Chiefkan
Comache	0.9147							Y	Oro (selection from Turkey889) / Tenmarq (Marquis/Crimean)
Blackhull	0.9147	SW		Y				Y	selection from field of Turkey

SWW= soft white winter wheat, HRW = hard red winter wheat, HWW = hard white winter wheat, SRW=soft red winter wheat, SW =soft white wheat. Y= Containing the same alleles as Cayuga. The marker name in the dark letters was linked with fine mapping PHS OTLs

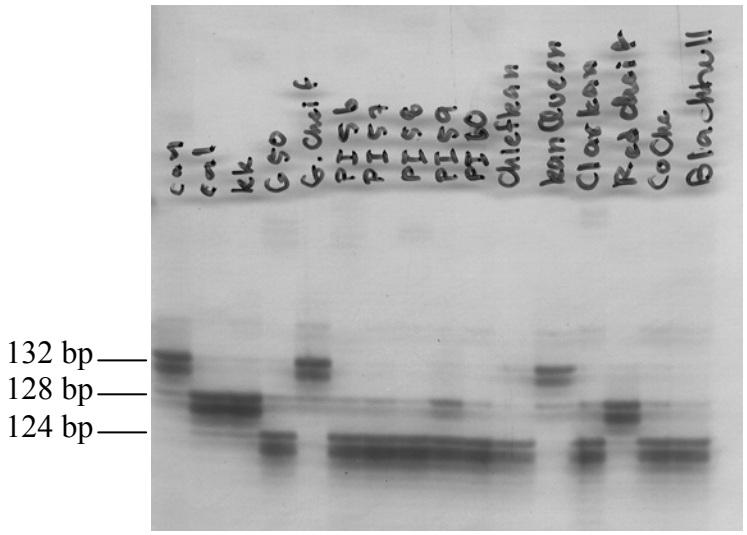


Figure 3.8. Genotype of Clark's Cream-related lines at Barc55

The pedigrees of PI 520756, PI520757, PI520758, PI 520759 and PI 520760 lines contained Clark's Cream. Genotyping also showed that these lines share the same alleles as Clark's Cream at Barc55. The association of allele at Barc55, Wmc474 and rCaPK at PHS QTL interval was found to have PHS resistance in CC-related lines (Table 3.5). There were two associated alleles at Barc55 and Wmc474. The first allele was the same as Cayuga allele. The second allele was the same as Clark's Cream allele. However, association of Cayuga allele with PHS resistance was more than Clark's Cream allele at Barc55 and Wmc474 (p-value = 0.0095 and p-value = 0.0279, respectively).

Haplotyping at Barc55

After separating the wheat lines according to Barc55 alleles (Table 3.6), The PHS mean score of allele at 132 bp and allele at 124 bp was around 1.67-1.68 lower than allele at 128 bp (PHS mean score = 2.09). This could suggest that allele that was the same as Cayuga (allele size=132 bp) and alleles that was the same as Clark's Cream (allele size = 124 bp) associated with PHS resistance than allele that was the same as Caledonia/Geneva (allele size =128 bp). Even though the PHS mean was different, non significantly different was found in those alleles

Table 3.5 Association of four flanking markers linked with two PHS QTLs for Clark's Cream-related lines

Flanking markers	Number of Alleles	Association alleles with PHS resistance (Alleles size)	P-value
Wmc453c	2		NS
Barc55	3	Cayuga Allele (~size 132 bp)	0.0095
		Clark's Cream Allele (~size 124bp)	0.0223
Wmc474	5	Cayuga Allele (~135bp)	0.0279
		Clark's Cream Allele (~145bp)	0.0403
rCaPK	2	Cayuga/ Clark's Cream allele (~355bp, present/ absent)	0.0008

NS= Not significant

Table 3.6. Association of Barc55 alleles with PHS score for 166 winter wheat lines.

Barc55 alleles	Number of lines	PHS mean	Std Dev	Std Err Mean	PHS
124 bp	61	1.68	0.152077	0.01947	MR
128 bp	86	2.09	0.176273	0.01901	MR
132 bp	19	1.67	0.133367	0.0306	R

MR = moderately PHS resistance
R = PHS resistance

DISCUSSION

Fine mapping population

In this study selected homozygous recombinant BC1F4 and BC1F5 families were used for fine mapping the PHS QTL effect of the 2B.1 region. These populations contained 81.75%-86.3% Caledonia background outside the PHS interval on 2B.1. These populations contained more of the recurrent parent (Caledonia) background than a normal BC1 because doubled haploid individuals containing a high recurrent parent background were backcrossed to the recurrent parent. The proportion of recurrent parent background was close to that expected for a BC2. This strategy saves a considerable amount of development time in a winter cereal crop. The recombinants were identified in all three generations, BC1F3, BC1F4 and BC1F5 generations resulting in many recombinants in a small population for fine mapping. Advantages to this strategy are shorter development time than for NILs and advanced backcrossed populations and fewer families required for phenotyping. Population development and fine mapping occurred simultaneously rather than sequentially. Although, the populations are not quite as uniform for recurrent parent background, the heterogeneity was offset by using more families and by sampling within families to increase phenotyping accuracy. In addition, fine mapping was augmented by comparing phenotypic means among the recombinant and non-recombinant families for the PHS QTL interval. It is likely that both groups had a similar background so the phenotypic difference results mostly from the PHS QTL interval.

Most of fine mapping works were successful by using NILs and advanced backcrossed populations (Paterson et al. 1990; Takeuchi et al. 2003; Nichols et al. 2006; Röder et al. 2008; Han et al. 1999; Alpert and Tanksley 1996; Yamamoto et al. 2000; Yamamoto et al. 1998; Vladutu et al. 1999). However, there were some studies were also successful by using BC1 and large segregating populations for fine

mapping. The Diageotropica (*Dgt*) gene in tomato was fine mapped using 1,308 BC1 individuals. Ten recombinants narrowed the region to 0.8 cM (Oh et al. 2002). Another study used 1,050 BC1F2 individuals to arrow the hybrid embryo sac sterility (*S32(t)*) locus in rice to 1.9 cM. Based on the recombinant events, it was located in a 64 kb region containing 6 recombinants and seven predicted open reading frames (Li et al. 2007). In another study, fine mapping used a large segregating population of 1,849 (1,256 F2 and 593 BC1F2) to fine map a frost resistance locus (*Fr-H2*) in barley. It was narrowed to 0.81 cM, containing 6 recombinations over 0.03 to 0.32 cM (Francia et al. 2007). These studies as well as our study showed that fine mapping can be done in any populations that contain enough recombinants in the targeted region.

In our study revealed that low polymorphism between two hexaploid wheat parents that used to develop fine mapping populations delimit the progress to narrow the region as small as possible like other diploid grass species such as in durum wheat, around 64 kb of rice collinear region (Distelfeld et al. 2004), in rice, around 64 to 75 kb (Gu et al. 2009; Li et al. 2007), in tomato (Oh et al. 2002) and in soybean, around 33 kb (Kim et al. 2010). There were many factors in wheat influent the progress of fine mapping. First, wheat has no physical map yet. The available sequences were some ESTs, transcripts and some STSs. Those sequences do not represent for the whole wheat genomes. Second, wheat is a self-pollinated species that limits the recombinant event between wheat varieties. Finally, the Cayuga and Caledonia wheat lines used in fine mapping populations shared the same parent; Geneva variety. They both shared around 75% of Geneva background. This might the main factor that it still be our challenge and limited the number of polymorphism at PHS interval on 2B.1.

Fine mapping versus doubled haploid genetic mapping

In previous work, PHS QTL on 2B.1 was the major QTL on 2B.1 in 209 individuals from a Cayuga x Caledonia doubled haploid population. The 2B.1 QTL explained 24% of the phenotypic variation for PHS. The QTL interval spanned 14 cM (Munkvold et al. 2009), and was delimited by three flanking markers, Gwm429, Barc55 and Wmc474. In this study, fine mapping in the BC1F4 (359 individuals) and BC1F5 (456 individuals) recombinants revealed that the PHS QTL resulted from two closely linked QTLs. One PHS QTL interval was 4.2 cM between Wmc453c and Barc55. The second PHS QTL interval was 1.2 cM between Wmc474 and rCaPK. This PHS QTL contributed two-thirds of the phenotypic variation whereas the first PHS QTL contributed one-third. The alleles for PHS resistance from Cayuga are in coupling. Although the SD phenotype was only scored in one environment, the SD QTL was collocalized in the 4.2 cM interval with the PHS QTL between Wmc453c and Barc55.

There are many factors that may cause the single broad QTL in one population to be two QTLs in another population. Those factors include missing data of a marker at the middle of the peak, dominant markers used in a heterogeneous population, and wrong order of the markers and double crossovers in some recombinants. However in this study, those factors are offset. For the first factor, there was almost no missing data (only 0.01%) on those middle markers between two PHS QTLs. For the second factor, homozygous recombinants were used for fine mapping so genotypes based on dominant markers were not different from those based on co-dominant markers. For the third factor, the order of both previous and newly designed primers was first determined using the Cayuga x Caledonia doubled haploid population first and later, the BC1F4 recombinants to order the markers in the QTL interval. For the last factor, double crossovers can happen but at very low frequency. In BC1F5 recombinants,

some recombinant groups resulted from double crossover but the frequency was low accounting 5.7 % for all BC1F5 recombinants (26 out of 456 recombinants). Within the PHS interval PHS 2B.1, these recombinants contained two separated Cayuga fragments whereas Caledonia fragment was the middle.

In this study, the new markers were designed from the reported wheat ESTs, their contigs, transcripts and SD-affecting known-function genes on the comparative map. This may help determine the new closest marker to the gene or determine the possible known-function gene(s) that may be the candidate gene(s). The primary limitation in this study was low recombination in the target region. This problem can be resolved by developing a mapping population containing more recombinants in the targeted region. In our project, we developed two mapping populations by screening the recombinant individuals containing breaking points in the targeted region in the initial population. The recombinant individuals were screened by three flanking markers; Wmc474, Barc55 and Gwm429, underlying the PHS QTL on 2B1. The initial population was used to generate the BC1F4 and BC1F5 mapping populations. Marker-assisted selection was used to screen homozygous recombinants. However, the low polymorphism between parents, Cayuga and Caledonia, could not be avoided. Consequently, fine mapping helped to reveal more precise region(s) that could not be revealed by previous mapping in the DH population.

Candidate gene identification based on homologous region with rice

Fine mapping revealed that PHS QTL interval around 3 mb containing two closely linked QTLs. The PHS interval on wheat short arm of chromosome 2B was homologous with rice long arm of chromosome 7. The homologous region was inverted direction (Appendix Figure A2.1). The marker closely linked to Wmc453c was ss4.4 (2.52 cM) that was designed from EST BE606438 matching with TIGR rice

gene LOC_Os07g48100. The marker closely linked to Wmc474 was ss26.3 (1.26 cM) that was designed from EST BE494262 matching with TIGR rice gene LOC_Os07g43470. The distance between these two genes was 2.7 mb based on the rice physical map. Based comparative map, there were 120-128 kb per cM in rice. The distance between ss4.4 to the middle of PHS QTL interval between Wmc453c and Barc55 was 550-587 kb and the distance between ss26.3 to Wmc474 was 156 kb. So, the distance between two PHS QTLs was 2.11-2.26 mb based on the rice physical map.

For the first PHS QTL between Wmc453c and Barc55, there were around 83 rice genes on the homologous region from TIGR rice loci LOC_Os07g47230 to LOC_Os07g48100 (Appendix Figure A2.2). This homologous region was around 500 kb. Those genes included 45 known-function genes, 14 retrotransposon / transposon and 24 unknown-function genes. There were four genes known-function genes that might involve seed dormancy. Three genes were kinase genes (LOC_Os07g47230, LOC_Os07g47270 and LOC_Os07g47950) and one gene contained bHLH (basic helix-loop-helix) (LOC_Os07g47960). For the second PHS QTL between Wmc474 and rCaPK, there were around 71 rice genes in this PHS interval from TIGR rice loci LOC_Os07g42740 to LOC_Os07g43470 (Appendix Figure A2.3). This homologous region was also around 500 kb. Those genes included 43 known-function genes, 7 retrotransposon / transposon and 21 unknown-function genes. There were five genes known genes that might involve seed dormancy. Two genes; Calmodulin binding protein (LOC_Os07g42740) and Calmodulin dependent protein kinase (LOC_Os07g42770) were under rice SD QTL (Gu et al. 2004). The other three genes were serine/ threonine protein kinase SAPK2 (LOC_Os07g42940), Calmodulin-binding transcription activator1 (LOC_Os07g43030) and myb-DNA binding gene (LOC_Os07g43420).

Another SD QTL on rice chromosome 7 by Jiang et al. (2003) did not match with either wheat PHS QTL because it QTL was located between the comparative location of the two wheat PHS QTL. There were 2 rice SD QTLs on chromosome 3 that lie within the homologous chromosome region. One of these rice SD QTLs by Jiang et al. (2003) was close to first wheat PHS QTL and another by Cai and Morishima (2002) close to the second wheat PHS QTL. The distance between these rice QTLs on chromosome 3 was around 3.8 mb.

