The agronomic and molecular characterisation of *Rht8* in hexaploid wheat

Ania M. Kowalski

A thesis submitted for the Degree of Doctor of Philosophy

University of East Anglia

John Innes Centre

Norwich

September 2015

Abstract

Ania M. Kowalski, September 2015

The agronomic and molecular characterisation of *Rht8* in hexaploid wheat

Reduced height 8 (*Rht8*) is the main alternative to the GA-insensitive *Rht* alleles in hot and dry environments and reduces plant height without yield penalty. The potential of *Rht8* in northern-European wheat breeding remains unclear. In the present study, near-isogenic lines contrasting for the *Rht8*/tall allele in the UK-adapted and photoperiod-sensitive variety Paragon were evaluated in trials with varying nitrogen fertiliser (N) treatments and water regimes across sites in the UK and Spain.

Rht8 conferred a robust height reduction of 11% regardless of treatment and was more resistant to root-lodging at agronomically-relevant N levels. In the UK, the *Rht8* NIL showed a 10% yield penalty due to concomitant reduction in grain number and spike number whereas grain weight and harvest index were not significantly different to the tall NIL. The yield penalty was abolished at low N and in irrigated conditions in the UK and Spain. This indicates the utility of *Rht8* in reduced-input agriculture. Decreased spike length and constant spikelet number in *Rht8* compacted spikes by 15% independent of environment. The genetic interval of *Rht8* overlaps with the most recent mapping of the *compactum* gene on 2DS (Johnson et al., 2008) and future work with the markers found in this study is required to genetically dissect these loci.

Rht8 had been previously fine-mapped to a 1.29 cM interval (Gasperini et al., 2012). *Rht8* was further fine-mapped using an RNA-Seq enabled bulked segregant analysis method, as well as utilising SNP-platforms and emerging Triticeae genomic resources to identify molecular markers. *Rht8* was reduced to a 1.015 cM genetic interval and syntenic intervals of 1.34 Mb on rice chromosome 4, 1.36 Mb on Brachypodium chromosome 5, 2.9 Mb on barley 2H and 4.25 Mb on *Ae. tauschii* 2D. Disruption to micro-collinearity was found with Brachypodium and rice, with better but imperfect collinearity with *Ae. tauschii* and barley. *Rht8* was also anchored to a single IWGSC-2 POPSEQ bin and to a 2.3 cM region in the whole genome shotgun-ordered wheat scaffolds.

Acknowledgements

I would like to thank Simon Griffiths, whose unwavering enthusiasm and optimism was reassuring during the project, and whose wider perspective on wheat breeding I learnt a lot from. I valued having the freedom to explore my own ideas during the PhD. I was inspired by the esteem with which he held his family life, which is not universally encountered.

Working with Martin Trick and Ricardo Ramirez-Gonzalez was a highlight of my time. Martin Trick developed the UniGene and v3.3 cDNA (with Sarah Ayling, TGAC) wheat references used in this project and used pipelines developed in previous work to call SNPs in the parent NIL data (Chapters 5 and 6) and *de novo* assembly. He also produced the raw RPKM values for the basis of the expression analysis. Martin helped to troubleshoot the VarScan SNP-calling process I performed in Chapter 5. He advised me through many frustrations with running data on the JIC Cluster and always gave careful thought to the problems we identified. Ricardo was particularly helpful. Ricardo ran my SNPs on his PolyMarker pipeline before it was publically available, which greatly expedited marker design (Chapter 5). At that preliminary stage there were many curiosities we encountered with the idiosyncrasies of wheat, and I enjoyed thinking about those. He generously helped me with scripting problems I encountered at various stages of alignment and quality-control of the RNA-Seq reads in Chapter 5. Ricardo did the BSA analysis reported in Chapters 5 and 6, using pipelines he developed in related work, described in Chapter 1. He helped with other bioinformatic queries, was always polite and cheerful and most of all was a supportive friend.

Matthew Moscou and Burkhard Steuernagel assisted me with barley resources before they were publically available and Matt especially helped me to download the Morex assemblies for the *Rht8* region.

Cathy Mumford, Simon Orford and Dean Cole made it possible to collect and measure the field results presented in Chapter 3. Field trials are a massive undertaking which I could not do alone. I appreciated their advice and willing support in data collection, locating seed in the barn and managing bagging of ears. Cathy was always extremely helpful, cheerful and attentive and I greatly appreciated her attention to the smallest detail in the experimental design and treatment applications. I also thank Simon and Dean for arranging casual labour to meet my requests and 'keeping an eye' on my glasshouse plants. Many casuals helped to collect and measure data at various stages, bag, harvest and thresh grain for a quick turnaround to enable timely glasshouse experiments and field trials. In particular, Joanna Wolstenholme, Jack Raven and Rachel Mumford helped with tiller measurements and grain processing, and Rajani Awal and Dean in data collection for Chapter 7.

The work to assess the agronomic performance of *Rht8* benefited hugely from collaborations which added another dimension to the analysis. Mike Gooding kindly collaborated with us in running field trials at the University of Reading. Those results gave me interesting data which we would not have seen under the field design at Norwich. Richard Casebow oversaw the field trials and was helpful with queries I had during and after with the data collected. Gustavo Slafer also ran two field trials in Lleida, Spain. His post-doc Ariel Ferrante collected the data and also made extensive measurements of other physiological traits which do not feature in this work. I appreciated their commitment and reliability.

Liz Sayers (JIC) provided the seed for the *Xgwm261*-allele introgressions in Chapter 7 and Fernanda Gonzalez (INTA) provided the seed of Klein 49 and Klein 147.

Richard Goram (JIC) ran large batches of plates for me on alternative PCR machines due to frequent breakages of the PCR blocks. This enabled me to test more quickly the SSR markers described in Chapter 5.

Trevor Paterson (The Roslin Institute) responded promptly and generously to queries using ArkMap, which enabled me to produce the graphics for the synteny comparisons for the thesis (Chapter 5). Chris Groom (JIC) was invaluable in sourcing references, EndNote queries and scanned in many older publications from the library.

Madzia and Sam stepped in last-minute to help me harvest and photograph the spikes of hundreds of plants over the course of a weekend. Three years later, that made part of Chapter 4 possible.

Andrew Davis took great photographs and came out to the glasshouse and field to give me the high-quality images which I could use in the thesis. He also helped set up a camera rig on a trolley with a platform for samples and lent me a camera so that I could take the spike and peduncle photos at any time of day and night. I am grateful he was so amenable to helping with various requests.

Laura Dixon and Alba Farré were great friends and colleagues and I was fortunate to have them to share ideas with. Alba helped with JoinMap in Chapter 6. Laura reviewed my work for presentations and reports. She also proof-read the thesis. I'm grateful that she always had this time to give me. We were particularly well-suited to sharing office-space with our sports kits! Oscar Gonzalez wrote the ImageJ macro used in Chapter 3, and I want to thank him for his good-humour and warm hugs.

Viktor Korzun, Colin Law and the *Rht8* community at EWAC gave me a sense of the great genetic tradition of aneuploidy and its significance in the *Rht8* story, as well perspective on where my work fitted into this ongoing research. This enabled me to feel pride in my work, which was a moving time for me, coming at the end of the PhD. Feeling the sense of community and the interest and appreciation of my efforts at the conference was the highlight of my PhD experience. The questions and discussions at the conference also inspired the ideas for Chapter 4 and I greatly enjoyed that creativity, which was a welcome contrast to much of the technically-dense work where building a narrative was challenging.

Dedications and thanks

I thank my parents who have always given me freedom to make my own decisions and trust me to go my own way, without comparisons. I thank my twin sister, Madzia, for sharing my experiences and with whom I have a relationship which I am only beginning to understand is very unique. Thanks too to Sam, you've become my closest family. I made great friends with Neal Greenaway, Mark Ashwood, David Thurkettle and Chris Judge, who often checked in on me and I will remember the supportive emails from Rachel Goddard. Hester van Schalkwyk was most encouraging in the final months and uniquely understood the links between wheat, CrossFit and long-distance. All those people I met through Quest, such as Terry, Ruby, Mandy, Maria and Fr Kieran helped me find my place in the most important work I did, on myself and in the growing of my faith, during the PhD years.

To my beautiful Jen: thank you for holding my hand along the runway, your resolute love and firm belief in me and for being committed to making decisions that prioritised my needs with work ahead of yours and ours. I appreciated your understanding when our plans and time together were disrupted by or had to be moulded around my work. It still amazes me that you were able to listen to it all and always be there to support me and react positively, even when I couldn't. The pep talks that you gave helped me along the way and I admire how you delivered them tirelessly, even when the recipient was tired and grumpy and wanted to be outside running around. Your skills in empathy coupled with the switch to our language to make me laugh helped pull me out of the self-centredness that a PhD can encourage and I esteem you so much for finding that balance. The best part was that I knew even if it all went wrong, it wouldn't change what you thought and felt for me. There is a wonderful sense of security, hope and freedom this gave me and I will forever treasure and ponder that. The pull-up bar gave me a welcome diversion in the final months and helped me to make progress in areas I value so much, thank you for that. Most of all, I treasure your skill of listening to understand, rather than listening to reply, which made me feel less isolated in the writing process. The long-distance has meant I value and respect your constancy and dedication. Thank you for your incredible patience. I'm excited for the future.

Publications

Kowalski, A. M., Gooding, M., Ferrante, A., Slafer, G. A., Orford, S., Gasperini, D. & Griffiths, S. 2016. Agronomic assessment of the wheat semi-dwarfing gene *Rht8* in contrasting Nitrogen treatments and water regimes. *Field Crops Research*. *Manuscript accepted.*

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Wheat and food security

1.1.1 Origin and spread of wheat

The first cultivation of wheat was 10,000 years ago during the Neolithic period. This time saw a transition from hunter-gathering after food to a settled lifestyle reliant on agriculture and crops, including barley (*Hordeum vulgare*), pulses and wheat. The earliest cultivated wheats were the diploid *Triticum monococcum* (einkorn wheat, genome A^mA^m) and the tetraploid *Triticum turgidum* ssp. *dicoccoides* (wild emmer, AABB), originating from south-eastern Turkey and northern Syria (Salamini et al., 2002). The tetraploid arose from the hybridisation of two diploid grasses 150,000 – 500,000 years ago: *Triticum urartu* (A^uA^u) and an unknown, possibly extinct species related to the extant *Aegilops speltoides* (genome SS) (Charmet, 2011).

Emmer was domesticated from its wild relative and in its domesticated form was cultivated for thousands of years in the Fertile Crescent of the Near East due to its adaptability and high yields. Domesticated emmer is the progenitor of modern durum wheat (*T. durum*, genome AABB) (Feldman, 2001). Emmer spread towards the Caspian Sea and hybridised independently a small number of times with *Aegilops tauschii* 8,500 years ago to form *Triticum aestivum* (AABBDD) (Dvorak et al., 1998), perhaps where the wild goat grass was growing as a weed in Neolithic fields of *T. dicoccum* (Charmet, 2011). The hexaploid wheat had superior yield, viability and adaptability compared to the progenitor species and spread all over the world (Feldman, 2001).

In what is known as the domestication syndrome, farmers selected for traits that clearly differentiated cultivated wheat varieties from their wild ancestors. One altered trait was the reduction of spike-shattering which is determined by the *brittle rachis* (*Br1*) locus on the short arm of group-3 chromosomes (Li and Gill, 2006). Non-brittle spikes would have been advantageous where harvested grain was retained as seed for consumption and for the subsequent growing season (Charmet, 2011).

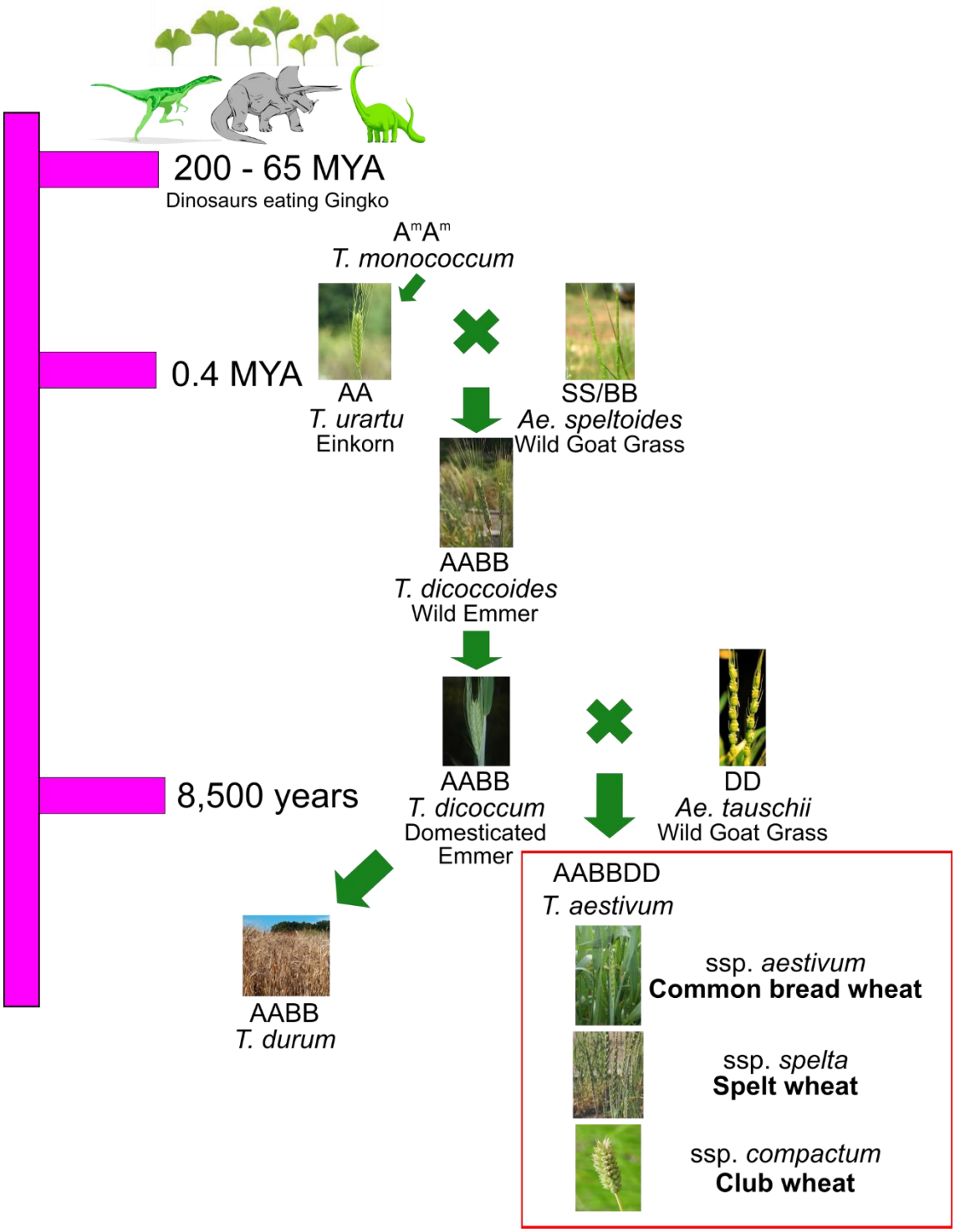


Figure 1.1: Evolutionary history of wheat. Wheat divergences based on Charmet (2011).

A second domestication trait was the conversion from hulled grain to free-threshing (naked) grain by the softening of glumes surrounding the grain. This reduced the labour required to separate grain from the spikelets and is also associated with shorter seed dormancy, which confers advantage over hulled

grain for germination in agriculture (Doebley et al., 2006). The genes associated with this free-threshing are *tenacious glume* (*Tg*) on chromosome 2D in *T. dicoccoides* (Taenzler et al., 2002) and *soft glume* (*Sog*) on chromosome 2A^{m}S in *T. monococcum* (Kerber and Rowland, 1974, Nalam et al., 2006). Free-threshing character was also conferred by the major domestication gene, *Q*, on chromosome 5A, which encodes an APETALA 2-like transcription factor and has pleiotropic effects on glume shape, glume tenacity, spike shape and rachis fragility (Simons et al., 2006).

1.1.2 Genetic bottlenecks due to domestication

Wheat domestication and later modern selection of wheat has had the effect of introducing genetic bottlenecks, especially into the wheat D-genome. In these genetic bottlenecks, the degree of genetic variability in extant tetraploid and hexaploid species has been found to be much lower than wild diploid and tetraploids (Reif et al., 2005). Analysis of nucleotide diversity found a loss of diversity of 75% during the domestication of tetraploid wheat, 55% for the A- and B- genomes of *T. aestivum* and 90% for the D-genome (Haudry et al., 2007). The D-genome is the least genetically diverse of the three sub-genomes due to the relatively recent hybridisation of the D-genome (Figure 1.1); a small number of these hybridisations occurring in the first place and then little inter-mating between bread wheat and the diploid progenitor (Dubcovsky and Dvorak, 2007, Dvorak et al., 1998). Conversely a substantial amount of inter-mating appears to have occurred between *T. aestivum* and tetraploid wheat (Dubcovsky and Dvorak, 2007, Luo et al., 2007). To boost diversity in pre-breeding programmes, methodologies to identify novel genetic variation from *Ae. tauschii* accessions have been described (Jones et al., 2013) and the hexaploidisation has been re-created to produce synthetic hexaploid wheat in Mexico and the UK.

1.1.3 Advances in wheat breeding in the 20th century

In the 20th century, two major breeding advances allowed substantial increases in wheat production. The work of Nazareno and Carlotta Strampelli in the early 20th century produced improvement in wheat varieties which many see as the first Green Revolution, or at least a precursor (Worland, 1999). The most significant

cross that introduced novel genetic variation into Europe was the crossing of the Japanese variety Akakomugi with an Italian land race and Dutch variety. The aim was to produce rust-resistant, early-flowering and lodging-resistant varieties (Borojevic and Borojevic, 2005).

Crosses with Akakomugi introduced the photoperiod (day length) insensitivity gene *Ppd-D1a* and the closely-linked *Rht8*. Prior to the introduction of *Ppd-D1a*, Italian wheat varieties were late-flowering due to their sensitivity to photoperiod, which meant grain fill occurred at a time when hot desiccating conditions restricted development. With earlier flowering, grain developed earlier in the season with higher soil moisture and lower temperatures. Early-flowering coupled with reduced height increased yields (Worland, 1999).

Strampelli's wheats became the main instruments of the *Battaglia del Grano* (the Wheat Battle), the propaganda slogan that the Fascist Regime used to define the campaign to reach wheat self-sufficiency in Italy. Strampelli himself joined the National Fascist Party in 1925 (Salvi et al., 2013). This agricultural aim was successful: by the late 1930s Italian wheat production doubled and the country was self-sufficient in wheat, with yields of 8 t ha^{-1} (Worland, 1999). From Italy, *Rht8* and *Ppd-D1a* were further transmitted to other high-yielding varieties and by the 1940s, wheats originating from the Strampelli cross were grown on 20 million ha in China and were the basis of wheat breeding programmes in other countries such as the USSR, Argentina and Norman Borlaug's Mexican-based programme (Salvi et al., 2013). In regions of the USSR, yield increased from 1.36 t ha^{-1} to 5.21 t ha^{-1} and in some places yields of 10 t ha^{-1} were recorded. This enabled some of the Soviet Bloc countries to turn from wheat importers to exporters (Borojevic and Borojevic, 2005).

The second major breeding advance was the Green Revolution of the 1960s and 1970s. One of the major foundation stones of this was the introduction of semi-dwarfing genes into wheat and the subsequent breeding of high-yielding wheat varieties. Greater yields were associated with improved lodging resistance and the resulting ability of wheat to tolerate higher levels of chemical fertilisers, as well as increased proportion of assimilate partitioned into the grain (Hedden, 2003). The genetic basis of the semi-dwarf varieties are the *Reduced height* (*Rht*) genes *Rht-B1b* (*Rht1*) and *Rht-D1b* (*Rht2*) (Gale and Youssefian, 1985).

These dwarfing genes first came to the US after World War II from the Japanese variety Norin 10 and previously originated from the Korean peninsula. Norin 10 was transferred to CIMMYT (International Maize and Wheat Improvement Centre in Mexico) where Norman Bourlag developed varieties with *Rht-B1b* and *Rht-D1b* (Borojevic and Borojevic, 2005). The CIMMYT wheat varieties were distributed all over the world. It is estimated that worldwide, over 70% of the wheat acreage contains at least one of these two genes (Evans, 1998) and that 90% of the semi-dwarf wheat varieties contain *Rht-B1b* and/or *Rht-D1b* (Worland et al., 1998b).

Together, these advances allowed for cereal crop production to triple globally in the last 50 years with only 30% increase in land area cultivated, at a time when the population has more than doubled (Pingali, 2012).

The success of the Green Revolution has not been enjoyed by everyone to the same extent. With the focus on intensification where returns would be high (in high-yield potential sites), marginal environments were left behind and areas with low population densities and poor market infrastructure such as Africa shared little of the Green Revolution success. Additionally, worldwide, female farmers and women-headed households were found to have gained less than their male counterparts (Pingali, 2012). Green Revolution-driven intensification has also had negative environmental impacts such as soil degradation and increased chemical runoff (Burney et al., 2010).

1.1.4 Production and use

Along with maize and rice, wheat is one of the three major cereal crops in the world and outcompetes the other two in terms of geographical range. The cultivation of wheat occupies the largest crop area on Earth, 215 million ha in 2012 (FAO, 2012). Over 25,000 wheat varieties are grown worldwide (Feldman, 2001) and these are the source of 20% of the world's calories. Currently, 95% of the wheat grown worldwide is hexaploid bread wheat with most of the remaining 5% being tetraploid durum wheat (Shewry, 2009). Small amounts of einkorn, emmer and spelt are still grown in some regions including Spain, Turkey, central and eastern Europe, Italy and India. The markets for some of the more ancient wheats are growing: as 'health foods' (spelt) and as less allogenic alternatives to hexaploid wheat (einkorn).

One subspecies of *T. aestivum* (Figure 1.1) is taxonomically distinguished by its compact spike and is called *compactum* (club wheat). The spikes of club wheat can be half the length of spikes in other hexaploid wheats with the same spikelet number (Zwer et al., 1995). This is believed to be principally due to the action of the dominant allele of the *compactum* (*C*) locus close to the centromere on 2D (Johnson et al., 2008, Rao, 1972), though QTLs affecting spike compaction have been reported on nearly every chromosome, including by Cui et al., 2012, Faris et al., 2014c, Jantasuriyarat et al., 2004, Ma et al., 2007, Manickavelu et al., 2011, and Sourdille et al., 2000.

Club wheat is a soft white wheat and is well-adapted to intermountain regions of the Pacific Coast States (including California, Oregon, Washington) because the strong, stiff culms are resistant to lodging and the firm spikes are shatter-resistant in the hot, dry and windy summers (U.S. Department of Agriculture, 1923). Club wheat was also reported to be adapted to dryland areas where stand establishment was difficult (Gul and Allan, 1972). Only about 1% of modern-day acreage planted in the Pacific Northwest is club wheat, much reduced on previous levels in the 1960s (75%) (Zwer et al., 1995). Owing to the soft texture and low protein content, most of the club varieties are not grown for bread. Instead, club wheat flour is used in blends with soft white wheat for export to Asia, especially for Japanese sponge-cake production, or used to make ‘cookies’ because bake-time is reduced and cake volumes are greater (http://cbarc.aes.oregonstate.edu/sites/default/files/usda-ars_club_wheat_breeding_in_oregon.pdf).

1.1.5 Future challenges and opportunities

Going into the future, there are intersecting challenges with wheat production. There is the challenge of feeding the 9-10 billion people expected by the middle of the century, in a world where already one in seven people have insufficient calories in their diet (FAO, 2009). If the answer to this challenge is to increase global crop production, then wheat production needs to double. In order to achieve this, a rate of 2.4% yield increase is required year-on-year. Yield trends from 1961 to 2008 showed an increase of 0.9% (Ray et al., 2013). Moreover, the rate of increase has slowed down in the last twenty years, particularly in the EU where major-producing countries of Germany, France and the UK have seen

yields stagnate (Charmet, 2011). Currently, the genetic progress is only sufficient to compensate the negative effects of abiotic and biotic stresses but not to increase yield overall. In the UK, increases of yields from 3 t ha^{-1} in the 1950s to 8.6 t ha^{-1} in 2014 have been achieved (DEFRA, 2015), though with sufficient water, nutrients and management of pests and pathogens, there is potential to achieve yields exceeding 10 t ha^{-1} (Shewry, 2009).

From another side, the strategy of unreservedly increasing production (including improving the toolkit in wheat breeding with dwarfing genes) to promulgate wheat in the human diet has been questioned. There is a small, but increasing, incidence of allergy (~0.3 – 3%) to cereals (Brouns et al., 2013) and intolerance in the form of coeliac disease (CD) (1% in EU and US populations) (Mustalahti et al., 2010, Rubio-Tapia et al., 2009). There is an even higher incidence of non-coeliac sensitivity in the population, self-reported or otherwise, with some estimates as high as 30% (Biesiekierski and Iven, 2015, Rona et al., 2007). Other concerns surround high-carbohydrate diets, especially those based on processed wheat-products such as bread and breakfast cereals include links with gastrointestinal discomfort and bloating (Weichselbaum, 2012), chronic inflammation and autoimmune diseases (de Punder and Pruimboom, 2013, Ruiz-Nunez et al., 2013), insulin sensitivity and metabolic syndrome (Gower and Goss, 2015) and cancer (Klement and Kaemmerer, 2011). This has led to a rise in nutritional and lifestyle plans which exclude or limit wheat products, such as the modern paleolithic (paleo) diet (Chauveau et al., 2013, Hwang et al., 2014, Klonoff, 2009, Pastore et al., 2015). The paleo-diet movement is growing: in 2013, ‘paleo’ was the most common word to precede the word ‘diet’ in the search engine Google (http://www.google.com/trends/topcharts?zg=full).

With so many stakeholders with different positions in the wheat and diet industries, polemic exists, including ardent objections from some wheat researchers, such as the authors of the article ‘Does wheat make us fat and sick?’ (Brouns et al., 2013). Despite these objections, there are ramifications on wheat production and emerging trends will challenge wheat production quality as well as quantity. The gluten-free market has been increasing 30% per year and is a billion-dollar industry. Breeding reduced CD-toxicity wheat varieties is now a target, through a combination of germplasm selection (e.g. mining variation in landrace collections) and/or genetic modification (Gilissen et al., 2014).

Future directions of wheat production remain to be seen. Increasing health-issues and morbidity in developed countries must be managed as well as the humanitarian need of feeding an expanding population. Overarching these challenges is the need to manage future strategies in an environmentally-sustainable way (Godfray et al., 2010).

1.1.6 Climate and resources

1.1.6.1 Climate

Wheat production is highly sensitive to climatic and environmental variations and therefore climate change represents a considerable challenge to increasing yields (Semenov et al., 2014). For example, a modelling study for the main wheat growing regions of Australia showed that an increase in growing-season temperatures of 2°C can reduce yield by up to 50%, most of which was attributed to increased leaf senescence due to temperatures above 34°C (Asseng et al., 2011).

Drought and heat stress often occur at the same time, for example in late summer in Europe. Worldwide, drought is the most significant environmental stress in wheat production and therefore, improving yields in water-limited environments is a major breeding goal (Cattivelli et al., 2008). In Europe, modelling predictions for 2050 suggest that climate change will not increase vulnerability of wheat due to drought stress. This was attributed to improvement of the current adaptation of wheat to areas with desiccating summers: quicker maturation. Instead, yield losses primarily due to an increase in the frequency and magnitude of heat stress at meiosis and anthesis were predicted, with northern European heat-sensitive varieties hit most severely (Semenov and Shewry, 2011, Semenov et al., 2014). To respond to climate change, improving varieties to be more tolerant to heat and drought stress will remain priorities for breeding.

1.1.6.2 Nitrogen

Nitrogen is a major macronutrient often limiting plant growth. The application of Nitrogen (N) fertilisers increased rapidly due to the impact of the Green Revolution. In the UK, N fertiliser inputs increased up to the 1980s, supporting

the increasing yields. The N fertiliser/yield relationship is not linear and levels off at around 200 kg N ha^{-1}, at which point increasing N input offers little yield increase, though exact levels differ for different wheat varieties (Barraclough et al., 2010). At very high levels of N application (350 kg N ha^{-1}), no further yield increase occurs, probably due to other limitations such as water or photosynthetic efficiency, although further N uptake is manifested in higher grain N content. Legislation in the UK has limited N application and rates have stabilised at under 200 kg N ha^{-1} (Hawkesford, 2014).

Inefficient use of N fertiliser is economically inefficient and environmentally damaging. There are a number of projects aimed at enhancing N use efficiency (NUE) of wheat to achieve greater yields with less input and balance this with grain quality attributes (Ortiz-Monasterio, 2012, Foulkes et al., 2009). Some researchers have called for the selection for NUE-traits in different wheat varieties at a range of N inputs to obtain greater trait differentiation and better dissect differences in components contributing to NUE (Hawkesford, 2014).

NUE has two components: N uptake efficiency, (NUpE) which is the amount of N taken up as a proportion of total N available, and N utilisation efficiency (NUtE), which is the proportion of N taken up which gets converted to grain yield. Fertiliser use efficiency is determined by NUpE, which is predominantly associated with root structure and functioning (Hawkesford, 2014). The percentage recovery of applied fertiliser has been estimated as 33% across all cereals (Raun and Johnson, 1999). Scientific knowledge of root architecture and the genetic control of root traits affecting nutrient acquisition, branching and anchorage is low compared with the progress made in understanding above-ground wheat adaptation. Since genetic variation in root architecture is associated with yield increase in low-fertility soils, there is a growing imperative to make progress in this relatively understudied area. The wide-reaching impacts, particularly on the poorest farmers on the poorest soils, has led some to call this the next Green Revolution (Lynch, 2007).

1.1.6.3 Organic agriculture

Synthetic nitrogen fertiliser production through the Haber-Bosch process uses natural gas. With pressure to reduce fossil fuel-based inputs, concerns over N

pollution and human health, organic agriculture has rapidly progressed in Europe since the 1990s. In 2010, 5% of the total agricultural area in the EU was cropped organically, doubling from 2000 (David et al., 2012). Organic wheat systems are diverse and all characterised by higher crop diversity and wider crop rotations than conventional agriculture. Soil fertility is maintained by rotations and organic matter. Wheat varieties in organic agriculture have to be more adaptable because there is little opportunity for immediate alleviation of abiotic and biotic stresses. For this reason, yield stability across varied environmental conditions is often more important than achieving maximal yield in an individual season (Wolfe et al., 2008). In organic systems, a greater proportion of N is available early on in development and therefore there is a greater need for early N uptake in wheat varieties than in conventional systems, where N input can be timed with crop demand (Gooding et al., 2012).

1.2 Stature in wheat breeding

1.2.1 The importance of controlling stature

Paintings in the late 19th and early 20th century of English summer landscapes show tall golden wheat, waving in the wind, at head-height or taller. In contrast, today, wheat is shorter and stockier and less amenable to such poetic imagery. This change was caused by an important breeding target of reducing the height of wheat during the Green Revolution.

Optimising wheat stature is important to maximise yield and this varies from 70 – 100 cm according to the yield potential of the environment (Fischer and Quail, 1990, Flintham et al., 1997). Shorter plants are more resistant to lodging (Berry et al., 2007). Further, reduced height of cereals is associated with a greater proportion of assimilates partitioned into the grain, resulting in further yield increases and higher harvest index (the ratio of grain weight to biomass above the ground) (Evans, 1998). However, reduced height is often associated with a reduction in yield (Law et al., 1978) thus understanding better genes which reduce height without yield penalty is important for wheat breeding.

Traditionally, the genetic control underlying height is assessed by measuring plant height at maturity. However, plant height is a dynamic trait (Wu and Lin,

2006) and changes throughout development. Height in triticale (a hybrid of wheat and rye) measured across three time points showed temporal dynamics for height QTL (Wuerschum et al., 2014). Such dynamic studies have been made possible by new technology. High-throughput phenotyping technologies are emerging and can measure various agronomic traits in a non-destructive way across a growing season (Busemeyer et al., 2013, Kjaer and Ottosen, 2015). These platforms have been found to be suitable to field conditions and in the near future will eliminate the phenotyping bottleneck and facilitate dynamic height measurement of wheat.

A total of 21 genes with major effect on wheat height have been identified and assigned *Reduced height* (*Rht*) symbols (McIntosh et al., 2013). These genes have been traditionally grouped into two categories, depending on response to application of exogenous gibberellins (GAs), namely GA-insensitive or GA-responsive. Gibberellins are a major class of plant hormones that regulate plant growth and development, from seed germination and stem-elongation to fruit-set and growth (Hedden and Kamiya, 1997). Mutants with impaired GA biosynthesis or response display GA-deficient phenotypes, which include dark green leaves, late-flowering and a dwarfed stature. Mutants deficient in GA biosynthesis can be rescued by exogenous GA application (Fleet and Sun, 2005).

In addition to these genes, height effects ascribed to QTLs have been reported even in elite panels of commercial wheat varieties (Griffiths et al., 2012, Wang et al., 2010). Therefore, there is still untapped genetic potential for optimising wheat stature in the future.

1.2.2 The Green Revolution genes Rht-B1 *and* Rht-D1

The most common sources of semi-dwarfism in wheat are *Rht-B1b* and *Rht-D1b.* These alleles are part of the *Rht-1* homoeoloci on the group four chromosomes and named according to their sub-genome location: *Rht-A1*, *Rht-B1* and *Rht-D1* (Gale and Marshall, 1976, McVittie et al., 1978). There are a series of alleles at these loci (Gale and Youssefian, 1985, Li et al., 2013, Wilhelm et al., 2013), but the most economically important and most common are the *b* alleles, formerly known as *Rht1* and *Rht2. Rht-B1b* and *Rht-D1b* are GA-insensitive meaning that the application of exogenous GA does not affect the dwarfing phenotype. The mutations in *Rht-B1b* and *Rht-D1b* disrupt their wild-type function as DELLA

proteins which is to act as negative regulators in the GA signalling pathway (Pysh et al., 1999). Both *Rht-B1b* and *Rht-D1b* have base substitutions which result in premature stop codons in the DELLA domain at the N-terminus. As a result, interaction with GA and subsequent degradation is inhibited, resulting in constant growth repression (Peng et al., 1999).

The GA-insensitivity of *Rht-B1b* and *Rht-D1b* causes decreased cell elongation but constant cell number, so smaller cells contribute to reductions in the internodes, without compacting the spike. The overall reduction in plant height is 15-36% (Gale and Youssefian, 1985, Trethowan et al., 2001). The effect of each of the genes is similar, but *Rht-D1b* has a slightly stronger effect than *Rht-B1b* according to Borner et al., 1993.

The reduced cell-size associated with *Rht-B1b* and *Rht-D1b* also decreases coleoptile length and seedling leaf area. This reduces overall seedling vigour and affects the capacity to emerge from deeper sowing. Deeper sowing is preferable in hot and dry conditions which increase seedling mortality, or to avoid animal seed-predation (Botwright et al., 2005, Brown et al., 2003, Mahdi et al., 1998, Rebetzke et al., 2001). Deep-sowing (>5 cm) of shorter-coleoptile *Rht-B1b* and *Rht-D1b* wheats can result in poor and delayed seedling emergence, small leaf area and decreased weed competitiveness (Hadjichristodoulou et al., 1977, O'Donovan et al., 2005, Rebetzke et al., 2007, Trethowan et al., 2001). In addition, though *Rht-B1b* and *Rht-D1b* have increased yield potential in high-input growing conditions, yield reductions have been reported in environments with low N inputs (Laperche et al., 2008) and under some water-limited conditions (Butler et al., 2005, Chapman et al., 2007).