Calmodulin binding protein and Calmodulin dependent protein kinase, serine/ threonine protein kinase SAPK were reported as genes involved seed dormancy (Reddy et al. 2002; Sheen 1996; Guo et al. 2002) and reported as eQTL genes for PHS on chromosome 3D, on chromosome 2A, on chromosome 4D, respectively (Munkvold 2007). The most potential candidate gene was Calmodulin binding protein and Calmodulin dependent protein kinase (CDPK or CaMK) because they was used to design markers for fine mapping and linked with PHS QTLs. Rice CaMK-35 genes (LOC_Os07g42770, 5861 bp) was matched with CDPK genes from other plant species such as *Zea mays*, *Arabidopsis thaliana*, *Gossypium hirsutum*, *Populus trichocarpa*, *Ricinus communis*, *Triticum aestivum*. For hexaploid wheat, it matched with BT009309 cDNA clone that is CDPK related protein kinase (E-value = 6e-55). This transcripts was a candidate for genome walking

Candidate gene identification based on homologous region with Brachypodium and Ae. tauschii

The PHS QTL region between Wmc453c and Barc55 was delimited by SSR markers and no ESTs were mapped in this interval. A comparative analysis with *Brachypodium* (from *Brachypodium* EST DV484967 to DV483654 and around 866 kb), revealed 43 genes with known function and three unknown. Under the

homologous region, one candidate gene is hypothesized to be involved SD. This gene was a zinc-finger protein like protein (*Brachypodium* EST DV475774). This gene was located the middle of all 46 genes. This gene was homologous with rice gene LOC_Os07g39310 on rice chromosome 7 and wheat transcript BQ609341 that can be used for designing new marker. The expression of this gene has been studied in Arabidopsis and was reported to be up-regulated by abscisic acid (ABA) (He and Gan 2004). Interruption of ABA-responsive elements (ABREs) prevented this gene from responding to ABA. Also, Arabidopsis gene *DAG1* encoding a zinc-finger transcription factor was showed to be involved in the control of seed germination (Papi et al. 2000). So, this gene is a candidate gene for the wheat PHS QTL between Wmc453c and Barc55 if its location can be verified. Unfortunately, designed 7 primers pairs based on this gene were still monomorphic. The distance between these two markers was 4.2 cM, still too large for genome walking. Genome walking may be possible for the PHS QTL between Wmc474 and rCaPK (1.2 cM) and ss26.3 (1.3 cM). Calmodulin/Ca^{2+} dependent protein kinase has been reported to be involved in SD, via ABA and GA responsiveness (Reddy et al. 2002; Sheen 1996; Li et al. 2008).

One of ESTs in the PHS interval between Wmc453c and Barc55 was mapped on *Aegilops tauschii* BAC contig1347 (467 kb). Unfortunately, four flanking markers, Wmc453c, Barc55, Wmc474 and rCaPK did not located on those BACs. The *Ae. tauschii* BAC sequence may reveal some potential known-function genes that could be involved in SD, however, the intergenic sequence for the D genome is unlikely to facilitate genome walking in the B genome of wheat.

Genome walking for candidate gene identification

Genome walking was conducted on this project described on Appendix 3. However, there was not successful in wheat to use this technigue to identify candidate genes because only short unknown sequences were identified.

The advantages of PCR-based genome walking over traditional genome walking were no needing BAC libraries and radioactive labeling. However this technigue still has limitation, so many strategies are needed to improve the progress. Those factors were many enzyme choices, partial or full DNA digestion, type of specific primers and how many round of PCR reaction. For the first factor, eight restriction enzymes with 6 base pair recognition with 3'overhang formed; *Apa*I, *Pst*I, *Sac*I and *Sph*I and blunt end formed; *Dra*I, *Eco*RV, *Pvu*II, and *Sca*I were used for genomic DNA cutting. As a result, the blunt end formed tended to give more products than 3'formed overhang. For the second factor, we used both partial and full digestion; we found that there were no different too much if we use many different enzymes for developing the libraries. For the third factors, the specific primers designed from genic region such as CaMK gene was successful to give a specific product than specific primers designed from non-genic region such as STS sequences or SSR marker. For the last factor, the nested PCR that use a nested primer at the second PCR tended to give more specific products. From our result, we are successful to isolate the unknown sequences surrounding Barc55 region and matching with wheat BACs containing candidate genes. However, the sequences were still short when comparing with the unknown sequences isolated from coding region CaMK gene, that longer more than one kb. However, these sequences were still very short to identify more genes comparing with direct BAC sequencing. The advantage was to isolate promoter, 5'UTR and 3' UTR where BAC sequencing can revealed gene content. Although those newly identified sequences can be used to find BACs on NCBI databases, it was

still less possible for wheat that have less BAC sequences than other species like rice. In our study, sequences surrounding Barc55 matching with a hexaploid wheat BAC and a diploid wheat BAC containing a kinase gene that located in both BACs. These two BACs were come from wheat chromosome 1A and 5A, respectively. These suggest that there is possible that duplicated region at Barc55 with other wheat chromosome 1 and 5. Also, Barc55 marker was also reported on chromosome 2B, 1B and 5B based on Graingene database. We still could not tell whether these BACs contain the homologous region with PHS interval on 2B or not because it is needed to prove more about the location of kinase genes on these two BACs on wheat chromosome 2B. For CaMK, we identify more sequence at upstream region and downstream region. This sequence can be used further for promoter isolation. Even though CaMK was the candidate gene, we still need to verify function first that there was any different between Cayuga and Caledonia. Genome walking by using genic sequences was high successful to isolated a specific product than using non-genic region that identified more non-specific products. However, this strategy investigates at small picture when comparing with comparative studies and traditional BAC identification. Also, wheat tremendously contains 90% of repetitive sequences leading to genome walking was difficult to identify candidate gene.

Origin of PHS Cayuga allele at PHS QTL on 2B

Genotyping showed that Caledonia alleles match those of Geneva on chromosome 2B, including the 2B PHS interval from marker gwm429 to rCaPK. Caledonia originated as an off-type of Geneva but chromosome 2B appears to be identical. Based on whole genome genotypes the relationship between Caledonia and Geneva suggests that Caledonia resulted from an outcross rather than a mutation (Long Xi Yu, personal communication). Cayuga was more closely related to Geneva

than was Caledonia. A recent study mapping QTL for PHS revealed that the major PHS QTL on chromosome 2B was contributed from the Cayuga parent (Munkvold et al. 2009). Although the resistance allele was believed to come from Clark's Cream, we could not identify the original Clark's Cream genotype that was used to develop Cayuga. However, the 132bp PHS allele at Barc55 was presented in both Golden Chief and Cayuga. Golden Chief was one of the parents that was used to develop Clark's Cream suggesting that a resistance allele at Barc55 may come from Golden Chief. It is possible that the original Clark's Cream that was used to develop Cayuga was heterogenous for alleles at Barc55.

Fine mapping has revealed that the PHS interval on chromosome 2BS contained two closely linked PHS QTL. The proximal QTL was between Wmc453c and Barc55 and the distal QTL was between Wmc474 and rCaPK. The PHS resistance allele at Barc55 that came from Golden Chief represented only the first QTL. The original parent that contributed the PHS resistance allele between Wmc474 and rCaPK is still unknown. However, association study of Clark's Cream related lines revealed both Cayuga alleles at Barc55 and Wmc474 associated with PHS resistance. Interestingly, KanQueen shares the most alleles with Cayuga at Wmc453c, Barc55, Wmc474 and rCaPK, which linked two closely linked QTL on the 2B PHS interval but it was not a parent of Clark's Cream.

There were three alleles at Barc55. The first allele (size 132 bp) was from Cayuga, Golden Chief and KanQueen, all PHS resistant lines. The second allele (size 128 bp) was from Caledonia and Geneva, both PHS susceptible lines. The third allele (size 124 bp) was from Clark's Cream (PHS resistant) and Clark's Cream-related lines PI 520756, PI520757, PI520758, PI 520759 and PI 520760 that were moderately PHS resistant. This suggests that the two alleles for Cayuga and for Clark's Cream are responsible for PHS resistance. Also, the association result of Clark's Cream-related

lines supported that these alleles for Cayuga and for Clark's Cream are responsible for PHS resistance. However, association in more white wheat lines will better increase statistical power of estimation. After evaluating of the association of Barc55 alleles with PHS score in 166 white wheat lines, we found that both PHS resistant and susceptible lines contained the same alleles. Based on BLUP mean values, the 132 bp Barc55 alleles was associated with the highest level of PHS resistance but was not significantly different from the other alleles. This suggests that alleles at Barc55 might contribute to PHS resistance but other loci are also important. However, this could be verified by haplotyping more markers at the PHS QTL on other chromosomes

Future direction

In this project, only the gene for Calmodulin/Ca^{2+} dependent protein kinase was revealed to be a candidate gene, but the role of this gene in SD and PHS will need to be verified. The other QTL candidate genes were not revealed from fine mapping due to a lack of polymorphic markers. Candidate genes based on comparative analysis will need to verify the location on wheat genome.

REFERENCES

Alfares W, Bouguennec A, Balfourier F, Gay G, Berges H, Vautrin S, Sourdille P, Bernard M, Feuillet C (2009) Fine Mapping and Marker Development for the Crossability Gene SKr on Chromosome 5BS of Hexaploid Wheat (Triticum aestivum L.). Genetics genetics doi:10.1534/genetics.109.107706

Alpert KB, Tanksley SD (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: a major fruit weight quantitative trait locus in tomato. Proc Natl Acad Sci USA 93:15503-15507

Anderson JA, Sorrells ME, Tanksley SD (1993) RFLP Analysis of Genomic Regions Associated with Resistance to Preharvest Sprouting in Wheat. Crop Sci 33:453-459

Cai H, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. Theor Appl Genet 104:1217-1228

Chung J, Babka HL, Graef GL, Staswick PE, Lee DJ, Cregan PB, Shoemaker RC, Specht JE (2003) The Seed Protein, Oil, and Yield QTL on Soybean Linkage Group I. Crop Sci 43:1053-1067

Distelfeld A, Uauy C, Olmos S, Schlatter A, Dubcovsky J, Fahima T (2004) Microcolinearity between a 2-cM region encompassing the grain protein content locus Gpc-6B1 on wheat chromosome 6B and a 350-kb region on rice chromosome 2. Funct Integr Genomics 4:59-66

Francia E, Barabaschi D, Tondelli A, LaidÃ² G, Rizza F, Stanca A, Busconi M, Fogher C, Stockinger E, Pecchioni N (2007) Fine mapping of a Hv CBF gene cluster at the frost resistance locus Fr-H2 in barley. TAG Theoretical and Applied Genetics 115:1083-1091

Gu XY, Liu T, Feng J, Suttle JC, Gibbons J (2009) The qSD12 underlying gene promotes abscisic acid accumulation in early developing seeds to induce primary dormancy in rice. Plant Mol Biol doi:10.1007/s11103-009-9555-1

Gu XY, Kianian SF, Foley ME (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (Oryza sativa). Genetics 166:1503-1516

Han F, Ullrich SE, Clancy JA, Romagosa I (1999) Inheritance and fine mapping of a major barley seed dormancy QTL. Plant Science 143:113-118

He Y, Gan S (2004) A novel zinc-finger protein with a proline-rich domain mediates ABA-regulated seed dormancy in Arabidopsis. Plant Mol Biol 54:1-9

Heyne EG (1956) Earl G. Clark, Kansas Farmer Wheat Breeder. Transactions of the Kansas Academy of Science (1903-) 59:391-404

Imtiaz M, Ogbonnaya FC, 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

Jiang L, Cao YJ, Wang CM, Zhai HQ, Wan JM, Atsushi Y (2003) Detection and analysis of QTL for seed dormancy in rice (Oryza sativa L.) using RIL and CSSL population. Acta Genetica Sinica 30:453-458

Kim D, Kim K, Van K, Kim M, Lee SH (2010) Fine mapping of a resistance gene to bacterial leaf pustule in soybean. Theor Appl Genet doi:10.1007/s00122-010-1266-0

Li AL, Zhu YF, Tan XM, Wang X, Wei B, Guo HZ, Zhang ZL, Chen XB, Zhao GY, Kong XY, Jia JZ, Mao L (2008) Evolutionary and functional study of the CDPK gene family in wheat (Triticum aestivum L.). Plant Mol Biol 66:429-443

Li D, Chen L, Jiang L, Zhu S, Zhao Z, Liu S, Su N, Zhai H, Ikehashi H, Wan J (2007) Fine mapping of S32(t), a new gene causing hybrid embryo sac sterility in a Chinese landrace rice (Oryza sativa L.). Theor Appl Genet 114:515-524

Lin SY, Sasaki T, Yano M (1998) Mapping quantitative trait loci controlling seed dormancy and heading date in rice, Oryza sativa L., using backcross inbred lines. Theor Appl Genet 96:997-1003

Manly KF, Cudmore, Jr.,Robert, H., Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. Mammalian Genome 12:930-932

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. Theor Appl Genet 111:1357-1364

Munkvold J, Tanaka J, Benscher D, Sorrells M (2009) Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. Theor Appl Genet 119:1223-1235

Nichols DM, Glover KD, Carlson SR, Specht JE, Diers BW (2006) Fine Mapping of a Seed Protein QTL on Soybean Linkage Group I and Its Correlated Effects on Agronomic Traits. Crop Sci 46:834-839

Ogbonnaya F, Imtiaz M, Ye G, Hearnden P, Hernandez E, Eastwood R, van Ginkel M, Shorter S, Winchester J (2008) Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. Theor Appl Genet 116:891-902

Oh K, Hardeman K, Ivanchenko MG, Ellard-Ivey M, Nebenfuhr A, White TJ, Lomax TL (2002) Fine mapping in tomato using microsynteny with the Arabidopsis genome: the Diageotropica (Dgt) locus. Genome Biol 3:research0049