Height reductions conferred by the single action of *Rht-B1b* or *Rht-D1b* can be insufficient to avoid lodging, especially in optimal conditions with high-fertiliser input and irrigation (Berry et al., 2007). Greater reductions in plant height through double-dwarfs with *Rht-B1b*+*Rht-D1b* result in lower biomass and slower seedling leaf area development, though lodging resistance improves (Butler et al., 2005, Flintham et al., 1997). Therefore, alternative dwarfing genes to optimise height in different environments are required in the wheat breeding toolkit, especially with climate change.

1.2.3 Other Rht *loci*

There are a further 18 *Rht* genes which differ from the *Rht1* homoeoloci in being classified as GA-responsive, labelled from *Rht4* to *Rht22* (with the exception of *Rht10*, which is *Rht-D1c* and GA-insensitive) (McIntosh et al., 2013). Further, *Rht23* was recently reported (Chen et al., 2015). The current classification is inadequate, because none of the GA-responsive genes have been cloned. Consequently, the molecular mechanisms of height-reduction remain unknown and roles in GA biosynthesis or signalling, if any, unclear. *Rht12* appears to be involved in GA biosynthesis (Chen et al., 2014) rather than signal transduction like the GA-insensitive genes. *Rht8* is reported not to be not involved in the GA pathway, but has reduced sensitivity to brassinosteroids in leaf tissues (Gasperini et al., 2012). *Rht23* is also reported to have no sensitivity to exogenous GA and have similar endogenous GA levels to its wildtype (Chen et al., 2015).

Rht8 is one of the few GA-responsive *Rht* alleles that reduce plant height and improve lodging resistance without yield penalty (Worland and Law, 1986) and is the main alternative to the GA-insensitive genes in agriculture. Previous evidence indicates that the GA-responsive genes *Rht4*, *Rht5*, *Rht12* and *Rht13* have more extreme height reduction than *Rht8* (Ellis et al., 2004, Flintham et al., 1997, Rebetzke et al., 2012b). Whether these alternative dwarfing genes can be used to improve yield, lodging resistance and seedling vigour in breeding programs is not yet fully established. However, the majority of the GA-responsive genes have a negative impact on yield which can be as severe as 30% (Chen et al., 2013, Daoura et al., 2014, Law et al., 1978, Wang et al., 2015b) and some also delay anthesis by one to five days (Chen et al., 2013, Daoura et al., 2014). Recently, there has been more interest in stacking some of the *Rht* genes together to see if some of the negative agronomic effects can be ameliorated in combination with other genes (Rebetzke et al., 2012a, Wang et al., 2014b, Wang et al., 2015b).

1.2.4 Rht8

Rht8 is prevalent in southern and eastern Europe, where it is likely to provide adaptation to the hot and dry conditions (Worland and Law, 1986) as it provides a semi-dwarf phenotype and improved lodging resistance with no effect on coleoptile length or seedling vigour (Ellis et al., 2004, Rebetzke and Richards,

2000). *Rht8* is also found in China, Australia and North America (Asplund et al., 2012). *Rht8* is not found in northern European germplasm, mainly due to the unfavourable linkage with *Ppd-D1a* (Worland et al., 1998a), estimated to be 22 cM away by Gasperini, 2010.

The *Ppd-D1a* allele contains a 2,089 bp deletion in its promoter region, which converts wheat from a long-day to a day-length insensitive plant (Beales et al., 2007). *Ppd-D1a* reduces time to flowering by early development of floral primordia, once vernalisation requirement has been satisfied, but without the need for long-day exposure. With *Ppd-D1a*, flowering is achieved around a week early in winter-sowing conditions in the UK and height is also reduced (Worland et al., 1998a). The height-reducing effect from *Ppd-D1a* is approximately 4 cm and is independent genetically to *Rht8* (Borner et al., 1993). For this reason, it is important to dissect away the effects of *Ppd-D1a* from *Rht8* to clarify genetic contributions.

Many previous agronomic assessments of *Rht8* have been confounded by the pleiotropic effects of *Ppd-D1a.* Some of these reports are conflicting. In one study in Australia, traits were investigated using recombinant inbred lines (RILs) in pots. *Rht8* decreased grain number per spike, biomass and grain yield, but increased grain weight (2%) and harvest index (6%) (Rebetzke et al., 2012b). Another study assessed agronomic traits in a Chinese winter-wheat variety with *Rht8* (+*Ppd-D1a*), measuring individual plants at F_2 and early generations ($F_{2:3}$ and $F_{3:4}$) in relatively small plots (three rows per plot with a plot length of 2 m). This study reported no difference in grains per spike or grain weight in *Rht8* compared with the tall variety, but found a 17% yield penalty and 10% increase in harvest index (Wang et al., 2015b). A report of *Rht8* in spring wheats in Montana and Washington in the USA specifically tested *Rht8* in a photoperiod sensitive background at the BC_5 generation. *Rht8* did not show any yield advantage over 10 sites studied and showed a penalty in three locations (Lanning et al., 2012). *Rht8* has also been combined with the GA-insensitive semi-dwarfing genes: *Rht8*+*Rht-B1b* or *Rht8*+*Rht-D1b* were 25% shorter and higher yielding (8%) than either dwarfing gene alone (Rebetzke et al., 2012a).

A comprehensive agronomic assessment of *Rht8* in a northern European climate in a commercially-relevant background but without confounding effects of other genes is lacking. This gap in knowledge was addressed in this PhD.

In order to study the effects of *Rht8* (and *Ppd-D1a*), precise genetic stocks were developed known as single chromosome recombinant lines. In these lines recombination is restricted to a single defined chromosome with an otherwise uniform genetic background. The 2D chromosome of the Italian variety Mara, descending from the Strampelli variety Ardito and carrying *Rht8* and *Ppd-D1a*, was substituted into the French photoperiod sensitive variety, Cappelle-Desprez (Worland, 1999). The substitution line was then used to develop chromosome recombinant lines for chromosome 2D. Around 90 lines were developed and genotyped with markers segregating on the recombined 2D chromosome (Law, 1966, Worland and Law, 1986).

Initially, the only way to detect *Rht8* in a variety was to compare the phenotype of chromosome 2D monosomic lines with the euploid parent. A 2D monosomic with *Rht8* typically had a 10% height reduction. From this, *Rht8* was described as a weak allele for height reduction on chromosome 2DS (Worland, 1999).

Subsequently, using the same 2D recombinant lines, the microsatellite marker *Xgwm261* was mapped 0.6 cM distally to *Rht8* on 2DS (Korzun et al., 1998). This marker is named after where it was developed: 'Gatersleben wheat microsatellite' (gwm261) and preceded by an 'X' to indicate a microsatellite marker. This marker could be used to rapidly screen varieties for the presence of *Rht8*. *Xgwm261* produces a number of allelic variants recognised by different lengths of microsatellite amplicons. The height-reducing allele of *Rht8* was associated with a 192-bp allele of *Xgwm261*, though more recently, the 192-bp allele was found not to be universally diagnostic for the height-reducing allele of *Rht8*, particularly from varieties not derived from Mara (Ellis et al., 2007). Our knowledge of the adaptive significance of variants at *Xgwm261* and the extent to which they correlate with variation at the *Rht8* locus is lacking. To address this, work to establish an allelic series of *Xgwm261* variants in a common background was started (Worland et al., 2001) but later suspended. This germplasm was recovered and developed further in Chapter 7.

Gasperini, 2010, used the 2D recombinant lines described above to develop a fine-mapping *Rht8* population in the Cappelle-Desprez background. This population was used in a comparative genomics approach to delimit *Rht8* to a 1.29 cM interval flanked by two single-strand conformation polymorphism (SSCP) markers, *DG279* and *DG371*. Further fine-mapping was prevented by the very low polymorphism between the parents to the fine-mapping population (4% of all markers tested).

To identify causal polymorphisms/genes for *Rht8*, new markers were produced and fine-mapped using these materials, as described in Chapters 5 and 6.

1.3 Mapping genes in wheat

1.3.1 Map-based cloning of genes in wheat

Map-based (positional) cloning is a strategy to isolate genes of interest without making any prior assumptions about the locus of interest. A prerequisite for map-based cloning is a fine-mapping population from a cross between two parents which differ for the trait of interest. Accurate scoring of the phenotype and molecular marker data are then also required to precisely locate the gene of interest on a genetic map. To translate this into physical information, the flanking markers are used to screen complete genome sequences, where available, or clone-based physical maps, such as Bacterial Artificial Chromosome (BAC) libraries. Variant identification in the target interval will lead to the identification of candidate genes which are then validated (Krattinger et al., 2009a).

Map-based cloning and sequencing the genome in wheat is challenging for four main reasons. First, wheat has a large genome at ~17-gigabase-pair (Gb) (Shewry, 2009), which is six times larger than the human genome and 125 times the size of the model plant Arabidopsis. Second, the wheat genome is highly repetitive, with repeat DNA content approximately 80% (Brenchley et al., 2012). This makes sequence assembly challenging with highly homologous stretches of sequence and the transposable element sequences break gene progression relative to related species (collinearity).

Third, as a result of its evolution (Figure 1.1), bread wheat (2n = 6x = 42) is a hexaploid species with an AABBDD genome. The three sub-genomes are

referred to as homoeologous and share sequence identities of ~96-98% (Dvorak et al., 2006), a figure which was found to be maintained across coding regions (Krasileva et al., 2013). Therefore differentiating and assigning genes from the sub-genomes is problematic. As a corollary, it has remained unclear whether the three sub-genomes contribute equally to wheat gene expression and therefore to wheat phenotypes. Advances in wheat genomics have made it possible to study this genome-wide, rather than on a small number of genes. One such study on ~10% of the total wheat gene content found that 45% of genes are expressed from all three sub-genomes and that most of the genes show expression that is dominated by a single sub-genome with very small contributions from the other two (Leach et al., 2014). Transcriptional silencing has been found to be involved in a third of genes which are expressed only from one of the sub-genomes, but this is dynamic in nature, changing temporally and spatially in different organs (Bottley et al., 2006). Taken together, variant discovery specific to one of the genomes is challenging, since the genomes are so similar. Variation between the wheat sub-genomes is called homoeologous variation. The hexaploid nature makes it challenging to distinguish this variation from differences in variation between wheat varieties (varietal variation).

The fourth reason complicating sequence assembly and cloning in wheat is that chromosomal rearrangements are relatively common within hexaploid wheat (Badaeva et al., 2007). The best characterised inter-translocations are between chromosomes 4AL, 5AL and 7BS (Devos et al., 1995, Liu et al., 1992, Nelson et al., 1995). The 3B:6B translocations are also found frequently in European wheats (Badaeva et al., 2007). There has been some attempt to quantify these: one estimate is that 13% of genes from 7BS have been translocated to 4AL (Berkman et al., 2012b). Previous studies have been based on cytology and molecular markers, but progress in genome sequencing has allowed genome-wide and sequence-based investigations into rearrangements. One study of 720 gene interchromosomal rearrangements in wheat reported that 40% were outside of these well-documented locations, scattered across chromosomes including inter-chromosomal translocations to 2DS (Ma et al., 2015a). A large number of intrachromosomal rearrangements has also been reported, including from chromosome 2DS to 2DL (Ma et al., 2014). The emerging extent of these rearrangements is an important consideration in mapping genes in wheat. First,

where comparative approaches are used to order sequence. Second, since translocations can alter levels of recombination and the chances of getting the desired recombinants as part of a map-based cloning strategy will be diminished (Law and Worland, 1997).

Despite considerable effort, only 16 targeted wheat genes have been positionally cloned (Table 1.1). The lack of high-density ordered sequences hinders marker development for high-resolution mapping. In most of the successful cases in Table 1.1, marker development was guided by good synteny with the sequenced genomes of rice or Brachypodium through comparative analysis, but this is not possible in all cases. Cloning wheat genes will become easier with the great advances in wheat genomics and the expansion of genetic resources in the last 5 years. New technologies, such as TILLING (Targeting Induced Local Lesions In Genomes) are being implemented for tetraploid and hexaploid wheat (Uauy et al., 2009) and will be publically available in late 2015. These will permit more precise and efficient characterisation of the function of candidate wheat genes. Many of these advances have occurred during the course of this PhD, which are outlined in 1.3.5.

Gene	Chr	Gene function	References
Gpc-B1	6BS	NAC transcription factor controlling senescence, grain protein, zinc and iron content	(Uauy et al., 2006)
Lr1	5DL	Leaf rust resistance CC-NBS-LRR	(Cloutier et al., 2007)
Lr10	1AS	Leaf rust resistance CC-NBS-LRR	(Feuillet et al., 2003)
Lr21	1DS	Leaf rust resistance CC-NBS-LRR	(Huang et al., 2003)
Lr34	7DS	Fungal resistance ABC transporter	(Krattinger et al., 2009b)
Ph1	5BL	Major chromosome pairing locus	(Griffiths et al., 2006)
PHS1	3AS	Resistance to Hessian fly, heat-shock protein	(Liu et al., 2013)
Pm3b	1AS	Powdery mildew resistance CC-NBS-LRR	(Yahiaoui et al., 2004)
Q	5AL	AP2 transcription factor influencing domestication traits	(Faris et al., 2003)
Sr33	1DS	Stem rust resistance CC-NBS-LRR	(Periyannan et al., 2013)
Sr35	3AL	Stem rust resistance CC-NBS-LRR	(Saintenac et al., 2013)
Tsn1	5BL	Disease resistance to toxins produced by tan spot fungus	(Faris et al., 2010)
VRN1	5AL	AP1-like MADS-box transcription factor controlling flowering	(Yan et al., 2003)
VRN2	5A	Dominant repressor of flowering, downregulated by vernalisation	(Yan et al., 2004)
VRN3	7BS	Vernalisation, orthologue of Arabidopsis FT	(Yan et al., 2006)
Yr36	6BS	Stripe rust resistance START kinase	(Fu et al., 2009)

Table 1.1: Genes which have been positionally cloned in wheat as of September 2015. CC-NBS-LRR = Coiled-coil, nucleotide-binding site, leucine-rich repeat.

1.3.2 NGS advances in sequencing the wheat genome

Because of the ploidy level, high-repeat DNA content and large genome, generating a high-quality reference genome sequence for wheat is a challenge. To reduce the sequencing and assembly complexity, several strategies have been undertaken in the wheat genome-sequencing community. Initially, researchers focused on coding sequences assembling large collections of expressed sequence tags (ESTs) into UniGenes (a collection of ESTs aligned to the same position on a genome, but with insufficient information to annotate as a gene) (Mochida et al., 2009). Most recently, since 2012, next-generation sequencing (NGS) technologies have revolutionised wheat genomics.

The first commercially available NGS system was developed by 454 and capable of sequencing over 20 million base pairs in four hours (Margulies et al., 2005). HiSeq2000 from Illumina can generate 600 Gbp of data per run, equal to more than 35 hexaploid wheat genomes (http://www.illumina.com). Although NGS technologies produce shorter reads and have greater error rates than Sanger sequencing, they made it feasible to generate the vast sequence data associated with the large wheat genome at lower cost and reduced timeframe (Berkman et al., 2012a).

NGS enabled a whole-genome shotgun (WGS) assembly of wheat to be published, which was based on low-coverage (5x), relatively long-read (454) shotgun sequences of the model wheat variety Chinese Spring. The assembly was fragmented and order was based on diploid progenitor genomes (Brenchley et al., 2012).

Technological advances of high-throughput chromosome isolation using flow cytometry enabled a chromosome-by-chromosome strategy to be adopted to sequence the wheat genome (Vrana et al., 2012). The International Wheat Genome Sequencing Consortium (IWGSC) was formed to construct the physical map and reference sequence in wheat using a chromosome-based approach. Flow sorting can reduce the sample size and complexity by separating chromosomes and, if the purity is high enough, avoid the complications of homoeologous sequences. Flow-sorting directly separated the largest chromosome, 3B, and a BAC library was constructed and assembled into a

physical map (Paux et al., 2008). To separate the other chromosomes from their homoeologues, aneuploid Chinese Spring genetic stocks were used and a 10.2 Gb draft (chromosome survey sequence, CSS) assembly was generated (IWGSC, 2014). Physical map construction of other chromosomes is at various stages of completion (http://www.wheatgenome.org/). The first IWGSC version 1.0 has been improved with more variation data from various sources and population sequencing (POPSEQ) (Mascher et al., 2013) data and released as IWGSC version 2.0 (IWGSC-2) (plants.ensembl.org).

In addition to the chromosome-based strategy, a whole-genome shotgun (WGS) approach has yielded scaffolds of each of the three homoeologous genomes, with better contiguity over coding regions, and covering new sequence space to the IWGSC CSS contigs (Chapman et al., 2015).

Recently, the same POPSEQ map was used to genetically anchor a proportion of both the IWGSC CSS contigs (4.5 Gb) and WGS scaffolds (7.1 Gb) (Borrill et al., 2015, CerealsDB, 2015a, Mascher, 2014, Mascher et al., 2013). These genetic maps, albeit relatively coarse due to the limited size of the POPSEQ population (Sorrells et al., 2011), allow a more targeted approach for gene discovery.

To expedite the bread wheat sequencing efforts, a further strategy is to leverage comparative analysis from the genome sequences of the three diploid ancestors (*T. urartu*, *Ae. speltoides* and *Ae. tauschii*, Figure 1.1). The A and D-genome progenitors have been sequenced using WGS (Jia et al., 2013, Ling et al., 2013). A physical map of *Ae. tauschii* has been generated (Luo et al., 2013) and the reference sequence is being produced (http://aegilops.wheat.ucdavis.edu/ATGSP/, 2015).

1.3.3 Comparative Genomics

Comparative genomics between wheat and more genetically-tractable diploid organisms within the related grass species (Poaceae family) has contributed greatly to the analysis of the more complex wheat genome. The grass species have diverged over the past 60 million years through whole genome duplications, chromosome rearrangements and deletions (Gale and Devos, 1998). A high-level

of conservation of gene content (synteny) and gene order (collinearity) has been reported between grass species (Moore et al., 1995).

The lineage with *Sorghum bicolor* (sorghum) and *Zea mays* (maize) diverged over 70 million years ago (MYA), followed with divergence between *Oryza sativa* (rice) and the Pooideae lineage (a subfamily within the Poaceae family) 50 MYA (Figure 1.2) (Middleton et al., 2014). *Brachypodium distachyon* (Brachypodium) and *Hordeum vulgare* (barley) have a higher conservation of synteny with wheat than rice and are more closely-related (Akpinar et al., 2015, Girin et al., 2014, Luo et al., 2013, Massa et al., 2011). The most closely-related species (other than diploid progenitors) to wheat is barley (Figure 1.2). A WGS assembly of barley has been published followed by a physical map (Ariyadasa et al., 2014, IBGSC, 2012). The barley resources are still not as complete or extensive as those for rice and Brachypodium, since a complete genomic reference has not been completed, but is anticipated soon. One difficulty accessing barley resources is that they are located in disparate locations without common identifiers, making it hard to compare between them. Very recently, all the genetic and physical resources have been integrated into one database with common identifiers in a web-based application called BARLEX (Colmsee et al., 2015).

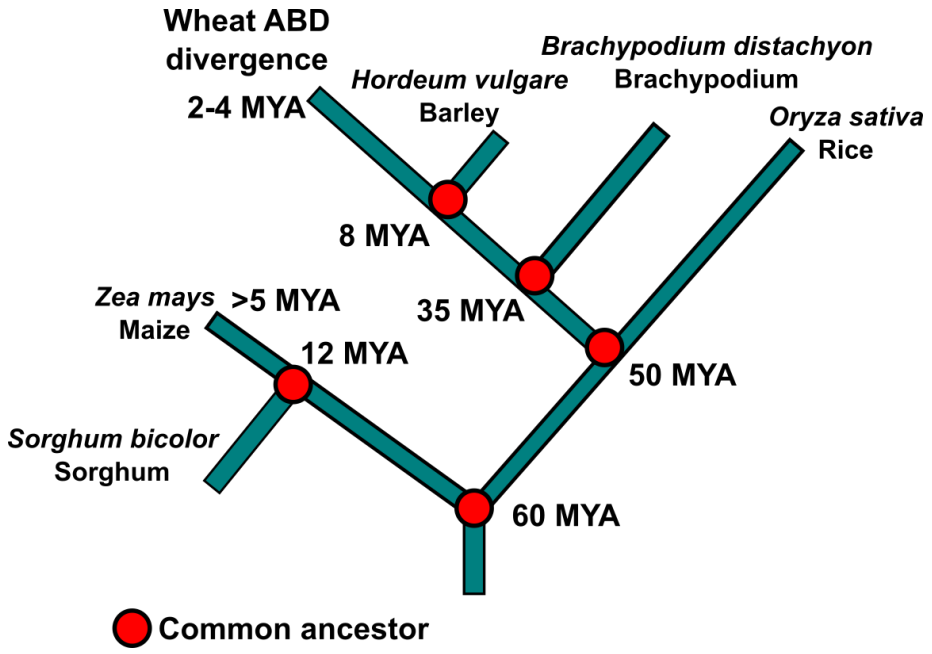


Figure 1.2: Evolutionary history of the Poaceae. Divergence based on Middleton et al., 2014.

Comparative analyses compiling annotations and genetic maps of different grass species have been combined into Genome Zippers. Genome Zippers allow ordering of genes and genetic maps based on physical data from syntenic species (http://aegilops.wheat.ucdavis.edu/ATGSP/, 2015, Mayer et al., 2011, URGI, 2015a). Recently, datasets of model species in the Pooideae and the Triticeae tribe (including barley and wheat) have been organised into an integrative viewer in Ense*mbl*Plants, which circumvents the requirement to perform manual iterations of comparative analyses between different resources stored in disparate locations (Bolser et al., 2015). These approaches have already been used to enrich *T. aestivum* gene annotations (Tulpan et al., 2015) and advance fine-mapping of genes on the wheat D-genome (Chen et al., 2015, Liang et al., 2015).

As the sequencing of bread wheat advances in a chromosome-by-chromosome approach, there has been a recent flurry of publications and those focusing on the D chromosome have assessed syntenic relationships between *T. aestivum*, *Ae. tauschii* and other related species. These have found that 26% of *Ae. tauschii* genes have no orthologues in collinear locations in rice, Brachypodium and sorghum (Luo et al., 2013). A Genome Zipper analysis on wheat chromosome 4DS found that 25% of genes were supported by orthologous relationships with all three of these reference species, and a majority of genes (55%) were supported by only one. The same study reported less than half (49%) of survey sequences from pyrosequencing of wheat chromosome 4 anchored to *Ae. tauschii* scaffolds on the homologous chromosome (Helguera et al., 2015). Another comparative analysis of *T. aestivum* and *Ae. tauschii* chromosome 5 found that orthologous genes matching *Aegilops* was lowest in barley compared with Brachypodium and rice (possibly a reflection on the incomplete barley reference), but that within this number, a significant number of *Aegilops* sequences matched with barley orthologues which had no similarity with any of the model grass genomes (Akpinar et al., 2015).

Taken together, the emerging picture is that the main limitation of the comparative genomics approach between wheat and related species is the mosaic of conserved synteny at the micro-collinearity level. This complicates the use of such comparisons for map-based cloning and marker discovery, but a manifold approach with different resources can circumvent some of these limitations.

1.3.4 Genetic mapping

A prerequisite for map-based cloning is a high-resolution genetic map which requires development of a population to then be saturated with molecular markers.

Earlier, restriction fragment length polymorphism (RFLP) (Chao et al., 1989), amplified fragment length polymorphism (AFLP) (Barrett and Kidwell, 1998), random amplified polymorphic DNA (RAPD) (Devos and Gale, 1992) and simple-sequence repeat (SSR) (Roder et al., 1998) markers were used for variety characterisation. SSR markers are PCR-based DNA markers which require only a small amount of template and can be efficiently used to screen large populations. Most recently, SSR markers have been superseded by single nucleotide polymorphism (SNP) based approaches (Cavanagh et al., 2013). SNPs occur in genomes at a much higher frequency than SSRs and have a lower error rate in detection (Duran et al., 2009). SNP variation can be detected in a much higher-throughput manner, for example using fluorescence-based genotyping technology such as the KASP (Kompetitive Allele Specific PCR) assay (He et al., 2014).

1.3.4.1 Identifying variation in wheat

Great progress has been made in generating wheat sequence from Chinese Spring and the WGS of synthetic wheat W7984 (Chapman et al., 2015, IWGSC, 2014). Re-sequencing whole genomes of specific wheat varieties of interest is not yet viable. For this reason, methods to reduce complexity such as focusing on transcriptomes or exomes have been employed to uncover variation, mainly SNP variation. Recent application of NGS has improved the throughput of SNP discovery. Thousands of SNPs have been uncovered from bread wheat transcriptomes (Allen et al., 2011, Cavanagh et al., 2013) and exomes (Allen et al., 2013). A large number of identified SNP variants have been converted into high-density SNP platforms which can genotype wheat populations. SNP platforms such as the iSelect array with 90,000 SNPs (Wang et al., 2014a) and Affymetrix Axiom® array with 820,000 SNPs (www.cerealsdb.uk.net/cerealgenomics) have driven down the cost per assay. Further, a proportion of SNPs in each platform have been genetically mapped by

combining different mapping studies (http://www.wgin.org.uk/, (Cavanagh et al., 2013). Bioinformatics pipelines such as PolyMarker now allow rapid conversion of array-based assays into genomic-specific KASP assays (Ramirez-Gonzalez et al., 2015). A limitation of these SNP platforms is that they rely on the pre-determined set of SNPs on the original discovery panel. Where variety- or population-specific variants are required, for example for fine-mapping, more targeted variant discovery is necessary.

1.3.4.1.1 Targeting variant discovery to genetic intervals

There are two main methods of direct (unbiased) variant detection in wheat, which both use a strategy of first reducing complexity before resequencing. The first method is genotyping-by-sequencing (GBS) (Poland et al., 2012). GBS reduces the genome complexity by digesting the template with restriction enzymes and size-selecting the fragments (Wang et al., 2015a). Downstream GBS bioinformatics analysis is currently complex and it was not clear at the start of this project in 2012 that SNP-calling would be accurate. The second method was used in this thesis, called bulked segregant analysis (BSA) (Michelmore et al., 1991). BSA is a technique that can be combined with NGS of mRNA (called RNA-Seq) to target SNP discovery to a particular genetic interval. Two pools (bulks) of individuals from a population segregating for a specific trait of interest are compared, allowing identification of allelic variation from one of the parents to the population which is enriched in the appropriate bulk. BSA has been combined with RNA-Seq to identify SNPs in mapping studies in tetraploid (Trick et al., 2012) and hexaploid wheat (Ramirez-Gonzalez et al., 2014).

1.3.5 Advances in wheat resources over the course of the thesis

At the beginning of this thesis, the IWGSC chromosome survey sequence had been made available to researchers (restricted by password access) via a BLASTable database hosted by the Unité de Recherche Génomique Info (URGI), a research unit in bioinformatics at Institut National de la Recherche Agronomique (INRA) (URGI, 2013). This first release version (IWGSC-1) was not curated as a set of contigs in their physical order. Instead, other than chromosome arm provenance, the contigs were unannotated. During the course of the project,

there was a rapid expansion in the Triticeae resources available and the most salient are shown in Figure 1.3. The main advances pertained to wheat gene models and accessibility of resources from syntenic species.

As part of the RNA-Seq BSA strategy, a genomic reference is required. With the absence of a reference genome, this was achieved in this project by using the best available gene models for wheat. Since the IWGSC-1 contigs were redundant and also unordered, these were not suitable in this state. At the start of the PhD, the available resource was a UniGene reference which was based on *de novo* assemblies of diploid progenitors which had been ordered using Brachypodium synteny and a coarse wheat genetic map (Harper et al., 2015). By the end of the project, more complete gene models were available. However, not all developments could be fully capitalised on due to time constraints.

Synteny between the Pooideae genomes had already been established as an extremely valuable resources prior to the start of this PhD. The emerging resources in barley and diploid progenitors throughout this PhD further contributed to this. However, datasets were deposited in disparate locations and could not always be unified using a common identifier, which made it difficult to navigate between them. By 2015, integration of these resources was much more comprehensive.

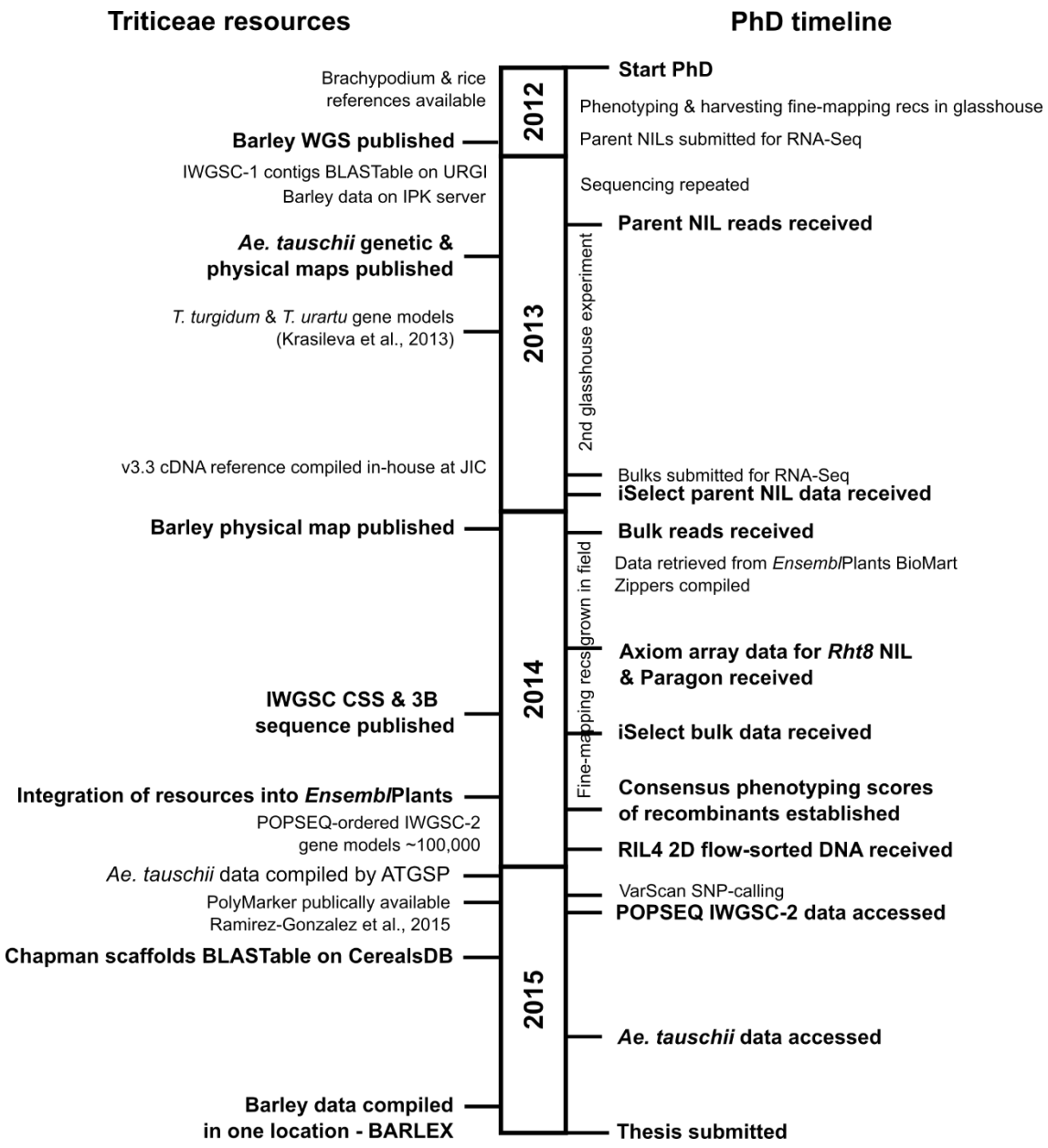


Figure 1.3: Developments in Triticeae resources over the course of the PhD (left) and a time-line of some of the work in Chapters 5 and 6 to fine-map Rht8 (right).

1.4 Aims of the thesis

Rht8 is one of the few *Rht* alleles that reduce plant height and improve lodging resistance without yield penalty and is the most prevalent alternative to the GA-insensitive genes in agriculture, mainly in southern and central Europe. *Rht8* is not found in northern-European germplasm and its potential for breeding in different N treatments and water regimes in such climates has not been thoroughly studied. Further, most agronomic assessments of *Rht8* have focused on height, leaving developmental traits, yield and the underlying components poorly understood. *Rht8* has also been anecdotally reported to have a compact spike in observations in the field, but the veracity of this has not been investigated further or quantified.

The first part of the thesis aims to address these questions. The first objective is to assess the influence of *Rht8*, without the confounding effects of *Ppd-D1a* or other major dwarfing genes, on height, height components and yield components in a wheat population adapted to northern European conditions and in a commercially-useful wheat background. To achieve this, the thesis will build on previously developed near-isogenic lines (NILs) contrasting for the *Rht8* allele from Mara and tall *rht8* allele from Cappelle-Desprez developed in the elite spring wheat, Paragon (Gasperini, 2010). The second objective is to investigate spike compactness in this material, to test the veracity of qualitative anecdotal evidence.

The second part of the thesis aims to further fine-map *Rht8* by an RNA-Seq enabled bulked segregant-analysis strategy, which had been applied at the time of starting this PhD in tetraploid wheat. This builds on previous efforts which mapped *Rht8* to a 1.29 cM interval on chromosome 2DS. To achieve this, the fine-mapping *Rht8* population has to be accurately phenotyped and wheat sequence from a variety of sources mined for useful variation. With emerging resources, this will be achieved by constructing Zippers of syntenic species.

The fine-mapping of *Rht8* will anchor the interval in the most up-to-date Triticeae resources and detect markers amenable to high-throughput genotyping which will be useful to breeders and the research community. This will be achieved by saturating the region with molecular markers, through manifold approaches from

the emerging resources in wheat. In doing so, the usefulness of different resources will be evaluated.

Chapter 2: Materials and Methods

2.1 Agronomic characterisation of *Rht8* in UK-adapted germplasm

2.1.1 Near Isogenic Lines

The material used in this project derived from previous work by Gasperini (2010). RIL28, from the 2D RIL population described in 2.3.1.1 was used as the Mara-derived *Rht8* donor (female parent) and crossed to Paragon. Paragon is a high-bread-quality commercial spring wheat variety in the UK and does not contain *Rht-B1b* or *Rht-D1b* and is photoperiod sensitive (*Ppd-D1b*). A series of backcrosses and marker-assisted selection with *Xgwm261* and *Xcfd53* (markers named after the locations where they were developed: Gatersleben wheat microsatellite and INRA Clermont-Ferrand, respectively) produced BC_3F_2 seed which was then multiplied in the field (Figure 2.1). The BC_3F_2 near-isogenic lines (NILs) contrasted for the *Rht8* allele from Mara (defined by marker-assisted selection for gwm261-192bp and cfd53-274bp) and tall *rht8* allele from Cappelle-Desprez. One (*Rht8*) short NIL and one tall NIL were selected at the start of this project (Figure 2.2) at the BC_3F_3 stage to be used in further trials.