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

Papi M, Sabatini S, Bouchez D, Camilleri C, Costantino P, Vittorioso P (2000) Identification and disruption of an Arabidopsis zinc finger gene controlling seed germination. Genes Dev 14:28-33

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine Mapping of Quantitative Trait Loci Using Selected Overlapping Recombinant Chromosomes, in an Interspecies Cross of Tomato. Genetics 124:735-742

Reddy VS, Ali GS, Reddy ASN (2002) Genes Encoding Calmodulin-binding Proteins in the Arabidopsis Genome. J Biol Chem 277:9840-9852

Röder M, Huang XQ, Börner A (2008) Fine mapping of the region on wheat chromosome 7D controlling grain weight. Funct Integr Genomics 8:79-86

Sheen J (1996) Ca2+-dependent protein kinases and stress signal transduction in plants. Science 274:1900-1902

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. Proc Natl Acad Sci USA doi:10.1073/pnas.0911965107

Takeuchi Y, Lin SY, Sasaki T, Yano M (2003) Fine linkage mapping enables dissection of closely linked quantitative trait loci for seed dormancy and heading in rice. Theor Appl Genet 107:1174-1180

Tan MK, Sharp PJ, Lu MQ, Howes N (2006) Genetics of grain dormancy in a white wheat. Aust J Agric Res 57:1157-1166

Tanksley SD (1993) Mapping Polygenes. Annu Rev Genet 27:205-233

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. Science 314:1298-1301

Upadhyay MP, Morris CF, Paulsen GM (1988) Characterization of preharvest sprouting resistance in Clark's Cream white winter wheat. Euphytica 38:85-92

Vladutu C, McLaughlin J, Phillips RL (1999) Fine mapping and characterization of linked quantitative trait loci involved in the transition of the maize apical meristem from vegetative to generative structures. Genetics 153:993-1007

Yamamoto T, Kuboki Y, Lin SY, Sasaki T, Yano M (1998) Fine mapping of quantitative trait loci Hd-1, Hd-2 and Hd-3, controlling heading date of rice, as single Mendelian factors. Theor Appl Genet 97:37-44

Yamamoto T, Yano M (2008) Detection and Molecular Cloning of Genes Underlying Quantitative Phenotypic Variations in Rice. Rice Biology in the Genomics Era, pp 295-308

Yamamoto T, Lin H, Sasaki T, Yano M (2000) Identification of Heading Date Quantitative Trait Locus Hd6 and Characterization of Its Epistatic Interactions With Hd2 in Rice Using Advanced Backcross Progeny. Genetics 154:885-891

Yano M (2001) Genetic and molecular dissection of naturally occurring variation. Curr Opin Plant Biol 4:130-135

Zanetti S, Winzeler M, Keller M, Keller B, Messmer M (2000) Genetic Analysis of Pre-Harvest Sprouting Resistance in a Wheat Spelt Cross. Crop Sci 40:1406-1417

APPENDIX 1

PCR PRODUCT VISUALIZATION OF NEWLY DEVELOPED MARKERS

Figure A1.1. New polymorphic markers that visualize on 4% polyacrylamide gels.

The first lane is the PCR product from the Cayuga parent (score 0) and the second lane is the PCR product from the Caledonia parent (score 1). The rest of the lanes are PCR products from progeny. Arrows represent the location of polymorphic bands. The location of marker loci is in parentheses.

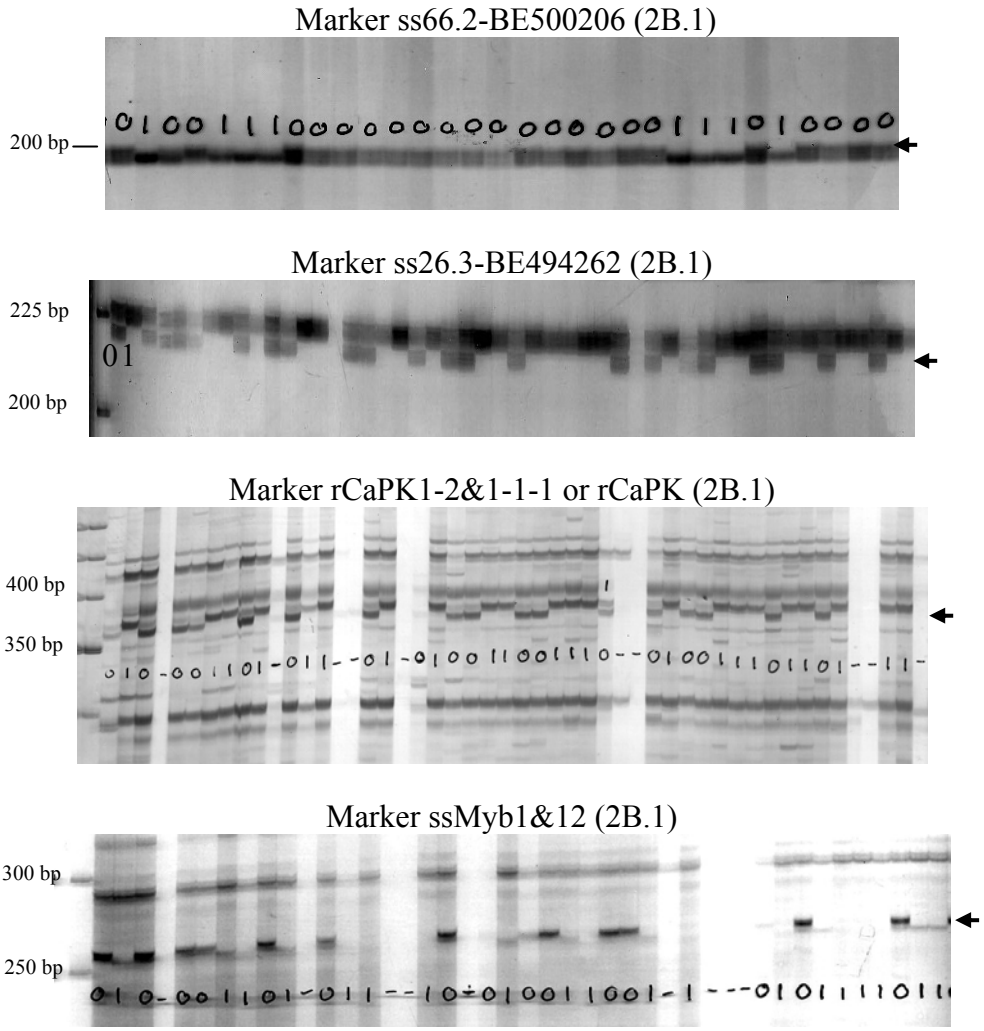


Figure A1.1. (Continued)

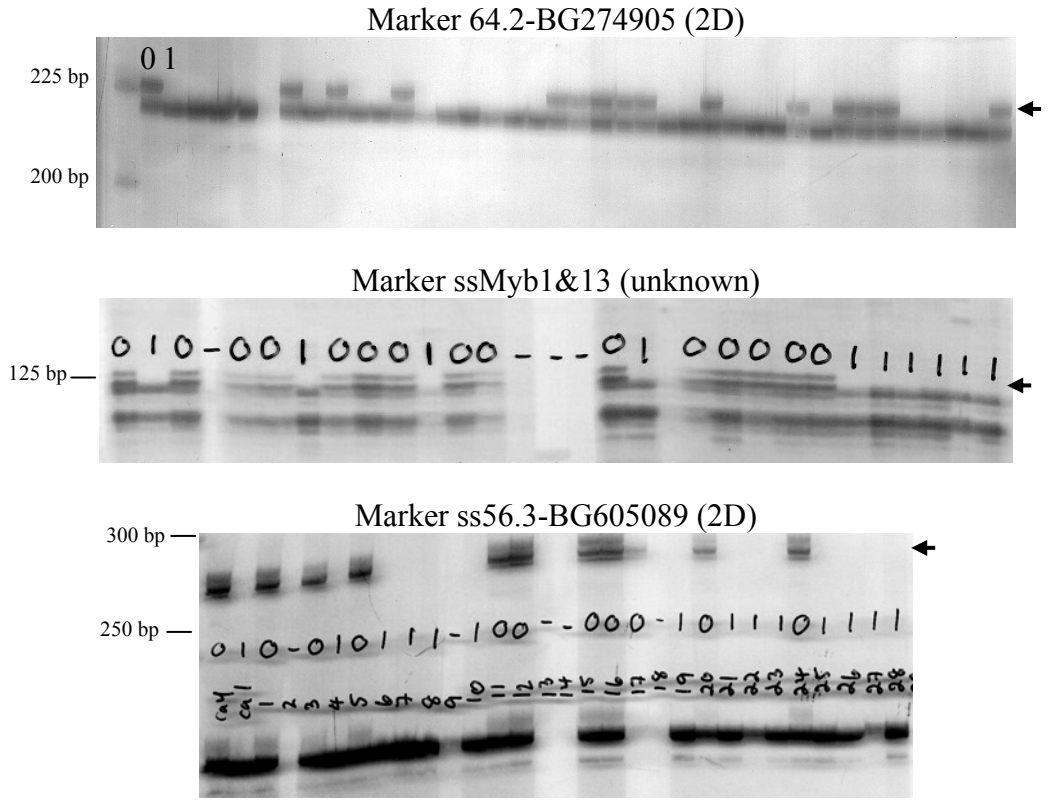


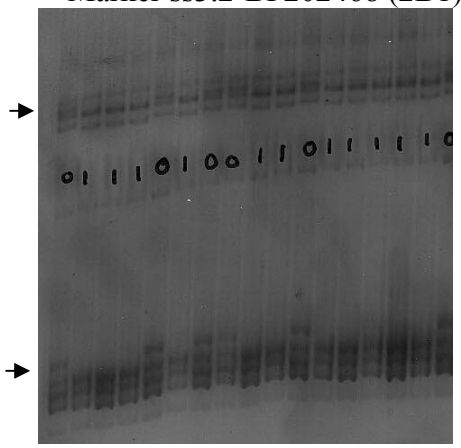
Figure A1.2. New polymorphic markers that visualize on 10% SSCP gels.

The first lane is the PCR product from the Cayuga parent (score 0) and the second lane is the PCR product from the Caledonia parent (score 1). The rest of lanes are the PCR products of progeny. Arrows represent the location of polymorphic bands.

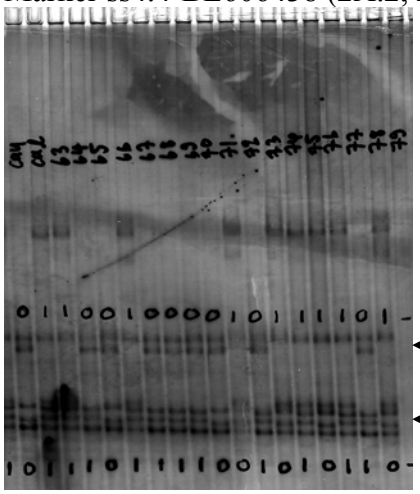
The location of marker loci is in parentheses.

Marker ss57.2-BE405569 (2B.1)

Marker ss3.2-BF202468 (2B1)



Marker ss4.4-BE606438 (2A.2, 2B.1)



Marker 44-BE636824 (2B.1)

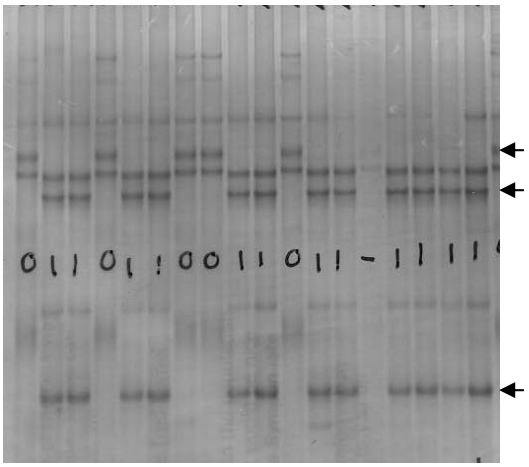
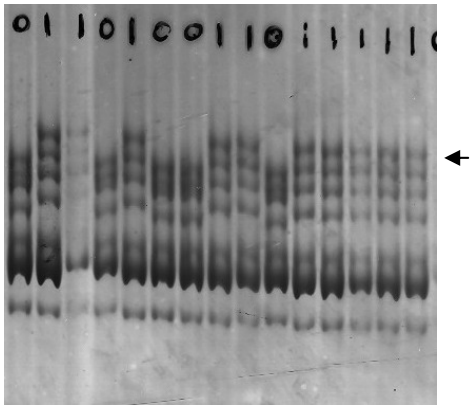
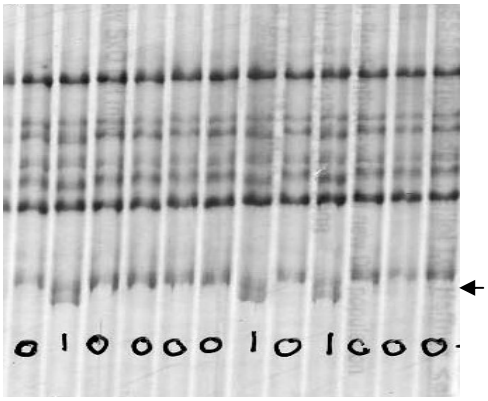


Figure A1.2. (Continued)

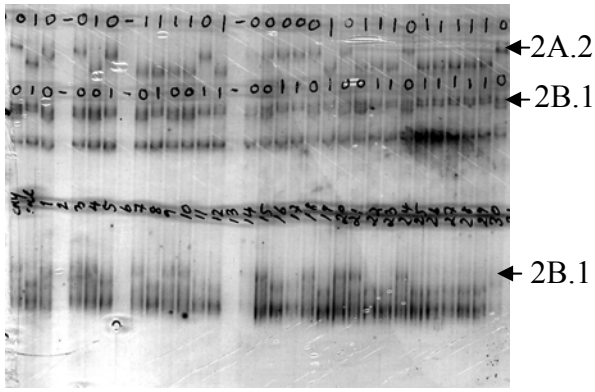
Marker ss47.5-BE498254 (2B.1)



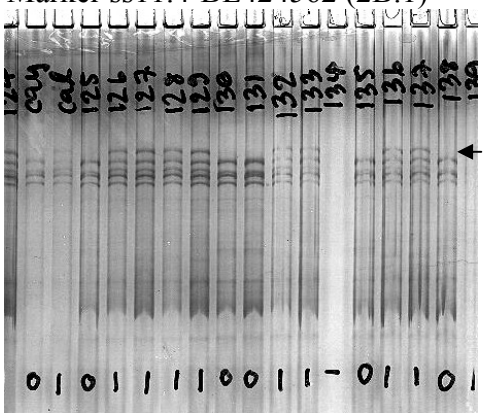
Marker ss31-BQ170428 (2B.1)



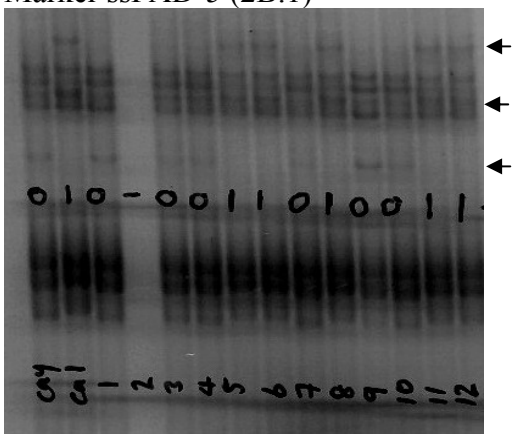
Marker ss20- BQ294702 (2A.2, 2B.1)



Marker ss11.4-BE424562 (2B.1)



Marker ssFAD-5 (2B.1)



Marker ss11.1-BE424562 (2B.1)

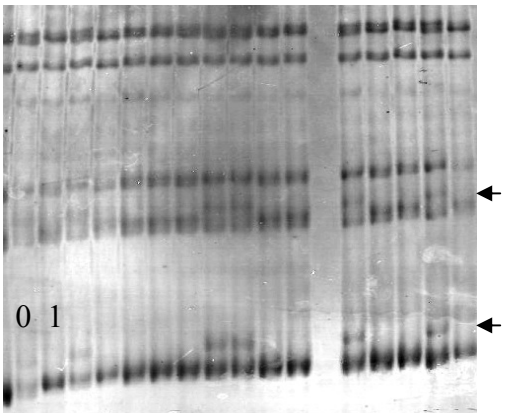
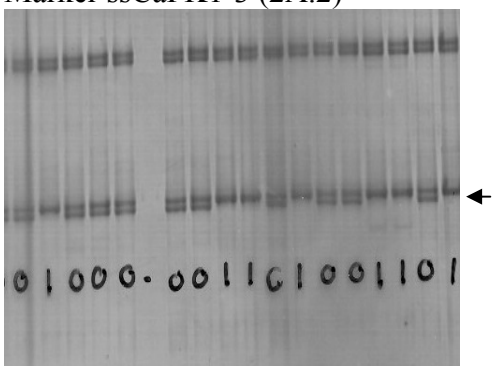
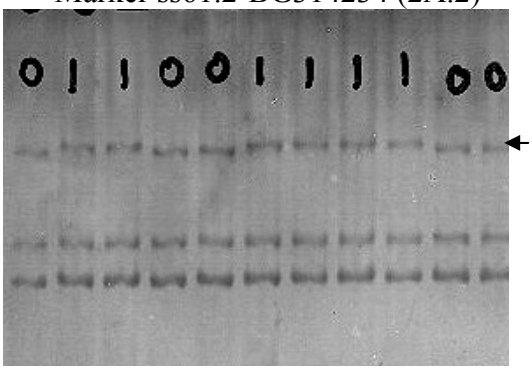


Figure A1.2. (Continued)

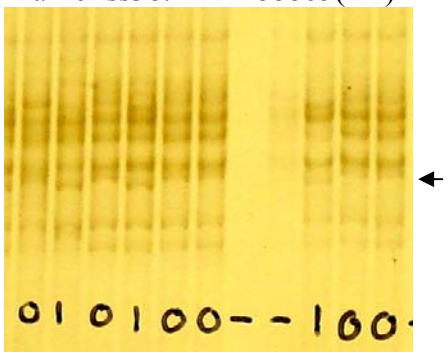
Marker ssCaPK1-3 (2A.2)



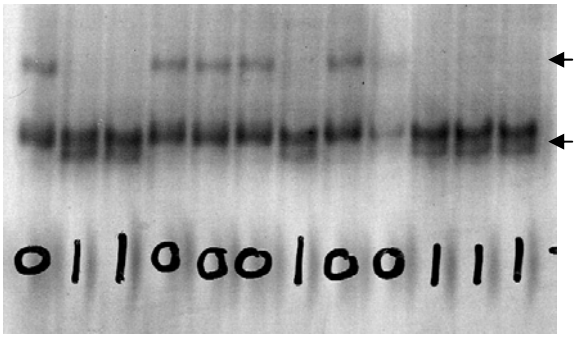
Marker ss61.2-BG314234 (2A.2)



Marker ss38.1-BE488865(2D)



Marker 56.2-BG605089 (2A.2)



APPENDIX 2

COMPARATIVE ANALYSIS OF THE HOMOLOGOUS REGION AT TWO PHS QTL WITH RICE CHROMOSOME 7

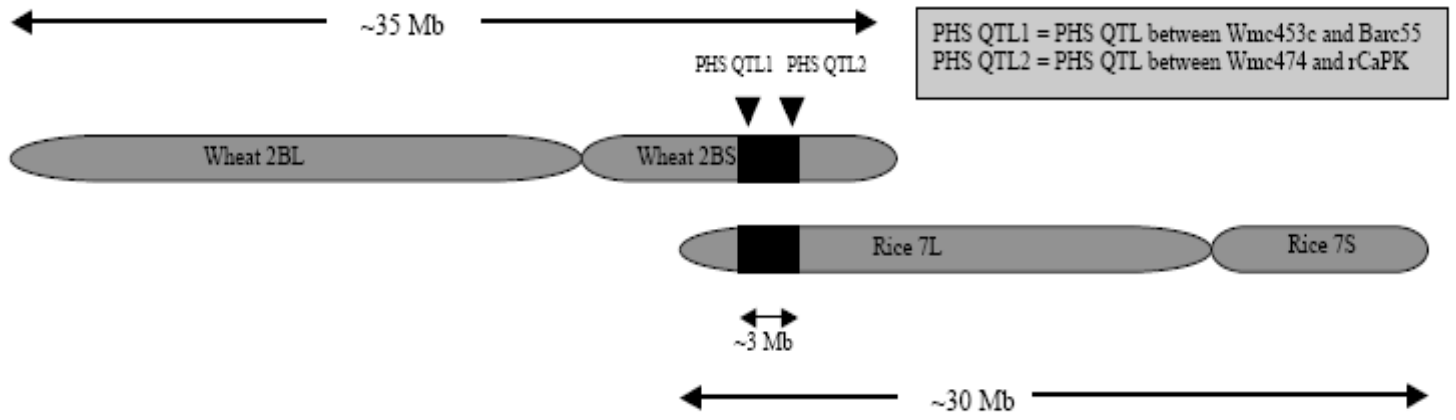


Figure A2.1. Illustration of the homologous region at PHS QTL interval on 2B (black area) on wheat short arm of chromosome 2B with rice long arm of chromosome 7.
The homologous region was around 3 Mb. The distance between two PHS QTLs was around 2.2 Mb, based rice physical map.

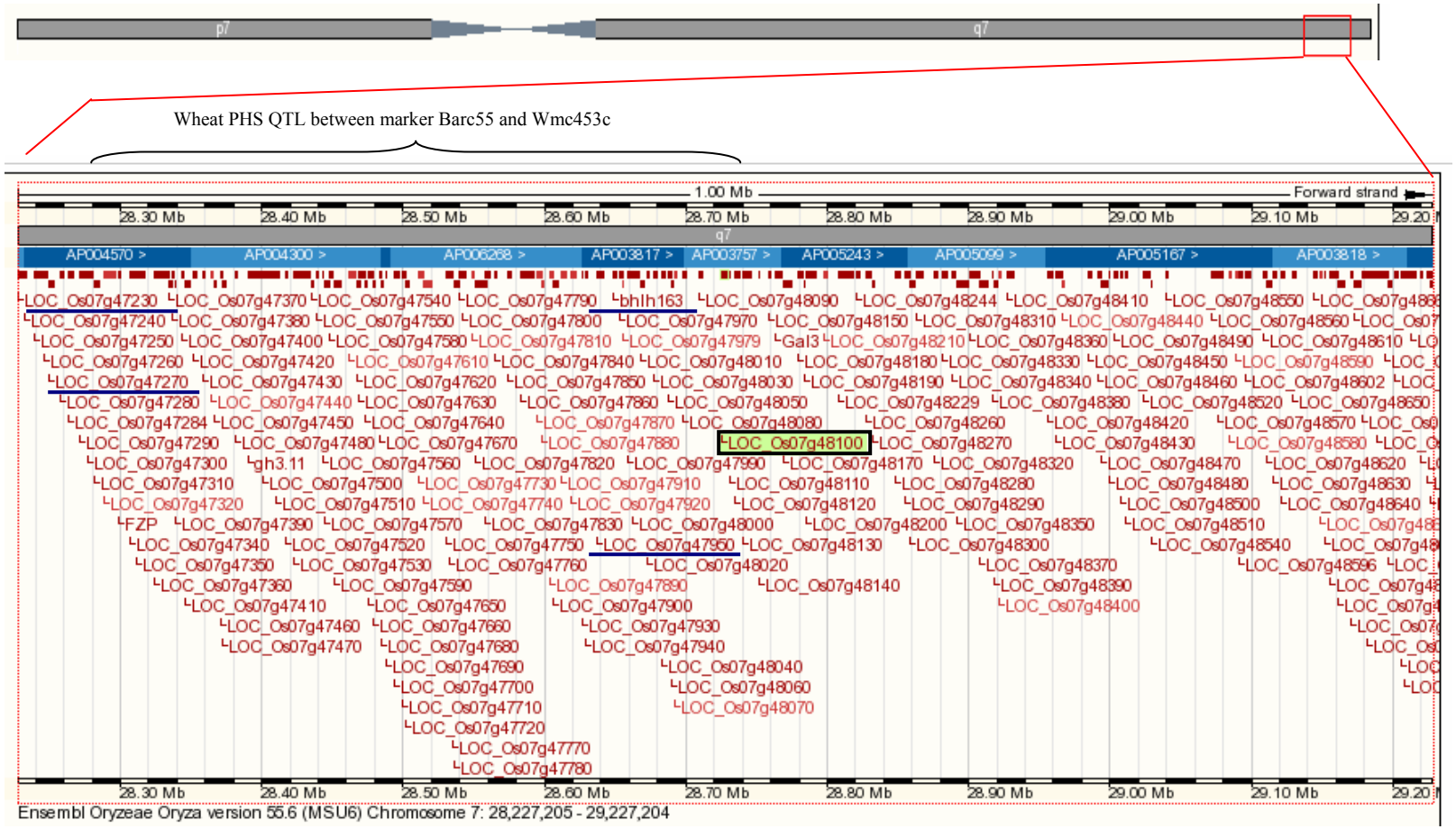


Figure A2.2. The homologous rice region with wheat PHS QTL between Barc55 and Wmc453c.
There were around 83 genes in this PHS interval (From TIGR loci LOC_Os07g47230 to LOC_Os07g48100) including 45 known-function genes, 14 retrotransposon / transposon and 24 unknown genes. There were four genes known genes that might involve seed dormancy (blue-underlined genes). Three genes were kinase genes and one gene was bHLH (beta helix-loop-helix) (LOC_Os07g47960). The TIGR rice loci on rectangular was the landmark of a homologous gene with maker ss4.4- BE606438