2.1.2 Sites and experimental design

The *Rht8* and tall NILs were grown along with Paragon in field trials across three locations: Church Farm, Bawburgh, Norfolk; the Crops Research Unit, Sonning, University of Reading and Lleida, Catalonia, north-eastern Spain. The site coordinates, soil, plot and drilling details are in Table 3.1, along with the specific traits measured at each location. The NILs were grown in Norwich over three years (2012-14) in a randomised complete block design (RCBD) with three replications in 2012/13 and five replications in 2014, though tiller samples were only taken from three out of five replications in 2014. The same design of RCBD with three replicates was used in Lleida over 2013-2014 and a split-plot design with five replicates was implemented in Reading. Drilling dates were third week

of November 2012 and 2013 in Lleida, 17th October 2013 in Reading, 13th October (nitrogen trial)/16th October (irrigation trial) 2012 and 19th October 2013 at Church Farm. In trials with contrasting nitrogen (N) treatments, 40 kg N ha^{-1} was applied at Zadoks growth stages GS30-31 (Zadoks et al., 1974) and a further dose of N applied at GS34-39 to make up to the required levels for N2 (total 100 kg N ha^{-1}) and N3 (total 200 kg N ha^{-1}). For the irrigation experiments, trickle irrigation was applied using a timer and piping between each row within a plot. In 2013, water was applied from 17th June – 25th July (after stem elongation) five days per week, receiving 15 litres m^{-2} day^{-1}. In 2014, irrigation was applied from 30th April to 23rd May on 14 occasions (no irrigation was supplied on rainy days), receiving approximately 14 litres m^{-2} day^{-1}, but on three occasions 10 litres m^{-2} day^{-1} due to leaks. Field trials were kept weed- and pest-free with products according to standard agronomic practice at each of the locations, with the exception that plant growth regulators (PGRs) were not applied.

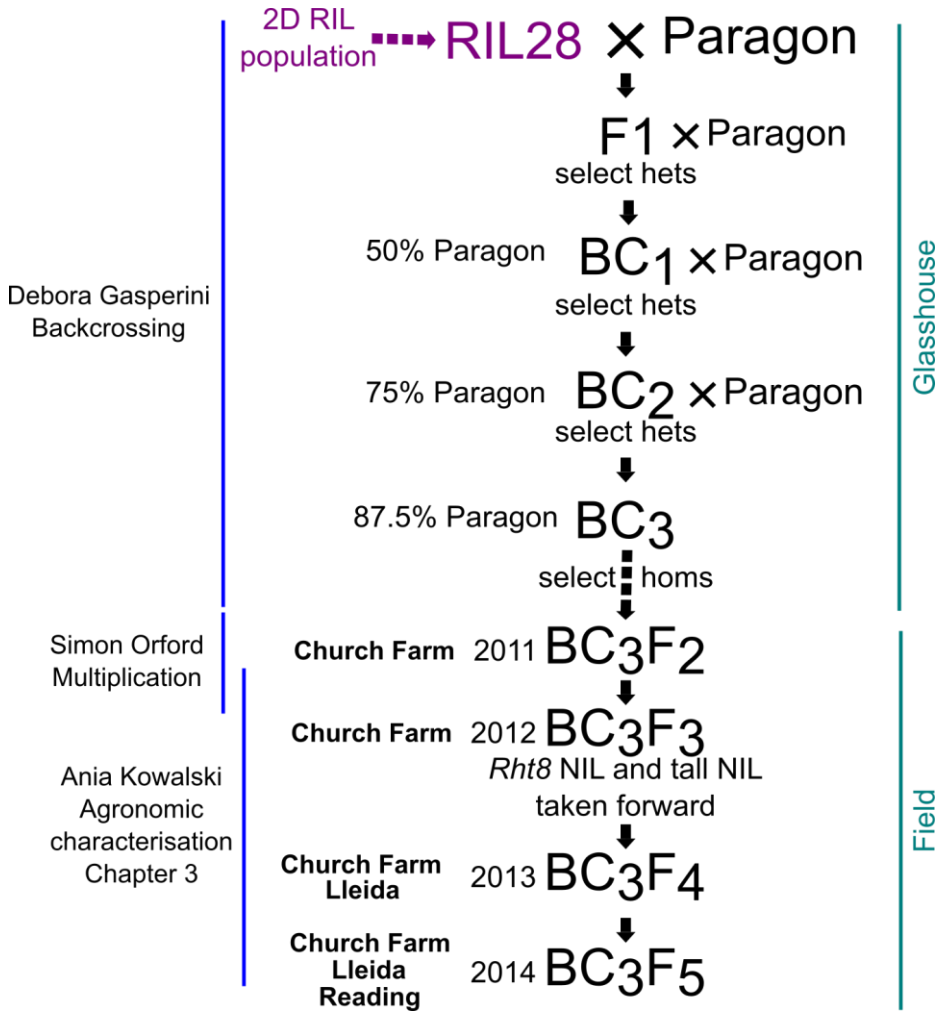


Figure 2.1: NILs carrying a Mara-derived Rht8 introgression in the spring variety Paragon used in this project. The material originated from a backcrossing programme by Debora Gasperini and then multiplied in the field by Simon Orford.

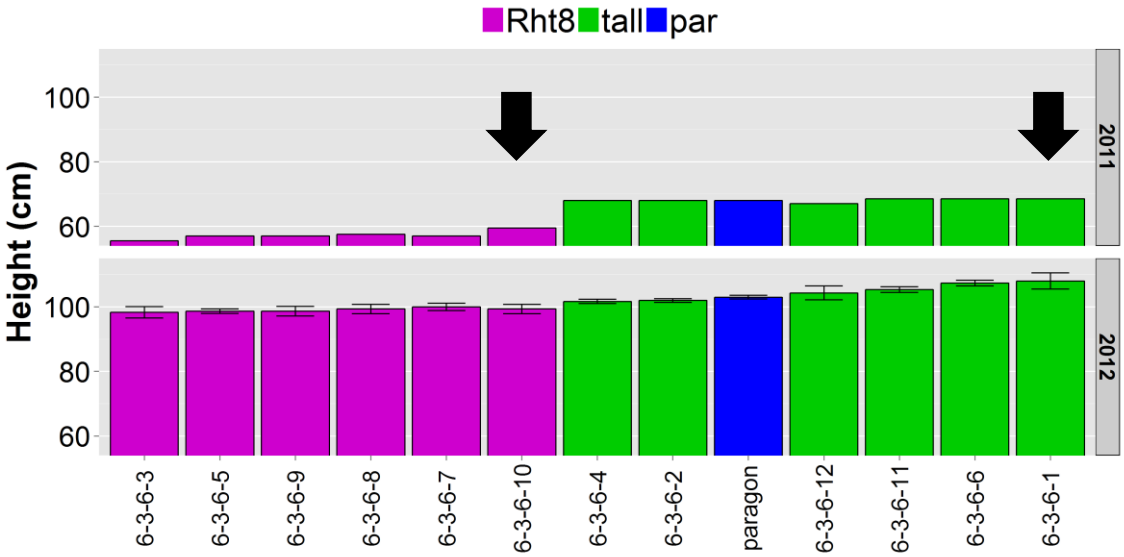


Figure 2.2: Selecting the Rht8 (short) and tall NILs at the BC_3F_3 stage for further field trials. The selected NILs are shown with arrows, Rht8 NIL (left,) and tall NIL (right).

2.1.3 Climate and day length

Weather data for Reading were recorded at an automated metrological station at the Sonning site; obtained for Lleida from Gustavo Slafer at the University of Lleida and from the Buxton Weather Station (http://www.buxton-weather.co.uk/weather.htm) for Norwich, ~15 km from the Norwich trial site in Bawburgh. The day-lengths for the 2013-2014 growing season for Norwich and Lleida were calculated using the maptools package (Bivand and Lewin-Koh, 2015) in R v3.1.1 (R Development Core Team, 2014). Maptools has functions for calculating sunrise and sunset using algorithms provided by the National Oceanic & Atmospheric Administration (NOAA). The R script using these functions to calculate day length is shown in Appendix 2.1.

2.1.4 Phenotyping and assessments

Grain yield was recorded per plot and adjusted for plot size. Grain m^{-2} was calculated from grain yield and thousand grain weight (TGW). Plant height was measured from soil level to the top of each wheat ear. This was measured at maturity from a visually-determined representative tiller per plot at Church Farm and in Reading using a rising disc of polystyrene (Peel, 1987). Plant height and height components were also measured from tillers taken from each plot (outlined below) and extended along a ruler. The spikes m^{-2} was calculated from plant

populations taken at Reading and Lleida using the mean of three 0.1m^2 circular quadrats in each plot.

Developmental stages of heading date at GS57 and anthesis at 50% emergence from the flag leaf were measured when 50% of the plants in each plot reached them. Senescence was measured at Church Farm in 2013 and assessed on a scale from 0 to 10 with 0 being peduncles completely green and 10 being loss of green colour in 100% of peduncles in a plot. Heading, anthesis and senescence were assessed in thermal time of °C days, by calculating the cumulative temperature from drilling to assessment date. The mean daily temperature was calculated from the minimum and maximum daily temperatures from the weather data described in 2.1.3. At Reading, the proportion of photosynthetically active radiation (PAR) and red:far-red reflectance ratios were measured at the base of the canopy on the dates shown in Appendix 3.8 using methodologies described elsewhere (Addisu et al., 2009b).

Prior to harvest, three main tillers from three plants were sampled from each plot in the UK sites only. Tillers were used for assessment of height, height components and yield components: harvest index (ratio of grain weight to above-ground biomass), spikelets spike^{-1} and TGW. Above-ground biomass was measured before threshing the grain. Morphometric measurements (grain width, length, area and TGW) were recorded from threshed grain using 300-400 grains per sample on the MARVIN grain analyser (GTA Sensorik GmbH, Germany). The internode below the spike was defined as the peduncle and the successive internodes as the first, second, third internodes, respectively. Each internode was measured from the mid-point of the subtending node.

Ground cover was measured at Church Farm on 25th March 2014 and 26th March at Reading. Images of plots were taken at waist-height and the proportion of green canopy in the plot was measured using an ImageJ (Abramoff et al., 2004) macro written by Oscar Gonzalez which calculates the proportion of the image with green cover. The script for the macro is in Appendix 2.2.

Lodging was measured in each plot where any degree of lodging had occurred at approximately GS70 in July 2014. Lodging score was calculated using the percentage of the plot area which had lodged multiplied by the angle of lodging (0 to 90°) (Fischer and Stapper, 1987). Lodging score ranged from 0 to 100, with

0 being no lodging and 100 being total displacement to horizontal across the whole plot.

2.1.5 Statistical analyses

Comparisons between NILs were carried out using genotype analyses of variance (ANOVAs) to assess the effects of genotype within treatment combinations. For Lleida, a two-way ANOVA was performed for data across both years, using a treatment structure of year*genotype with block as the random effect. For Church Farm, for the nitrogen and irrigation trials in 2013 and for nitrogen in 2014 a two-way ANOVA was performed with the treatment*genotype (treatment structure) and block as random effect. Residual Maximum Likelihood (REML) analysis was carried out for the irrigation experiment at Church Farm in 2014, where the fixed effects were N treatment*water treatment*genotype. The NILs in Reading were compared at different nitrogen treatments using a split-plot ANOVA where nitrogen treatment was the main plot and genotype the sub-plot. ANOVA, REML and correlations were performed using GenStat 16th edition (VSN International). Fisher's least significant difference (LSD) test was used to determine significant differences between means at the 0.05 level. The complete data from the analyses of each trial is shown in Appendix 3.

2.2 Compact spike morphology

2.2.1 Measuring spike compactness

Spike compactness was scored visually by assessing the percentage of spikes in the plot which were compacted. No distinction could be made visually to the degree of compactness. All five plots at each water regime at low nitrogen (40 kg N ha^{-1}) for the *Rht8* NIL showed a degree of spike compactness and all 15 plots with the *Rht8* NIL in Reading were also assessed to have spike compactness. All other plots and genotypes had no discernible compaction (0%).

Spike compactness was quantified using the same tiller samples taken before harvest as outlined in 2.1.4. Spike length and spikelet number spike^{-1} were used to calculate compactness as cm spikelet^{-1}. To measure compactness in the fine-mapping *Rht8* recombinants described in 2.3.1.2, a subset of 20 fine-mapping

recombinants typed short and 20 fine-mapping recombinants typed tall were selected and spike lengths measured from the images taken of developing spikes at as outlined in 2.3.4.1. Spikelet number was counted visually and compactness calculated from the two values.

Pearson's correlation coefficients were calculated between spike measurements and height. Analyses are as described in 2.1.5 and shown in full in Appendix 4.

2.3 Development of molecular markers within the *Rht8* interval

2.3.1 Plant material

2.3.1.1 2D RIL (coarse-mapping) population

The 89 2D RILs were initially obtained by developing a 2D substitution line in which the 2D chromosome of Mara, the Akakomugi-derived *Rht8* donor, was substituted into the Cappelle-Desprez background. This was achieved by back-crossing to the existing Cappelle-Desprez monosomic stock for 2D (Law, 1967, Law and Worland, 1973). The Mara 2D substitution line was crossed to the recipient parent and the F_1 further crossed to the Cappelle-Desprez 2D monosomic line. The progeny with 41 chromosomes were extracted from the hybrid progeny and self-fertilied for selection of disomic lines carrying different 2D chromosomes with homozygous recombination events in an otherwise homozygous Cappelle-Desprez background (Korzun et al., 1998). These recombinant lines are the 89 2D RILs.

2.3.1.2 Fine-mapping and medium-resolution populations

The fine-mapping population from which the fine-mapping (FM recs) and medium-resolution mapping (gwm recs) populations were selected were developed by Gasperini (2010). An outline of the population development is shown in Figure 2.3. RIL4, from the 2D RIL population described in 2.3.1.1, was used as a female parent and crossed into Cappelle-Desprez. The F_1 plants were self-fertilised to produce 3104 F_2 plants which were screened with the markers *Xgwm261* and *Xcfd53* for recombinants. Recombinants were self-fertilised and

the resulting 152 F_3 families were genotyped to identify homozygous *Rht8* recombinants (recombinant with respect to the original crossing parents and homozygous at both flanking-marker loci). The F_3 recombinants were self-fertilised and the F_4 seed was obtained as the start-point for the work in this project. Originally, of these recombinants, 79 were used to resolve *Rht8* to a 1.29 cM interval between *DG279* and *DG371*. A total of 69 wider recombinants which were recombinant between *Xgwm261* and *Xcfd53* but outside the *DG279*-*DG371*-defined interval were used as the medium-resolution mapping population (gwm recs). All recombinants were screened using *Xgwm261* and *Xcfd53* to verify previous genotyping. From the 79 FM recs, F4-1-1-10-5 and F4-1-1-2-9 were discarded as scoring errors in population development since they were heterozygous at one of the flanking-marker loci. F4-1-1-6-4 and F4-1-1-6-5 were completely sterile when grown in the first glasshouse experiment and were discarded. F4-3-1-6-4 and F4-3-1-1-8 had insufficient seed to be taken forward without staggering a generation and were discarded. This left a total of 73 FM recs. The FM recs were further genotyped with *DG279* and *DG371* prior to fine-mapping.

2.3.1.3 DNA extraction

Wheat seeds for the parent NILs, 2D RIL population, fine-mapping population and medium-resolution mapping population were germinated on wet filter paper in Petri dishes at 20°C for 24 hours following a cold treatment at 5°C in the dark for 48 hours. Germinated seeds were planted into individual cells (4 x 4 cm) of 96-well trays filled with a mixture of peat and sand. Two-week old leaf tissue was harvested into microtubes containing a 3mm tungsten bead (Qiagen, 699997) in a 96-well collection box (Qiagen, 19560) by folding a 5 cm section of leaf into a concertina. DNA was extracted according to the Somers and Chao protocol (http://maswheat.ucdavis.edu/PDF/DNA0003.pdf), adapted from Pallotta et al., 2003. DNA was quantified at a wavelength of 260 nm using a NanoDrop 2000 (ThermoScientific). Yields per extraction were 60 – 150 ng μl^{-1}.

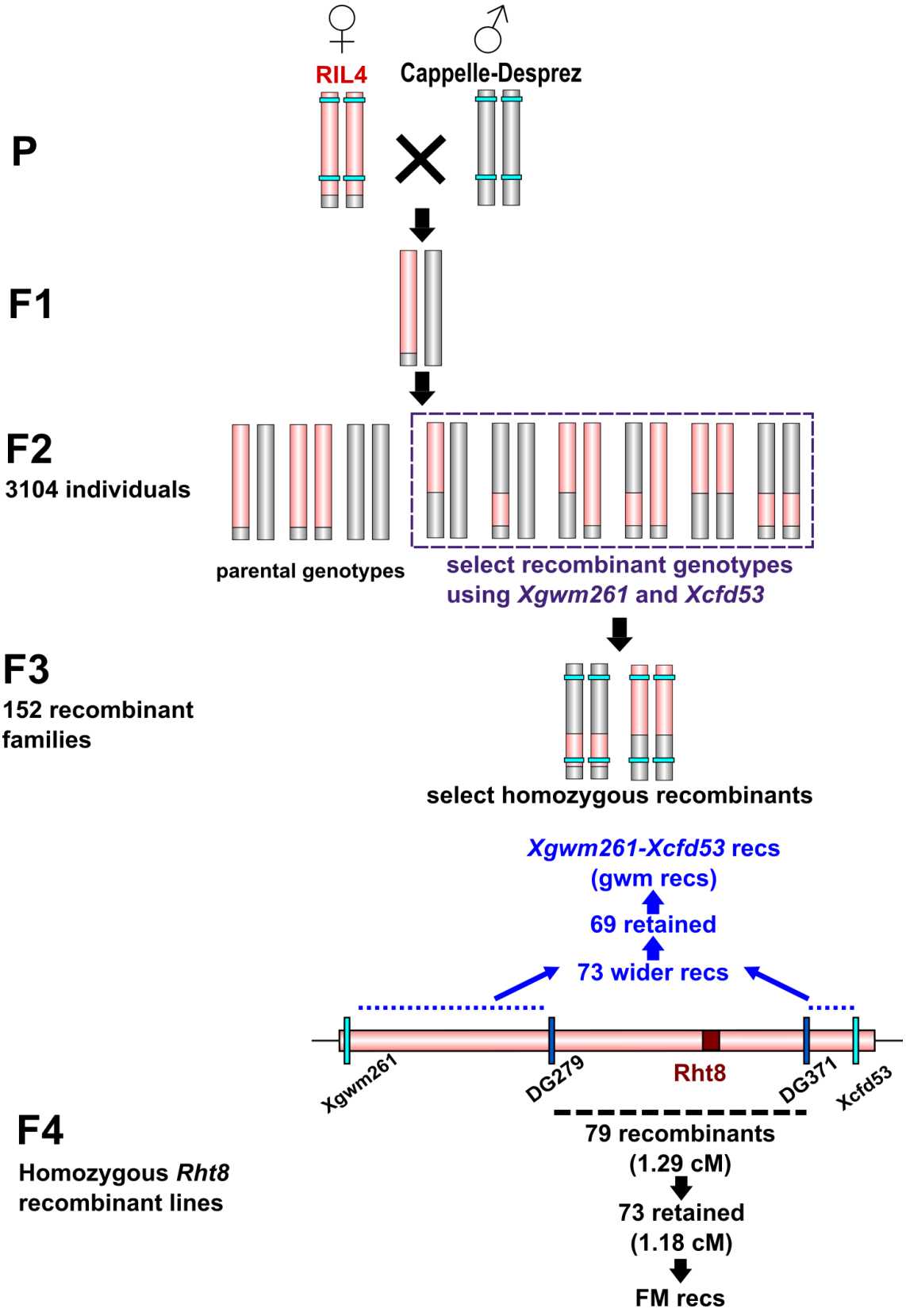


Figure 2.3: Background to Rht8 fine-mapping population development by Gasperini (2010) and the selections made for this project. Marker names and numbers are derived from laboratory designators and chronologically-ordered lists used by the scientists who first published them and preceded by an 'X' to indicate a microsatellite marker: 'Gatersleben wheat microsatellite' (gwm261) (Roder et al., 1998) and INRA Clermont-Ferrand (cfd53) (Paillard et al., 2003).

2.3.1.4 Screen with flanking markers

The FM recs and gwm recs were screened with *Xgwm261* and *Xcfd53* to verify previous genotyping. Primer sequences and amplification conditions were obtained from GrainGenes (http://wheat.pw.usda.gov/cgi-bin/graingenes). The FM recs were also screened with *DG279* and *DG371*, using the primer sequences and amplification conditions described by Gasperini et al., (2012). PCR reactions were conducted as described in 2.3.10.1 but LIZ1200 (Applied Biosytems) was used as the sizing standard.

2.3.2 Material for genetic dissection

2.3.2.1 Phenotyping for height in glasshouse experiments

The 73 fine-mapping recombinants described in 2.3.1.2 were grown in the glasshouse and in the field in order to obtain height measurements. First, the fine-mapping recombinants and the parent NILs were grown in a lit glasshouse in long days and phenotyped for height. Wheat seeds were germinated on wet filter paper in Petri dishes at 20°C for 24 hours following a cold treatment at 5°C in the dark for 48 hours. Germinated seedlings were grown in a mixture of peat and sand and vernalised at 5°C with 70% relative humidity for 10 weeks under short days (10 h light/14h dark). The light in the vernalisation chamber was provided by tungsten lamps to an intensity of 250±50 µmol m^{-2} s^{-1} at canopy level.

Plants were subsequently transplanted into 1L pots in September 2012 and transferred to a glasshouse. Plants were grown in long days (16h light/8h dark) under HPS 400 W lamps providing 400±50 µmol m^{-2} s^{-1} at canopy level. The temperature ranged from 15°C - 33°C (mean of 23°C) and relative humidity ranged from 35% - 81% with a mean of 56%. Temperature and humidity was recorded using a USB data logger (RS Components Ltd, 4801064) tied to a stake at canopy level which recorded data every 30 min.

Plants were arranged in a randomised-block design trial of eight blocks split equally over two central benches. Each block contained one replicate of each genotype along with controls of the tall parent (Cappelle-Desprez) and short parent (RIL4) (Figure 2.4). No plant growth regulators were applied. Final plant

height was measured from the soil to the tip of the spike of the main tiller. Where spike length was measured, the height to the bottom of the spike was subtracted from the total height. The ears of each plant were bagged and hand-threshed.

Bench 1

Block 1	F4-3-7-9-1	F4-2-7-3-6	F4-1-1-12-1	F4-1-1-10-5	F4-1-2-4-3	RIL6
	F4-3-8-5-2	F4-3-8-3-3	F4-1-7-1-1	RIL6	F4-2-1-16-3	F4-1-7-4-1
	F4-1-6-12-1	F4-1-7-17-1	F4-3-2-5-1	F4-1-7-7-1	F4-3-1-1-6	F4-3-2-13-1
	F4-3-1-3-6	F4-2-7-9-2	F4-2-2-7-1	F4-1-6-19-2	F4-1-2-2-1	F4-3-3-15-1
	F4-3-2-8-1	F4-2-8-4-1	F4-3-2-15-1	F4-1-6-17-1	F4-1-9-1-1	F4-1-6-13-2
	F4-2-1-8-4	F4-2-2-10-1	F4-1-2-9-1	F4-3-7-8-2	F4-2-3-8-1	F4-3-7-14-3
	RIL4	F4-2-7-12-2	F4-3-7-6-1	F4-1-1-2-9	F4-1-1-9-7	F4-1-9-3-1
	F4-3-7-7-2	F4-2-1-11-4	F4-2-2-2-3	CD	F4-3-8-2-2	F4-1-6-11-1
	F4-1-7-18-3	F4-2-2-6-1	F4-3-2-7-2	F4-3-2-2-1	F4-2-7-4-1	F4-3-1-2-6
	F4-1-7-15-1	F4-3-2-16-1	F4-1-6-9-1	F4-2-3-7-1	F4-3-8-6-3	F4-3-7-1-2
	F4-2-7-6-2	F4-2-7-2-1	F4-2-3-2-1	F4-2-8-5-2	F4-3-7-10-1	FILLER
	RIL4	F4-2-1-4-1	F4-3-7-13-3	F4-1-6-3-1	CD	F4-2-8-6-1
	F4-2-8-1-2	F4-1-2-7-1	F4-1-9-2-1	F4-3-2-12-1	F4-1-6-16-1	RIL4
	F4-2-1-12-1	F4-3-8-1-1	F4-1-1-11-1	F4-3-7-5-3	F4-1-1-7-3	1-1-6-5
Block 2	F4-2-2-10-1	F4-3-8-6-3	F4-2-8-5-2	F4-3-2-15-1	F4-1-1-10-5	F4-1-2-7-1
	F4-2-1-8-4	F4-3-7-8-2	F4-3-1-1-6	F4-3-2-2-1	F4-2-1-11-4	RIL4
	F4-2-7-6-2	F4-1-6-3-1	F4-3-2-7-2	F4-3-7-1-2	F4-2-1-16-3	F4-2-8-6-1
	F4-2-8-4-1	F4-1-6-16-1	F4-3-7-13-3	F4-3-1-2-6	F4-3-7-9-1	F4-3-2-8-1
	FILLER	F4-1-6-11-1	F4-2-3-2-1	RIL4	F4-1-9-2-1	F4-1-1-11-1
	F4-2-8-1-2	F4-1-7-17-1	F4-1-2-9-1	F4-1-9-1-1	F4-2-2-7-1	F4-3-7-6-1
	F4-1-7-15-1	F4-1-1-7-3	F4-3-1-3-6	F4-2-3-7-1	F4-1-6-12-1	F4-2-1-12-1
	F4-1-1-2-9	F4-2-7-9-2	F4-3-3-15-1	F4-1-6-17-1	F4-3-7-7-2	F4-2-7-4-1
	F4-1-6-9-1	F4-2-1-4-1	F4-3-8-1-1	CD	F4-1-6-19-2	F4-3-2-12-1
	F4-2-2-6-1	F4-1-7-4-1	F4-3-2-5-1	F4-1-1-12-1	F4-3-7-14-3	F4-3-7-10-1
	F4-1-6-13-2	F4-2-2-2-3	F4-1-9-3-1	F4-1-2-2-1	F4-1-2-4-3	FILLER
	F4-3-2-16-1	F4-3-7-5-3	F4-1-7-7-1	F4-3-8-5-2	F4-2-3-8-1	1-1-6-4
	F4-3-2-13-1	RIL6	F4-1-7-1-1	F4-2-7-2-1	F4-2-7-3-6	CD
	F4-3-8-2-2	F4-1-1-9-7	F4-2-7-12-2	RIL6	F4-1-7-18-3	F4-3-8-3-3
Block 3	F4-1-9-3-1	F4-2-1-4-1	F4-3-2-12-1	F4-1-2-7-1	F4-3-8-6-3	CD
	F4-3-8-2-2	F4-1-1-12-1	RIL6	F4-1-2-2-1	F4-3-8-3-3	F4-1-7-7-1
	F4-1-7-4-1	F4-2-7-6-2	F4-3-7-10-1	F4-3-2-7-2	F4-2-2-2-3	F4-3-7-5-3
	CD	F4-2-3-2-1	F4-3-7-13-3	F4-2-1-8-4	F4-1-6-13-2	F4-3-7-7-2
	F4-2-8-4-1	F4-2-7-4-1	F4-3-2-13-1	RIL4	F4-2-8-6-1	F4-3-1-3-6
	F4-2-8-1-2	F4-1-1-9-7	F4-1-1-10-5	F4-3-1-2-6	F4-2-7-12-2	F4-2-1-12-1
	F4-1-6-11-1	F4-3-2-2-1	F4-3-2-5-1	F4-1-7-1-1	F4-2-7-2-1	F4-1-2-4-3
	F4-1-1-11-1	F4-1-1-7-3	F4-1-7-15-1	F4-3-8-1-1	F4-3-2-8-1	F4-2-3-7-1
	F4-1-6-19-2	F4-1-6-17-1	F4-1-6-16-1	F4-3-2-15-1	F4-1-6-9-1	F4-1-6-3-1
	F4-2-2-7-1	F4-3-7-6-1	F4-2-1-16-3	F4-1-7-17-1	F4-1-7-18-3	F4-1-2-9-1
	F4-2-1-11-4	RIL6	F4-2-2-10-1	F4-2-3-8-1	CD	F4-3-3-15-1
	F4-3-7-9-1	F4-3-7-8-2	F4-2-8-5-2	F4-3-8-5-2	F4-2-7-9-2	1-1-6-5
	RIL4	F4-1-1-2-9	F4-2-2-6-1	FILLER	F4-1-9-1-1	F4-3-2-16-1
	F4-3-7-14-3	F4-3-1-1-6	F4-1-6-12-1	F4-2-7-3-6	F4-3-7-1-2	F4-1-9-2-1
Block 4	F4-1-1-10-5	F4-1-2-4-3	F4-1-6-16-1	F4-2-1-8-4	F4-2-7-12-2	F4-3-7-14-3
	F4-1-2-7-1	F4-1-6-3-1	RIL6	F4-3-2-15-1	F4-2-3-7-1	F4-1-1-2-9
	F4-2-8-6-1	F4-1-7-4-1	F4-2-1-16-3	F4-3-2-12-1	RIL4	F4-2-2-7-1
	F4-3-7-10-1	F4-1-1-11-1	F4-2-8-4-1	F4-3-8-3-3	F4-3-7-9-1	F4-3-7-8-2
	RIL6	F4-2-1-11-4	F4-3-2-5-1	F4-1-7-7-1	F4-3-7-7-2	F4-1-6-17-1
	F4-3-7-1-2	F4-3-2-16-1	F4-3-2-8-1	F4-1-6-13-2	F4-2-2-6-1	F4-3-2-2-1
	F4-1-7-17-1	F4-3-7-13-3	F4-2-1-12-1	F4-1-7-15-1	F4-1-2-2-1	RIL4
	CD	F4-1-1-12-1	FILLER	F4-1-9-2-1	F4-1-1-7-3	F4-1-9-3-1
	F4-3-7-5-3	F4-2-7-4-1	F4-1-6-9-1	F4-3-8-5-2	F4-3-1-2-6	F4-2-2-10-1
	F4-2-7-9-2	F4-2-1-4-1	F4-2-8-1-2	F4-2-3-8-1	F4-1-6-11-1	F4-2-7-2-1
		F4-3-8-6-3	F4-3-1-3-6	F4-1-7-1-1	F4-1-7-18-3	F4-3-1-1-6
	F4-1-2-9-1	F4-1-9-1-1	F4-3-8-1-1	F4-3-3-15-1	F4-1-1-9-7	1-1-6-4
	F4-3-2-7-2	CD	F4-2-7-3-6	F4-1-6-12-1	F4-1-6-19-2	F4-3-2-13-1
	F4-2-2-2-3	F4-2-3-2-1	F4-3-8-2-2	F4-2-8-5-2	F4-3-7-6-1	F4-2-7-6-2

Bench 2

1-1-6-4	F4-2-1-8-4	F4-2-3-8-1	F4-3-2-12-1	F4-3-8-6-3	F4-1-1-7-3	Block 5
F4-2-7-3-6	F4-3-2-7-2	RIL4	F4-2-1-16-3	CD	F4-3-2-13-1	
F4-1-7-7-1	F4-3-7-5-3	F4-1-2-2-1	F4-1-6-19-2	F4-3-8-2-2	F4-3-7-13-3	
F4-3-1-1-6	F4-2-3-7-1	F4-1-1-12-1	F4-2-7-9-2	F4-1-1-9-7	F4-3-8-3-3	
F4-1-7-17-1	F4-2-1-12-1	F4-1-6-17-1	F4-1-2-7-1	F4-1-7-1-1	F4-1-1-10-5	
F4-1-6-9-1	F4-2-8-5-2	CD	F4-2-8-4-1	RIL6	RIL4	
F4-2-2-7-1	F4-1-6-13-2	F4-3-8-1-1	F4-1-9-2-1	F4-3-2-5-1	F4-1-7-18-3	
CD	F4-3-3-15-1	F4-3-7-14-3	F4-1-7-15-1	F4-1-1-2-9	F4-1-6-16-1	
F4-2-8-6-1	RIL6	F4-1-6-11-1	F4-2-7-6-2	F4-3-7-1-2	F4-2-7-2-1	
F4-1-2-4-3	F4-3-7-8-2	F4-2-1-11-4	F4-1-6-12-1	F4-3-8-5-2	F4-3-7-6-1	
F4-3-2-15-1	F4-1-2-9-1	F4-1-9-1-1	F4-1-9-3-1	F4-1-7-4-1	F4-2-8-1-2	
F4-2-3-2-1	F4-3-1-2-6	FILLER	F4-2-7-4-1	F4-2-2-2-3	F4-2-2-6-1	
F4-3-2-8-1	F4-2-1-4-1	F4-1-1-11-1	F4-3-1-3-6	F4-2-2-10-1	F4-3-7-9-1	
F4-3-2-2-1	F4-2-7-12-2	F4-3-7-10-1	F4-3-2-16-1	F4-3-7-7-2	F4-1-6-3-1	
F4-2-2-2-3	F4-2-7-6-2	F4-2-2-6-1	F4-1-6-3-1	F4-1-1-11-1	F4-3-3-15-1	Block 6
F4-3-8-6-3	F4-3-7-1-2	F4-1-6-11-1	F4-1-2-7-1	F4-2-7-12-2	CD	
F4-1-7-15-1	RIL4	F4-3-7-6-1	F4-3-7-13-3	F4-3-2-7-2	F4-3-2-15-1	
F4-1-6-13-2	F4-2-2-10-1	F4-1-9-2-1	F4-3-7-10-1	F4-1-1-2-9	F4-3-2-5-1	
1-1-6-5	F4-3-1-2-6	FILLER	F4-3-7-7-2	F4-1-1-9-7	F4-2-1-11-4	
F4-3-2-8-1	F4-3-1-3-6	F4-2-7-4-1	F4-3-2-12-1	F4-1-1-7-3	F4-3-7-9-1	
F4-2-1-16-3	F4-2-3-8-1	F4-1-2-9-1	F4-1-9-3-1	F4-1-7-1-1	F4-2-7-3-6	
F4-2-8-4-1	F4-1-6-12-1	RIL6	F4-3-7-5-3	F4-1-6-9-1	F4-2-3-7-1	
F4-3-8-1-1	F4-1-2-2-1	F4-1-9-1-1	F4-3-2-16-1	F4-2-8-1-2	F4-3-8-3-3	
F4-1-6-19-2	F4-3-2-2-1	F4-3-7-14-3	F4-3-7-8-2	RIL6	RIL4	
CD	F4-2-7-9-2	F4-2-1-4-1	F4-1-1-12-1	F4-2-3-2-1	F4-2-8-5-2	
F4-3-1-1-6	F4-2-7-2-1	F4-1-1-10-5	F4-1-2-4-3	F4-1-7-7-1	F4-1-7-4-1	
F4-1-7-17-1	F4-3-8-2-2	RIL6	F4-1-6-16-1	F4-2-1-12-1	F4-1-7-18-3	
F4-1-6-17-1	F4-2-2-7-1	F4-2-1-8-4	F4-2-8-6-1	F4-3-8-5-2	F4-3-2-13-1	
F4-2-8-5-2	F4-1-6-3-1	F4-1-6-13-2	RIL4	F4-3-2-2-1	F4-3-1-1-6	
F4-2-1-11-4	F4-1-1-11-1	F4-2-7-4-1	F4-3-7-14-3	F4-3-1-3-6	1-1-6-5	
F4-3-8-3-3	F4-1-9-2-1	F4-3-7-8-2	F4-1-1-9-7	F4-1-1-2-9	F4-2-8-1-2	
F4-3-8-6-3	FILLER	F4-3-8-2-2	F4-2-1-12-1	F4-1-6-9-1	F4-2-7-3-6	
F4-2-7-6-2	F4-2-7-12-2	F4-3-2-15-1	F4-2-3-7-1	F4-3-7-10-1	F4-3-8-1-1	
RIL6	F4-1-6-17-1	F4-3-2-8-1	F4-2-3-8-1	F4-2-2-6-1	CD	
F4-1-7-17-1	F4-3-8-5-2	F4-1-7-18-3	F4-2-1-4-1	F4-1-6-12-1	F4-3-7-6-1	Block 7
F4-2-2-10-1	F4-1-9-3-1	RIL6	F4-2-3-2-1	F4-2-8-6-1	F4-3-2-16-1	
F4-3-7-9-1	F4-3-2-5-1	F4-2-8-4-1	F4-1-7-7-3	RIL6	F4-3-7-5-3	
F4-3-7-7-2	F4-3-7-1-2	F4-3-7-13-3	F4-3-1-2-6	F4-1-7-15-1	F4-2-2-2-3	
RIL4	F4-1-2-2-1	F4-2-1-16-3	F4-3-3-15-1	F4-2-7-2-1	F4-1-7-1-1	
F4-2-7-9-2	F4-1-2-7-1	CD	F4-1-6-16-1	F4-1-2-9-1	F4-1-1-12-1	
F4-1-7-4-1	F4-1-2-4-3	F4-1-1-10-5	F4-3-2-7-2	F4-3-2-12-1	F4-1-7-7-1	
F4-2-2-7-1	F4-1-9-1-1	F4-1-6-11-1	F4-3-2-13-1	F4-1-6-19-2	F4-2-1-8-4	
F4-2-2-2-3	F4-1-6-16-1	F4-1-9-1-1	F4-3-1-2-6	F4-1-7-7-1	F4-3-7-10-1	Block 8
F4-3-7-7-2	F4-3-8-1-1	F4-1-7-4-1	F4-3-1-3-6	FILLER	F4-2-7-3-6	
CD	F4-2-2-10-1	RIL4	F4-1-6-17-1	F4-1-7-18-3	F4-3-7-13-3	
F4-3-7-9-1	F4-3-8-2-2	F4-3-2-15-1	F4-2-7-6-2	F4-3-8-3-3	F4-3-2-12-1	
F4-3-7-5-3	F4-1-7-17-1	F4-1-2-9-1	CD	F4-1-7-1-1	F4-2-8-5-2	
F4-2-8-4-1	F4-1-6-13-2	F4-3-7-14-3	F4-2-7-2-1	1-1-6-5	F4-2-1-8-4	
F4-1-1-2-9	F4-1-1-7-3	F4-3-2-7-2	RIL4	F4-1-7-15-1	F4-1-6-9-1	
F4-3-7-1-2	F4-3-7-8-2	F4-1-1-11-1	F4-2-1-16-3	F4-2-7-4-1	F4-1-6-3-1	
F4-2-2-7-1	F4-1-9-2-1	F4-3-2-8-1	F4-3-8-5-2	F4-2-1-4-1	F4-3-7-6-1	
F4-3-2-5-1	F4-3-8-6-3	F4-1-6-12-1	F4-2-2-6-1	RIL6	F4-2-8-1-2	
F4-1-1-12-1	RIL4	F4-1-6-19-2	F4-1-6-11-1	F4-2-1-11-4	F4-3-2-2-1	
F4-1-2-7-1	F4-2-3-2-1	F4-3-2-13-1	F4-1-2-4-3	F4-1-2-2-1	F4-3-3-15-1	
F4-1-1-9-7	F4-2-1-12-1	F4-2-7-12-2	F4-1-1-10-5	F4-2-3-8-1	F4-1-9-3-1	
F4-3-2-16-1	F4-2-7-9-2	F4-3-1-1-6	RIL6	F4-2-3-7-1	F4-2-8-6-1	

Figure 2.4: Design of glasshouse experiment in autumn 2012. Each cell represents a 1L pot.

Recombinants from the extremes of the height distributions of the initial phenotyping (Figure 5.2) were selected for a further glasshouse experiment in summer 2013 to determine with confidence short and tall recombinants for RNA-Seq. The experiment was conducted in the same way as described in the preceding paragraph, with the exception that pots were further spaced to avoid mildew which had affected the first experiment.

The recombinants were randomised in 24 blocks across four benches (Figure 2.6), with additional replicates for some of the recombinants. Some replicates were lost due to disease. Parent NILs were replicated to N=63 with two or three replicates of each parent per block. The temperature ranged from 18°C - 54°C

(mean of 25°C) and humidity ranged from 11% - 85% with a mean of 54%. The plant height was measured and selections made on the basis of phenotype for short and tall bulks for RNA-Seq and the iSelect SNP array as shown in Appendix 2.3 and Figure 5.2.

Figure 2.5: Spacing of 1L pots in the glasshouse experiment in 2013. Photograph taken in June 2013.

Figure 2.6: Design of glasshouse experiment in summer 2013. Each cell represents a 1L pot.

2.3.2.2 Material for iSelect 90K SNP array

Parent NILs along with short and tall bulks were genotyped on two separate runs of the iSelect 90K SNP array (Wang et al., 2014a). Samples were prepared for both runs in the same way. The parent NILs and Mara were genotyped first. For the bulks, three individuals were selected from each of the short (S1-S3) and tall (T1-T3) recombinant types (Appendix 2.3). The individuals were also pooled following DNA extraction to make the short bulk (SB) and tall bulk (TB).

2.3.2.3 DNA extraction

Harvested spike tissue of each parent NIL and recombinant was ground to a fine powder in a liquid-nitrogen cooled mortar using a pestle. DNA was extracted from approximately 100 mg of ground tissue using a DNeasy® Plant Mini Kit (Qiagen, 69104) according to manufacturer's instructions. DNA was quantified at a wavelength of 260 nm using the NanoDrop 2000 spectrophotometer (ThermoScientific) and a 15 µl sample of 60 ng μl^{-1} DNA per genotype was submitted to the University of Bristol Genomics facility. For the bulks (SB and TB), 5 µl of DNA from each individual (S1-3; T1-3, Appendix 2.3) was combined into one sample.

2.3.2.4 Material for Affymetrix Axiom® 820K SNP array

Seed from the *Rht8* NIL in the Paragon background and Paragon was sent to the University of Bristol Genomics facility to be genotyped on the Axiom® 820K SNP array (www.cerealsdb.uk.net/cerealgenomics).

2.3.3 Targeting genome-specific allelic variation

2.3.3.1 Flow-sorted 2D DNA from the short parent NIL

2.3.3.1.1 Plant material sent for flow-sorting

Approximately ~12,000 seed (700g) from the short parent NIL, RIL4, was sent to the Institute of Experimental Botany (IEB), Prague, Czech Republic for flow-sorting and 2D BAC library construction. A pilot experiment confirmed that 2D

could be successfully sorted from a small quantity of DNA prior to flow-sorting DNA from the full seed collection for library construction. The chromosomes were flow-sorted as described in IWGSC, 2014. The chromosome 2D DNA was amplified using the Illustra GenomiPhi DNA amplification kit, as outlined in Simkova et al., 2008. Samples were pooled from three independent amplifications and lyophilized.

2.3.3.1.2 Assessing purity of flow-sorted 2D DNA

A total of 10.22 µg of amplified DNA from chromosome 2D of RIL4 was received from the IEB on 27 October 2014. The purity of the sorted fraction was estimated as 94.44% with contamination mainly from chromosome 7D. The DNA was dissolved in 20 µl of dH_20 over 24 hours at ambient temperature and further diluted to 10 – 15 ng μl^{-1}. DNA concentration was measured at 260 nm using the NanoDrop 2000 spectrophotometer (ThermoScientific).

The 2D flow-sorted DNA was tested with markers mapping across a range of chromosomes from previous mapping projects using the JIC core collection of KASP markers (marker spreadsheet provided by Michelle Leverington-Waite). KASP markers were selected on the basis that they showed good cluster separation in previous genotyping and were polymorphic between Mara and Cappelle-Desprez. An optimal cycle number for each marker was determined, since at the highest cycle number (40), most of the markers amplified the 2D DNA and negative control. Out of a total of 15, three markers amplified the 2D DNA (shown in Table 2.1 and Figure 5.3). KASP assays were performed as described in 2.3.10.2.

SNP ID	Mara	Cappelle Desprez	Identifier (core collection)	Chr	AXC	FAM	VIC	amplifies 2D?	Optimal cycle number
BS00062738	A:A	C:C	Bristol 18:B06	7D	22	A	C	Yes	25
BS00023159	G:G	C:C	Bristol 10:H07	7D	40.4	C	G	Yes	27
BS00033613	C:C	T:T	Bristol 15:F06	7A	178.1	C	T	No	30
BS00027942	G:G	A:A	Bristol 13:F09	6B	25.2	A	G	No	30
BS00022157	A:A	G:G	Bristol 03:D02	5D	73.4	A	G	No	30
BS00021939	G:G	A:A	Bristol 03:D12	5A	17.6	A	G	Yes	30
BS00023431	G:G	T:T	Bristol 13:F01	4B	51.1	G	T	No	30
BS00036493	A:A	C:C	Bristol 20:B08	4A	8.8	A	C	No	25
BS00040001	T:T	C:C	Bristol 20:F04	3B	149.9	C	T	No	30
BS00070870	G:G	A:A	Bristol 28:C12	3A	75.9	A	G	No	30
BS00022946	T:T	C:C	Bristol 03:D09	2B	95	C	T	No	25
BS00090234	A:A	G:G	Bristol 20:D06	2B	87.2	A	G	No	30
BS00022332	C:C	T:T	Bristol 05:B11	2A	49.2	C	T	No	30
BS00022260	C:C	T:T	Bristol 05:D08	2A	-	C	T	No	25
BS00062783	C:C	G:G	Bristol 24:C07	1A	-	C	G	No	25

Table 2.1: Assessing the purity of the 2D flow-sorted DNA using KASP markers from the JIC core collection.

2.3.3.1.3 Ensuring genome-specificity with 2D flow-sorted DNA and nulli-tetrasomics

Genome specificity for 2D was tested initially by validating markers with Chinese Spring nulli-tetrasomic (NT) (Sears, 1966) DNA (for SSRs). The complete set of chromosome 2 NT DNA was used: AAAABB (N2DT2A), AAAADD (N2BT2A), BBBBAA (N2DT2B), BBBBDD (N2AT2B), DDDDAA (N2BT2D) and DDDDBB (N2AT2D). The SSRs tested did not amplify the nulli-tetrasomics for 2D (N2DT2A and N2DT2B), whereas the markers did amplify the nulli-tetrasomics with the D-genome present (all N2A~ and N2B~). Later, once the 2D flow-sorted DNA was received, specificity to the 2DS genome for both SSR and KASP markers was tested (as reported in Figure 5.11).

2.3.3.2 PolyMarker

PolyMarker (Ramirez-Gonzalez et al., 2015) was used to increase the likelihood of generating homoeologue-specific assays on putative SNPs. PolyMarker is a primer-design pipeline for SNP assay development which generates a multiple alignment between the target SNP sequence and the IWGSC CSS for each homoeologous genome. A mask is then generated with informative positions (further detail in 2.3.7.4). This indicates whether the SNP is varietal or homoeologous and whether the designed primers are specific to the target

genome (the SNP is only present in the target genome), semi-specific (polymorphism found on two out of three genomes) or non-genome specific. The web-based interface for PolyMarker (http://polymarker.tgac.ac.uk) was used for primer design and SNPs identified as homoeologous based on the IWGSC CSS alignments were not considered for marker validation (with the exception of cases investigated in 2.3.7.4).

Primers for KASP assays were ordered from Sigma-Genosys Ltd, UK, with forward primers carrying standard FAM or VIC tails at the 5' end with the SNP at the 3' end. (FAM tail: 5' GAAGGTGACCAAGTTCATGCT 3'; VIC tail: 5' GAAGGTCGGAGTCAACGGATT 3').

2.3.4 Sample preparation for RNA-Seq

2.3.4.1 Plant material

Plant material for RNA-Seq was harvested from the glasshouse-grown plants described in 2.3.2.1 .Tissue from the spike and elongating peduncle during stem elongation (GS 30 – 39) (Figure 5.2A) of the parent NILs and recombinants to the fine-mapping population was dissected destructively from individual plants, photographed, snap frozen in liquid nitrogen and then stored at -80°C. The spike and peduncle lengths were measured using imageJ software (Abramoff et al., 2004) and samples were selected from the middle of the distribution of lengths (Figure 5.2B) in order to use tissue at the same developmental time-point. For the parent NILs, two biological replicates per tissue per genotype were selected, as shown in Table 5.1.

For the bulked segregant analysis, spike tissue from a total of 18 recombinants was selected. Nine short and nine tall recombinants were selected on the basis of height distribution, as shown in Appendix 2.3. The mean height of the recombinants for the short bulk was ~9 cm less than the mean height of the tall recombinants (Appendix 2.3).

2.3.4.2 RNA extraction

Harvested spike/peduncle tissue was ground to a fine powder in a liquid-nitrogen cooled mortar using a pestle. Total RNA was extracted from approximately 100

mg of ground tissue using an RNeasy® Plant Mini Kit (Qiagen, 74903) according to manufacturer's instructions. RNA samples were treated with DNase I using the RNase-Free DNase Set (Qiagen, 79254) DNA and then RNA cleanup was performed using the RNeasy® Kit. The RNA concentration was measured at 260 nm using a NanoDrop 2000 (ThermoScientific) and was approximately ~250 ng μl^{-1}. The RNA quality was assessed using the A260/280 and A260/230 ratios (>2.0 in all cases) and samples were frozen at -80 °C. Prior to sequencing, RNA samples were diluted using nuclease-free water to achieve 5 µg at a minimum concentration of 20 ng μl^{-1}.

2.3.4.3 Library construction and sequencing

RNA-Seq samples were submitted to The Genome Analysis Centre (TGAC) for library construction and sequencing. All samples passed the QC checks by TGAC, which used Total RNA Analysis pg sensitivity for Eukaryotes (Agilent Technologies). The parent NIL samples (P1-8, Table 5.1) were sequenced two per lane across four lanes, with each parent NIL represented in each lane to avoid lane bias. The six bulk samples (B1-6, Table 5.1, Appendix 2.3) were randomised for short/tall and multiplexed three per lane across two lanes. In all cases one Illumina TruSeq RNA v2 library was constructed per sample and the libraries were barcoded. Sequencing was carried out on the Illumina HiSeq2000 with 100 bp paired-end reads. The Illumina 100-bp reads were received as FASTQ compressed files.

2.3.5 References used for alignment

2.3.5.1 Customised UniGene reference

The customised UniGene reference is described by Harper et al., 2015 and was developed using *de novo* transcriptome assembly, a SNP genetic linkage map and comparative genomics approaches. *De novo* leaf transcriptome assemblies from *T. urartu*, *Ae. speltoides* and *Ae. tauschii* transcriptomes (representing the A, B and D genomes, respectively) were assembled into UniGenes using the Trinity package (Grabherr et al., 2011). Since the B-genome diploid which was sequenced was dissimilar to the B-genome in hexaploid wheat, the B-genome

assemblies were specifically adjusted ('cured') using a *de novo* transcriptome assembly from the tetraploid *T. dicoccoides*. The UniGenes were used as query sequences in BLASTN homology searches of the Brachypodium genome and the hit with the greatest sequence similarity was retained. This anchored the UniGenes to the Brachypodium chromosomes and provided a physical position for each gene. A Chinese Spring x Paragon mapping population was transcriptome-sequenced and SNPs in the UniGenes were identified. The SNP linkage map was used to fine order the UniGenes (previously in Brachypodium-like order) into a wheat-like order. The same order was used for the A, B and D genome since the transcriptome sequencing of the Chinese Spring x Paragon mapping population was not genome-specific and all three homoeologues were collapsed together. UniGenes which were monomorphic in the Chinese Spring x Paragon population were assigned a position according to the Brachypodium physical order. To eliminate redundancy due to alternative splice forms, the longest UniGene was retained where multiple UniGenes mapped to the same location. The resulting reference comprised 147,411 UniGenes (47,160 for the A genome, 59,663 for the B-genome and 40,588 for the D genome).

2.3.5.2 v3.3 cDNA reference

The v3.3 cDNA reference is described in full in Borrill (2014) and was provided by Martin Trick (JIC). The UniGenes described in 2.3.5.1 were used as queries in BLASTN homology searches of the IWGSC CSS contigs and the best-scoring hit (above e-value 1E-30) selected. These IWGSC CSS contigs corresponding to the UniGenes were called pseudomolecules v3. Pseudomolecules v3 consisted of genomic sequence.

A file with annotation of the IWGSC CSS contigs was generated by Sarah Ayling (TGAC) with predicted mRNA features. These mRNA features were predicted using a *de novo* assembly pipeline at TGAC and were also combined with the best tetraploid and diploid wheat gene models at the time (Krasileva et al., 2013).

The annotation file was used to extract the gene models from pseudomolecules v3 and these comprised the v3.3 cDNA reference. The mRNA features from the IWGSC CSS contigs were ordered using the Chinese Spring x Paragon map. Gene models and contigs which could not be anchored due to lack of

polymorphism in the Chinese Spring x Paragon mapping population were assigned into the 'unordered' bin (remaining assigned to chromosome arm only). Hence the reference contained an ordered and unordered section. Since there was redundancy in the mRNAs (splice isoforms), only the longest mRNA was retained to achieve a non-redundant set of 75,419 gene models. Of the gene models, 42% were in the ordered section and 58% were unordered.

2.3.5.3 2D v3.3 cDNA interval

As marker development progressed during the course of Chapter 5, an interval was demarcated in the ordered section of the v3.3 cDNA reference. This was done by anchoring the *Rht8* flanking markers as shown in Table 5.4 and then taking a conservative region either side. The reference comprised 59 gene models totalling 65,564 bp in length, shown in Appendix 2.4.

2.3.5.4 *De novo* spike transcriptome assembly

The *de novo* spike assembly of Cappelle-Desprez was built by Martin Trick using the Trinity package, using the RNA-Seq reads P1 and P3 (Table 5.1). The longest splice isoform was selected to remove redundancy and any assembly that matched the UniGenes (from leaf transcriptomes) was removed. A total of 82,762 spike-specific unordered assemblies were retained.

2.3.6 Read mapping

Reads were mapped to the UniGenes directly from the compressed reads by Martin Trick using Maq (Mapping and Assembly with Quality) (Li et al., 2008) to map reads and call variants using mapping quality scores. The methodology is described in Trick et al., (2012). Reads were mapped to the v3.3 cDNA reference, spike *de novo* assembly and 2D v3.3 cDNAs in three separate alignments, all following the same methodology and using the read aligner bowtie2 (v2.1.0) (Langmead and Salzberg, 2012) with the default parameters for read pair libraries. Mapped reads were subsequently filtered using SAMtools (v0.1.19) (Li et al., 2009) and mapping statistics were checked using SAMStat (v1.0) (Lassmann et al., 2011). The steps are shown in full in Appendix 2.5 for the read mapping to the v3.3 cDNAs and 2D v3.3 cDNAs, with reads being mapped to the

spike assembly using the same method. The BAM files were then processed to identify variant candidates using bulk frequency ratios by Ricardo Ramirez-Gonzalez.

2.3.6.1 Coverage statistics

Coverage statistics were obtained from Maq, bowtie2 and VarScan LSF output considering only properly paired reads and are shown in Appendix 2.6 and Appendix 5.9.

2.3.7 SNP-calling

2.3.7.1 SNPs between the parent NILs in the UniGenes

Varietal SNPs, representing allelic variation (as opposed to inter-genome SNPs between homoeologous genomes or varietal SNPs between Cappelle-Desprez and the Chinese Spring reference) were called between each sample (P1-P8, Table 5.1) and the UniGene reference by Martin Trick, as described previously (Trick et al., 2012). Briefly, in a two-step process, first Maq (Li et al., 2009) (default parameters) was used to call SNPs between the reference and each parent NIL, generating two SNP sets. In the second step, a custom Perl script was used to derive the difference between the sets. SNPs were filtered with a minimum depth threshold of 10x. The SNP-calling process identified a total of 60,454 putative SNPs between any of the eight samples and the reference, across 32,663 unique UniGenes. These were arranged in a spreadsheet to aid further inspection and sorting. Ancillary synteny data for each UniGene was added, including the best hit for the UniGene from BLASTN analysis against Brachypodium and rice gene models (E-value cut-off 1E-50).

The 638 concordant SNPs in the parent NILs described in 5.5.1.1 and shown in Figure 5.9 were normalised to account for the relative under-representation of the D-genome in the reference (Table 5.2). The SNPs on each chromosome arm were presented as a percentage of total SNPs on that genome, so that each homoeologous genome represented 100%.

2.3.7.2 SNP identification in v3.3 cDNAs

SNP variants were identified by Ricardo Ramirez-Gonzalez using methodology first described in Trick et al., (2012) and extended to work with BAM files and to allow detection of SNPs where a variant is completely absent from one of the parental sequences (BFR of infinity), as outlined in Ramirez-Gonzalez et al., (2014). The objective was to identify SNPs that were highly enriched for the parental allele in the corresponding bulk i.e. SNPs found in the short parent also present in the short bulks and vice versa for the tall parent/bulk. A total of 15 different combinations of parent NIL and bulk BAM files were compiled, since the phenotyping had not been verified at this point and there was uncertainty as to whether there would be any biases in the SNP calling as a result. The different *in silico* mixes are shown in Appendix 2.7.1 with the number of SNPs identified in each mix. Once the phenotyping had been verified, all the samples within each parent NIL and bulks were pooled to increase coverage (mix number 2, Appendix 2.7.1).

The SNP calling process involved first identifying varietal SNPs between the parents. A consensus from the two parents was identified and a varietal SNP was called. SNPs were called at bases which had a minimum coverage of 20.

In addition to the threshold for coverage, a second parameter was the threshold at which to accept a varietal SNP: initially this parameter was set to 100% to capture only the most stringent varietal SNPs (i.e. all of the bases at that position differed from the reference in one variety) (Appendix 2.8). However, crucially, no SNPs on the short arm of chromosome 2 (2AS, 2BS or 2DS; group 2S) which harboured the *Rht8* introgression were identified. For this reason, the parameter was adjusted and varietal SNPs were called where a minimum of 20% of the bases differed from the reference in one of the parents. Lowering this parameter allowed for sequencing error but primarily was a cautious approach to account for missing sequence from one or more homoeologues and potential genome misassignment of the IWGSC CSS contigs (discussed further in 5.3.2).

Subsequently, in the second step, for each bulk the frequency of the base at each SNP position was calculated and the bulk frequency ratio (BFR) between bulks determined. In this way, the BFR provided a relative measure of SNP enrichment

in both bulks, which was normalised by coverage to eliminate bias over regions with a greater representation of reads. A high BFR indicated that the allelic variation was contributed from one bulk and absent in the other bulk.

2.3.7.2.1 Bulk frequency ratios

In order to set the minimum BFR threshold, the relative proportion of SNPs from group 2S was calculated as the BFR increased. As the BFR increased, the number of SNPs from 2S decreased markedly from >100 at BFR = 6 to single counts at BFR = 18 (Appendix 2.7.2). The relative 2S enrichment when considering density peaked at BFR = 7 SNPs. For this reason the BFR threshold was set to 6. The vast majority of SNPs with BFR = infinity were being called due to a small ratio of reads calling a SNP in one of the parent NILs relative to the coverage of reads at that position. In order to eliminate this large number of potentially low-confidence SNPs, the ratio threshold for the informative parent of SNPs with BFR = infinity was set to ratio ≥ 0.2. Setting both the BFR and ratio thresholds retained a total of 7666 putative SNPs across 2055 unique genes genome-wide. The SNP frequencies on each chromosome arm shown in Figure 5.9 were normalised as described in 2.3.7.1.

2.3.7.3 SNP identification in narrowed 2D v3.3 cDNA interval

The narrowed interval on 2D consisted of 59 gene models from the 'ordered' section of the full v3.3 cDNA set, from mrna126380 (2D: 716490) to mrna057019 (2D:1386885). VarScan 2.0 (Koboldt et al., 2012) was used to call SNPs, which is open software designed to detect variants from multiple pileup files and filters variants by coverage, read depth, variant frequency and base quality. The process had to be customised since the software was originally designed to detect human tumour variants. SNP were called from the parent NIL and bulk BAM files by piping an mpileup2snp output from SAMtools (Li et al., 2009) directly to VarScan. The commands for this are shown in full in Appendix 2.5. Files were output from VarScan in two formats: one viewable in IGV (VCF format) and another which was human readable and could be opened in a spreadsheet which enabled further SNP filtering (an example of this output is shown in Table 2.2).

Source	Chrom	Position	Ref	Var	Cov	Reads1	Reads2	Freq (%)	StrandFilter	R1+	R1-	R2+	R2-
CD	mrna126380	84	A	T	1141	551	589	51.62	Pass	1	529	22	576
CD	mrna064977	606	C	T	133	67	66	49.62	Pass	0	3	64	2
CD	mrna066573	1	T	C	24	2	15	62.5	Pass	1	2	0	15
CD	mrna066573	135	G	T	279	0	271	97.13	Pass	0	0	123	148
CD	mrna066573	146	A	C	262	0	238	90.84	Pass	0	0	100	138
CD	mrna105132	221	A	G	738	0	737	99.86	Pass	0	0	437	300
CD	mrna105132	261	G	C	789	0	789	100	Pass	0	0	387	402
RIL4	mrna004763	1461	T	C	183	70	113	61.75	Pass	0	7	63	5
RIL4	mrna007148	30	C	T	49	29	20	40.82	Pass	1	25	4	19
RIL4	mrna014279	89	G	A	80	41	39	48.75	Pass	28	13	27	12
SHORT	mrna105132	1374	C	G	262	127	135	51.53	Pass	54	73	50	85
SHORT	mrna009588	850	T	C	11	3	8	72.73	Pass	1	2	4	4
TALL	mrna105132	532	T	C	219	0	219	100	Pass	0	0	112	107
TALL	mrna096393	2154	G	A	33	7	26	78.79	Pass	6	1	21	5
TALL	mrna106738	210	G	T	188	141	47	25	Pass	73	68	21	26

Table 2.2: Exemplar VarScan SNP-calling output. Columns left to right: Source = the pooled reads from which the SNP was called relative to the reference (either CD, RIL4, SHORT or TALL); Chrom: gene in reference; Position = base position at which the SNP is located; Ref = the base call on the reference; Var = variant base (SNP) call; Cov = total depth of coverage; Reads1 = number of reads supporting the reference; Reads2 = number of reads supporting the SNP; Freq = the SNP frequency from the read count (Reads2/total count); StrandFilter = Ignores SNP with >90% support on one strand; R1+/- = reference-supporting reads on forward/reverse strand; R2+/- = SNP-supporting reads on forward/reverse strand (http://varscan.sourceforge.net).

SNPs in the 2D interval were prioritised for marker validation by first removing likely homoeologous SNPs and second by determining the highest-frequency variant calls. The two steps are shown in Appendix 5.10.

First, in order to identify likely homoeologous SNPs, datasets between the parent NILs were considered separately from the bulks. Between each of these datasets, shared SNPs found with respect to the reference were discarded, since these were most likely either homoeologous SNPs or varietal SNPs between Cappelle-Desprez and Chinese Spring (the reference). This retained a total of 401 putative SNPs across 51 unique genes between the parents and the reference and 388 putative SNPs across 47 unique genes between the bulks. Taking the putative varietal SNPs identified, the overlap between the parent NILs and bulk datasets was examined in order to determine which SNPs from the parents were enriched in the corresponding bulk, similar to the BFR approach. A total of eight SNPs were common to the short parent and short bulk and 22 SNPs common to the tall parent and tall bulk (Appendix 5.11). In both datasets, there was more overlap in SNPs between the converse parent/bulk (Appendix 5.11).

In the second step, the frequencies of the variant call (Frequency column in Appendix 5.12 and 5.15) were considered in the putative SNP parent and bulk

datasets. This step was necessary to ensure higher-confidence calls that would normally be reported by VarScan. Personal correspondence with the developer suggested that due to the high depth of coverage (minimum average over the whole reference >200 x, Appendix 5.9), the in-built VarScan quality scores were all high thus all SNPs passed the statistical tests (e.g. Fisher's Exact Test p-value was low – data not shown). The majority of the putative SNPs (~ 80% in both parent and bulk datasets) had a frequency <50%, (distribution of frequencies shown in Appendix 5.14) meaning that in fewer than half the SNP calls, most of the reads at the base supported the reference rather than the variant call.

The putative SNPs which were considered for validation by developing markers from the overlap between parent NILs and bulks are shown in Appendix 5.13 and those from the high-frequency prioritisation are shown in Appendix 5.15.

2.3.7.4 Troubleshooting v3.3 cDNA and IWGSC CSS alignments

To ensure that SNPs were truly varietal rather than inter-homoeologous, the 2D cDNA sequences were used as query sequences in BLASTN sequence homology searches (Altschul et al., 1997) of the v3.3 cDNAs in a manual alignment step. SNPs were retained if the sequence around the SNP had significant (e-value above 1E-05) BLAST hits to the 2A and 2B genomes and the alignment of these using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) unambiguously found the SNP to be varietal. These were then aligned again to the CSS contigs using PolyMarker (Ramirez-Gonzalez et al., 2014). Those SNPs found to be varietal again, this time from the contig alignments, were retained, as shown in Appendix 5.10. In some cases, the alignments to the v3.3 cDNAs and the CSS contigs yielded conflicting results since some SNPs that were found to be homoeologous in one alignment were varietal in another alignment. This was surprising, since the mRNAs comprising the v3.3 cDNA reference originated from gene models predicted using the IWGSC 1.0 CSS contigs. This was examined in detail by considering several case-studies, which revealed the limitations associated with an unassembled reference genome. These case-studies fell into two main classes, and examples from each class are presented here:

1) **Redundant IWGSC CSS contigs with different base calls at the SNP position.**

SNP *mrna007148_169* was identified as a varietal SNP between RIL4 and *Cappelle-Desprez* (G/A) and tested as marker vcf_11 which was monomorphic (Appendix 6.6).

Step 1: BLASTing the 2D gene model to retrieve the best-hit homoeologues to the 2A and 2B homoeologues in the v3.3 cDNAs:

Sequences producing High-scoring Segment Pairs:		High Score	Sum Probability P(N)	N
mrna007148	mRNA(UCW_Tt-k41_contig_43937) **2D**:1366856..1364433	10915	0.	1
mrna007149	mRNA(UCW_Tt-k41_contig_43937) **2A**:1749073..1752126	10051	0.	1
mrna071326	mRNA(UCW_Tt-k61_contig_46909) **2B**:2092121..2087167	9525	0.	1
mrna043642	mRNA(UCW_Tt-k35_contig_2226) 4D:10009008..1001...	1013	9.9e-90	3
mrna043643	mRNA(UCW_Tt-k35_contig_2226) 4B:11040469..1103...	1004	7.9e-88	3
mrna043641	mRNA(UCW_Tt-k35_contig_2226) 4A:11696865..1170...	1022	5.4e-67	2
mrna137635	mRNA(MLOC_11551.2) 4B:1123679..1120274	386	1.5e-08	1
mrna137636	mRNA(MLOC_11551.2) 4D:4150795..4154079	338	2.5e-06	1

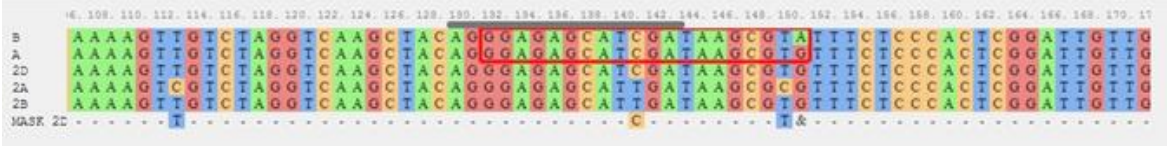
In this first step, all three homoeologues were retrieved in the BLAST search with a high score.

Step 2: BLASTing the 2D gene to retrieve the best-hit to the IWGSC CSS contigs:

Sequences producing High-scoring Segment Pairs:		High Score	Sum Probability P(N)	N
2DS 5381947	13777 606514 4821150+,...,1671179-	6901	0.	2
2AS 5306358	7274 158566 5163351-,...,3566353+	6424	0.	4
2AS 5182464	8700 186985 2670+,...,5163351+	6406	0.	4
2BS 5226042	8809 283565 4206288+,...,962013-	6349	0.	3

In the second step, two 2AS contigs were returned in the BLAST search on the Unité de Recherche Génomique Info (URGI) BLAST webpage (https://urgi.versailles.inra.fr/blast/blast.php). When *mrna007148*, *mrna007149* and *mrna07136* were checked to verify which CSS contigs they were assigned to in the v3.3 cDNA reference, the 2DS and 2BS contigs and mRNAs corresponded. The 2A *mrna007149* was assigned to 2AS_5182464.

Step 3: PolyMarker alignment:



The alignment (a graphical output of the mask that PolyMarker generates) shows the VIC and FAM primer in the red box, with the G/A varietal SNP at the 3' end.

The alignment shows that the SNP is varietal, since a G is present in the 3' position of the 2A, 2B and 2D CSS contigs. The best-hits to the CSS contigs from PolyMarker agreed with the contigs that the mRNAs has been assigned to, with the exception of 2AS. PolyMarker used the best-hit to 2AS in the BLAST in step 2 (2AS_5306358), whereas the v3.3 cDNA reference used the other 2AS contig (labelled B and A in the alignments below, respectively).

Step 4: Aligning the v3.3 cDNAs and CSS contigs.

```
(2D) mrna007148   TACAGGGAGAGCATCGATAAGCGTGTTTCTCCCACTCGGATTGTTGTCTTGAAACCATGC 204
(2A) mrna007149   TACAGGGAGAACATCGATAAGCGCATTTCTCCCACTCGGATTGTTGTCTTGAAACCATGC 204
(2B) mrna071326   TACAGGGAGAGCATTGATAAGCGTGTTTCTCCCACTCGGATTGTTGTCTTGAAACCATGC 960
     2DS_5381947  TACAGGGAGAGCATCGATAAGCGTGTTTCTCCCACTCGGATTGTTGTCTTGAAACCATGC 186
(A)  2AS_5182464  TACAGGGAGAACATCGATAAGCGCATTTCTCCCACTCGGATTGTTGTCTTGAAACCATGC 497
(B)  2AS_5306358  TACAGGGAGAGCATTGATAAGCGCGTTTCTCCCACTCGGATTGTTGTCTTGAAACCATGC 186
     2BS_5226042  TACAGGGAGAGCATTGATAAGCGTGTTTCTCCCACTCGGATTGTTGTCTTGAAACCATGC 186
                  ********** *** ********  **********************************
```

Conclusion: Aligning the v3.3 cDNA homoeologues with the corresponding CSS contigs indicates that there is a different base-call at the SNP position (highlighted in yellow and red) between the redundant 2AS contigs. The alignment around the SNP is robust so presumably this is a sequencing difference as opposed to alignment error. The different base call explains why the PolyMarker assignment (B) returns a varietal SNP whereas the v3.3 cDNA assignment (A) indicates a homoeologous SNP between the A genome and B/D. Clearly redundancy of the CSS contigs can in some cases (quite frequently, from all the SNPs which were manually aligned in this same way) lead to a difference in calling the SNP homoeologous or varietal. It is difficult in these cases to judge which alignment is correct, since the base difference is probably sequencing error in the original CSS contigs and a limitation of working with an unassembled reference, polyploidy genome. For this reason, the SNPs described in 2.3.7.3 were filtered (steps shown in Appendix 5.10) to ensure both the v3.3 cDNA alignments and the CSS contigs unambiguously returned a varietal-SNP verdict.

2) CSS contig alignments are more reliable than low-scoring homoeologues in v3.3 cDNAs.

SNP *mrna026970_384* was identified as a varietal SNP between RIL4 and Cappelle (C/T), and tested as marker Freq_2 based on primers from PolyMarker CSS alignments, which was polymorphic.