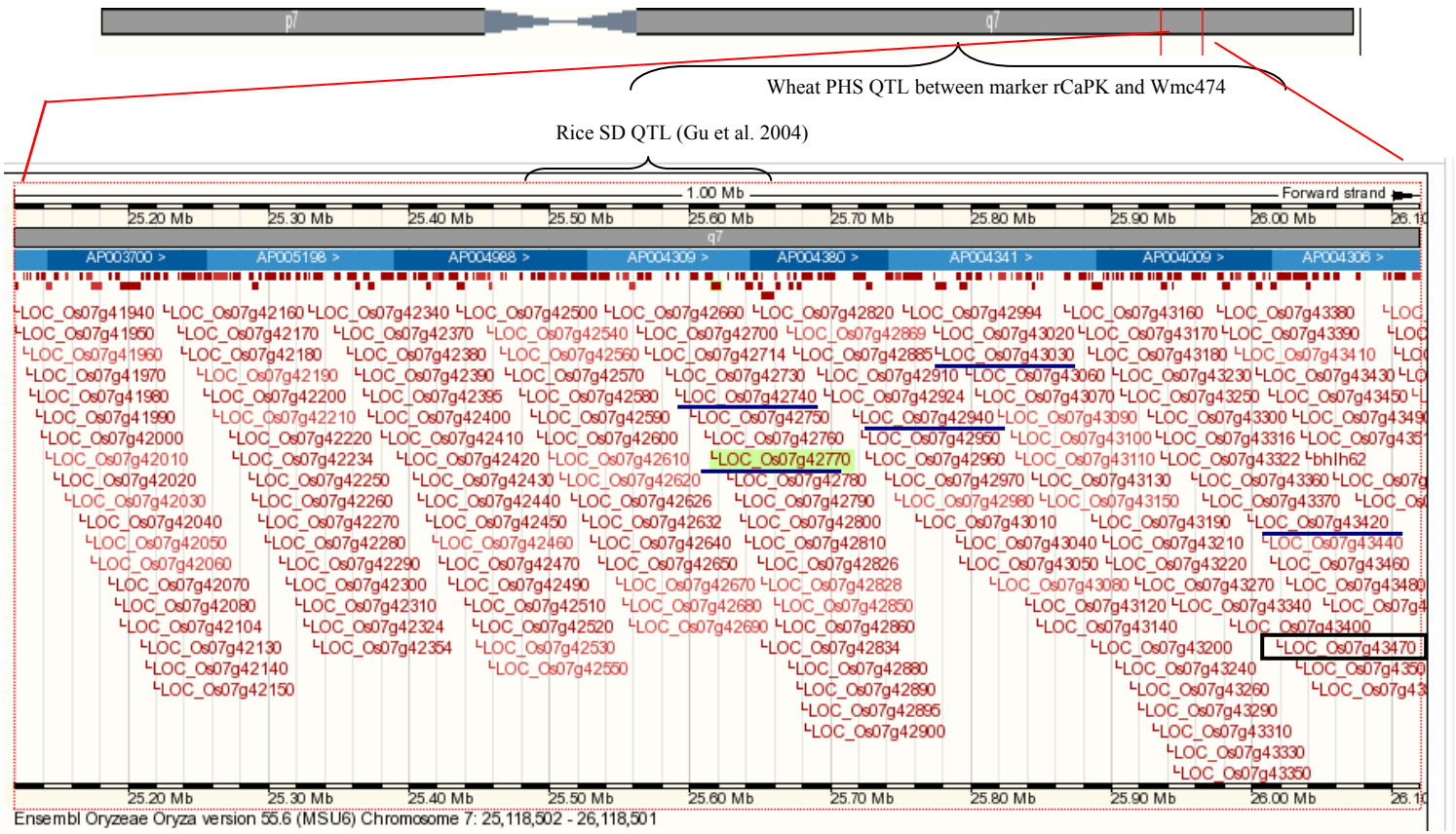


Figure A2.3. The homologous rice region with wheat PHS QTL between rCaPK and Wmc474.
There were around 71 genes in this PHS interval (From TIGR loci LOC_Os07g42740 to LOC_Os07g43470) including, 43 known-function genes, 7 retrotransposon / transposon and 21 unknown genes. There were five genes known genes (blue-underlined genes) that might involve seed dormancy. Two genes; Calmodulin binding protein (LOC_Os07g42740) and Calmodulin dependent protein kinase (LOC_Os07g42770) matched with rice SD QTL (Gu et al. 2004). The TIGR rice loci on rectangular was the landmark of a homologous gene with maker ss26.3- BE494262.

APPENDIX 3

GENOME WALKING FOR THE PREHARVEST SPROUTING INTERVAL ON CHROMOSOME 2B

INTRODUCTION

Genome walking could be one of strategies used to identify genes. It was used to isolate unknown sequences from known regions. The traditional approach was to probe on libraries, a time-consuming process that requires radioactive labeling. The technique used here is a PCR-based method that does not require the BAC libraries. Hexaploid wheat has no public BAC libraries or physical map of chromosome 2B. So, knowing more sequences around flanking markers might help us to identify unknown sequences around flanking markers. PCR-based genome walking can be used to identify promoters, 5' untranslated regions (UTR) and 3' UTRs of the genes and walk bi-directionally from sequence tagged sites (STSs) or expressed sequence tags (ESTs). The identified sequences may be useful for gene prediction or NCBI searching for BACs containing those sequences. There has been successful cloning of promoter sequences by the PCR genome walking technique. For example, promoters involved in auxin transport were identified in pea (Chawla and DeMason 2003). In wheat, a 579 bp fragment of TaPT2 (wheat high-affinity phosphate transporter) promoter was isolated using the Universal Genome Walking Kit (Tittarelli et al. 2007). However, genome walking has limitations. It amplifies small fragments and sometimes fails to amplify fragments. This can be overcome by using different restriction enzymes and partially digesting genomic DNA. A study in poplar reported isolation of a fragment up to 4 kb by using partial digestion whereas no clear amplified products resulted by using full digestion (Rishi et al. 2004).

The objective of this work were to use PCR-based genome walking to identify the

unknown sequence bi-directionally surrounding flanking markers and coding sequence of a candidate gene, Calmodulin/Ca^{2+} dependent protein kinase (CDPK or CaMK)

MATERIAL AND METHODS

Three flanking markers, Barc55, ss26.3, Wmc474 and a candidate gene, Calmodulin dependent protein kinase gene (CaMK) were the targets that were used for genome walking.

1) Large scale clean DNA extraction

Four grams of leaf tissue from Cayuga and Caledonia were used for large scale extraction, separately. This method was modified from general large scale DNA extraction. Liquid nitrogen was used to freeze the leaf tissues. The frozen tissues were ground in a well chilled mortar and pestle. The fine powder was transferred to a new 50 ml tube and then added the extraction buffer (0.5 M NaCl, 0.1 Tris-HCL pH 8.0, 0.05 M EDTA and 1.25% (w/v) SDS). Before adding the extraction buffer to the fine tissue power, 0.38 g. sodium bisulfite was added per100 ml extraction buffer pH 7.8-8.0 (with 4N NaOH), and heat to 65°C. Extraction buffer (18 ml) was added to the fine powder and then well mixed. The mixing solution was incubated at 65 °C for 20-30 min and inverting tubes every 5-10 min. Next, 15 ml of phenol: chloroform: isoamyl alcohol (25:24:1 ratio) was added to the solution, mixed vigorously and then inverted by hand for 15 min. After centrifuging for 15 min at 4,000 rpm by Table top centrifuge (Eppendorf Centrifuge 5810), the upper phase was poured into a new 50 ml tube through one layer of Miracloth (Calbiochem Lot. 407592). Pipetting the upper phase was not recommended because the genomic DNA might be broken. The equal volume of chloroform: isoamyl alcohol (24:1) was added the supernatant, mixed vigorously and then inverted by hand for 15 min. After centrifuging for 15 min by Table top centrifuge, the upper phase was transferred with large hole tips. RNase (5µl

of a 10 μg/ml stock) was added to remove RNA and then incubated at 37 °C for 15-30 min. To precipitate DNA, 2 volume of 95% ethanol was added and mixed well. Due to the large amount of DNA, DNA was scooped with u-shaped Pasteur pipette and transferred to a new 1.5 ml microcentrifuge tube. The DNA was washed with 70% ethanol and repeated with fresh 70% ethanol until discolored and then let it air dry. DNA was dissolved by 1 ml of TE and incubate at 65 °C until dissolved with gentle tapping every 10 min and do not leave over 40 min. The dissolved DNA was centrifuged at 12,000 rpm for 10 min and then transferred to a new 1.5 ml tube.

2) Genomic DNA digestion, purification and adaptor ligation.

There were 8 restriction enzymes with 6 base pair recognition with 3'overhang formed; *Apa*I, *Pst*I, *Sac*I and *Sph*I and blunt end formed; *Dra*I, *Eco*RV, *Pvu*II, and *Sca*I used for genomic DNA cutting. Adapter sequences and restriction enzyme selection followed the genome walking protocol done by Clontech, USA (Ashoub and Abdalla 2006; Siebert et al. 1995). Two different adapter sequences were used. First adapter was used to ligate with digested genomic DNA after treatment with 3'overhang formed enzymes; *Apa*I, *Pst*I, *Sac*I and *Sph*I. Second adapter was used to ligate digested genomic DNA after treatment with blunt end formed enzymes; *Dra*I, *Eco*RV, *Pvu*II, and *Sca*I. Both adapter sequences were shown on Table A3.1.

For genomic DNA digestion, 10 μg DNA from each parent and 50 units of enzyme were used for full digestion and 3μg DNA and 20 units of enzyme was used for partial digestion. The reaction volume was 50 μl. Amount of a digestion enzyme and time length for each enzyme was different. Partial digestion time with *Apa*I was 1.5 hr, with *Pst*I was 2.5 hr, with *Sac*I was 0.5 hr and with *Sph*I was 1 hr. Full digestion time with *Dra*I, *Eco*RV, *Pvu*II and *Sca*I was 18 hrs or overnight.

Table A3.1. Adapter and adapter primer sequences.

First Adapter type (Adapter I) for 3'overhang formed enzymes	
Primer name	Sequences
Adapter I-*Apa*I Adapter I -*Pst*I Adapter I -*Sac*I Adapter I-*Sph*I Adapter primer I (AP1)	5'GAATTCGAGCTCGCCCGGGATCCTCTAGA**GGCC- 3'** 5'GAATTCGAGCTCGCCCGGGATCCTCTAGA**TGCA- 3'** 5'GAATTCGAGCTCGCCCGGGATCCTCTAGA**AGCT- 3'** 5'GAATTCGAGCTCGCCCGGGATCCTCTAGA**CAGT- 3'** 5'GAATTCGAGCTCGCCCGGGAT-3'
Second Adapter type (Adapter II) for blunt end formed enzymes	
Primer name	Sequences
Adapter II- *Dra*I Adapter II- *Eco*RV Adapter II- *Pvu*II Adapter II- *Sca*I	5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCC**GGGCAGGT**-3' 3'-H2N-**CCCGTCCA**-P-5'
Adapter II primer (Adp1) Adapter II primer (Adp2)	5'-GGATCCTAATACGACTCACT**ATAGGGC**-3' 5'-A**ATAGGGC**TCGAGCGGC-3'

After digestion, 30 µl of sterile distilled water was added to each tube. The digested DNA was purified by adding 80 µl of phenol: chloroform: isoamyl alcohol (25:24:1). After mixing by vortex 3-5 times and centrifuging at 4,000 rpm for 10 min, the supernatant was transferred to new PCR tubes. The purification was performed more by adding 80 µl of chloroform: isoamyl alcohol (24:1) and mixed by vortex for 3-5 times. After centrifuge at 4,000 for 10 min, the supernatant was transferred to new 1.5 ml tubes, the digested DNA was precipitated by adding 150 µl cold 95% ethanol and incubate at -20 °C for 1 hr. After centrifuge at 12,000 rpm for 10 min, the DNA was washed with 70% ethanol. Dry digested DNA was dissolved in 20 µl of TE.

For adaptor ligation, the purified-digested DNA was diluted 10 times. 20 µl of ligation contained 16 µl of diluted digested DNA , 1 µl of 100 µM of Adaptor I or Adapter II , 2µl of T4 ligase buffer and 1 µl of T4 ligase. The ligation reaction was incubated overnight at 16 °C.

3) PCR reaction with adaptor primers and specific primers

Two types of primers used for PCR-based genome walking were specific primers and adapter primers binding with an adaptor site. For Adapter I binding, adapter I primer (AP1) was used for all digested genomic DNA libraries with 3' overhang formed with *Apa*I, *Pst*I, *Sac*I and *Sph*I whereas NAP1 was used for second PCR. Adapter II primers (Adp1, Adp2) were used for digested genomic libraries with blunt end formed with *Dra*I, *Eco*RV, *Pvu*II, and *Sca*I. Adp1 was used for first PCR whereas Adp2 was used for second PCR or nested PCR. The adapter sequences were shown on Table A3.1. Specific primer sequences for flanking marker Barc55, ss26.3 and Wmc474, and a candidate gene (CaMK) detail were shown on Table A3.2. The PCR reaction extension time was longer than normal time up to 5 min that should amplify up to 3-5 kb. Taq DNA polymerase was used. The PCR product was cloned directly to a vector by TOPO TA cloning kit.

Table A3.2. Specific primers designed based on marker flanking sequences and a candidate sequence.