Step 1: BLASTing the 2D mrna to retrieve the best-hit homoeologues to the 2A and 2B homoeologues in the v3.3 cDNAs:

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P(N)	N
mrna026970 mRNA(UCW_Tt-k45_contig_18925:tt-k35_contig_501...	7020	0.	1
mrna026969 mRNA(UCW_Tt-k45_contig_18925:tt-k35_contig_501...	6693	9.0e-298	1
mrna067755 mRNA(UCW_Tt-k45_contig_54572) 4Du:483240..477506	4341	1.5e-191	1
mrna076822 mRNA(UCW_Tt-k61_contig_46681:tt-k45_contig_653...	1320	3.6e-104	2
mrna018605 mRNA(UCW_Tt-k61_contig_4463) 2B:20713826..2071...	1351	7.4e-104	2
mrna053115 mRNA(UCW_Tt-k51_contig_3527) 2A:16618210..1661...	1229	1.5e-98	2
mrna053114 mRNA(UCW_Tt-k51_contig_3527) 2D:12077440..1207...	1212	3.7e-96	2
mrna083035 mRNA(UCW_Tt-k55_contig_76821) 3B:8530333..8533169	1281	1.3e-94	2
mrna075192 mRNA(UCW_Tt-k51_contig_14240) 7Au:4796015..479...	2022	8.3e-87	1
mrna056212 mRNA(UCW_Tt-k55_contig_8411) 4B:2645086..2642090	1948	1.8e-83	

A low-scoring 2A mRNA was returned (mrna053115).

Steps 2 & 3: BLASTing the 2D mrna to retrieve the best-hit to the IWGSC CSS contigs and comparing this to CSS contigs in PolyMarker alignments:

The contig assignments from BLASTing the 2D mRNA sequence against the CSS contigs retrieved the same hits as the contigs used in the PolyMarker alignments. The PolyMarker alignments indicate a varietal SNP, with a C being called at the SNP position in all the homoeologues.

Step 4: Aligning the v3.3 cDNAs and CSS contigs.

```
(2D)  mrna026970     AGGCCGCGA-----------------GCGTCC----------TCTTT-------GTCGA 386
(2A)  mrna053115     AGGTTGCGC-----------------ATATCC----------TCTTC-------GTTGA 407
(2B)  mrna026969     AGGCCGCGA-----------------GCGTCC----------TCTTC-------GTCGA 389
      2DS_5358861    AGGCCGCGA-----------------GCGTCC----------TCTTT-------GTCGA 514
      2AS_5305357    GGATGGTGACATATTAGGATGGAAGTCGTATTTAGGCATATAGTCTTTAAATGGAGTCTT 868
      2BS_5228991    AATCGGTAGCTTGC--------------CGTCCA-----------TTT----GGAGCCAA 842
                        *                 *            **       *
```

The alignments were gapped and had a poor overall match. The SNP is homoeologous according to the v3.3 cDNAs , but varietal according to the CSS contigs. Since the 2A mRNA had a relatively low BLAST score, this was used as a query in a BLASTN homology search against the CSS contigs and found to be anchored to 2AL_6432943, which is not the same contig reported as the best 2A hit when the 2D mRNA is used in a BLASTN search of the contigs. The 2B and 2D mRNAs match the CSS contigs reported by PolyMarker.

Conclusion:

The gene models in the v3.3 cDNA reference are shorter in length than the CSS contigs and in many cases, one gene model does not have the other two homologues present. Thus taking the ‘best’ 2A/2B hit might not be reliable, where

the score of one is outside the top three reported. This is confounded where the 2D mRNA is relatively short, compared to the longer CSS contigs. A KASP marker (Freq_2) developed on this SNP was found to be varietal, using the 2D flow-sorted DNA. Hence the varietal call from the CSS contigs was reliable in this case.

2.3.7.5 SNP-calling in the iSelect SNP array data

Data from the second run with the bulked segregant analysis (BSA) was added to the first run of the parent NILs. The data was received as AA/AB/BB/NC (NC is 'no call') calls from genotyping using the polyploid version of GenomeStudio (Wang et al., 2014a) and orthologue annotation, as well as map position on the Avalon x Cadenza map (http://www.wgin.org.uk/) and Akhunov genetic map (Cavanagh et al., 2013). SNPs were considered between homozygous and heterozygous calls, as described in 5.3.4. The SNPs between parent NILs on the iSelect array shown in Figure 5.9 were normalised for the number of pre-defined variants captured on the array for each chromosome arm (data from http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/iselect_mapped_snps.php).

2.3.7.5.1 Mapping SNPs and marker position on the iSelect SNP array

SNPs between the parent NILs were considered in the context of all the mapped markers on the iSelect array. A total of 9800 markers on the array had a genetic position on the Avalon x Cadenza map and 38,832 markers were mapped using the Akhunov genetic positions. From these totals, redundant markers which mapped to the same genetic position and had the same genotype in both parent NILs were removed for clarity to reduce the linkage group size. SNPs being called due to missing data (NC) were also removed. The remaining markers were mapped using the chromosome arm and genetic position along the chromosome (in cM) using MapChart v2.2 (Voorrips, 2002).

2.3.8 SNP-calling in the 820K Affymetrix Axiom® SNP array data

Affymetrix data was received as two csv files with the same calls as the iSelect data, described in 5.3.4. From the total SNPs identified between the *Rht8* NIL and Paragon, NC calls were excluded to leave a total of 6089 SNPs (Table 5.2) and the 2DS contigs which the SNPs mapped to were considered for microsatellite variation.

2.3.9 In silico SSR discovery on wheat 2DS sequence

Sequence from 2DS IWGSC CSS contigs was mined for microsatellite variation using the online web-interface for WebSat (http://wsmartins.net/websat/) (Martins et al., 2009). Sequence was entered in FASTA format 150,000 characters at a time with parameters set to identify motif length from mono- to hexa-nucleotide with a minimum repeat number of six. Default parameters were used for primers designed on identified SSRs (WebSat uses primer3), but SSRs that were based on mono-nucleotide repeats were omitted where longer nucleotide lengths were available. Overlapping SSRs were also omitted. Primer sequences were downloaded as csv files.

2.3.10 Validating variants with markers

2.3.10.1 SSR validation

Primers were ordered from Sigma-Genosys Ltd, UK, with the forward primers tailed at the 5' end with 5' TGTAAAACGACGGCCAGT 3'. A multiplexed PCR was set up with one of four dyes (Applied Biosystems Standard Dye Sets) which were 6-FAM (blue), VIC (green), NED (yellow but visualised as black on GeneMapper v4) and PET (red). Each PCR assay was in a 6 µl volume and contained 20 ng of genomic DNA, 3.125 µl HotStar Taq Master Mix (Qiagen, 203443) and 0.625 µl of primer mix (18.75 µl of dye + 18.75 µl of reverse primer (100 µM) + 1.25 µl of tailed forward primer (100 µM) + 211.25 µl of dH_2O). Amplification was carried out on a G-Storm thermal cycler using the following programme: initial denaturation at 95°C for 15 min then 35 cycles of [94°C for 1

min, a primer-pair annealing temperature for 1 min and 72°C for 1 min], then final extension at 72°C for 10 min. All annealing temperatures were 60°C unless otherwise specified. Following amplification, 1 µl of each sample (up to four PCR samples with different dyes) was diluted in 25 µl of dH_2O and 1 µl of the dilution was added to 8.9 µl of Hi-Di Formamide (Applied Biosystems) and 0.1 µl of the size standard LIZ500 (Applied Biosystems). LIZ500 was used for allele sizing up to 500 bp and this was the case for all SSRs unless otherwise specified.

Products were separated by capillary electrophoresis on an ABI 3730 DNA Analyzer (Applied Biosystems) with a POP-7™ polymer column and manual SSR allele sizing was performed using GeneMapper v4 software (Applied Biosystems). Each SSR marker was first tested on the parent NILs and four types of polymorphism could be identified, outlined in Figure 5.5. Polymorphic markers were then used to genotype the mapping populations. The markers tested are shown in Appendix 2.9. Polymorphic markers are shown in Appendix 2.10.

2.3.10.2 Validating SNPs with KASP assays

KASP assays were conducted in a 384-well format using an optically clear 384-well plate (Framestar, 4titdue Ltd). An aliquot of 2 µl of DNA at ~10 ng μl^{-1} was dried at 65°C for 30 min. KASP assays were carried out using 0.056 µl of primer mix (12 µl FAM primer (100 µM) + 12 µl of VIC primer (100 µM) + 30 µl of common primer (100 µM) + 46 µl of dH_2O) and 2 µl of KASP V4 Mastermix (LGC group, UK). PCR cycling was performed on the Eppendorf Mastercycler pro, using the same program with an optimal cycle number for each marker (as indicated in Appendix 2.11 and Appendix 2.12): 94°C for 15 min, 10 cycles of 94°C for 20 s, 65°C for 1 min and 94°C for 20 s, then 30 – 40 cycles of [94 °C for 20 s and 57 °C for 1 min]. Fluorescence was measured by a Tecan Safire plate reader at ambient temperature. Results were analysed manually using the KlusterCaller software (version 2.22.0.5; LGC group, UK). Blue clusters on the y-axis were FAM-labelled and red clusters on the x-axis were VIC-labelled, with the no template control labelled black and 2D DNA labelled pink. If the genotyping clusters were not sufficiently separated after initial amplification, additional cycles were added in groups of five to a maximum of 40 cycles and re-analysed.

2.3.11 Anchoring of the Rht8 *interval in Triticeae resources*

2.3.11.1 *Ensembl*Plants and barley resources

The *Rht8* interval was anchored in the most recent resources using the flanking markers *DG279* and *DG371*. This was updated as resources become available during this project. The final update to these resources was in March 2015 for *Ensembl*Plants (release 26) (http://plants.ensembl.org), May 2015 for population sequencing (POPSEQ) data for IWGSC-2 (Mascher, 2014) and Chapman assemblies (CerealsDB, 2015a) and August 2015 for the *Ae. tauschii* resources.

The EST sequence of *DG279* was used as a query in a BLASTN homology search against the IWGSC CSS hosted on the URGI server (URGI, 2013) to retrieve two overlapping CSS 2DS contigs shown in Figure 2.7. The Bd21 Genome Annotation v1.0 (International Brachypodium Initiative, 2010) was updated using Brachypodium Munich Information Center for Protein Sequences (MIPS) gene models in the *Ensembl*Plants 2013 release (which first included the IWGSC arms). The Brachypodium orthologue was confirmed as *Bradi5g03460*. The rice locus identifiers used previously from MSU (Ouyang et al., 2007) by Gasperini (2010) became obsolete in the *Ensembl*Plants IWGSC v1 release, which instead used the International Rice Genome Sequencing Project (IRGSP-1.0) assembly (Kawahara et al., 2013). Therefore, for consistency the new identifiers were used (Figure 2.7). *DG279* mapped to the new IRGSP-1.0 locus name of *Os04g0209200*.

Initially, barley resources were fragmented and were made available on the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) barley BLAST server (Deng et al., 2007) as ‘high-confidence’ (HC) and ‘low confidence’ (LC) MIPS gene models, as well as the whole-genome shotgun assembly of Morex. Prior to the barley data being published, the Morex assembly data was provided by Burkhard Steuernagel and Matthew Moscou. The corresponding HC barley gene was *MLOC_5957* on 2H: 15451125, which mapped to a Morex contig in the 12.11 cM bin on chromosome 2.

Later, *DG279* was mapped onto the v3.3 cDNAs. This reference contained ordered and unordered cDNAs, but only the ordered genes were of use for anchoring, since the unordered cDNAs could only be sorted according to chromosome arm provenance. *DG279* mapped to a 2A cDNA on the ordered cDNAs. In order to delimit the 2D interval on the v3.3 cDNAs, the Morex assembly data was used to find Morex contigs within the same genetic bin (12.11 cM) which did map to an ordered cDNA in the wheat reference. Brachypodium and rice synteny was used to locate the appropriate genetic bins in the Morex assemblies and then these were used as queries in BLASTN homology searches of the HC barley genes. Of the 43 Morex contigs in the 12.11 cM bin, only two could be anchored to ordered 2D v3.3 cDNAs. The most proximal (conservative) of these was taken to anchor the *DG279*-end of the 2D interval.

An analogous process was used to anchor the distal end of the *Rht8* locus with *DG371*. The EST retrieved a highest hit by BLASTN sequence homology search to a 2BS CSS contig. The highest level of homology (by e-value) to Brachypodium and rice genes was *Bradi5g04710* and *Os04g0261400* in the new rice annotation. The gene in barley was much more difficult to anchor, presumably due to the 2BS localisation in wheat chromosome arm sequence. *DG371* was mapped to *MLOC_58453* which was not a HC gene, but mapped to contig_42684 in the 14.38 cM bin on chromosome 2 according to the IPK server. However, this contig could not be found in the Morex assembly data. Brachypodium synteny was used to consider the closest gene which had a strong identity hit to a wheat 2DS contig which could be anchored to the HC barley data and Morex assemblies. Accordingly, *Bradi5g04673* was used (via 2DS_5366894) to anchor *DG371* to the 15.44 cM bin as *MLOC_12182*. Since *DG371* mapped to a 2B v3.3 cDNA, the Morex contigs in the 15.44 cM bin were considered to find a distal *Rht8* interval anchor on the 2D ordered v3.3 cDNAs. The most distal of these was taken as a conservative estimate of the v3.3 cDNA position.

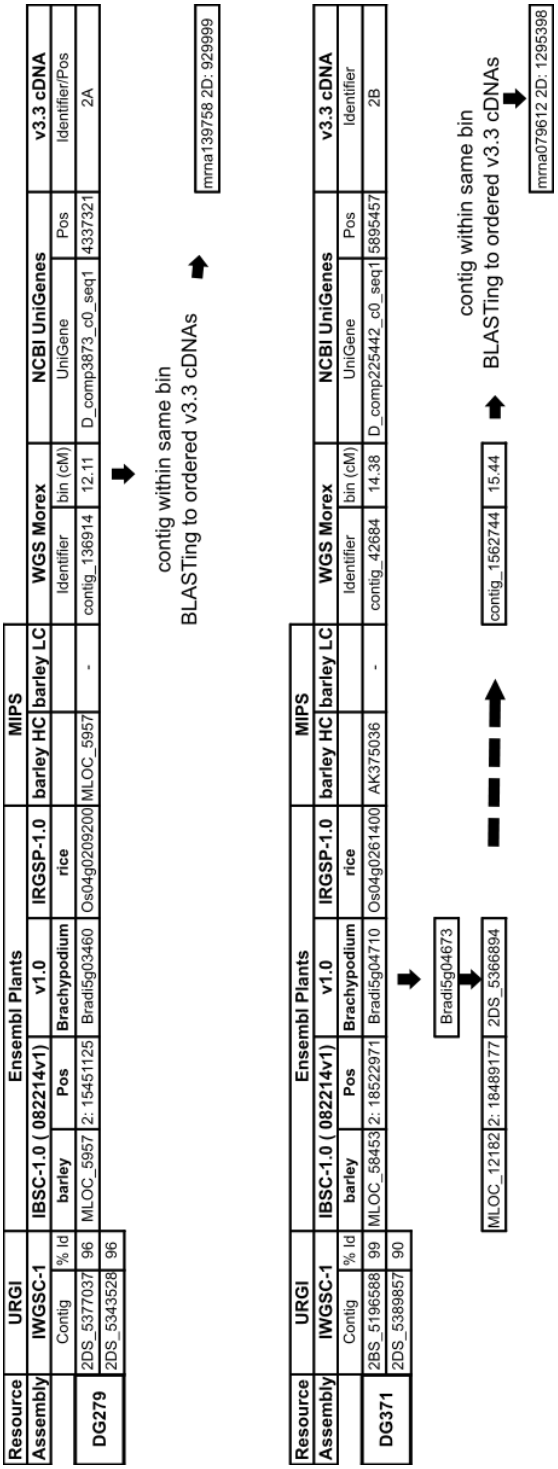


Figure 2.7: Anchoring the Rht8 interval in Triticeae resources and wheat references, updated during the PhD.

2.3.11.2 Constructing zippers

The Plant Mart menu within BioMart (Kasprzyk, 2011) in *Ensembl*Plants (release 25) was used to export iteratively assemblies shown in Figure 2.7 for chromosome 2H for barley, chromosome 5 for Brachypodium and chromosome 4 for rice. For each dataset, orthologues were selected from the attributes menu and each of the other species was selected, as well as the IWGSC CSS contigs. The data was exported as a csv file and modified further. For each dataset, the 2DS CSS contig for each gene was retained and this formed the basis of the zipper. Each zipper was annotated with orthologous information and subsequently with marker information. Anchoring of the wheat resources in the zippers was achieved by using the IWGSC CSS contig to each gene as an identifier in searches between datasets.

2.3.11.3 IWGSC-2 POPSEQ bins

The wheat chromosome 2 data from IWGSC-2 contigs arranged into genetic bins according to POPSEQ was downloaded from the URGI server (Unité de Recherche Génomique Info), a research unit in bioinformatics at Institut National de la Recherche Agronomique (INRA) (URGI, 2015b). Subsequently the IWGSC CSS contig which new markers anchored to was used as an identifier in searches in the spreadsheet to retrieve the genetic bin corresponding to that contig.

2.3.11.4 Chapman assembly

In May 2015, the WGS Chapman assembly was hosted on CerelasDB with an interface for a MegaBLAST search (Morgulis et al., 2008). A file ordering the Chapman scaffolds by their identifier into genetic bins ordered by POPSEQ was also made downloadable in 2015 from IPK (Mascher, 2014). The FASTA sequences of 2DS contigs from the genome zippers described in 2.3.11.2 were used as queries against the Chapman scaffolds in a MegaBLAST search (E-value cut-off 1E-05) on the CerelasDB interface. MegaBLAST is able to handle longer DNA sequences (as is the case with the Chapman scaffolds) than the BLASTN program of the BLAST algorithm. Chapman scaffolds matching with a minimum 99% sequence identity were retained. FASTA sequences of the Chapman scaffolds corresponding to the syntenic intervals of the zippers were mined for

SSRs as described in 2.3.9. Markers in Table 6.2 were annotated with the genetic bin the Chapman scaffold anchored to, using the 2DS contig to the marker as the identifier between datasets.

2.3.11.5 Constructing synteny maps

The synteny maps were constructed using ArkMAP (Paterson and Law, 2013), which at the time of writing incorporated *Ensembl*Genomes release 25. The barley 2H map was downloaded first (IBSC-1.0; 2H: 15,200,000-20,000,000). The context menu was used to show conserved synteny with Brachypodium (v1.0) and rice (IRGSP-1.0). The threshold for defining an orthologous relationship was set to 125/200 combined similarity threshold. A total of 48 orthologous genes were found on Brachypodium chromosome 5, 16 on chromosome 3 and 15 on chromosome 1. For rice, 38 orthologous genes were found on chromosome 4, 11 on chromosome 11 and 9 on chromosome 10. For the wheat synteny map, the wheat interval was downloaded first (IWGSC-2; 2D: 6478405 – 10885088). A total of 71 orthologues were identified on barley 2H (only 1 or 2 genes were found outside of 2H, data not shown on map), for Brachypodium: 57 on chromosome 5, 17 on chromosome 4 and 15 on chromosome 3 and for rice: 39 on chromosome 4, 14 on chromosome 11 and 10 on chromosome 7. For clarity, not all the orthologous relationships were shown on the map. Instead, the number of orthologues shown on the map was filtered to 1 in 10 for barley, 1 in 10 for Brachypodium and linked genes in rice.

2.3.12 Genotyping and mapping with the 2D RIL population

The polymorphic markers shown in Appendix 2.10 and Appendix 2.11 were used to genotype the 2D RIL population (described in 2.3.1.1) with SSR and KASP assays (2.3.10). The conditions in terms of optimal cycle number (for KASP) and detail about the type of polymorphism identified (SSR) are shown in Appendices 2.10-2.12. Each polymorphic marker was also tested with the flow-sorted 2D DNA as outlined in 2.3.3.1, to ensure 2D specificity (shown in Figure 5.11). The markers were arranged into classes according to the graphical genotypes (Appendix 5.16). The genotypic information (Appendix 5.17) was combined with

the genotypic scores of markers mapping close to the *Rht8* interval used by Gasperini (2010) (shown in black in Figure 5.11). The linkage maps were created using the Haldane mapping function in JoinMap® version 3, using a log-of-odds (LOD) threshold of 3.0. Genetic distances were not adjusted using recombination frequencies at this coarse-mapping stage. MapChart v2.2 (Voorrips, 2002) was used to draw the linkage map.

2.3.13 Costings for marker development

The cost per marker shown in Table 5.6 were calculated on the basis of the costings in Table 2.3.

Calculating costs:

	Axiom:	2 samples	£600
	iSelect	10 samples	£500
4 lanes, 8 libs	NCBI	parent NILs	£7,580
6 lanes, 14 libs	BSA parent NIL and bulks		£11,550

1 lane	£1,715
1 library	£90

For each marker: allow 300 reactions to map all populations

For SSR:	labelled adapter	£96	800ul	
	For 300 reactions:	£5	38ul	
	Liz500	£360	500ul	
		£20	27ul	
	HotStar Taq	£327	10200ul	
		£31	300 reactions	950 µl
	Plates	£3		
	primer set	£6		

For KASP:	KASP mix	£450	25ml	
		£11	300 reactions	600 µl
	Plate	£1.50		
	primer set	£12		

Table 2.3: Costings used to calculate the cost of developing an individual marker outlined in Chapter 5.

2.4 Fine-mapping and further characterisation of the *Rht8* interval

2.4.1 Phenotyping the fine-mapping population at the Rht8 locus.

The phenotyping of the fine-mapping recombinants in the glasshouse is described in 2.3.2.1. Sterility in the glasshouse was measured by assessing the grain content of each spike. A scale of 0-5 was used to assess sterility, as shown in Appendix 6.3, with 5 being totally sterile (no grain in the spike). A linear model was fitted in R v3.1.1 (R Development Core Team, 2014) to measure the effect of sterility on the spike length, with genotype as a fixed effect and with block and bench as random terms (Appendix 6.2). Block and bench were subsequently removed from the model since they did not affect the spike length at the $P<0.05$ level.

The fine-mapping recombinants were further grown in monodrill trials in the field at two locations (Church Farm and Morley) in the 2013-2014 season, with the exception of six recombinants not grown at Morley due to insufficient seed (details in Appendix 6.1). The monodrill was set up to drill 10 plants per row with a spacing of 12.5 cm between plants. There were four rows per plot, 10 plots per block and five blocks within a replicate (Figure 2.8). The outer rows were drilled with the wheat variety Soissons (drilled as discard), which is an easily visually-identifiable early-flowering, short and awned wheat (Figure 2.9). This was used to eliminate edge effects that might affect height and also to shelter the inner rows from lodging in the event of extreme weather. Soissons was also included in the field design as a marker for orientation in the field (Figure 2.8). Each recombinant was grown in five replicates. The parent NILs were grown at least one per block. The recombinants were randomised within each replicate across the blocks in a complete block design.

Field trials were kept weed- and pest-free with products according to standard agronomic practice at each of the locations, with the exception that plant growth regulators (PGRs) were not applied. Trials were drilled in October 2013.

Plant height was measured in the field upon maturity (Figure 2.10). A height measurement was made per row (each cell in Figure 2.8). A main tiller was selected from a representative plant within a row, avoiding the edges of rows, therefore (pseudo-)replication was in reality higher than the five replicates of the field design. Height was measured using an extended ruler from soil level to the tip of the spike.

The height distributions of recombinants were plotted in histograms and a bimodal split identified in each of the locations shown in Figure 6.2, which was calculated as the middle of the bin which separated the two bimodal distributions. The bimodal split was 76.75 cm in the glasshouse, 88.75 cm at Church Farm and 104.25 cm at Morley. Recombinants in the distribution below the split were typed as short (‘b’) and above the split were typed as tall (‘a’). The full data is shown in Appendix 6.1.

A consensus score of short or tall for each recombinant was based on a minimum match of two out of three locations by comparing phenotype scores at each location. In one case with a contrasting score between glasshouse and Church Farm (and no data at Morley), the field score was used since it was deemed more reliable (Appendix 6.1).

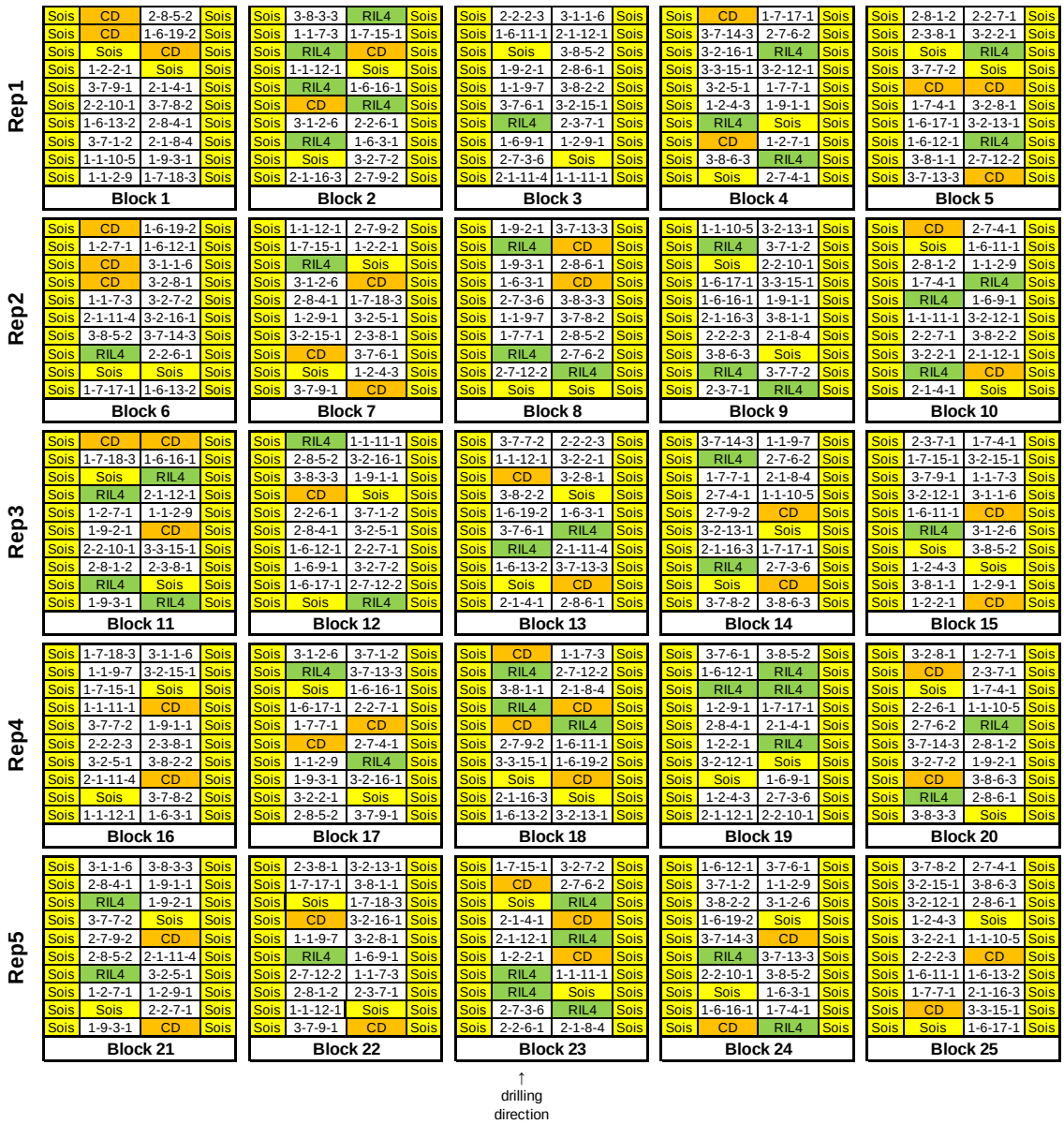


Figure 2.8: Trial design to phenotype the fine-mapping recombinants in the field using the monodrill. The same design was implemented in each field location (this particular design is for Morley), but blocks were arranged with two replicates adjacent to three replicates at Church Farm due to field constraints. Each cell represents a row. One plot is each horizontal section of each block. Highlighted yellow = Soissons (discard); Highlighted green = RIL4 (short parent NIL); highlighted orange = Cappelle-Desprez (tall parent NIL). Trials were drilled in October 2013.

Figure 2.9: Field layout with the outer Soissons rows visible. Photograph taken at Church Farm on 8th July 2014.

Figure 2.10: Method used to phenotype recombinants in the field. A ruler was used to measure from soil level to the tip of the ear. The ruler in the image is for scale only (20 cm intervals), a more precise extended ruler was used to record heights to the nearest half centimetre.

2.4.2 Mapping Rht8

2.4.2.1 Fine-mapping in stages

The first mapping stage using the coarse-mapping population (described in 2.3.1.1) is outlined in 2.3.12. From Figure 5.11, the 62 markers grouping with the flanking *Rht8* markers *DG279/DG371* were carried forward for fine-mapping. In the medium-resolution mapping, the *Xgwm261-Xcfd53* recombinants (described in 2.3.1.2) were genotyped with the 62 markers. Markers were arranged into classes according to co-segregating graphical genotypes (Appendix 6.5). The *Rht8* phenotype score was obtained from Gasperini's (2010) height data for these recombinants and used as a graphical genotype, as shown in Appendix 6.5. The graphical genotype of *Rht8* was combined with the marker class genotypes and marker class order around *Rht8* was determined using JoinMap® version 3 for linkage analysis as described in 2.3.12. The marker classes that mapped within the *DG279-DG371* interval (classes 13-29) contained a total of 33 markers, which were retained for the final fine-mapping step.

The 73 fine-mapping recombinants (described in 2.3.1.2) were used to genotype the 33 markers and subsequently marker classes A-G were established on the basis of co-segregating markers (Appendix 6.6). The consensus a/b scores for the fine-mapping population were used as a graphical genotype at the *Rht8* locus. Linkage between marker classes A-G and the *Rht8* locus was established using Joinmap version 3.0 (Van Ooijen and Voorrips, 2001), with default settings and the Kosambi mapping function. The linkage to *Rht8* was determined using a LOD threshold score of 3.0. Genetic distances were computed based on recombination frequency. The recombination frequency between marker classes was calculated from the numbers of recombination events between classes divided by 6208, which was derived from the formula: recombination frequency = [(recombinant heterozygotes + 2x recombinant homozygotes)/2x F_2 plants], where F_2 plants = 3104. The linkage map was drawn using MapChart v2.2 (Voorrips, 2002).

2.4.2.2 Aligning the genetic map of the *Rht8* region with physical maps

2.4.2.2.1 Syntenic species

Marker information from Table 6.2 was used to anchor the markers and marker classes onto the physical maps of barley, Brachypodium and rice. This was done using the resources outlined in section 2.3.11. The alignment of the *Rht8* genetic map with the physical maps was done using the physical positions of each marker to construct the individual physical maps for each species. Distances between markers were calculated using the physical positions and these were drawn to scale, with the lines between markers shown. The map was constructed manually using the vector graphics software Inkscape version 0.91 (www.inkscape.org).

2.4.2.2.2 Ae. tauschii *and* T. aestivum *resources*

To identify *Ae. tauschii* BAC contigs (Luo et al., 2013) orthologous to the *Rht8* region, the FASTA sequences of the 2DS CSS contigs corresponding to the markers in classes A-G were used to perform BLASTN analysis against the *Ae. tauschii* SNP marker sequence database (http://probes.pw.usda.gov/WheatDMarker/phpblast/blast.php). The best-hit SNP marker identifiers were retrieved, with parameters of an E-value cut-off 1E-10 and overall bit score >200. The comparative map, gene list and genome zipper were downloaded from the 'Sequencing the *Aegilops tauschii* Genome' project, (UC Davis Plant Science and USDA, 2015) then compiled and annotated using the SNP marker identifiers. Extant markers that had already been tested during the project were annotated onto the *Ae. tauschii* resources using the Brachypodium genes as identifiers between datasets.

To anchor the *Rht8* linkage map in the IWGSC-2 and Chapman POPSEQ bins, the resources described in 2.3.11.3 and 2.3.11.4 were inspected and the 2DS CSS contig corresponding to the marker was used as the identifier between datasets.

The *Ae. tauschii* genetic map was constructed by using the linkage bins in the gene list described above and mapping the SNP markers in the appropriate cM bins using MapChart v2.2 (Voorrips, 2002). This was then annotated with the

BAC contig onto which the SNP markers mapped to, using the comparative map file. The genetic maps of the IWGSC-2 and Chapman POPSEQ bins were drawn to scale according to the number of contigs/scaffolds within each bin. Finally, the maps were compiled in Inkscape version 0.91 (www.inkscape.org) and lines between maps drawn manually.

2.4.3 Gene content of Rht8 *interval*

2.4.3.1 Differential expression analysis

The sorted and indexed BAM files from the v3.3 cDNAs and UniGenes were processed to measure transcript abundance by Martin Trick, using the methodology outlined in Trick et al., (2012). Transcript abundance was expressed as reads per kilobase per million mapped reads (RPKM values) for the parent NILs in the UniGenes and for both the parent NILs and bulks in the v3.3 cDNA gene models. Since there was no reference group against which to identify functional enrichment, the data was analysed using an adapted workflow based on the R package sRAP (v1.8.0) (simplified RNA-Seq Analysis Pipeline) (Warden et al., 2013). RPKM values were normalised by setting an RPKM cutoff in order to eliminate false discovery of high fold-changes between genes with very low absolute expression levels (<0.1). Additionally, genes with an expression of zero were rounded to a small number (0.001) to avoid logarithms of zero. Expression values were then log2 transformed and genes were defined as differentially expressed if they showed an absolute fold-change>1.5. Conservative parameters (according to Warden et al., 2013) were used to define differentially expressed genes (DEGs) in both the normalisation step and the fold-change step. The analysis is shown in full in the R script in Appendix 2.13.

The GO Analysis Toolkit and Database (agriGO) (http://bioinfo.cau.edu.cn/agriGO/) was used compare the molecular function and biological processes for the 1735 DEGs between the parent NILs and a reference library. At the time of doing this work, the *Triticum aestivum* gene models from the IWGSC-2 data (gene models in Appendix 6.9 and Appendix 6.10) were unavailable on agriGO therefore the Brachypodium homologues to the DEGs were used as identifiers. The singular enrichment analysis (SEA) tool was used with the Bradi genome locus (JGI) set as the reference library with 25,219

Brachypodium genes (Du et al., 2010). Significant GO terms were identified using a p-value <0.05 from the SEA.

2.4.3.2 Gene content of *T. aestivum* genetic bin and *Ae. tauschii* physical interval

The IWGSC-2 data was represented in *Ensembl*Plants as contiguous sequence, despite there being no greater resolution to order individual IWGSC CSS contigs within genetic POPSEQ bins. In order to identify the gene content of the 17.3 cM bin and surrounding bins, a physical interval had to be identified *Ensembl*Plants. To do this, 2CS CSS contigs from the flanking POPSEQ bins were used as an approximate estimate (with the knowledge that the intervening sequence would contain some sequence space from each of the flanking bins). The 2DS CSS contig was used as a query for homology searching in the *Triticum aestivum* BLASTN menu, and the pseudo physical position retrieved. For the 17.3 cM bin, this was achieved using a contig from the flanking 17.0 and 18.1 cM bins. The physical intervals used were 2D: 6478405-8959961 for the 17.3 cM bin and 2D: 8745143 – 10885088 for the 33.1 cM bin.

Using these physical intervals, BioMart (Bolser et al., 2015) was used to export the peptide sequences and the corresponding syntenic information as outlined 2.3.11.1. In order to functionally annotate the genes, the peptide sequences were used as queries in BLASTP searches of non-redundant protein sequences of the National Center for Biotechnology Information (NCBI), using 'flowering plants' (taxid 3398) as a filtering parameter in the 'organism' menu (http://blast.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome). The best-hit by query-coverage percentage and peptide identity was retrieved, however where the best hit was an uncharacterised protein, lower hits were used where annotations existed. BLASTP searches which returned retrotransposons were discarded, but these were infrequent. The annotation was also checked using the GenBank identifier. The GenBank identifier was used for further literature searches on the NCBI and elsewhere (for example in www.uniprot.org) to investigate putative biological roles.

2.5 Germplasm development to study rare alleles at the *Xgwm261* locus

2.5.1 Plant material

The original introgressions of the rare *Xgwm261*-allele donors into the Mercia background were performed by Liz Sayers (JIC) in 2006 in multiple streams. Four streams produced heterozygous offspring which were fertile. Seed was obtained for these and 48 seeds for each stream were germinated and planted into individual cells in a 96-well tray as described in 2.3.1.3. Leaf material was used for DNA extraction (described in 2.3.1.4) to screen with *Xgwm261* (described in 2.3.1.4) and allelic variation scored as shown in Appendix 7.1.

2.5.2 Germplasm development

The homozygous individuals for the parent and donor *Xgwm261* allele identified in 2.5.1 were transplanted into 1L pots and grown to maturity in glasshouse conditions in the summer of 2015. Plants were bagged in order to bulk seed.

2.5.3 Height measurements

Measurements for height and height-components were made as described in 2.3.2.1.

2.5.4 Statistical analyses

The Student's t-test was used to compare height and height components between the donor and parent alleles within each cross. The deviation of the observed segregation patterns from the Mendelian 1:2:1 ratio were tested by $\chi 2$ using the Microsoft Excel function.

Chapter 3: Agronomic characterisation of *Rht8* in UK-adapted germplasm.

3.1 Introduction

Crop height is a key trait to optimise the performance of wheat (*Triticum aestivum* L.). Decreasing height is important in order to reduce lodging (Berry et al., 2007). However, reduced height is often associated with a reduction in yield (Law et al., 1978) thus understanding better genes which reduce height without yield penalty continues to be a prominent breeding target.

The widespread deployment of the semi-dwarfing genes *Rht-B1b* and *Rht-D1b* was a key part of the Green Revolution in reducing plant height by 15-36% (Gale and Youssefian, 1985, Trethowan et al., 2001) and increasing yields, as well as improving lodging resistance to high nitrogen (N) application (Hedden, 2003). *Rht-B1b* and *Rht-D1b* are gibberellin (GA)-insensitive and inhibit cell elongation, with subsequent height reduction. The reduced cell-size also causes a reduction in coleoptile length (Trethowan et al., 2001). Shorter coleoptiles and decreased leaf area of seedlings reduce early vigour and impede emergence in deeper sowing due to dry conditions (Botwright et al., 2005, Rebetzke and Richards, 1999).

Despite the prevalence of *Rht-B1b* and *Rht-D1b* in over 70% of wheat varieties worldwide (Hedden, 2003), the genes perform less well in lower yield-potential environments. *Rht-B1b* and *Rht-D1b* increase sensitivity to drought and temperatures stress (Gale and Youssefian, 1985) and are associated with reduction in yield in more marginal environments such as low rainfall (Chapman et al., 2007, Butler et al., 2005) and low nitrogen fertiliser input (Laperche et al., 2008). Circumstantially, this is further evidenced with reduced prevalence of *Rht-B1b* and *Rht-D1b* in southern/central European gene pools where there are higher temperatures associated with arid summers (Šíp et al., 2009).

The main alternative to the GA-insensitive dwarfing alleles found in agriculture is the GA-responsive *Rht8* on 2DS, recognised by a 192-bp allele at the closely-linked microsatellite locus *Xgwm261* (Korzun et al., 1998). *Rht8* is well-adapted to dry, Mediterranean-like environments (Worland and Law, 1986) as it provides semi-dwarf stature with the benefits of early seedling vigour and a longer coleoptile (Ellis et al., 2004). *Rht8* is found extensively in southern Europe and parts of eastern/central Europe, as well as China and Australia (Asplund et al., 2012). *Rht8* is not found in northern European germplasm and has not been tested extensively in the UK, principally due to the tight linkage with *Ppd-D1a* (Worland et al., 1998a).

Rht8 was introduced together with the closely-linked photoperiod-insensitive *Ppd-D1a* into European wheats in the 1930s from the crossing programmes with the Japanese variety Akakomugi by the Italian breeder Strampelli (Borojevic and Borojevic, 2005, Lorenzetti, 2000). Photoperiod-insensitive wheat flowers rapidly in both short and long days, whereas photoperiod-sensitive wheat is delayed in short days, flowering rapidly in long days. *Ppd-D1a* is advantageous in climates (such as southern Europe) where earlier flowering avoids late-season drought stress (Kato and Yokoyama, 1992) and high-temperatures at grain fill (Bennett et al., 2012). In northern Europe, breeders likely selected against the tightly-linked *Ppd-D1a*, because in a climate with relatively cool summers, a long vegetative phase coupled with late flowering is favourable to maximise yield (Kato and Yokoyama, 1992). In order to test the behaviour of *Rht8* without the disadvantageous (to the UK) *Ppd-D1a* allele, a population was developed in the Griffiths' group in a photoperiod sensitive (*Ppd-D1b*) background (first described in Gasperini, 2010).

Near-isogenic lines (NILs) contrasting for the *Rht8* allele from Mara (defined by marker-assisted selection for *Xgwm261*-192bp and *Xcfd53*-274bp) and tall *rht8* allele from Cappelle-Desprez were developed in the elite spring wheat, Paragon (Gasperini, 2010). Paragon does not contain *Rht-B1b* or *Rht-D1b* but has reduced height probably due to accumulation of several minor genes and is photoperiod sensitive. From the BC_3F_3 stage, one short NIL and one tall NIL were selected and grown alongside Paragon in two sites in the UK and a high-temperature site in Lleida, Spain. These results are described in this Chapter.

Previous agronomic assessments of *Rht8*, many confounded by the pleiotropic effects of *Ppd-D1a*, have mainly focused on height with limited investigation into yield, yield components or developmental traits without the earliness conferred by photoperiod insensitivity. Dissecting the genetic and physiological effects of *Rht8* away from *Ppd-D1a* was a key aim of this study.

Optimum plant height for maximising yield varies according to the yield potential of the environment (Fischer and Quail, 1990, Flintham et al., 1997) and ranges from 70-100cm (Flintham et al., 1997). In high-yield potential environments (high fertiliser input and irrigation), the dwarfing conferred by the single action of *Rht-B1b* or *Rht-D1b* is insufficient to avoid lodging (Berry et al., 2007). Hence there is growing interest in combining *Rht8* with other dwarfing genes to 'fine-tune' height. Double-dwarfs with *Rht-B1b+Rht-D1b* confer maladaptive traits of poor establishment and low biomass (Butler et al., 2005, Flintham et al., 1997) whilst other *Rht* genes such as *Rht3*, *Rht10* and *Rht12* are too extreme in their height reduction to be of commercial value (Ellis et al., 2004, Flintham et al., 1997). Instead the more subtle height reduction conferred by *Rht8* makes the combination of *Rht8* with *Rht-B1b/RhtD1b* to create 'sesqui-dwarfs' a more attractive target and this was first studied in a high-yielding Australian semi-dwarf wheat background (Rebetzke et al., 2012a). The combination of dwarfing alleles reduced lodging and increased grain yields relative to the single dwarfs. The work presented in this Chapter is the prelude to work the Griffiths' group is carrying out to obtain a similar stacking of *Rht8* with other *Rht* genes in Paragon, a high-yielding UK spring wheat, and testing performance in northern European climates.

Molecular-marker studies for *Rht8* in Akakomugi-derived, Mara progeny found a height reduction of 10-15% (Gasperini et al., 2012, Korzun et al., 1998). In the field, *Rht8* has been found to decrease height by a mean of 6.5% across a range of environments in Colorado (Lanning et al., 2012). In the BC_3F_2 NILs used in the current work, an initial assessment of the heights of the *Rht8* NILs reported a 20% height reduction in the glasshouse (Gasperini, 2010). A modest number of studies have characterised the effect of *Rht8* in terms of yield and yield components; none have used morphometric measurements (presented in this Chapter) to report on grain size. For lines carrying *Rht8*, yield increases of 9.7% (Rebetzke and Richards, 2000) and 3.8% (Borner et al., 1993) have been

reported. However, in more recent agronomic assessments, the *Rht8* allele did not confer yield advantage over 10 sites studied and instead showed a penalty in three of the 10 environments (Lanning et al., 2012). The *Rht8* allele has been shown to increase carbon-partitioning to the grain to increase grain number and yield in Australia in a relatively low-yield environment (2.5 – 4.6 t ha^{-1}) (Rebetzke and Richards, 2000). In studies with *Rht8+Ppd-D1a*, *Rht8* had little effect on grain number (-1%), but this was linked with earlier flowering (Addisu et al., 2009a, Rebetzke et al., 2012b). With need to clarify these conflicting reports, an extensive assessment of yield and yield components of *Rht8* was carried out across multiple sites in the work presented here.

There is growing need to reduce or curtail the use of synthetic inputs including N fertiliser. In the UK, N fertiliser inputs are already limited by legislation and it is anticipated that a more severe reduction will be enforced by EU-wide legislation in 2016. For this reason, there is imperative to understand how the action of semi-dwarfing genes differs at contrasting N applications. The use of *Rht8* in alternative management systems, such as organic agriculture with low N inputs, has not been extensively tested. This is despite the highly promising increase in early crop vigour reported with this gene (Ellis et al., 2004), a trait which has been identified as particularly useful in organic contexts in order to promote early nutrient uptake (Wolfe et al., 2008). To this end, the agronomic performance of *Rht8* was assessed here at N inputs below that of conventional agriculture.

There has been limited work to contrast the performance of *Rht8* in irrigated and non-irrigated systems. In one experiment, an *Rht8+Ppd-D1a* NIL in the Mercia background was found to have increased drought tolerance at booting, resulting in increased grain per spikelet relative to *Rht-B1b* and *Rht-D1b* even at temperatures as high as 36°C (Alghabari et al., 2014). Since the plants in those experiments were potted, the translation of these results to the field is uncertain. Additionally, the considerable G x E interactions, as well as the photoperiod insensitivity, meant that effects could not be unambiguously ascribed to the semi-dwarfing *Rht8* allele. In the field, *Rht8* NILs in a *Ppd-D1b* background yielded less than those with *Rht-B1b* and *Rht-D1b*, in conditions with late-season drought and temperature stress (Lanning et al., 2012). Experiments with the *Rht8* and tall NILs were conducted in irrigated conditions in the UK growing season, to test whether these findings could be extended to relatively cooler summers.

In this Chapter, the agronomic performance of *Rht8* is described for the first time in a UK-adapted spring wheat background (Paragon) in terms of height, yield, yield components and developmental traits. The trait responses of *Rht8* in contrasting irrigation treatments and a range of N inputs were examined as a preliminary study into whether the gene could be usefully deployed in lower input management systems typical of organic agriculture.

3.2 Inter-site comparison

Near-isogenic lines (NILs) contrasting for the *Rht8* allele from Mara (defined by marker-assisted selection for *Xgwm261*-192bp and *Xcfd53*-274bp) and tall *rht8* allele from Cappelle-Desprez were developed in a Paragon background (Gasperini, 2010). At the BC_3F_3 stage, one short NIL (herein *Rht8* NIL) and one tall NIL were selected and grown along with the recurrent parent to the population, Paragon. Three growing environments were used for this study (Table 3.1), two in the UK and a high-temperature site (that was irrigated to field capacity) in Lleida. The UK sites had shorter days in winter and longer days in summer relative to Lleida (Figure 3.1).

A range of temperature was encountered in UK and Lleida (Figure 3.2A) with Lleida being higher throughout the reproductive and grain-filling phases (including booting and anthesis). Lleida had a higher range (with a low of 4°C and high of 24°C in July, when harvest was completed) than the UK (low of 4°C and high of 18°C in Reading). The two UK sites had similar climates, differing only slightly in temperature at the end of stem elongation/beginning of grainfilling. Reading was 2°C hotter than Norwich in June and 1°C hotter in July. There was more rainfall at the start of the season in Reading compared to Norwich during the vegetative phase followed by a drier latter half of the season in Norwich (Figure 3.2C). UK sites had markedly lower levels of solar radiation, with half the levels in Lleida in some months (Figure 3.2B). Church Farm in Norwich was the highest yield-potential site: though variable, the highest average yield was close to 11 t ha^{-1}, compared to 9 t ha^{-1} in Reading and only 7 t ha^{-1} in Lleida (Table 3.1).

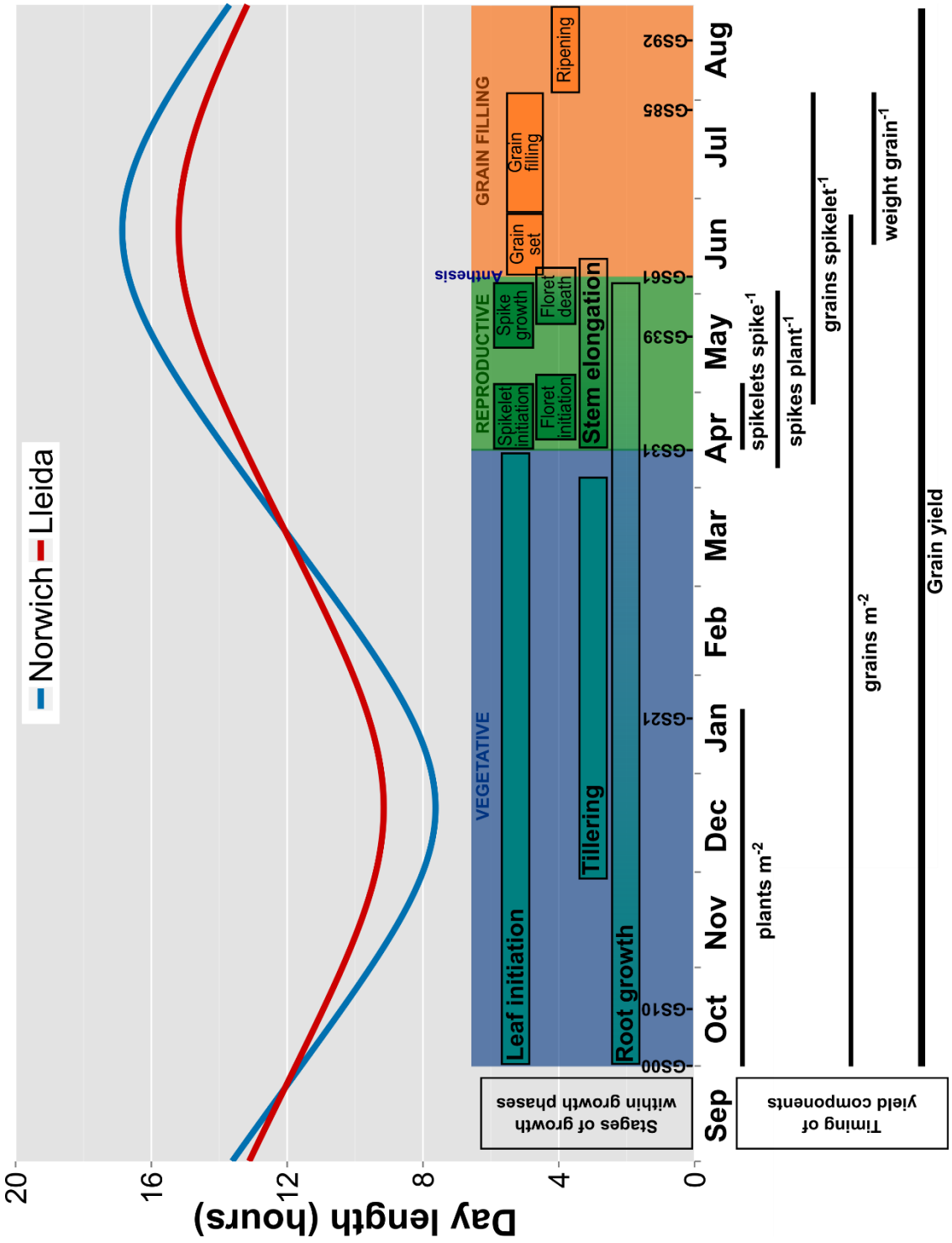


Figure 3.1: Day-length over the wheat growing season (sowing to harvest) overlayed with growth stages and timing of yield components. The growth phases were based on a UK autumn drilling season, and adapted from HGCA, 2008. Growth season for Lleida is shifted 4-6 weeks earlier. Day-length was calculated by using the latitude and longitude coordinates to determine sunrise and sunset over the year 2013-2014 and finding the numeric difference, as described in 2.1.3. The yield components are overlayed on the growing season and adapted from Slafer, 2012.

Site			Soil		Drilling		Experimental Setup					Measured traits					Yield
Environment	Location	Coordinates	Type	Composition	Nominal depth (mm)	Density (seeds m^{-2})	Plot size	Year(s)	Treatment	Reps	Design	Analysis	Height	Yield	Tiller	Grain	Developmental	Mean (t DM/ha)
Church Farm	Norwich UK	52°37'40"N, 1°10'46"E	Sandy clay loam	-	50	260	1.5 x 4m	2011	-	1	-	-	H	-	-	-	-	-
								2012	-	3	RCBD	One-way ANOVA	H	Y, HI, GN, SS	S, I	TGW, GA	-	8.7
								2013	N (N2, N3)	3	RCBD	Two-way ANOVA	H	Y, HI, GN, SS	S, I	TGW, GA	HD, AN	8.4
								2013	Irrigation (✓/✗)	3	RCBD	Two-way ANOVA	H	Y, HI, GN, SS	S, I	TGW, GA	HD, AN	8.3
								2014	N (N1, N2, N3)	3	RCBD	Two-way ANOVA	H, L	Y, HI, GN, SS	S, I	TGW, GA	GC, HD, SEN	10.6
								2014	Irrigation (✓/✗) & N1/N3	3	RCBD	REML	H, L	Y, HI, GN, SS	S, I	TGW, GA	GC, HD, SEN	10.7
Sonning	Reading UK	51°29'32"N, 0°56'19"W	Sandy loam	9% clay 19% silt 72% sand	50	260	1.92 x 5m	2014	N (N1, N2, N3)	5	Split-plot	Split-plot ANOVA	H	Y, HI, GN, SS, SA	S, I	TGW, GA	PAR, R:FR	9.0
Lleida	Spain	41°37' 50" N, 0° 35" E	Loam	25% clay 43% silt 32% sand	40	300	1.26 x 3.5m	2013 & 2014	Fully irrigated	3/year	RCBD	Two-way ANOVA	H	Y, HI, GN, GS, SA	-	TGW	-	7.0

Table 3.1: Experimental details of sites used for comparing Rht8 and tall NILs and traits measured. Y=yield, H=height, S=spike length, I=internode lengths, TGW=Thousand grain weight, GN=grain m^{-2}, GS=grains $spike^{-1}$, HI=harvest index, SA=spikes m^{-2} ,SS=spikelets $spike^{-1}$, GA=grain area; also length, width), HD=heading, AN=anthesis, GC=Ground cover, SEN=senescence, PAR=photosynthetically active radiation, R:FR=red:far red ratio.

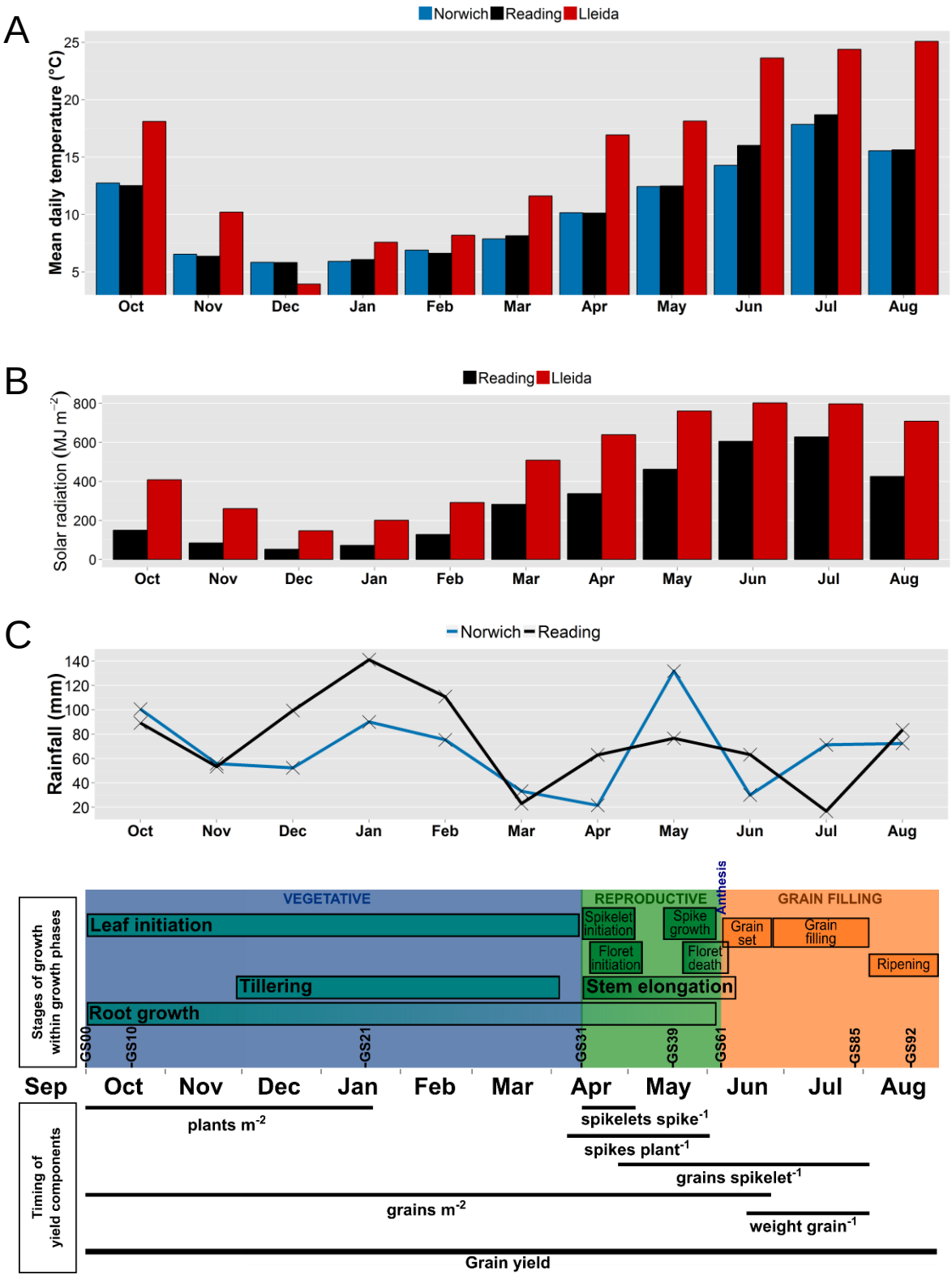


Figure 3.2: Monthly weather data at experimental sites over the 2013-2014 growing season with growth stages and timing on yield components. (A) Mean daily temperature each month based on the mean of daily maximum and minimum temperatures (B) Total solar radiation measured each month based on the total mean daily levels (from maximum and minimum levels) (no data for Church Farm was available) (C) Total rainfall received each month (Lleida experiments were irrigated to field capacity so comparison not pertinent). The growth phases were adapted from HGCA, 2008. Timing of yield components was adapted from Slafer, 2012.

3.3 Plant height and height components

In order to assess the extent of height reduction conferred by *Rht8*, total plant height (PH) (Figure 3.3A) was taken upon maturity and internode components were measured from tiller samples (Figure 3.3B).

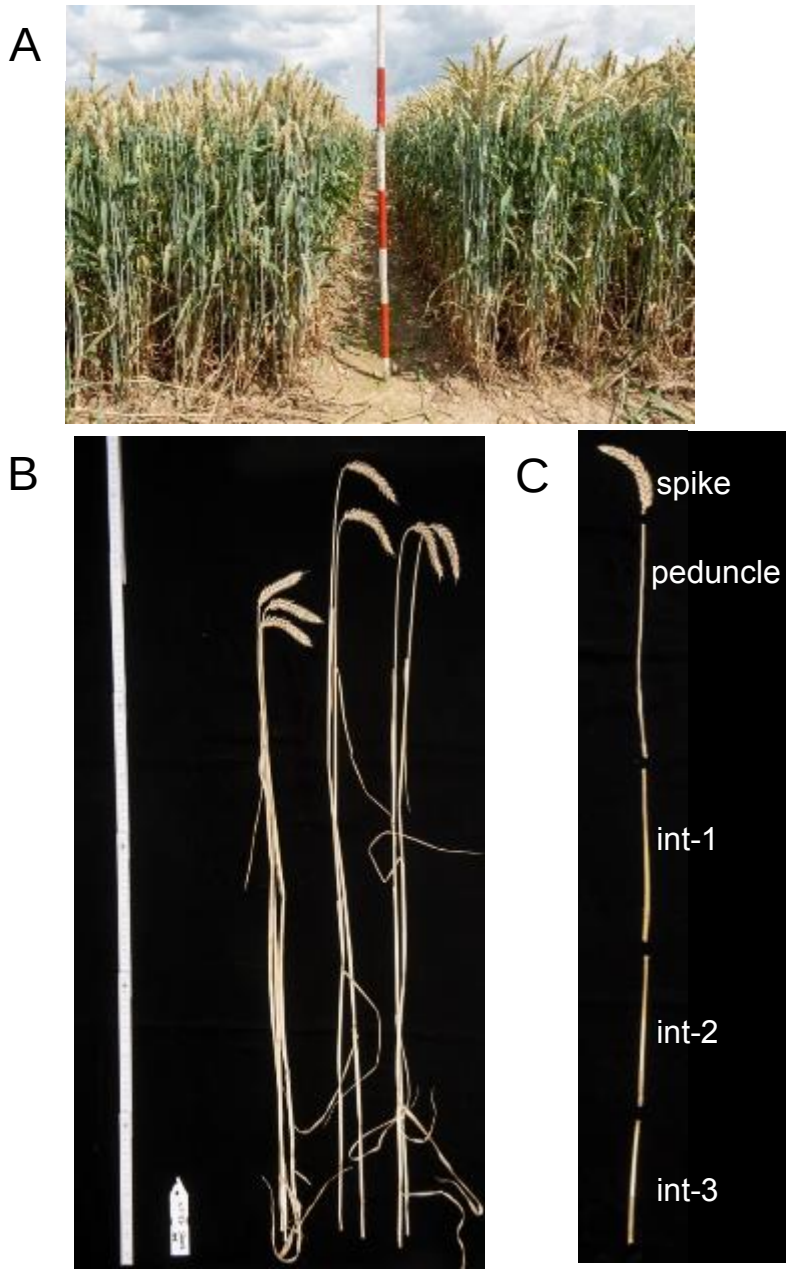


Figure 3.3: Measuring total height and height components. (A) Stature at maturity was measured in the field (left = Rht8, right = Paragon control) (B) Tiller samples taken from field plots (Rht8, tall, Paragon, left to right) (C) internode components measured from tiller samples.

Comparison of the *Rht8* NIL, tall NIL and Paragon showed that there was a highly significant difference between PH in all seven environments ($P<0.05$ in 2012, $P<0.001$ in all other environments) (Table 3.2). PH of the *Rht8* NIL ranged from 88 – 115 cm (excluding plot to bulk seed in 2011) and was on average 11% shorter than the tall NIL across environments. The height reductions were

relatively consistent across environments (Table 3.2), with the exception of the two seasons in Lleida which were the most variable within a single location, with the highest and lowest percentage difference relative to the tall NIL reported (-1 to -22%). The tall NIL was taller than Paragon in seven of the environments, with the exception of Lleida in 2014 (Table 3.2 and Figure 3.4). This indicated that there was some genetic background, distinct from the *Rht8* locus since this had been genotyped during population development (described in Chapter 2), which was making the tall NIL taller than Paragon, and, speculatively, the *Rht8* NIL taller than it might otherwise be. Interestingly, this background effect had not been prominent in the field in 2011 and 2012 (described in Chapter 2), where all the tall NILs developed were of the same height as Paragon, implying a G x E interaction. In order to mitigate these background effects, trait responses of the *Rht8* NIL were considered relative (in percentage terms) to the tall NIL.

Correlation analysis across environments between PH and height components revealed that PH was very positively and significantly correlated mostly with the length of the spike and peduncle (r = 0.63, 0.47; P<0.01) and also top internode (r = 0.35, P<0.05) (Table 3.3). There were no significant correlations observed between PH and the lower internodes. This was in contrast to the tall NIL and Paragon, where only the bottom two internodes showed a positive and significant correlation with PH.

Using analysis of variance (ANOVA), the effect of N and irrigation treatment on PH was measured (Figure 3.5 and Figure 3.6). Increased N treatment had a significant effect in increasing PH (P=0.004 Church Farm, P<0.001 Reading). Irrigation also increased PH, but only significantly differently to unirrigated (rainfed) treatment in 2014 (P=0.03). In 2013, where the irrigation was applied relatively late, there was no significant increase in height due to treatment (P=0.096). Crucially, there was no G x E interaction in any of the environments. This means that genotypes were affected in the same way by all treatments considered. In other words, the height reduction conferred by *Rht8* was maintained regardless of N or irrigation treatment.

Taken together, the data indicated that *Rht8* conferred a stable and significant height reduction of ~11% relative to the tall NIL, across environments of varying yield potential and climatic conditions. The magnitude of the height difference

between the *Rht8* and *tall* NIL was proportionately unaffected by N or irrigation treatment and the total height reduction was principally contributed by the spike, peduncle and top internode.

3.3.1 Genotyping NILs

In order to assess the genotypic differences between the NILs, the *Rht8* NIL selected from BC_3F_5 was genotyped along with Paragon using the 820,000 (820K) feature Axiom® SNP array (www.cerealsdb.uk.net/cerealgenomics) (described in 2.3.8). A total of 6088 SNPs were found between the *Rht8* NIL and Paragon (discarding SNP calls due to missing data in one of the genotypes), indicating a 99.4% Paragon background, well above the expected theoretical 87.5% at BC_3 (population development described in Chapter 2). Within the identified SNPs, 2% were located on 2DS. However, since the mapping data for the markers had not been released at the time of writing, the precise background contribution could not be further assessed.

	Church Farm									Reading			Lleida	
	2011	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)	2013	2014
par	68.0	93.4	107.7	103.7	106.8	108.7	111.2	115.4	102.8	97.5	104.9	108.2	111.0	119.0
Rht8	59.5	87.8	98.3	94.3	97.3	100.7	102.8	104.2	96.5	93.0	100.7	103.9	93.3	114.7
tall	68.5	99.0	113.7	107.3	112.7	114.0	115.6	113.7	110.5	105.1	109.7	112.2	119.0	115.3
P-value	-	*	***		***		***		***		***		***	
L.S.D.	-	7.6	3.2		3.3		2.2		5.0		2.9		7.1	
Rht8 (% of tall)	87	89	87	88	86	88	89	92	87	88	92	93	78	99
difference (%)	**-13**	**-11**	**-13**	**-12**	**-14**	**-12**	**-11**	**-8**	**-13**	**-12**	**-8**	**-7**	**-22**	**-1**

lowest highest

*Table 3.2: Total plant height at maturity. Data shown as mean values. The p-value refers to significant differences in height between genotypes within each experiment determined by the least significant difference (L.S.D.) test. 2011 data was based on one replicate. *P<0.05, **P<0.01, ***P<0.001. N1=40kg N ha^{-1}, N2=100kg N ha^{-1}, N3=200kg N ha^{-1}.*

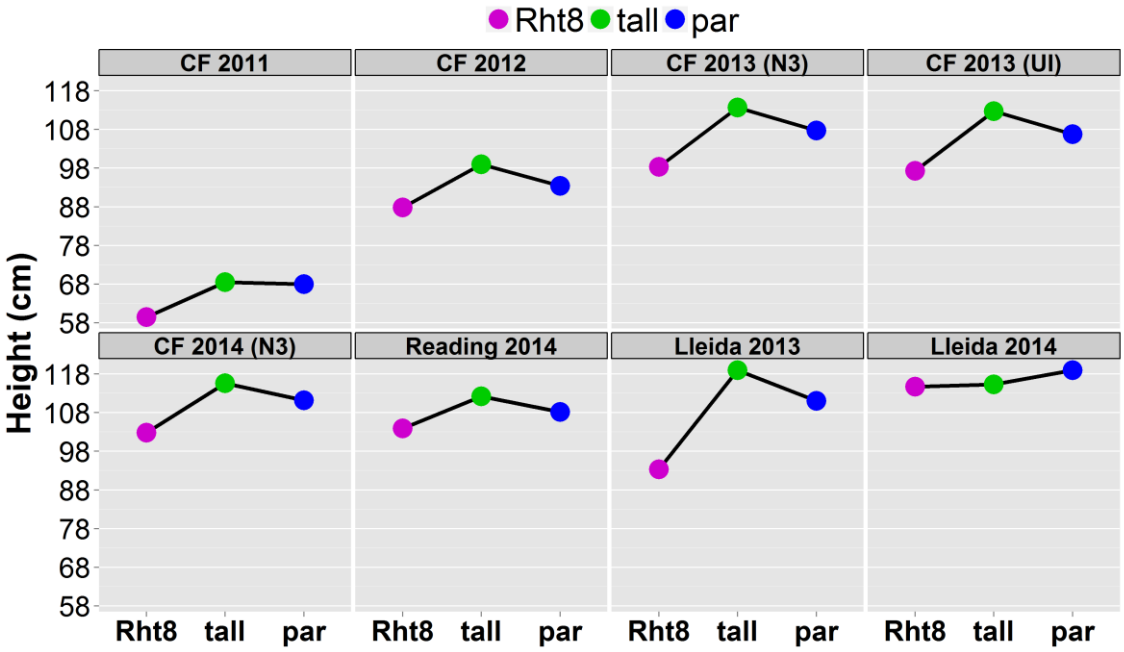


Figure 3.4: Mean plant heights of the Rht8 NIL, tall NIL and Paragon at standard agronomic conditions. Nitrogen treatment is indicated only where alternative treatments to standard agronomic conditions were present. N3=200kg N ha^{-1}, UI=unirrigated (rainfed) (indicated only where there was a contrasting irrigation regime).