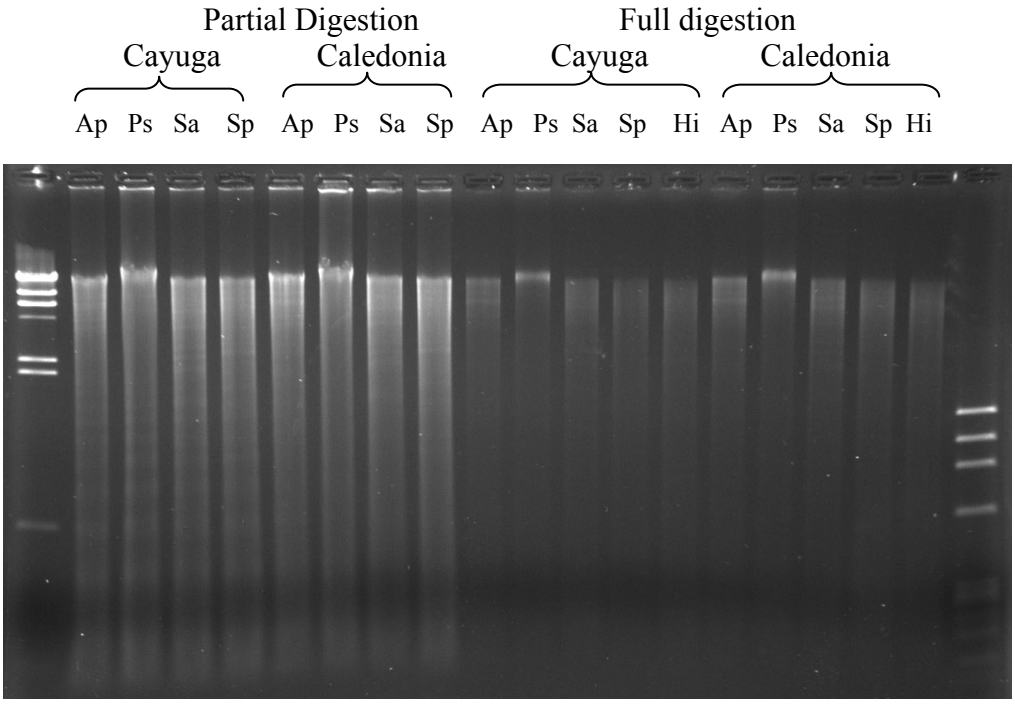
Flanking marker name or gene (STS, EST)	Type of Adapter primer	Specific primer name	Primer sequence
Barc55 (BV211787)	AP1 (First PCR)	Barc55Fv Barc55Rv Barc55IRv1	5'- gcggtcaacacactccactcctctctc-3' 3'- cgctgctcccattgctcgccgtta-3' 5'- cggagccaccagatcctcca;3'
	NAP1 (Second PCR)	Barc55iF2	5'-gaggagatcaccactctagagaagacta-3'
	AdP1(first PCR)	Barc55-iR1 Barc55-iF1	5'-atagatagatggagagaggagtggagt-3' 3'-ctatctatctctcttccccctctctt-3'
	AdP2 (Second PCR)	Barc55-iR2 Barc55-iF2	5'-gtgttgactttcgtctagggttcatt-3' 5'-gaggagatcaccactctagagaagacta-3'
ss26.3 (BE494262)	AdP1(first PCR)	GTP-BP-iR1 GTP-BP-iF1	5'-aagttgtaaactcataggacgcaacc-3' 5'-ctggagttattatgtacaaaggggcta-3'
	AdP2(Second PCR)	GTP-BP-iR2 GTP-BP-iF2	5'-cctttggttaagaaagtgagcaactc-3' 5'-atgtacaaaggggctaaagtacagctt-3'
Wmc474	AdP1(first PCR)	Wmc474-iR1 Wmc474-iF1	5'-agtggaaacatcattcctggtaagt-3' 5'-gctattaaactagcatgtgtcgtgt-3'
	AdP2(Second PCR)	Wmc474-iR2 Wmc474-iF2	5'-cgtatacatacacgacacatgctagtt-3' 5'-gtatgtatacggtgcgtgtactcatgt-3'
CaMK (BT009309)	AdP1(first PCR)	CAMK36-iR1 CAMK36-iF1	5'-ctttacaaagtcagacaagccaaagtc-3' 5-gtccaaacaagagtggtttaatctcc-3'
	AdP2(Second PCR)	CAMK36-iR2 CAMK36-iF2	5'-agtctatgaccttcaagggtgagtttt-3' 5'-ggtatgcacccttcagtacagaaaat-3'

The insert clones were screened by PCR with M13 primers. The colony was picked and shaking in PCR reaction that has M13 primers. The PCR condition was normal at 55 °C annealing temperature but the beginning denaturing time before entering the PCR cycle is 10 min. Some PCR products were run on 1% agarose gel. The expected PCR products were sent for sequencing. Before sequencing, the M13 primers was removed by using ExoSap (Exonuclease I from BioLabs, Shrimp Alkaline Phosphatase from Promega) because in sequencing only one primer M13 was used either forward or reverse primer. The sequencing was performed at Cornell University Life Science Core Laboratories center. The sequences were analyzed by Vecter NTI (Lu and Moriyama 2004), NCBI Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and ClustalW (http://www.ebi.ac.uk/Tools/clustalw/). Before, sequences were modified to be complementary and inverse sequences by using sequence editor http://www.fr33.net/seqedit.php.

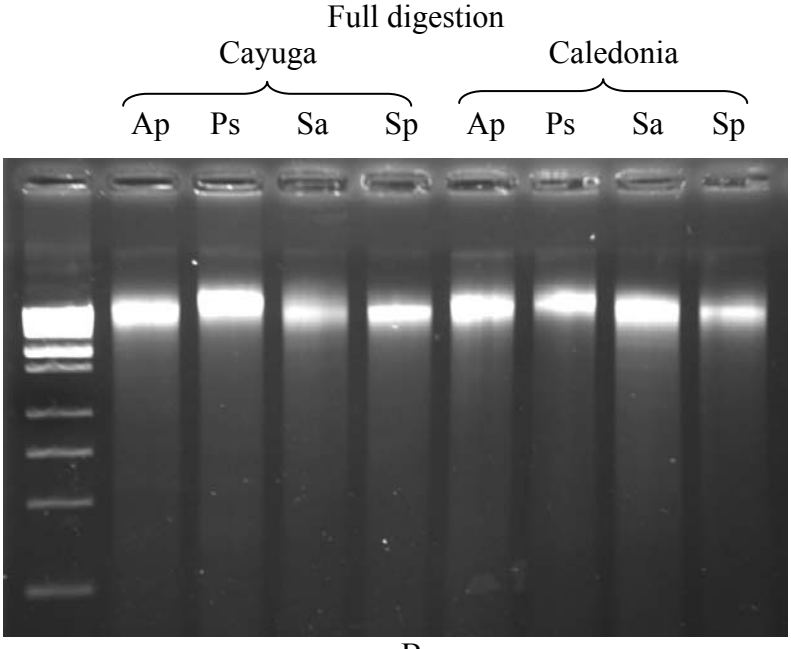
RESULTS

Genome walking for Barc55 with adapter I primer for 3'overhang formed genomic libraries

After extracting four grams of leaf tissues from Cayuga and Caledonia, four grams of leaf tissue yielded 900 μg of genomic DNA. Three micrograms of genomic DNA each parent was used for partial digestion (Figure A3.1A) and 10 μg of genomic DNA each parent was used for full digestion (Figure A3.1B). Due to the large amount of genomic DNA for full digestion result, it still was partial digestion. The purified partial digested DNA was used for adaptor ligation.



A.

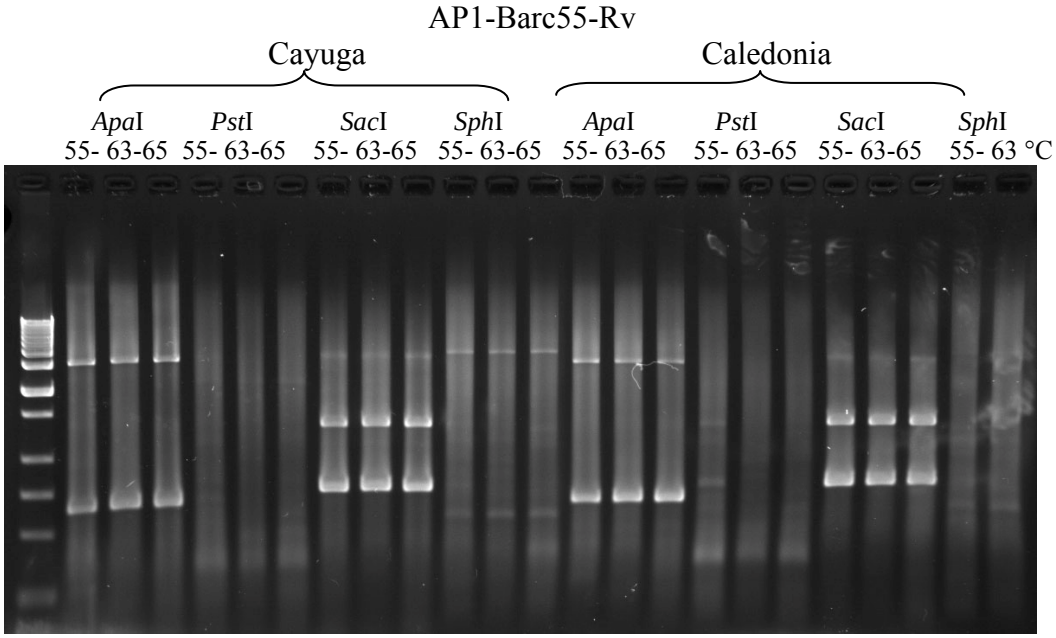


B.

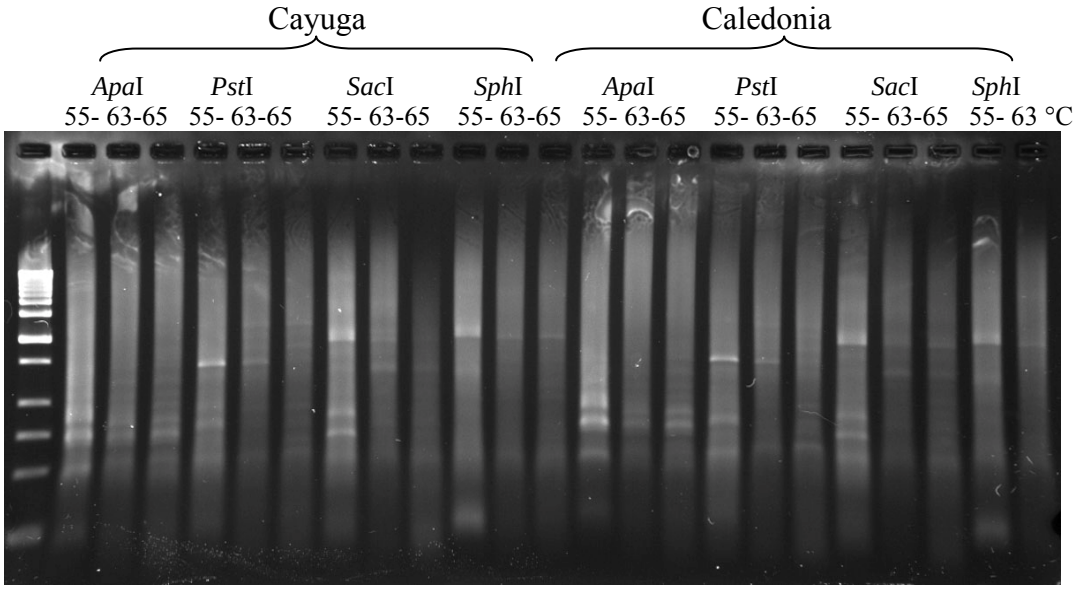
Figure A3.1. Wheat genomic DNA digestion with 3'overhag end formed enzymes. Partial digestion of 3 µg DNA with 20 unit enzyme for 1.5, 2.5, 0.5 and 1 hr (L--> R) Full-digestion of 1.5 µg DNA with 100 unit enzyme for 18 hours (A). The full digestion of 10 µg DNA with 50 unit enzyme (B). Ap = *Apa*I, Ps =*Pst*I, Sa = *Sac*I, Sp=*Sph*I, Hi = *Hind*III

After adding with adaptor, the digested DNA that added with 4 different adaptors was used for PCR reaction at 3 different annealing temperatures; 55, 63 and 65°C. As the result, only AP1-Barc55-Rv and AP1-Barc55IRv1 give the products (Figure A3.2). The product from AP1-Barc55-Rv was distinct and still was the same product in all three annealing temperatures (Figure A3.2 A). However, the products from AP1-Barc55IRv1 were not clear band and different in each annealing temperature (Figure A3.2B). The other specific primers didn't amplify any product. The products amplified from AP1-Barc55-Rv have been successful for cloning whereas the intensity of the product from AP1-Barc55IRv1 was very low so it was not successful for cloning. After sequencing some those sequences from products of AP1-Barc55-Rv, it contained AP and Barc55-Rv primer sequences but did not contain Barc55 product. So, only one specific primer site either Barc55 forward or reverse primer was not specific enough to amplify Barc55 product. However, the Barc55 product that located on 2B was sequenced again to confirm, it was found that the sequence was the same sequence as reported as GrainGene database. So, it is possible that the restriction enzymes might not suitable enzymes for non-genic region amplification. So, second PCR might help identify more specific bands and new choice of restriction enzyme might help to make suitable libraries for genome walking.

After doing second PCR with NAP1 adapter primer and nested specific primer (Barc55-iF2), more additional sequences around 415 bp at downstream was identified (Figure A3.3). NCBI blast revealed this sequence showed identity (85% identity) with T.monococcum clone BAC 116F2 and 115G1 gene. The BAC was 215 kb on chromosome 5A, containing 4 genes; putative mitochondrial carrier protein, putative cleavage stimulation factor subunit1, putative serine/threonine kinase and putative ABC transporter (Table A3.3).



A.



B.

Figure A3.2. PCR amplification with adaptor I primer (AP1) and specific primers, Barc55-Rv and Barc55-IRv1.

The distinct large fragments were used for sequencing.

Figure A3.3. Genome walking sequence of Barc55 downstream region.

Sequence of Barc55-F2 (618 bp) walking showing the forward primer site and adapter site (A) and an alignment with original sequence BV211787 that used to design specific primer (97 % identity). The sequence was identified more at downstream region that on black line box (415 bp) (B).

Barc55-iF2 primer

GAGGAGATCACCACTCTAGAGAAGACTAGGCATGGCGTTTGTGCTCGGGTCCGTCCCTG
CAGCGTCGATAAACCCAACGGCTGAGCTCTGTCTGGCGGCCGGCCCCGTCCACGTCACCTC
CTGGGGTCATTGGTGTCTGGCCTTCTCATGCAATAGCCTCAGGAGACTCATTCGACACAGG
CTGCCTAACGATCTTGTAGGACTAGTACAAGCGCTCCCAATCAAGTTCATGTTGGGGATCG
TTGCAGAAATTAAAAAATTTCTACGCATCACCAAGATCAATCTATGAAGTAACTAGCAACG
AGAGAGGGGGAGTGCATCTTCATACCCTTGAAGATCGTGAGACGGAAGCGTTGCAAGAAC
GCGGTTGGAGGAGTCGTACACGTAGCGATTCAGATCGCGGCCGAATCCGATCTTAGCACC
GAACAACGGTGCCTCCGCGTTCAACACACGTGCAGCCCGATGACGTCTCCCGCGCCTTGAT
CCAGCAAGGAGGAGGGAGAGGTTGAGGAAGAGGGCTCCAACAGCAGCACGACAGCGTGG
TGGTGATGGAGTGGCAGTTCTCCGGCAGGGCTTTGCTAAGCTCACGCGGAGGAGGAGAGG
TGTTGGGGAGGGGAG**TCTAGAGGATCCCGGGCGAGC**

NAP-Adapter primer site

A.

```
Barc55-iF2-5-2          ------------------------------------------------------------
Barc55-iF2-5-3          ------------------------------------------------------------
Barc55-iF2-1-2          ------------------------------------------------------------
BV211787-Barc55-STS     TCTCTCTTCCCCCTCTCTTCCCATCCCATCCATTTTCGATTAACGGCGAGCAATGGGAGC 60


Barc55-iF2-5-2          ------------------------------------------------------------
Barc55-iF2-5-3          ------------------------------------------------------------
Barc55-iF2-1-2          ------------------------------------------------------------
BV211787-Barc55-STS     AGCGCTGTCTCCGGTGGAGGATCTGGTGGCTCCGATGACTAGGAGAGCCTGCTGCGCGAT 120


Barc55-iF2-5-2          ------------------GAGGAGATCACCACTCTAGAGAAGACTAGGCATGGCGTTTGT 42
Barc55-iF2-5-3          ------------------GAGGAGATCACCACTCTAGAGAAGACTAGGCATGGCGTTTGT 42
Barc55-iF2-1-2          ------------------GAGGAGATCACCACTCTAGAGAAGACTAGGCATGGCGTTTGT 42
BV211787-Barc55-STS     GCAAAGCCGCGGGACGGGGAGGAGATCACCACTCTAGAGAAGACTAGGCATGGCGTCTGT 180
                                          ************************************ ***

Barc55-iF2-5-2          GCTCGGGTCCGTCCCTGCAGCGTCGATAAACCCAACGGCTGAGCTCTGTCTGGCGGCCGG 102
Barc55-iF2-5-3          GCTCGGGTCCGTCCCTGCAGCGTCGATAAACCCAACGGCTGAGCTCTGTCTGGCGGCCGG 102
Barc55-iF2-1-2          GCTCGGGTCCGTCCCTGCAGCGTCGATAAACCCAACGGCTGAGCTCTGTCTGGCGGCCGG 102
BV211787-Barc55-STS     GCTCGGGTCCGTCCCTGCAGCATCGATAAACCCAACGGCTGAGCTCTGTCTGGCGGCCGG 240
                        ******************** ************************************

Barc55-iF2-5-2          CCCCGTCCACGTCACCTCCTGGGGTCATTGG-TGTCTGG-CCTTCTCATGCAATAGCCTC 160
Barc55-iF2-5-3          CCCCGTCCACGTCACCTCCTGGGGTCATTGG-TGTCTGG-CCTTCTCATGCAATAGCCTC 160
Barc55-iF2-1-2          CCCCGTCCACGTCACCTCCTGGGGTCATTGG-TGTCTGG-CCTTCTCATGCAATAGCCTC 160
BV211787-Barc55-STS     CCCCGTCCACGTCACCTCCTGGGGTCATTGGGTGTCTGGGCCTTCTCATGCAATAGCCTC 300
                        ******************************** ******* ********************

Barc55-iF2-5-2          AGGAGACTCATTCGACACAGGCTGCCTAACGATCTTGTAGGACTAGTACAAGCGCTCCCA 220
Barc55-iF2-5-3          AGGAGACTCATTCGACACAGGCTGCCTAACGATCTTGTAGGACTAGTACAAGCGCTCCCA 220
Barc55-iF2-1-2          AGGAGACTCATTCGACACAGGCTGCCTAACGATCTTGTAGGACTAGTACAAGCGCTCCCA 220
BV211787-Barc55-STS     AGGAGACTCATTCGACACAGGCTGCCTAACGATCTAGCAGGACTAGT------------- 347
                        ********************************** * *********
```

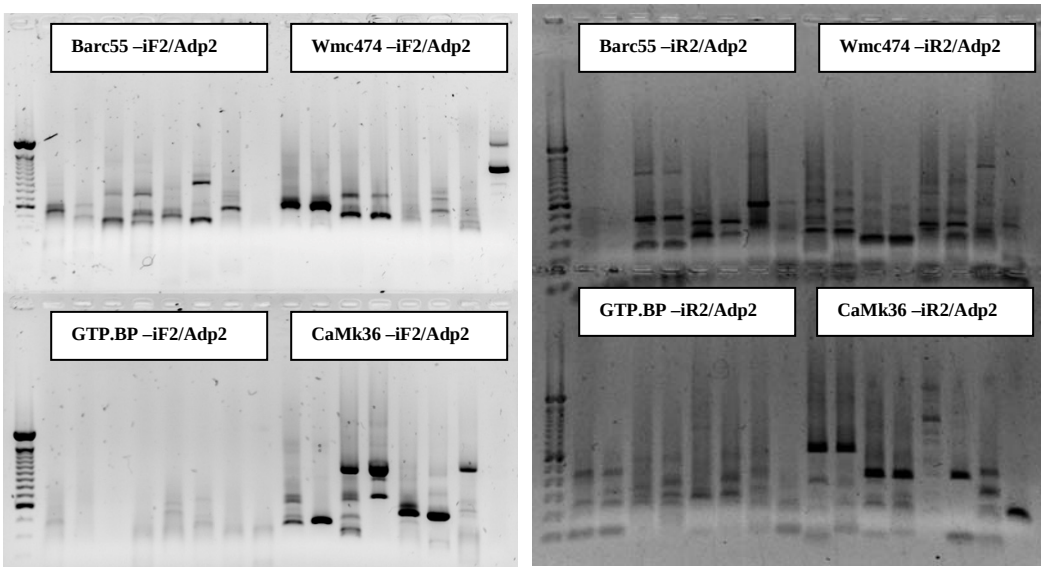
Figure A3.3. (Continued)

B.

Genome walking for flanking markers; Barc55, ss26.3 and Wmc474 and a candidate gene (CaMK) with adapter II primer for blunt end formed genomic libraries.

The libraries used for this step was different from previous step. There were four enzymes used to generate the libraries; DraI, EcoRV, PvuII, and ScaI. These enzymes generate blunt end libraries. There were three flanking targets were used; Barc55, ss26.3 and Wmc474. Also, conserved coding region between rice CaMK and wheat BT009309 cDNA clone was a target. Before performing PCR-based genome walking, the libraries were confirmed that the libraries contained the PCR product amplified by all markers Barc55, ss26.3, wmc474 and rCaPK first. This verifies that these libraries were suitable for PCR-based genome walking. As the result of genome walking (Figure A3.4), there were many products for Barc55, Wmc474 and CaMK but not for ss26.3. Those bands were subjected for sequencing revealed that only Barc55 and CaMK contained the target sequences whereas the others were not specific bands. The summarize data is described on Table A3.3. The identified sequences were shown on Figure A3.5-A3.7. For Barc55 walking, the original sequence BV211787 that use to design marker was a sequence-tagged site (451 bp) including Barc55 SSR repeat. The sequence from walking was 227 bp including specific primer site and adapter II site. This sequence showed 96% identity with the original sequence (Figure A3.5). After blasting on NCBI, the sequence matched with T. aestivum clone BAC 1031P08 (E-value = 3e-31). This clone BAC 1031P08 was 292 kb from chromosome 1A, containing 3 genes on this BAC; Globulin 1, X-type HMW glutenin and protein kinase whereas the rest of sequences were repeat sequence such as transposons and LTR-retrotransposons. For CaMK walking, both sequences were match with our target wheat BT009309 cDNA (1498 bp) clone at 99-100% identity. One sequence (1051 bp) covered downstream sequence of BT009309 spanning to 3' UTR (251 bp) (Figure

A3.6) and another sequence (631 bp) covered upstream sequence of BT009309 and isolated more additional sequence around 124 bp (Figure A3.7). Both sequences contained both exons and introns. All sequences were used for TAGdb (http://flora.acpfg.com.au/tagdb/cgi-bin/index) to identify more new sequences surrounding these sequences but none of any new sequences was identified.



A. B.

Figure A3.4. Band products for genome walking for downstream direction (A) and upstream direction (B) of known sequences.

Barc55-F2 primer

GAGGAGATCACCACTCTAGAGAAGACTAGGCATGGCGTTTGTGCTCGGG
TCCGTCCCTGCAGCGTCGATAAACCCAACGGCTGAGCTCTGTCTGGCGGCC
GGCCCCGTCCACGTCACCTCCTGGGGTCATTGGTGTCTGGCCTTCTCATGC
AATAGCCTCGGCAGACTTCAAGCGAAAGCCGACGTGGGTGTAGACCACGA
CCTGCCCGG**GCCGCTCGAGCCCTATT**

Adp2 primer site

A.

```
BV211787-Barc55-STS      CCACATATATCCCACATTTAAATGAACCCTAGACGAAAGTCAACACACTCCACTCCTCTC 60
BiF2-6-1F-ConIn          ------------------------------------------------------------

BV211787-Barc55-STS      TCCATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTCTCTTCCCCCTCT 120
BiF2-6-1F-ConIn          ---------------------------------------------------------- 

BV211787-Barc55-STS      CTTCCCATCCCATCCATTTTCGATTAACGGCGAGCAATGGGAGCAGCGCTGTCTCCGGTG 180
BiF2-6-1F-ConIn          ------------------------------------------------------------

BV211787-Barc55-STS      GAGGATCTGGTGGCTCCGATGACTAGGAGAGCCTGCTGCGCGATGCAAAGCCGCGGGACG 240
BiF2-6-1F-ConIn          ------------------------------------------------------------

BV211787-Barc55-STS      GGGAGGAGATCACCACTCTAGAGAAGACTAGGCATGGCGTCTGTGCTCGGGTCCGTCCCT 300
BiF2-6-1F-ConIn          --GAGGAGATCACCACTCTAGAGAAGACTAGGCATGGCGTTTGTGCTCGGGTCCGTCCCT 58
                           ************************************ ******************

BV211787-Barc55-STS      GCAGCATCGATAAACCCAACGGCTGAGCTCTGTCTGGCGGCCGGCCCCGTCCACGTCACC 360
BiF2-6-1F-ConIn          GCAGCGTCGATAAACCCAACGGCTGAGCTCTGTCTGGCGGCCGGCCCCGTCCACGTCACC 118
                         ***** ********************************************************

BV211787-Barc55-STS      TCCTGGGGTCATTGGGTGTCTGGGCCTTCTCATGCAATAGCCTCAGGAGACTCATTCGAC 420
BiF2-6-1F-ConIn          TCCTGGGGTCATTGG-TGTCTGG-CCTTCTCATGCAATAGCCTCGGCAGACT--TCAAGC 174
                         *************** ******* ******************** * *****  *    *

BV211787-Barc55-STS      ACAGGCTGCC-TAACGATCTAGCAGGACTAGT---- 451
BiF2-6-1F-ConIn          GAAAGCCGACGTGGGTGTAGACCACGACCTGCCCGG 210
                           * ** * * *     *  * ** ***  *
```

B.

Figure A3.5. Genome walking sequence of Barc55 region.
Sequence of Barc55-F2 (227 bp) walking showing the forward primer site and adapter site (A) and an alignment with original sequence BV211787 that used to design specific primer (96 % identity). The underline sequence was Barc55 SSR region (B)

Figure A3.6. Genome walking sequence of a candidate gene CaMK downstream region.