	Rht8		par		tall	
	r	p-val	r	p-val	r	p-val
spike	0.63	***	0.17	NS	0.26	NS
peduncle	0.47	***	0.19	NS	0.00	NS
int-1	0.35	*	0.03	NS	0.50	***
int-2	0.23	NS	0.87	***	0.85	***
int-3	0.27	NS	0.70	***	0.81	***

lowest highest

*Table 3.3: Simple correlation coefficients (r) between total plant height and height-related traits from tiller samples, across all environments. NS=not significant at P<0.05, *P<0.05, **P<0.01, ***P<0.001.*

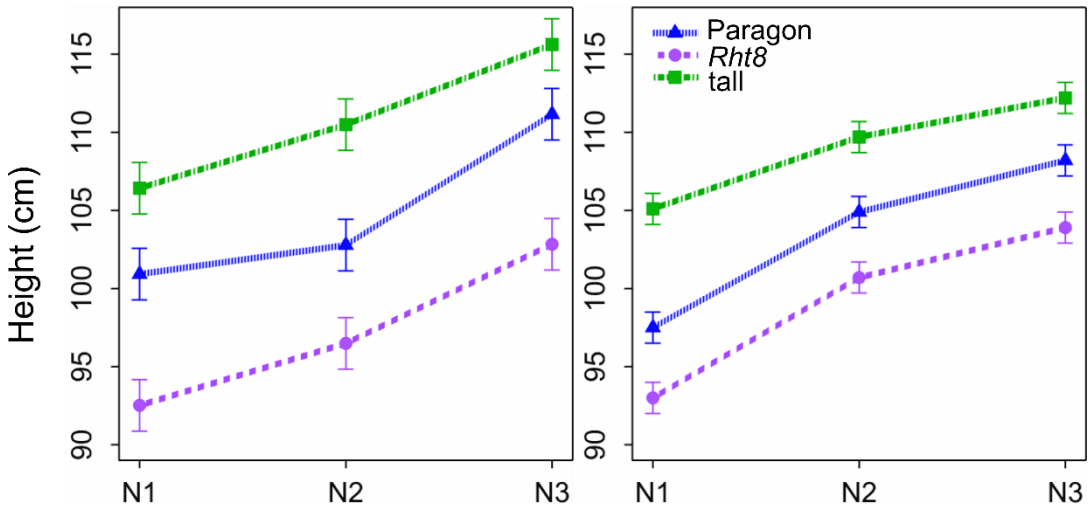


Figure 3.5: Mean plant height at different N treatments in Church Farm (left) and Reading (right). Data from 2013-2014 season. Error bars represent standard error. N1=40kg N ha^{-1}, N2=100kg N ha^{-1}, N3=200kg N ha^{-1}.

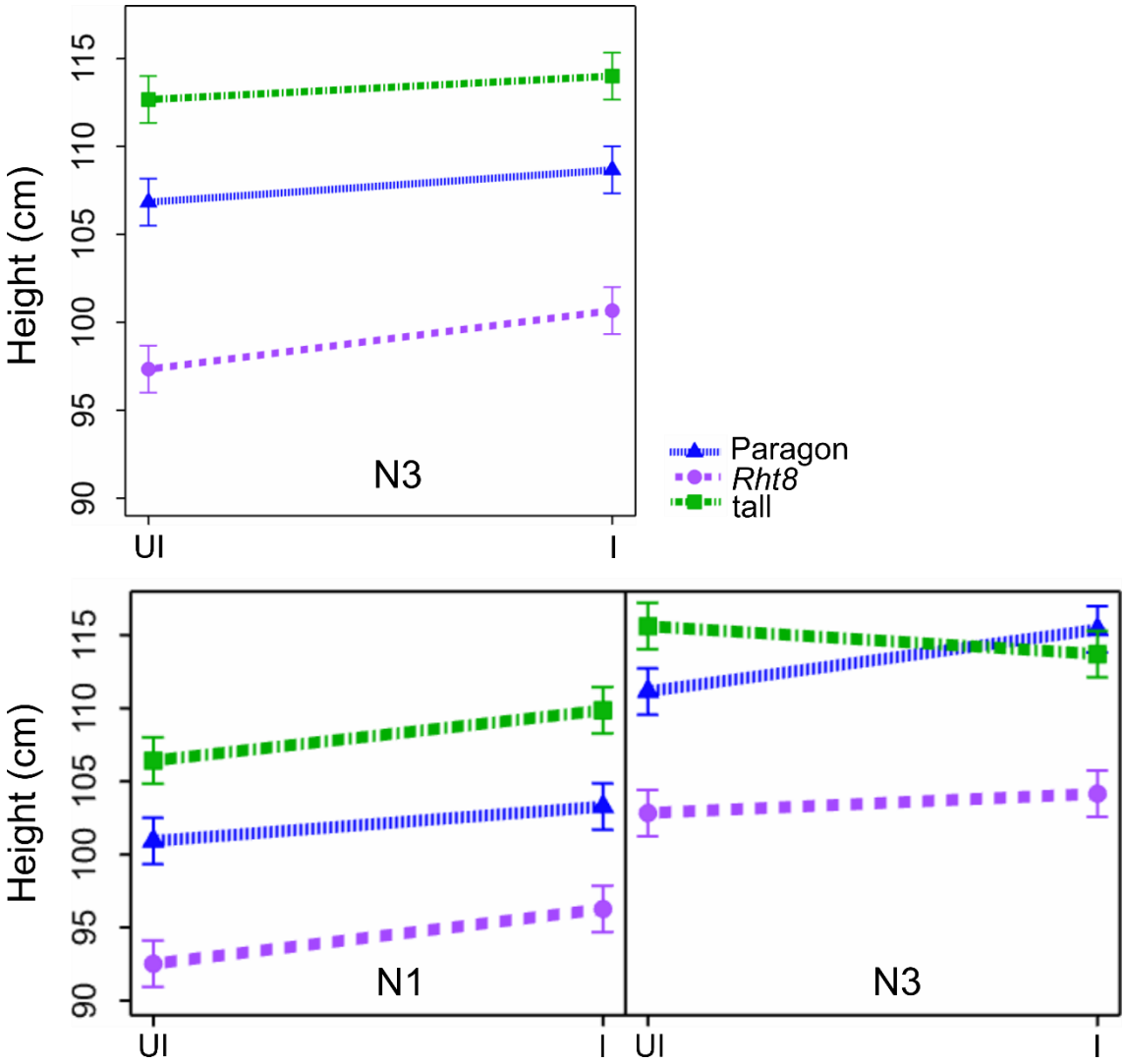


Figure 3.6: Mean plant height in irrigated (I) and unirrigated (UI; rainfed) conditions at Church Farm. Data from 2012-2013 season (top) and 2013-2014 in contrasting N treatments (bottom). Error bars represent standard error. N1=40kg N ha^{-1}, N3=200kg N ha^{-1}.

3.4 Grain yield and yield components

Yield is a complex polygenic trait, determined by genetic and environmental interactions throughout the growing season of wheat (Figure 3.1). To facilitate studying this complex trait, yield is dissected into yield components (Figure 3.7). At the highest level, wheat yield is a product of grains per unit area (GN) and grain weight (TGW). Generally, increases in yield have been achieved by increased GN (Peltonen-Sainio et al., 2007). There is a trade-off between the increase in GN and reduction in TGW (Acreche and Slafer, 2006). However, some exceptions to this exist, with high GN and high TGW (Griffiths et al., 2015). For this reason, increasing both GN and TGW to maximise yield has been proposed as a target 'ideotype' (Ma et al., 2015b). TGW can be broken down into grain length (GL) and grain width (GW), with GL believed to be the key component and most responsive (Gegas et al., 2010). In order to study the contribution of *Rht8* on yield as fully as possible, key yield components were also measured (shaded in Figure 3.7).

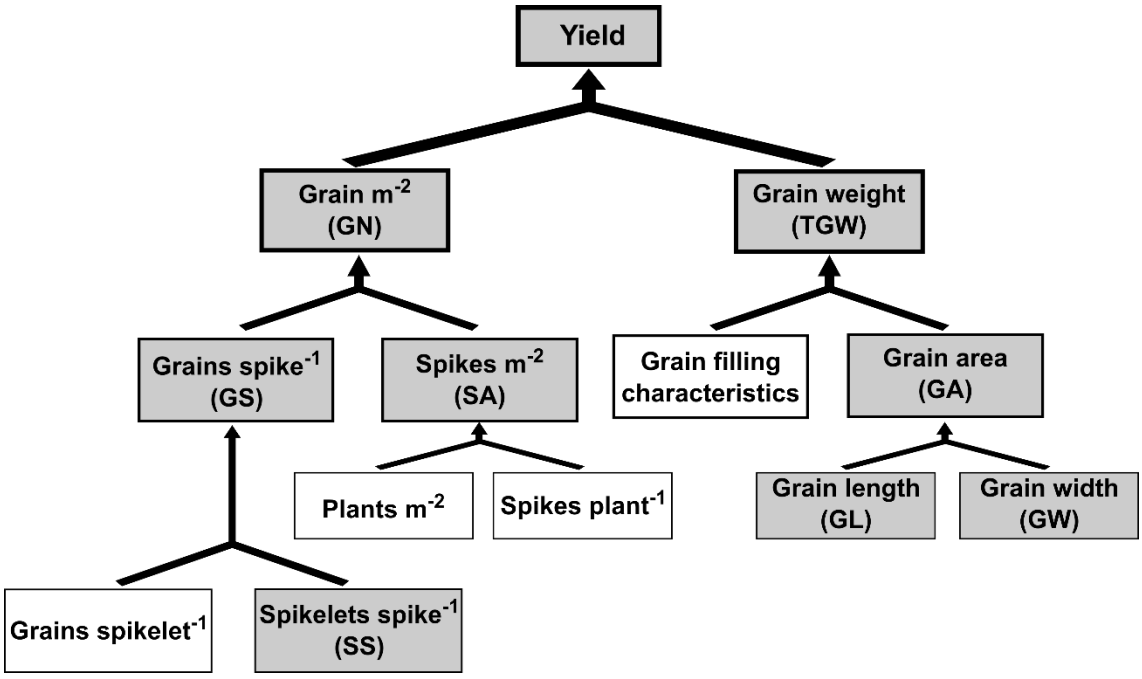


Figure 3.7: Diagram of wheat yield illustrating the contribution of yield components commonly measured in agronomic trials. Components measured in this study are highlighted in grey. Abbreviations are shown in brackets and correspond to the description in Table 3.1. Adapted from Slafer, 2003.

Grain yield varied significantly between the NILs based on means over six of the seven environments (full data in Appendix 3.1). Within the environments, in standard agronomic conditions (rainfed and high N (N3)), *Rht8* conferred a yield penalty of a mean -8%, ranging from -4 to -26% (Figure 3.8A). Exceptions to the yield penalty within the environments were at non-standard agronomic conditions (Figure 3.8B), at the lowest N treatment (N1) and in irrigated conditions. At these treatments, the yield of the *Rht8* NIL was either comparable to or had a higher mean to that of the tall NIL. The yield penalty observed in the UK standard agronomic conditions was abolished in the high-temperature site in Lleida. There was a borderline non-significant difference (P=0.07) in yield between the NILs, with the mean yield of the *Rht8* NIL 10 and 16% higher than the tall NIL in 2013 and 2014, respectively (Figure 3.8C). The highest overall yields were observed in 2013 in Church Farm under irrigated conditions (~12 t ha^{-1}), whereas the lowest observed were in the lowest N treatment in Reading (~5.5 t ha^{-1}) (Appendix 3.1).

GN correlated most strongly to yield out of the components measured (Table 3.4). There were highly positive (r = 0.85-1) and significant interactions between yield and GN across all environments. There were also significant (P<0.05 and P<0.01) differences in GN between the NILs in most of the environments (full table in Appendix 3.1). The negative impact of *Rht8* on yield was closely mirrored by a concomitant decrease in GN (Figure 3.8A), averaging -7% in standard agronomic conditions. Where the yield penalty was abolished, in Lleida and at low N levels, the decrease in GN was also eliminated (Figure 3.8B&C). The difference in GN between the NILs was not significant in Lleida (P=0.1) with the mean GN of the *Rht8* NIL 2% (2013) and 10% (2014) higher than the tall NIL.

Since GN is a product of grains $spike^{-1}$ (GS) and spikes m^{-2} (SA) (Figure 3.7), these sub-components were investigated. There was limited data for these components: GS was only measured in Lleida and SA data was obtained in Lleida and Reading. SA had a positive correlation with yield (Table 3.4) in Reading (r = 0.53) and Lleida (r = 0.43) but the correlation was only significant in Reading (P<0.05). There were no significant differences between the NILs observed in GS in Lleida (P=0.6). There was a highly significant (P<0.001) reduction in SA in Reading (a mean of -15%) across all N treatments. In Lleida, this reduction was reversed. There was no longer a significant difference in SA between NILs in

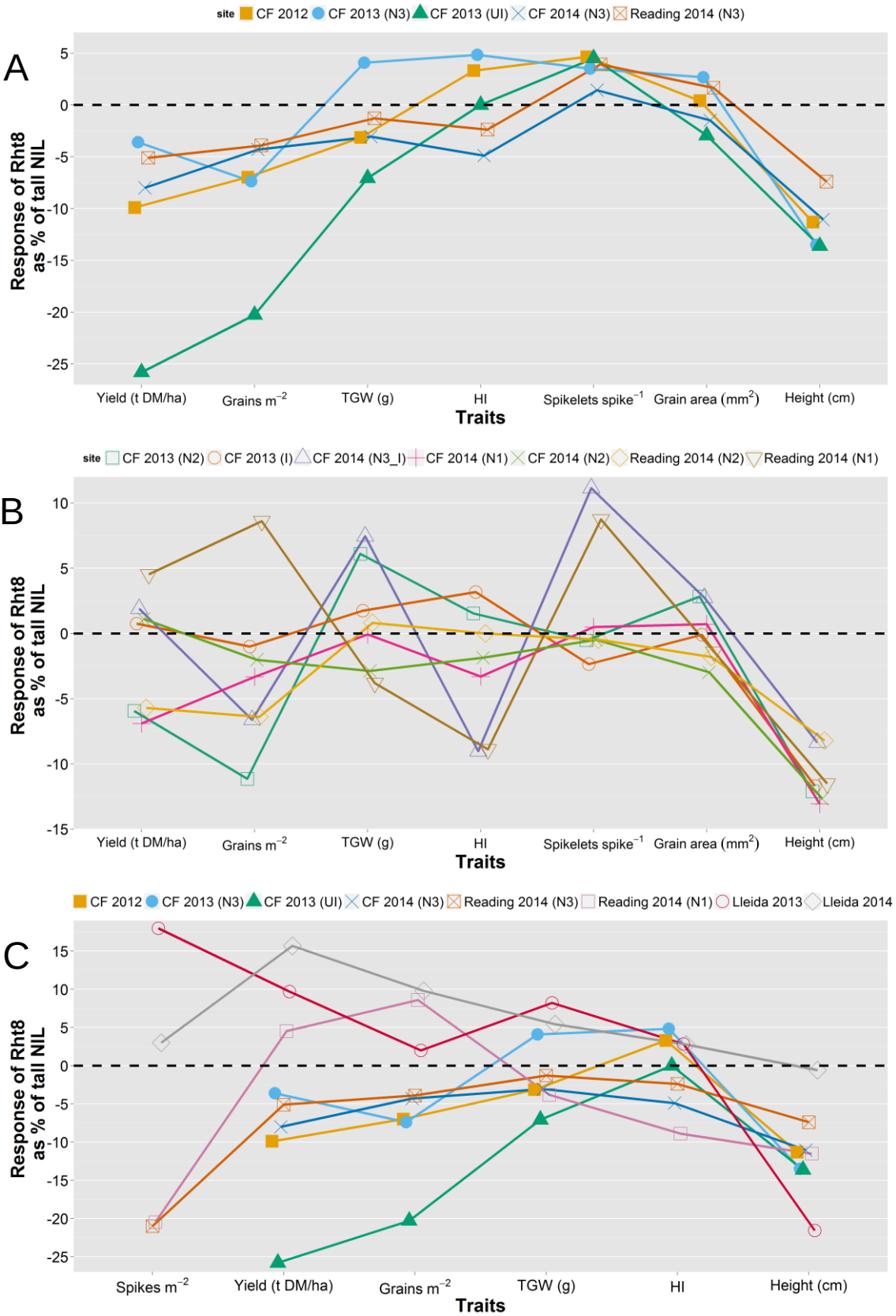


Figure 3.8: Yield and height trait responses of the Rht8 NIL relative to tall NIL. (A) In standard agronomic conditions in Church Farm and Reading, unirrigated (UI; rainfed) and N3=200kg N ha^{-1} (B) in non-standard conditions in Church Farm and Reading (irrigated and N1=40kg N ha^{-1}, N2=100kg N ha^{-1}) (C) in standard agronomic conditions at all three sites (reduced dataset to show comparison of traits measured across all three sites, with the exception of the additional trait of spikes m^{-2} which was only measured in Reading and Lleida).

Yield/Grains m^{-2}		Rht8		tall		par	
		r	p-val	r	p-val	r	p-val
Church Farm	2013 nitrogen	1.00	***	0.92	**	0.97	**
	2014 nitrogen	0.95	***	0.96	***	0.92	***
	2013 irrigation	0.95	***	0.96	***	0.92	***
	2014 irrigation	0.91	***	0.86	***	0.93	***
Reading	2014	0.99	***	1.00	***	0.99	***
Lleida	2013 & 2014	0.85	*	0.92	**	0.83	*

Yield / TGW		Rht8		tall		par	
		r	p-val	r	p-val	r	p-val
Church Farm	2013 nitrogen	-0.98	***	0.29	NS	-0.52	NS
	2014 nitrogen	0.83	**	0.81	**	0.59	NS
	2013 irrigation	0.95	**	0.83	*	0.99	***
	2014 irrigation	0.09	NS	0.12	NS	0.21	NS
Reading	2014	0.22	NS	0.28	NS	0.02	NS
Lleida	2013 & 2014	-0.12	NS	-0.11	NS	-0.24	NS

Yield/Spikes m^{-2}		Rht8		tall		par	
		r	p-val	r	p-val	r	p-val
Reading	2014	0.53	*	0.61	**	0.43	NS
Lleida	2013 & 2014	0.43	NS	0.51	NS	0.25	NS

Grains $spike^{-1}$/Spikes m^{-2}		Rht8		tall		par	
		r	p-val	r	p-val	r	p-val
Lleida	2013 & 2014	-0.58	NS	-0.53	NS	-0.88	*

*Table 3.4: Simple correlation coefficients (r) between yield and yield components across environments. NS=not significant at P<0.05, *P<0.05, **P<0.01, ***P<0.001.*

Lleida, with the mean of the *Rht8* NIL showing an 11% increase across both years (Figure 3.8C).

Other yield trait responses of *Rht8* relative to the tall NIL were smaller (Figure 3.8). TGW did not correlate stably and strongly with yield to the extent of GN, having a significant positive and negative correlation with yield in *Rht8* in consecutive years at Church Farm (Table 3.4), but no other significant correlations in other environments. There were significant differences between the NILs in TGW in three out of seven environments (data in Appendix 3.1). In 2013 at Church Farm, the *Rht8* NIL had a significant (P<0.05) mean 5% increase in TGW across the N treatments compared to the tall NIL. In Lleida, there was a significant (P<0.05) increase of 7%. In the Reading environment, there was a significant (P<0.05, L.S.D) decrease of 4% but only at the lowest N treatment.

Responses in Grain area (GA) only varied significantly between NILs (P<0.01) in the 2013 Nitrogen trial at Church Farm (Figure 3.8A; full data in Appendix 3.1). The *Rht8* NIL had a 3% increase in GA in both N treatments. In the same trial, Grain Length (GL) and Grain Width (GW) also increased significantly (P<0.001 GL; P<0.05 GW) by 1 and 3%, respectively. Thus the increase in GA was due to an approximately concomitant increase in both grain dimensions.

The harvest index (HI) was significantly (P<0.05) different in the *Rht8* NIL compared to the tall NIL in two environments. In one year at Church Farm (2013), the *Rht8* NIL had a 5% increase in HI, but only at N3 (LSD test). However, in the subsequent year, *Rht8* conferred a reduction in the mean HI relative to tall NIL 7%. Thus this trait did not have a robust, extensive response.

There was no difference in the number of spikelets spike^{-1} (SS) between the NILs in any of the environments (P>0.05).

3.5 Yield response to irrigation, contrasting N and high temperature

In order to determine if *Rht8* conferred any adaptation at lower input conditions, the NILs were grown in trials with contrasting N treatments and water regimes. The trials in Lleida were fully irrigated, and this provided opportunity to observe increased adaptation to high temperature, which is notoriously difficult to dissect from drought stress. The temperatures in the Lleida growing season were high relative to the UK (Figure 3.2), but were below the 27° – 30°C at anthesis range which has been used to define 'heat stress' (Semenov et al., 2014).

Soils in arable rotations in typical agronomic conditions supply enough N for wheat to fulfil approximately half its yield potential. The remaining yield potential can be realised with applied N fertiliser (HGCA, 2008). The NILs were grown in three contrasting N fertiliser regimes. At the standard agronomic treatment (200 kg N ha^{-1}), the *Rht8* NIL had a yield penalty in UK sites. At lower N treatments, the yield penalty was abolished (Figure 3.9A). A higher resolution of the effects of N was available in Reading, since the split-plot experimental design allowed the genotype to be analysed as a sub-plot effect nested within N as a main-plot

factor. In the Reading data, the yield penalty was abolished at the lowest N input only (N1), whereas at Church Farm, the penalty was not observed at N1 or N2 (Figure 3.9A). Dwarfing alleles in wheat have been shown to affect Nitrogen Use Efficiency (NUE) (Gooding et al., 2012). There is a complex relationship between the effect of different soil N levels and the components of NUE, nitrogen uptake (NUpE) and utilisation efficiency (NUtE) (Ortiz-Monasterio, 2012). NUpE is the major contributor to NUE at low N and NUtE at higher N (Hawkesford, 2014). The abolishment of the yield penalty at standard N levels in the high-temperature conditions in Lleida (Figure 3.8C) suggested that at increased temperatures, *Rht8* conferred adaptive advantage which was not due to 'escape' by earlier heading or flowering (personal communication). Additionally, the *Rht8* NIL did not show reduced TGW or reduced GN, which has been reported in wheat under heat stress (over 30°C) (Semenov et al., 2014). Taken together, the results here offer a preliminary indicator that *Rht8* has a penalty in NUE at higher N levels (typical of standard N inputs), but this disadvantage is overcome at higher temperatures, and even errs toward an almost significant yield advantage. At lower N, the yield penalty is overcome, speculatively due to improved NUpE in *Rht8* at levels where soil N is much lower than standard agronomic fertiliser input levels.

Irrigation treatments were conducted in 2013 and 2014 growing seasons at Church Farm. At 200 kg N ha^{-1} (N3), the yield penalty of *Rht8* observed in rainfed conditions was abolished by providing irrigation, since there was no difference in yield between NILs. This result was observed across both years (Figure 3.9B). In 2014, at N1, irrigation increased yield across all genotypes, but in the same proportion, so that the yield penalty in the *Rht8* NIL was maintained (Figure 3.9B).

Although the UK growing climate does not subject wheat to the same drought stress as southern Europe, the different timings of the irrigation in consecutive years offer some contrast in temporal application of drought stress. In 2013, irrigation at Church Farm was supplied after GS61, when stem extension was complete (Figure 3.2), whereas in 2014, irrigation was supplied throughout the reproductive phase. The timing of irrigation did not greatly affect the yield of the *Rht8* NIL at N3 relative to tall NIL (1% and 2% increase relative to the tall NIL in 2013 and 2014, respectively). Though speculative, this reflects similar findings in the agronomic performance of *Rht-B1* and *Rht-D1* genes. In water stressed

environments, the yield seemed more closely related with the right plant height than with the combination of dwarfing alleles (Butler et al., 2005).

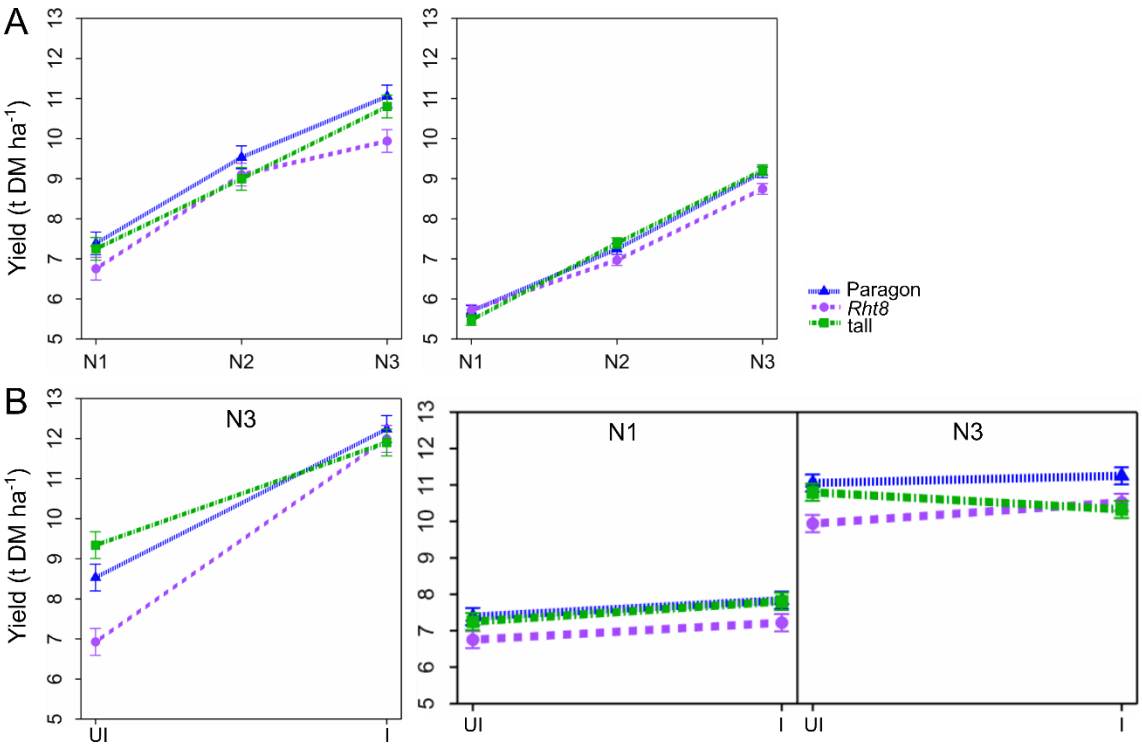


Figure 3.9: Yield of the Rht8 NIL, tall NIL and Paragon in contrasting N treatments and irrigation regimes. (A) 2013-2014 season at Church Farm (left) and Reading (right) (B) Church Farm in 2012-2013 (left) and 2013-2014 (right). N1=40kg N ha^{-1}, N2=100kg N ha^{-1}, N3=200kg N ha^{-1}, I=irrigated, UI=unirrigated (rainfed).

3.6 Interplay between yield, grains m^{-2} and spikes m^{-2}

The interplay between yield, GN and SA is shown in Figure 3.10. As already described in 3.4, yield was tightly correlated with GN. In the temperate UK climate, the yield penalty observed at N2 and N3 was abolished at N1. This was reflected in the GN, which was reduced at N2 and N3 relative to the tall NIL, but no differences between NILs were observed at N1. SA was reduced at all levels of N, which was presumably offset by increased grain spike^{-1} at N1, the other component of GN (Figure 3.7), though this data was not collected. At the high-temperature site, the *Rht8* NIL had no penalty and instead had a borderline significant (P=0.07) increase in yield. GN tightly mirrored this: there was no reduction in GN and instead higher GN in the *Rht8* NIL compared to the tall NIL

(Figure 3.10). There was also no significant reduction in SA in the *Rht8* NIL which had been observed under conditions with the yield penalty.

At the highest N input (N3), GN was highest in the high-temperature site (ranging from 18,000 – 22,000 grains m^{-2} across NILs in Lleida) compared to the UK site (18,000 – 19,000 grains m^{-2}). SA was also increased in Lleida, ranging from 564 – 627 spikes m^{-2} compared to 383 – 487 spikes m^{-2} in Reading (data in Appendix 3.1). This was unsurprising, given that SA is determined by tiller production which was likely limited by the solar radiation received during the vegetative phase in the UK but not in Lleida (Figure 3.2).

In sum, the data suggested that in UK field conditions, the yield penalty at N2 and N3 was due to decreased GN and not TGW. This is in accordance with studies which have shown that in temperate conditions (such as the UK), with an absence of stress during grain fill, GN is the dominant component influencing yield (Peltonen-Sainio et al., 2007). Moreover, it has been shown that introgression of semi-dwarfing genes increases juvenile spike formation which enhances the responsiveness of GN (Miralles et al., 1998). It seems likely that a combination of these effects was acting in the *Rht8* NIL in the UK climate, making GN dominate in determining yield.

The reduction in GN observed in the UK data was in turn due to decreased SA. Conversely, in Lleida, where no yield penalty was observed, there was also no reduction in either GN or SA. It is interesting to consider the hierarchy of influence of yield components reported here with previous findings examining relationships between yield components and environmental modulation of yield responses (Slafer et al., 2014). In that study, a large database of wheat yield components from published literature was examined. A 'hierarchy of plasticities' was reported, where within the GN components, SA was more dominant in determining GN than GS, particularly when driven by environmental factors. This was in part related to the investment required to produce a tiller compared to a floret primordium. Furthermore, there were no trade-offs reported between GS and SA where there was a large change in yield (>50%) due to environmental factors, but a strong trade-off was present for large changes in GN driven by genetic factors (Slafer et al., 2014). Since there was no GS data for Reading, these findings could not be unambiguously verified. However, GN did appear to be driven by SA for the

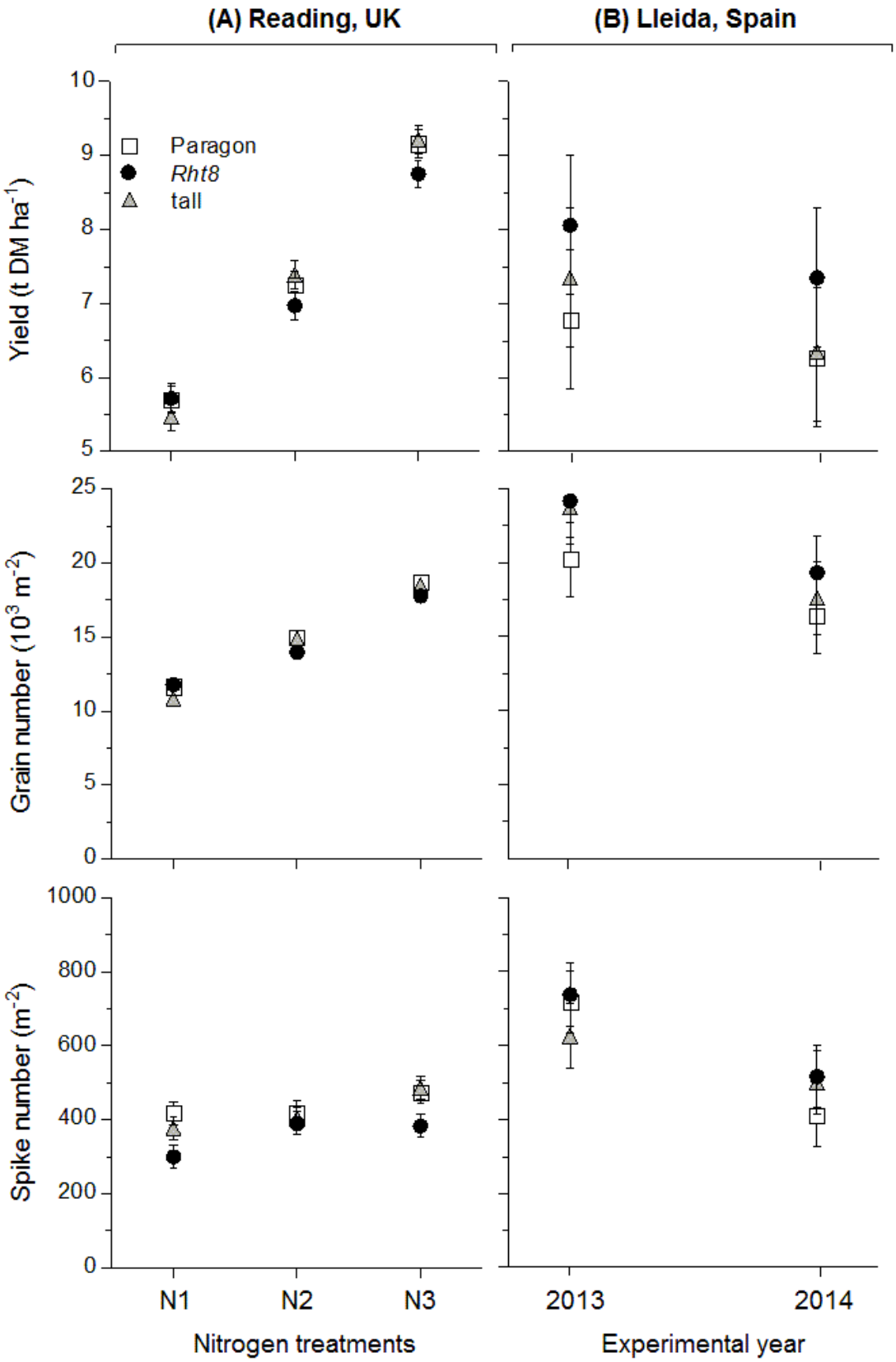


Figure 3.10: Yield, grains m^{-2} and spikes m^{-2} (top to bottom) in the Rht8 NIL, tall NIL and Paragon in Reading (column A) and Lleida (column B). Error bars represent standard error. N1=40kg N ha^{-1}, N2=100kg N ha^{-1}, N3=200kg N ha^{-1}.

available data. Interestingly, in Lleida, there were strong negative correlations between SA and GS (indicative of a trade-off), but these were only significant in one of the genotypes (Paragon) (Table 3.4). From this, it can be reasoned the genetic (*Rht8*) factors were stronger than environmental, but not overwhelmingly so as described in cases of a 50% yield differential by Slafer et al. 2014. In light of this, it can be recommended that further investigations into the agronomic performance of *Rht8* in different treatments should be moderate rather than extreme (which would tip the balance into yield relationships driven by environmental rather than genetic contributions).

3.7 Lodging

Lodging in cereals is the permanent displacement of the culms from the upright vertical position. Lodging limits yield particularly in high-yield potential environments, such as western Europe. In the UK, severe lodging occurs in UK cereal crops every three-four years, when 15-20% of the wheat growing area lodges (Berry et al., 2004). Lodging can reduce grain yield by up to 50% (Fischer and Stapper, 1987). In wheat, there are two types of lodging: stem lodging, caused by buckling of the stem, and root lodging, caused by over-turning of the anchorage system. In the 2013-2014 growing season at Church Farm, a storm with high winds caused lodging in early July, at approximately GS70. Root lodging was the only type observed (no buckling of stems) and lodging only affected plots at higher N levels (N2 and N3) (Figure 3.11).

Lodging has been found to decrease both GN and TGW (Acreche and Slafer, 2011). Correlation analysis was performed in order to assess which yield and height components were linked to increased lodging across all genotypes. Yield was highly positively and significantly correlated with lodging (r = 0.75, P<0.001), as was GN (r = 0.70, P<0.001). There were no significant correlations between lodging and TGW or harvest index. Lodging was also significantly positively correlated with overall height (r = 0.73, P<0.001) and some height components (Table 3.5). The lack of negative correlation between lodging and yield might be due to the relatively moderate lodging found here: in the work by Acreche and Slafer 2011, lodging was artificial and to 80° displacement from the vertical, whereas the lodging here was to an average 45° displacement.

Figure 3.11: Lodging at Church Farm in July 2014. Lodging was severe and caused 90° displacement of culms from the vertical position in a few plots (left). The predominant type of lodging observed was root lodging (right).

		r	p-val
	Yield	0.75	***
	Grains m^{-2}	0.70	***
	TGW	0.10	NS
	Harvest Index	-0.07	NS
	Height	0.73	***
ratios	Spike	0.16	NS
ratios	Peduncle	0.27	NS
ratios	Int-1	0.53	*
ratios	Int-2	0.26	NS
ratios	Int-3	0.47	*

lowest highest

*Table 3.5: Simple correlation coefficients (r) between lodging and yield and height components. NS=not significant at P<0.05, *P<0.05, ***P<0.001.*

It should also be noted that stems were re-erected by leaning to prepare for harvest prior to GS75, at which stage the greatest yield losses have been reported (Berry and Spink, 2012).