Sequence of CaMK-F2 (1051 bp) walking showing the forward primer site and adapter site (A) and Alignment with original sequence BT009309 that used to design specific primer (99 % identity) showing more additional sequence was isolated at 3'UTR (251bp) as black line box (B)

CaMK-F2 primer

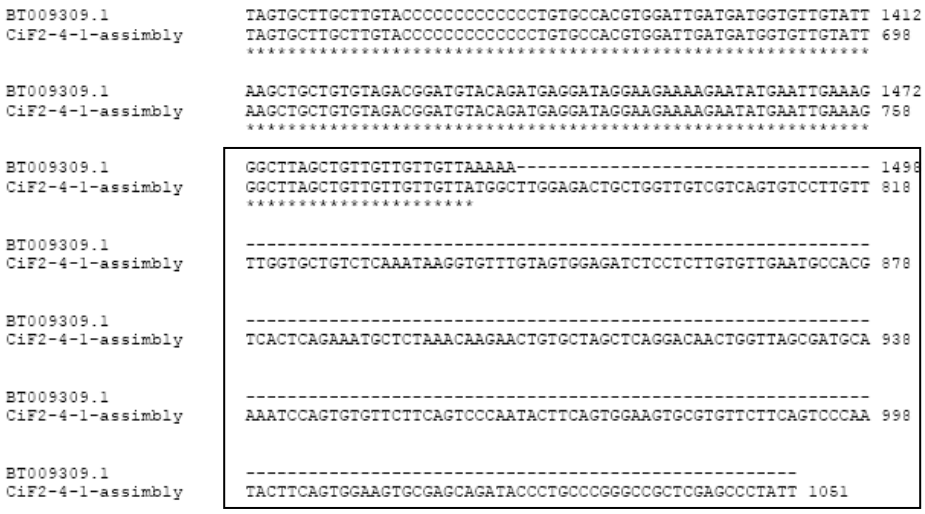
GGTATGCACCCTTCAGTACAGAAAATTGGATTTTGAAGAGTTCGCCGCTTCCGCCATCAG
CGTATACCAGATGGAAGCTCTGGAGACCTGGGAACAGCATGCTCGGCGCGCGTACGAGCT
GTTCGACAAGGAGGGCAACCGGCCTATCGTGATCGAAGAACTCGCGTCGGTACGAGTCTG
AACTGTAGTCTGAACAGTTAATTTATGCGTGATCACTTGACATTTCACTGATCACTGTCACG
TCTTTGGTCCATCAGGAACTCGGCCTCGGCCCGTCGGTGCCCCTCCACGTCGTCCTCCAGG
ACTGGATCAGGCACGCCGACGGGAAGCTCAGCTTCCTCGGGTTCATAAAGCTCTTGCACGG
TGTTTCTTCCCGGTCGATCCCGAAAGCCTGAGACGGAAACAAGCATTATTATTCTTGCATC
AAACCTAGGAGATGACTGATGGCTAAGCCACTGCCAATCTCTCGCTGGTGTCTGTGTGCGT
GGACGACAACTGTTTGAGGTGGCTGGAGGAGGACGATGGCTTGCGACCGCGACCGTAACG
GAAGAGAAGCCGGCAGTCTTTGTGTAAGAGGTAAGAAGAAGATAAGAAGAAGGTTGTTGC
TTGATGCTTATGACTACTACAGTAGTAGCTACTAGTGCTTGCTTGTACCCCCCCCCCCCCTG
TGCCACGTGGATTGATGATGGTGTTGTATTAAGCTGCTGTGTAGACGGATGTACAGATGAG
GATAGGAAGAAAAGAATATGAATTGAAAGGGCTTAGCTGTTGTTGTTGTTATGGCTTGGA
GACTGCTGGTTGTCGTCAGTGTCCTTGTTTTGGTGCTGTCTCAAATAAGGTGTTTGTAGTGG
AGATCTCCTCTTGTGTTGAATGCCACGTCACTCAGAAATGCTCTAAACAAGAACTGTGCTA
GCTCAGGACAACTGGTTAGCGATGCAAAATCCAGTGTGTTCTTCAGTCCCAATACTTCAGT
GGAAGTGCGTGTTCTTCAGTCCCAATACTTCAGTGGAAGTGCGAGCAGATACCCTGCCCGG
GCCGCTCGAGCCCTATT

Adapter site

A.

```
BT009309.1          GGGTTATTGACTTTGTTAACACGGTATGCACCCTTCAGTACAGAAAATTGGATTTTGAAG 840
CiF2-4-1-assimbly   ----------------------GGTATGCACCCTTCAGTACAGAAAATTGGATTTTGAAG 38
                                           ************************************

BT009309.1          AGTTCGCCGCTTCCGCCATCAGCGTATACCAGATGGAAGCTCTGGAGACCTGGGAACAGC 900
CiF2-4-1-assimbly   AGTTCGCCGCTTCCGCCATCAGCGTATACCAGATGGAAGCTCTGGAGACCTGGGAACAGC 98
                    ************************************************************

BT009309.1          ATGCTCGGCGCGCGTACGAGCTGTTCGACAAGGAGGGCAACCGGCCTATCGTGATCGAAG 960
CiF2-4-1-assimbly   ATGCTCGGCGCGCGTACGAGCTGTTCGACAAGGAGGGCAACCGGCCTATCGTGATCGAAG 158
                    ************************************************************

BT009309.1          AACTTGCGTCGG------------------------------------------------ 972
CiF2-4-1-assimbly   AACTCGCGTCGGTACGAGTCTGAACTGTAGTCTGAACAGTTAATTTATGCGTGATCACTT 218
                    **** *******

BT009309.1          ----------------------------------------AACTCGGCCTCGGCCCGTCG 992
CiF2-4-1-assimbly   GACATTTCACTGATCACTGTCACGTCTTTGGTCCATCAGGAACTCGGCCTCGGCCCGTCG 278
                                                            ********************

BT009309.1          GTGCCCCTCCACGTCGTCCTCCAGGACTGGATCAGGCACGCCGACGGGAAGCTCAGCTTC 1052
CiF2-4-1-assimbly   GTGCCCCTCCACGTCGTCCTCCAGGACTGGATCAGGCACGCCGACGGGAAGCTCAGCTTC 338
                    ************************************************************

BT009309.1          CTCGGGTTCATAAAGCTCTTGCACGGTGTTTCTTCCCGGTCGATCCCGAAAGCCTGAGAC 1112
CiF2-4-1-assimbly   CTCGGGTTCATAAAGCTCTTGCACGGTGTTTCTTCCCGGTCGATCCCGAAAGCCTGAGAC 398
                    ************************************************************

BT009309.1          GGAAACAAGCATTATTATTCTTGCATCAAACCTAGGAGATGACTGATGGCTAAGCCACCG 1172
CiF2-4-1-assimbly   GGAAACAAGCATTATTATTCTTGCATCAAACCTAGGAGATGACTGATGGCTAAGCCACTG 458
                    ************************************************************ *

BT009309.1          CCAATCTCTCGCTGGTGTCTGTGTGCGTGGACGACAACTGTTTGAGGTGGCTGGAGGAGG 1232
CiF2-4-1-assimbly   CCAATCTCTCGCTGGTGTCTGTGTGCGTGGACGACAACTGTTTGAGGTGGCTGGAGGAGG 518
                    ************************************************************

BT009309.1          ACGATGGCTTGCGACCGCGACCGTAACGGAAGAGAAGCCGGCAGTCTTTGTGTAAGAGGT 1292
CiF2-4-1-assimbly   ACGATGGCTTGCGACCGCGACCGTAACGGAAGAGAAGCCGGCAGTCTTTGTGTAAGAGGT 578
                    ************************************************************

BT009309.1          AAGAAGAAGATAAGAAGAAGGTTGTTGCTTGATGCTTATGACTACTACAGTAGTAGCTAC 1352
CiF2-4-1-assimbly   AAGAAGAAGATAAGAAGAAGGTTGTTGCTTGATGCTTATGACTACTACAGTAGTAGCTAC 638
                    ************************************************************
```

Figure A3.6. (Continued)



B.

Adapter Adp2

AATAGGGCTCGAGCGGCCCGGGCAGGTAAATTATAATTAGTCAAATTATGATCAATTCCT
ATCAAAATTGACCATGCATATCAAATGGCGCTTACTGTAACTTCTATCACTTATCATGTTTG
TTTCTTTTCAGATTGTGTAAAGGAGGCGAACTACTTGATAAGATATTGGCGAGGTACAATC
CTACCAGCGTTCGCAAATTGGAACAATTCTTATGATTATATCTGCTGGGGTTCTAACTATAT
CCAATCTATTGTTTTGCAGAGGTGGAAAGTATTCTGAAGAGGATGCAAAGGTTGTTATGCT
GCAAATTTTGAGTGTAGTATCATTTTGCCATCTTCAAGGTGTTGTTCATCGGGATCTGAAAC
CAGAGGTTAGTTTTTTTTCACACAGTTAGCAGCTAATATTAATTCATCGTGCTGTCTTTTAG
TTACTATTATCAACCTGGGAGTTTCTTAGTACCATACCATATTTTCTTCGATTGTTGGTTCTT
TATATTTATCCAAAGCACTTCTGTTCAGTCAGCAGATGCAATACAAAGATTTTGCAACTAA
ATACATTTGTGTTTCTTTATCCCAGAATTTTCTATTCTCATCGAAG
GAGG**AAAACTCACCCTTGAAGGTCATAGACT**

CaMK-R2 primer

A.

```
 Score = 51.8 bits (56),  Expect = 2e-10
 Identities = 28/28 (100%), Gaps = 0/28 (0%)
 Strand=Plus/Plus

Query  7    GCGAACTACTTGATAAGATATTGGCGAG  34
            ||||||||||||||||||||||||||||
Sbjct  131  GCGAACTACTTGATAAGATATTGGCGAG  158

 Score =  201 bits (222),  Expect = 2e-55
 Identities = 111/111 (100%), Gaps = 0/111 (0%)
 Strand=Plus/Plus

Query  33   AGAGGTGGAAAGTATTCTGAAGAGGATGCAAAGGTTGTTATGCTGCAAATTTTGAGTGTA  92
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  246  AGAGGTGGAAAGTATTCTGAAGAGGATGCAAAGGTTGTTATGCTGCAAATTTTGAGTGTA  305

Query  93   GTATCATTTTGCCATCTTCAAGGTGTTGTTCATCGGGATCTGAAACCAGAG  143
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  306  GTATCATTTTGCCATCTTCAAGGTGTTGTTCATCGGGATCTGAAACCAGAG  356

 Score = 98.7 bits (108),  Expect = 2e-24
 Identities = 54/54 (100%), Gaps = 0/54 (0%)
 Strand=Plus/Plus

Query  142  AGAATTTTCTATTCTCATCGAAGGAGGAAAACTCACCCTTGAAGGTCATAGACT  195
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  561  AGAATTTTCTATTCTCATCGAAGGAGGAAAACTCACCCTTGAAGGTCATAGACT  614
```

Query = BT009309
Sbject =sequence from CaMK genome walking

B.

Figure A3.7. Genome walking sequence of a candidate gene CaMK upstream region

Sequences of CaMK (631 bp)genome walking at upstream containing both adapter II site and specific primer site (A) and Alignment with BT009309 revealing 100% identity located at upstream of BT009309 and match with 3 exons. Genome walking contained more sequence 124 bp at 5' upstream (B).

Table A3.3. Summary of PCR-genome walking result.

Flanking marker name or gene	Type of Adapter primer	Specific primer name	PCR band product	PCR containing real sequence	Real band product	NCBI Blast
Barc55	AP1	Barc55F Barc55R	YES YES	No No		
	NAP1	Barc55iF	YES	YES	618bp	*T.monococcum* clone BAC 116F2 and 115G1 gene * (85%identity, 3e-119)
	AdP2	Barc55-iR2 Barc55-iF2	Yes Yes	In progress Yes	227bp	*T. aestivum* clones BAC 1031P08 ** (E-value = 3e-31)
ss26.3	AdP2	GTP-BP-iR2 GTP-BP-iF2	No No			
Wmc474	AdP2	Wmc474-iR2 Wmc474-iF2	Yes Yes	Not clear Not clear		
CaMK 36	AdP2	CAMK36-iR2 CAMK36-iF2	Yes Yes	Yes Yes	~1051bp ~631bp	putative calcium/calmodulin-dependent protein kinase (CaMK)*** (E-value = 3e-88)

* *T.monococcum* clone BAC 116F2 and 115G1 gene was 215 kb on chromosome 5A, containing 4 genes; putative mitochondrial carrier protein, putative cleavage stimulation factor subunit1, putative serine/threonine kinase and putative ABC transporter.
* **T. aestivum* clones BAC 1031P08 was 292 kb on chromosome 1A, containing 3 genes on this BAC; Globulin 1, X-type HMW glutenin and protein kinase whereas the rest of sequences were repeat sequence such as transposons and LTR-retrotransposons.
*** It matched with CaMK gene of *Oryza sativa* chromosome 7, BAC clone:OJ1340_C08 that was 192 kb.
This BAC contained around 19 known function genes including CaMK and 10 unknown gene.

REFERENCES

Ashoub A, Abdalla K (2006) A primer-based approach to genome walking. Plant Mol Biol Rep 24:237-243

Chawla R, DeMason D (2003) Genome Walking in pea: an approach to clone unknown flanking sequences. Pisum Genet 35:3-5

Lu G, Moriyama EN (2004) Vector NTI, a balanced all-in-one sequence analysis suite. Briefings in Bioinformatics 5:378-388

Rishi AS, Nelson ND, Goyal A (2004) Genome walking of large fragments: an improved method. J Biotechnol 111:9-15

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 23:1087-1088

Tittarelli A, Milla L, Vargas F, Morales A, Neupert C, Meisel L, Salvo-G H, Penaloza E, Munoz G, Corcuera L, Silva H (2007) Isolation and comparative analysis of the wheat TaPT2 promoter: identification in silico of new putative regulatory motifs conserved between monocots and dicots. J Exp Bot 58:2573-2582