Interestingly, unlike the GA-insensitive *Rht-B1* or *Rht-D1* genes, no *Xgwm261* allele linked with the *Rht8* locus has been found to have a significant effect on lodging resistance (Šíp et al., 2009). In order to further explore this, lodging score between the NILs was analysed by ANOVA. There was a significant difference in lodging between the NILs (P<0.001), with the *Rht8* NIL having half the mean lodging score of the tall NIL at N3 (38% vs 74%, 0%=no lodging) (Figure 3.12). There was a significant N*allele interaction (P<0.01), which was expected since at N2, the *Rht8* NIL was completely resistant to lodging. The irrigation treatment at both N1 and N3 had no effect on the lodging score (P=0.4). In sum, the findings here indicate for the first time that *Rht8* confers lodging resistance at agronomically-relevant N treatments.

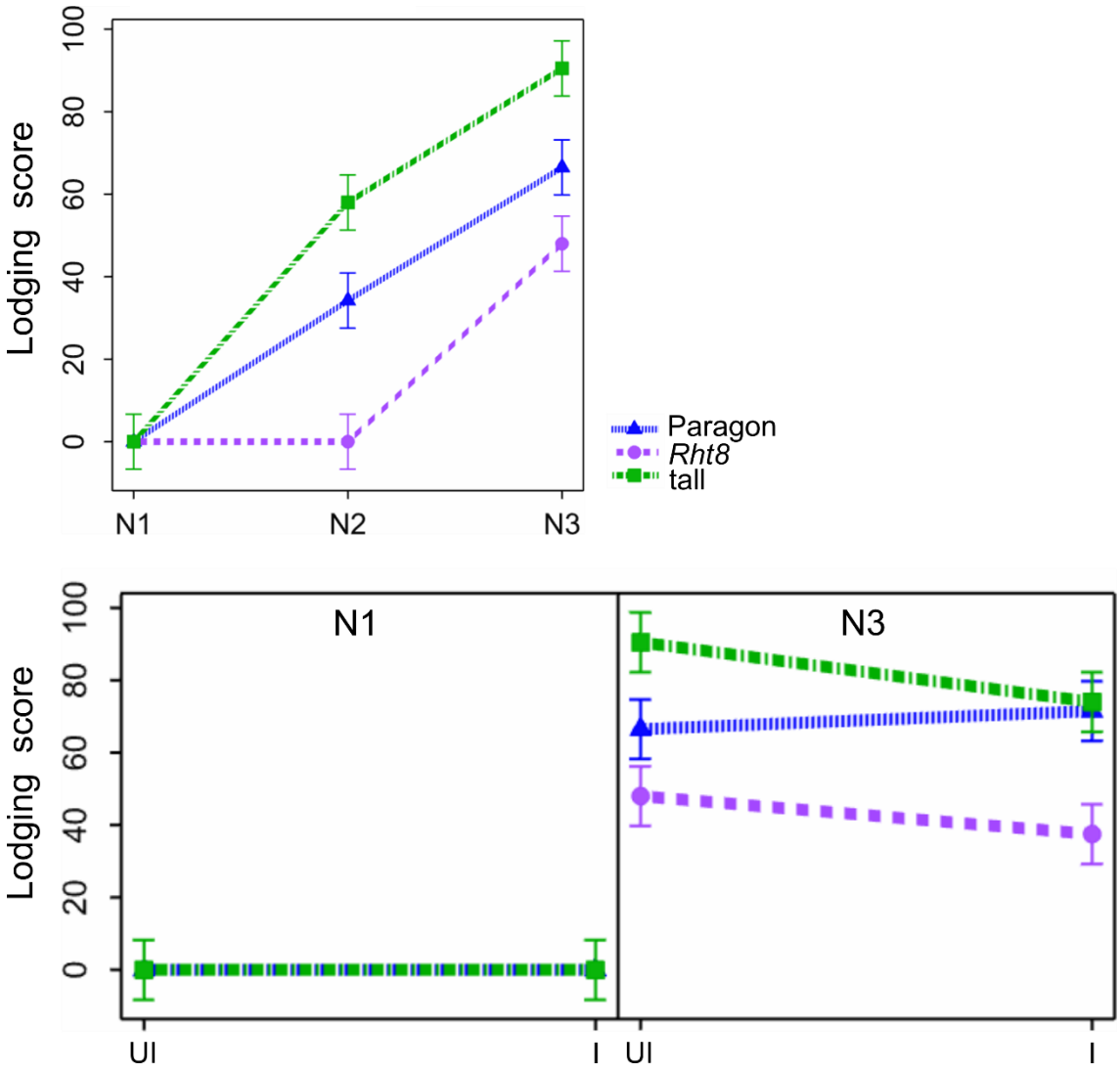


Figure 3.12: Lodging score of the Rht8 NIL, tall NIL and Paragon in contrasting N treatments and irrigation regimes. Error bars represent standard error. N1=40kg N ha^{-1}, N2=100kg N ha^{-1}, N3=200kg N ha^{-1}, I=irrigated, UI=unirrigated (rainfed).

3.8 Developmental traits

The effects of *Rht8* on developmental traits has usually been reported in a *Ppd-D1a* background. Here, developmental traits could be analysed without the earliness conferred by photoperiod insensitivity. One study which did measure heading in a spring wheat background in Colorado found no difference between *Rht8* and wild-type (Lanning et al., 2012), though no other developmental traits were measured.

In work from Australian trials, *Rht8* has been proposed as a way to provide semi-dwarf stature with the benefits of early seedling vigour and a longer coleoptile, leading to improved emergence (Ellis et al., 2004). This was tested in UK conditions by measuring ground cover as a proxy to establishment. Ground cover was measured at both UK sites at the end of March 2014, towards the end of the vegetative phase, using images of plots from which proportion of green canopy was measured (using an ImageJ macro developed by Oscar Gonzalez) (Figure 3.13). There was no significant difference (P=0.8) between NILs in ground cover estimated using this method.

Figure 3.13: Ground cover estimated by calculating percentage of green canopy. Photo taken at waist height before (left) and the same image output from analysis (right).

Heading date was measured across two growing seasons, 2013 and 2014, at Church Farm. Contrasting N treatment had no effect in either year on heading (P>0.05) (Table 3.6A). Irrigation significantly (P<0.001) delayed heading in all NILs, by approximately 1 calendar day, but only in the 2014 season. Though not significant, the mean heading thermal time was reduced by 0.25% in the *Rht8* NIL in 2013 and 0.4% in 2014 across treatments when compared to the tall NIL (Table 3.6A). This equates to half a calendar day.

Anthesis was measured in 2014. Overall, different N treatment or water regime had no significant effect on anthesis (P>0.05). Significant (P<0.05, LSD) differences between the *Rht8* and tall NIL were observed in two of the four treatments (Table 3.6B). At N2 and in rainfed conditions the *Rht8* NIL had a mean delay of 1 – 1.4% in thermal time to anthesis compared to the tall NIL, which equates to 2 – 2.5 calendar days.

Heading (^{0}C d)

A

	2013				2014				
	N2	N3	UI	I	N1 UI	N1 I	N2	N3 UI	N3 I
par	1498	1471	1456	1458	1766	1777	1769	1766	1779
Rht8	1505	1493	1478	1465	1779	1782	1782	1779	1812
tall	1515	1488	1482	1471	1782	1800	1782	1782	1807
P-value allele	*		*		***				
P-value treatment	NS		NS		***(I)/NS(N)				
L.S.D.	13		14		13(I)/12(N)				

Rht8 (% of tall)	99.4	100.3	99.7	99.6	99.8	99.0	100.0	99.9	100.3
difference (%)	-0.6	0.3	-0.3	-0.4	-0.2	-1.0	0.0	-0.1	0.3

Anthesis (^{0}C d)

B

	2013			
	N2	N3	UI	I
par	1581	1564	1551	1547
Rht8	1592	1574	1547	1556
tall	1608	1579	1569	1566
P-value allele	*		*	
P-value treatment	NS		NS	
L.S.D.	13		14	

Rht8 (% of tall)	99.0	99.6	98.6	99.4
difference (%)	-1.0	-0.4	-1.4	-0.6

lowest highest

*Table 3.6: Heading and anthesis dates in 2013 at Church Farm shown as thermal time. Data shown as mean values. The p-value refers to significant differences between genotypes determined by the least significant difference (L.S.D.) test. NS=means not significantly different at P<0.05, *P<0.05, ***P<0.001. P-values are shown separately for N and I in the factorial experiment. N1=40kg N ha^{-1}, N2=100kg N ha^{-1}, N3=200kg N ha^{-1}, I=irrigated, UI=unirrigated (rainfed).*

Senescence at the plot level was measured visually using a 1–10 score in the 2013 season at Church Farm. Irrigation increased senescence and no differences were observed in contrasting N treatments (Figure 3.14). There was

high variation within genotype and no significant differences were observed between NILs (overlapping error bars), with the exception of more rapid early senescence under irrigation in the *Rht8* NIL. This difference disappeared by 2000°C d (Figure 3.14A).

In sum, there was no strong effect of *Rht8* on the developmental traits studied, which is in line with the common belief that the earliness from the *Ppd-D1* locus was dominant in previous studies where *Rht8* was linked with *Ppd-D1a*.

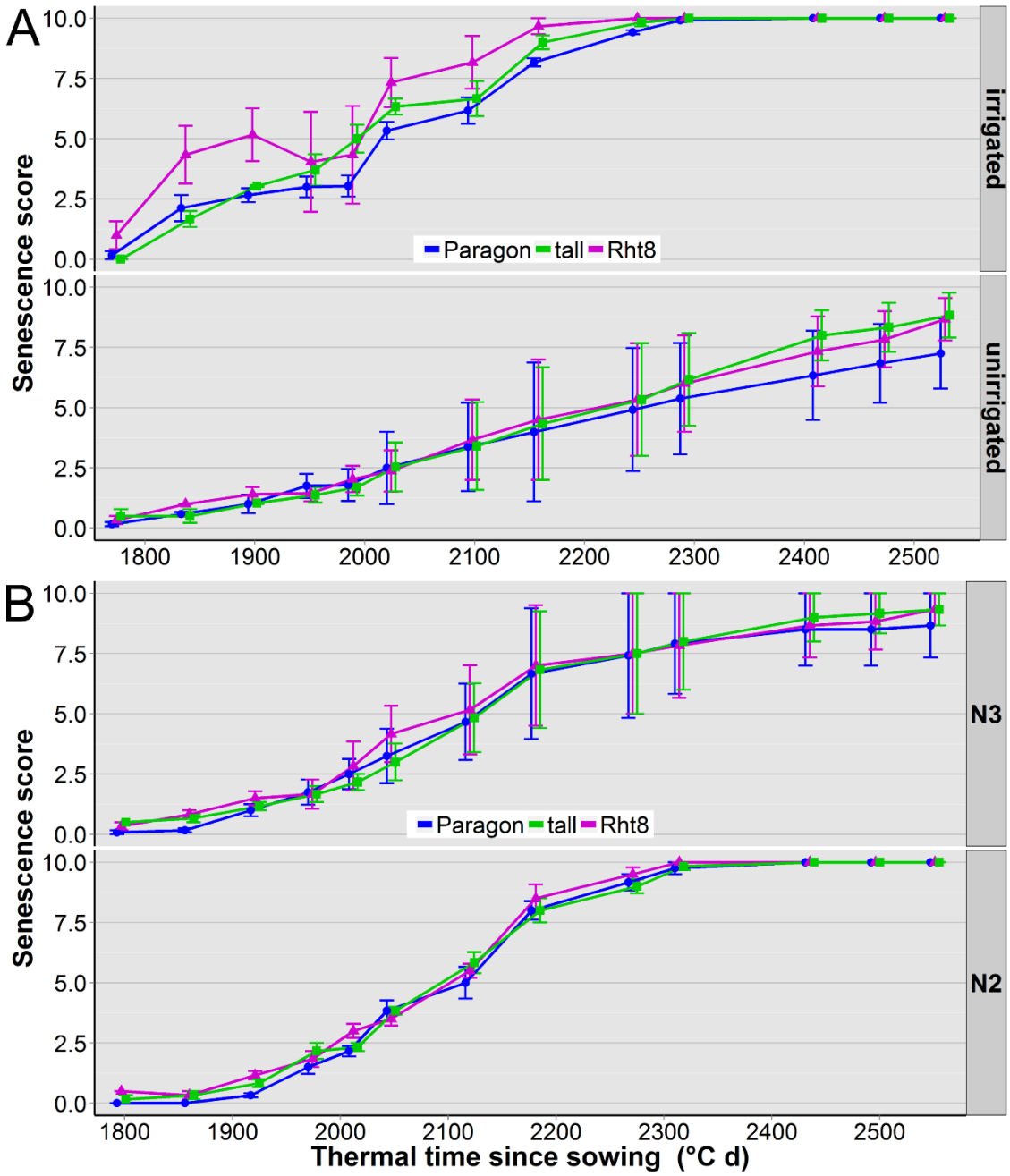


Figure 3.14: Senescence estimated on plot-level shown as thermal time at Church Farm in 2013 (A) in contrasting irrigation regimes (B) in contrasting N treatments. Data recorded from initial senescence (a score of 0) to complete senescence of plot (a score of 10). Data points are mean values. Error bars represent standard error. N2=100kg N ha^{-1}, N3=200kg N ha^{-1}.

The earlier results presented in 3.6 indicated that spikes m^{-2} (SA) modulated grains m^{-2} (GA), which in turn affected yield, and that the response was different between the NILs. Tiller dynamics determine SA to a large extent (Sreenivasulu and Schnurbusch, 2012). In wheat, a low red:far red (R:FR) ratio reduces tillering (Casal et al., 1987, Kasperbauer and Karlen, 1986) and also has a role in modulating root anchorage to increase lodging (Sparkes and King, 2008). In order to elucidate the interplay between tiller dynamics and some of the yield trait responses described in this Chapter, light quantity in terms of photosynthetically active radiation (PAR) and light quality (R:FR) at base canopy level were measured.

Pre-anthesis, there was no difference in R:FR between the NILs, which is in accordance with the green canopy findings (Table 3.7). From October to March, the R:FR halved from ~1 to 0.5, reflecting the canopy growth and increase in density which promoted mutual shading among plants. Post anthesis, R:FR increased with time, reaching ~0.85 by GS85 (end of grain filling). N treatment only had a significant (P<0.001) effect on R:FR in early – mid July, and increased N reduced R:FR across all genotypes in equal proportion (treatment*allele P=0.8). The only significant (P<0.01, LSD test) difference between the *Rht8* and tall NILs occurred on 7th July, with an 18% increase in R:FR in the *Rht8* NIL at treatment. Over all treatments at that first time-point post anthesis, the *Rht8* NIL had a mean 14% increase compared to the tall NIL.

lowest highest

R:FR	pre-anthesis						post-anthesis								
	30/10/2013			04/03/2014			07/07/2014			14/07/2014			24/07/2014		
	N1	N2	N3	N1	N2	N3	N1	N2	N3	N1	N2	N3	N1	N2	N3
par	1.04	1.08	1.10	0.49	0.50	0.48	0.57	0.54	0.40	0.76	0.65	0.61	0.84	0.85	0.83
Rht8	1.06	1.06	1.08	0.49	0.51	0.49	0.59	0.53	0.40	0.73	0.69	0.54	0.86	0.84	0.83
tall	1.06	1.04	1.09	0.50	0.48	0.52	0.54	0.45	0.36	0.71	0.65	0.58	0.82	0.84	0.82
P-value allele	NS			NS			**			NS			NS		
P-value treatment	NS			NS			***			***			NS		
L.S.D.	0.05			0.05			0.07			0.04			0.04		
Rht8 (% of tall)	99.5	102.2	98.9	98.8	105.4	93.5	110.3	117.8	112.9	102.0	105.7	94.1	105.3	100.5	100.7
difference (%)	-0.5	2.2	-1.1	-1.2	5.4	-6.5	10.3	17.8	12.9	2.0	5.7	-5.9	5.3	0.5	0.7

*Table 3.7: Red: Far Red ratios at canopy level at Reading in 2014. Data shown as mean values. N=5. The p-value refers to significant differences between genotypes determined by the least significant difference (L.S.D.) test. NS=means not significantly different at P<0.05, **P<0.01, ***P<0.001, N1=40kg N ha^{-1}, N2=100kg N ha^{-1}, N3=200kg N ha^{-1}.*

PAR was measured post anthesis at three time points in July 2014, one week apart (Figure 3.15). At all time-points, there was a significant (P<0.001) effect of N treatment: PAR increased ~5% between incremental N treatments, such that there was ~10% greater reduction in PAR at N1 than at N3 (full data in Appendix 3.7). There was a significant (P<0.05) difference in PAR between NILs, with the mean of the *Rht8* NIL lower than the tall NIL at each time-point. Furthermore, the differential between the *Rht8* and tall NIL increased with time (-1.4%, -1.5% and -2.3%).

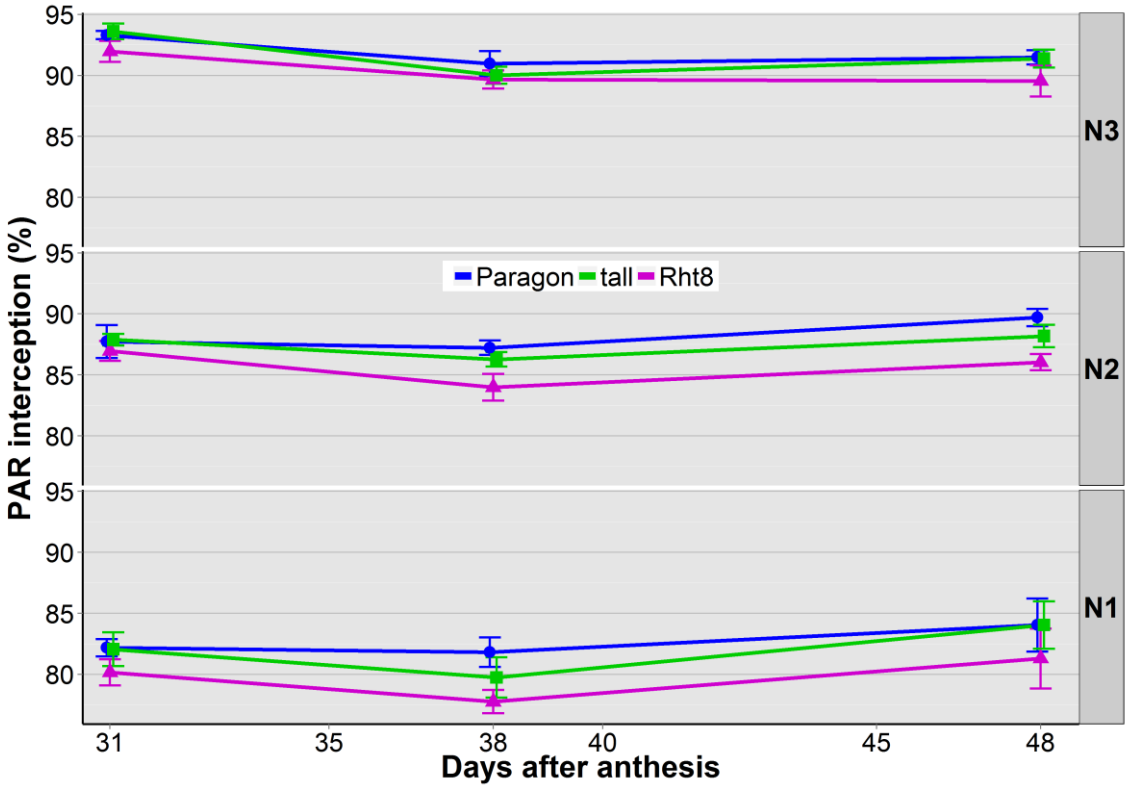


Figure 3.15: PAR interception at canopy level at Reading in 2014 in contrasting N treatments. Data points are mean values. Error bars represent standard error. N=5. Values recorded on same dates as R:FR ratio, displayed as days after anthesis (6/6/14). N1=40kg N ha^{-1}, N2=100kg N ha^{-1}, N3=200kg N ha^{-1}.

3.9 Discussion

The work in this Chapter assessed the agronomic performance of *Rht8* in the field in a photoperiod sensitive (*Ppd-D1b*) background (first described by Gasperini, 2010) for the first time in UK-adapted, spring wheat. Previous reports on the trait responses of *Rht8* have been confounded by the presence of the tightly linked *Ppd-D1a* allele. In particular, it was unclear in many studies whether the reduction in height was due to quicker development from earlier flowering. This has led to an 'adaptation vs escape' paradigm (described in Semenov et al., 2014): wheats grown in southern Europe are not necessarily more tolerant of heat and drought stresses than wheats grown in more temperate conditions e.g. northern Europe. Instead, agronomic practices and photoperiod insensitivity are used to 'escape' terminal desiccating conditions. This suggests untapped potential for breeding. Without *Ppd-D1a* in this study, it was assessed whether *Rht8* confers adaptive advantage. Furthermore, outside of height, yield and in particular yield component traits in *Rht8* remain poorly studied.

There was no significant premature development of the *Rht8* NIL in any of the developmental traits examined (heading, anthesis or senescence), therefore, reduction in height could be ascribed to the genetic effects of *Rht8*. The genetic control underlying plant height was studied here by assessing the trait at final height maturity. Height was reduced by ~11% and was principally correlated with the spike, peduncle and top internode. The height reduction was relatively stable, remaining unperturbed by environmental factors of N or water regime. The consistent reduction in height adds further merit to the work of 'stacking' *Rht8* with other dwarfing genes e.g. work by Rebetzke et al., 2012. This finding differed to a study of *Rht8* in a spring wheat background in Colorado (Lanning et al., 2012), where the magnitude of height reduction tended to be smaller in lower-yielding environments. However, the sites varied greatly for precipitation, whereas here, the lowest-yielding environment was irrigated to field capacity.

Plant height, despite being measured terminally in this work, is, however, a dynamic trait (Wu and Lin, 2006). Additionally, dynamic analysis of QTLs for plant height in wheat discovered several conditional QTLs affecting height at distinct temporal phases (Wang et al., 2010). An attempt to measure height dynamics over time was attempted in a controlled environment room (CER) experiment, but

was foiled by rampant mildew which severely stunted growth. Understanding the dynamics of height in future work is particularly important in order to understand the molecular function of *Rht8* and to identify the key temporal stages for expression analysis work, which was also undertaken in the fine-mapping work of this project.

In high yield-potential sites, *Rht8* conferred a ~10% yield penalty in standard agronomic conditions of high N and without irrigation. This is in accordance with findings of Lanning et al,, 2012, who also reported a negative impact of *Rht8* on yield in standard agronomic conditions. The yield penalty at high N was mitigated in irrigated conditions, and was also abolished at the lowest N input. Additionally, in the lower yield-potential, high-temperature site in Lleida, the *Rht8* NIL had no yield penalty. It appears that in conventional agriculture, *Rht8* is not advantageous in terms of yield. However, there is growing pressure to reduce fertiliser input and PGRs, typical of organic agriculture. In these conditions, breeders are more focused on varieties with stable yield, rather than highest possible yield, due to more heterogeneous conditions across growing seasons. The work in this Chapter suggests that *Rht8* could be an attractive proposition in these environments, where semi-dwarfing stature can be achieved without yield penalty and offer further agronomic advantages by reducing lodging.

UK data showed that the yield penalty found in rainfed conditions was mitigated upon irrigation. It is notoriously difficult to separate heat and drought stress, but this was achieved by irrigating in Lleida to field capacity. Since there was no difference in rate of development between the NILs, these initial findings would suggest that *Rht8* might confer ‘adaptation’ to high temperature, rather than ‘escape’. Although the *Rht8* NIL did not appear to have good drought tolerance, the relatively wet vegetative phase in the UK precludes the sort of environmental conditions which lead to early drought stress (e.g. Siberia) where *Rht8* is found in commercial germplasm. Future collaboration with groups working in these conditions would be beneficial to assess further the performance of *Rht8* in different types of drought. It is interesting to speculate on the molecular mechanism underlying the apparent reduced drought tolerance of *Rht8*. *Rht8* has reduced sensitivity to brassinosteroids (Gasperini et al., 2012). Brassinosteroid signalling is important for the development of stomata (Casson and Hetherington, 2012) and thus perhaps drought tolerance. An unsuccessful attempt to measure

the effect of drought more precisely and over time was made in the aforementioned CER experiment. Future work at contrasting temperatures and irrigation treatments (early and late onset) would further elucidate the interplay between these two variables which could then be translated to the field. This is particularly important with a more erratic growing season in the UK in face of climate change, where drought might not always be terminal.

The yield penalty was driven by concomitant reduction in grain number per unit area (GN), and spikes m^{-2} (SA). Conversely, where the yield penalty was abolished, these two yield trait responses were no longer reduced. SA is determined by tiller dynamics early on in development, although relatively little is known about the molecular basis of tiller formation in wheat (Sreenivasulu and Schnurbusch, 2012). Tillering is the emergence of side shoots at leaf-stem junctions up to the point when the Green Area Index (GAI) reaches ~1, which is typically just before GS31 (HGCA, 2008). Before GS31, high N uptake affects canopy size by promoting tillering and encourages tiller survival. A study of durum wheat in response to N and water availability found that increasing N during this early phase of growth increased GN by increasing both the number of fertile florets and the proportion of them setting grains (Ferrante et al., 2013). Further, tillering responses are affected by R:FR and low R:FR reduce GN. The story is further intertwined with the findings that R:FR regulates yield components in order to adjust to the availability of limited resources such as N (Cecilia Ugarte et al., 2010). The initial findings in the yield components pointing to *Rht8* modulating tiller dynamics was strengthened by two observations in light quantity and quality at the base of the canopy. First, PAR was reduced in the *Rht8* NIL and the reduction was more severe as the canopy matured. This suggests that overall the *Rht8* NIL was using less solar radiation in a useful way, thus compromising yield, at least in conditions where PAR was limiting (i.e. UK). In Lleida, where PAR was in excess, this deleterious effect on yield was compensated. Second, though not highly significant, the data showed a trend in increased R:FR in the *Rht8* NIL, however there was no difference reported between N treatments.

Tiller number is the limiting component of yield in northern European environments with high N input and temperate climates. In these conditions, the *Rht8* NIL had reduced SA, as a result of decreased tillering. Tiller number is established early on in the vegetative phase. It can be surmised that the tall NIL,

with more tillers than the *Rht8* NIL, had plentiful resources in high N conditions for all tillers to be productive and consequently had a yield advantage over the *Rht8* NIL. Conversely, in low N, advantage from increased tiller number (in the tall NIL) was eradicated since the yield potential could not be fully achieved due to the limitation of N uptake efficiency (NUpE). Thus the yield penalty of *Rht8* was abolished in low N conditions. This hypothesis points to the study of floret generation and tiller dynamics to understand further the mechanisms modulating GN, SA and the differences in canopy conferred by *Rht8.*

Large variation has been observed for anchorage and stem strength in UK wheat varieties (Berry et al., 2003). It has been suggested that breeders are unlikely to have exploited this variation since improved understanding of the importance of these traits has only recently been established (Berry et al., 2007). The development of root and stem traits associated with lodging resistance continues until anthesis (Crook and Ennos, 1995) and competes with resources determining GN and well as stem reserves. Therefore, the ideal wheat 'ideotype' has been described as one with the best combination of lodging resistance with the least investment in biomass in order to minimise conflict with yield potential (Berry et al., 2007). In the work presented in this Chapter, *Rht8* conferred decreased GN and yield in the same conditions in which it promoted root lodging resistance, with no difference in harvest index. This strongly points to *Rht8* acting in the root system as opposed to stem biomass, specifically in promoting root plate spread and root plate depth, since these are characters associated with improved root lodging risk (Berry et al., 2007). Further investigation into the root characteristics and how they vary at high temperatures where the yield/GN penalty is reversed is required to determine whether mechanisms other than height reduction are conferring lodging resistance. Based on these findings, in UK growing conditions, *Rht8* can be proposed as an attractive candidate to provide lodging resistance in certain environments. First, where lodging is extensive and the risk to yield obliteration outweighs the ~10% penalty. Second, where yield stability is preferred over absolute maximal yield; and/or ecological systems (namely organic agriculture) where the N fertiliser and the use of PGRs to reduce lodging risk are not permitted.

In the global context of fertiliser reduction, the ability to identify genetic control of NUE-related traits and implement this in breeding programs is an important part

of future genetic gain. Though speculative, the results in this Chapter indicate that the *Rht8* NIL had improved (or at worse adaptively neutral) NUpE at low N input. This is particularly welcome in organic systems, where a greater proportion of N is available earlier on in the growing season as well as much reduced N levels in the soil. NUpE is a trait predominantly associated with the root structure. High NUpE is associated with early root proliferation and shallow proliferation of roots to capture applied N and then later, longer roots to access deeper N reserves (Hawkesford, 2014). The reduction in lodging was not due to differences in stem biomass (since harvest index was not significantly different between NILs). It can be hypothesised that deeper or increased lateral roots that make the *Rht8* NIL better anchored and more resistant to lodging also support a higher NUpE, which is only evident when the N input is reduced to such levels that efficiency cannot be overcome by increased availability in the soil (such as at the N3 treatment). Work in this Chapter clearly calls for a close examination of the spread and depth of roots in the *Rht8* NIL, and to determine whether root traits confer resistance to lodging in a distinct mechanism to just reducing stature, which has been well-established in *Rht8* and in the semi-dwarfing genes of the Green Revolution (Worland et al., 1998b, Hedden, 2003). This is exciting since breeders and scientists (e.g. Lynch, 2007) have already identified that overturning our relatively poor knowledge of wheat adaptation below the soil (compared to above-ground knowledge) could signal the next Green Revolution.

Background effects in the selected NILs were seen consistently across environments with respect to height. Linkage drag might have introduced negative alleles at other loci during the backcrossing process. To compensate for background differences between the NILs, there is scope to use the remaining NILs at BC_3F_2, from which the *Rht8* and tall NIL used here were selected. This would offer a comparison between multiple NILs with the same *Rht8* genotype. This strategy has been used in the past for assessing the effects of dwarfing genes (Chen et al., 2013, Wang et al., 2014b).

Finally, the findings presented in this Chapter would benefit to being extended to a direct comparison of *Rht8* relative to *Rht-B1b* and *Rht-D1b*. This could easily be obtained in future, since our group has developed NILs in Paragon with both these genes and a meaningful comparison could be made in the same genetic background.

Chapter 4: Compact spike morphology caused by *Rht8*

4.1 Introduction

Spike compactness in the glasshouse and field in wheat with the Mara-derived, *Rht8* allele had been reported anecdotally before the start of this project. Further observation of spike compactness in the material grown in Chapter 3 led to a closer inspection of this trait. This is described in this Chapter.

Spike characteristics determine the number of grains per spike and contribute to yield. In addition, variations in spike morphology are widely-used criteria for species determination. Subsequently, the genes and the underlying mechanisms controlling spike morphology are important to taxonomists, breeders and scientists.

In bread wheat, there are three major genes which affect gross morphology of the spike: *Q*, which determines whether a spike is square-headed or spear-like (speltoid); *S*, which controls grain and glume roundness, and *C*, which determines how compact the spike is.

Q is one of the most important genes in the domestication of wheat because it confers the free-threshing character and a square-spike phenotype (Muramatsu, 1963). More primitive wild (spelt) wheat with the *q* allele has a speltoid spike with an elongated rachis and adherent glumes, which make the wheat difficult to thresh (non-free threshing). The mutation to *Q* resulted in the free-threshing character, along with reduced rachis fragility and reduced glume tenacity. This had a profound effect on agriculture, allowing large-scale, efficient harvesting of grain (Simons et al., 2006). The cloning of *Q* on chromosome 5AL revealed that the gene encoded an AP2-class transcription factor involved in plant development (Simons et al., 2006, Zhang et al., 2011). The *q*-to-*Q* mutation resulted in a single amino-acid substitution (gain-of-function) (Simons et al., 2006). A putative miRNA172 binding site in exon 10 of *Q* further points to the involvement of miRNA regulation (Zhang et al., 2011). Consistent with its role as

a transcription factor, Q pleiotropically affects spike length and shape, plant height and spike emergence time (Muramatsu, 1963, Muramatsu, 1986, Sears, 1952, Simons et al., 2006, Zhang et al., 2011). Further, it has been shown that there is co-regulation and complex interactions among the Q/q homoalleles on 5DL and 5BL with 5AL, which result in phenotypic differences in spike morphology (Zhang et al., 2011).

A gene which modulates *Q* expression to control threshability and rachis fragility is *Tenacious glumes* (*Tg*). *Tg1* is on 2DS and coincident with *Xgwm261* in QTL studies (Jantasuriyarat et al., 2004, Nalam et al., 2006). Further, *Tg1* has a homoeologue on 2B (*Tg2*) (Faris et al., 2014b, Simonetti et al., 1999) and putative homoeologue on 2A (*Tg3*) (Faris et al., 2014a). *Tg* is a semi-dominant gene that inhibits expression of *Q*, though the mechanism remains unknown (Jantasuriyarat et al., 2004, Kerber and Rowland, 1974). Therefore a dominant *Q* and recessive *tg* allele must be present for the free-threshing phenotype.

The recessive *s* allele on chromosome 3DL confers sphaerococcoidy (round grains and glumes) (Rao, 1977). This allele defines the sub-species of *T. aestivum*, called *sphaerococcum*, or shot wheat, which has short, dense spikes (Sears, 1947).

The gene *compactum* (*C*) determines spike compactness and defines a subspecies of hexaploid wheat known as *T. aestivum* ssp. *compactum*, or club wheat. Club wheat is characterised by the dominant *C* allele which results in a compact “club” spike. It is generally accepted that that the origin of club wheat is from a mutation at the *C* locus in *T. aestivum*, and not from a tetraploid or diploid ancestor (Johnson et al., 2008). The gene c*ompactum* was mapped to 2D (Rao, 1972), and since then, relatively few studies have investigated this gene. Notably, Johnson et al., 2008 mapped *C* to two bins, either side of the 2D centromere, though the precise location could not be determined. In that study, prior cytogenetic work which localised *C* to the long arm (Unrau, 1950) was cited to corroborate the localisation of *C* to the 2DL bin rather than the bin the other side of the centromere. Conversely, a more recent publication stated that *C* was located on 2DS, based on personal communication with a researcher (Faris et al., 2014c). Intriguingly, in a different study, a spike compactness QTL was reported on 2DS, close to the 2DS bin reported earlier (Manickavelu et al., 2011).

Since the *Rht8* introgression in the *Rht8*/tall NILs in Paragon was in this region, and there was anecdotal evidence of spike compaction, the map data between these studies was examined in relation to the markers used to select for the *Rht8* allele in the Paragon back-crossing population. The compact spike phenotype associated with the *Rht8* allele was, for the first time, quantified in the work presented in this Chapter.

Two spike-compaction genes in other Triticeae species have been suggested as orthologues to *C*. The first is the *soft glume* (*Sog*) gene, found in a compactoid-spike variety (called *sinskajae*) of diploid *T. monococcum* (Taenzler et al., 2002). Johnson and co-workers (2008) placed *Sog* on $2A^{m}S$, close to the centromere, in an approximately homoeologous location to *C*, which was mapped to two bins either side of the centromere separating the chromosome arms on 2D. In light of the unresolved location of *C* (Faris et al., 2014c, Johnson et al., 2008), the relationship between *Sog* and *C* remains uncertain. Second, the barley *zeocriton* (*Zeo*) gene, which confers a dense spike in barley, was also investigated as a possible *C* orthologue by Johnson et al., 2008. *Zeo* has been isolated and shown to be an AP2-like gene (*HvAP2*) on 2HL, the homoeologous chromosome to wheat 2D (Houston et al., 2013). The mRNA turnover of *HvAP2* was found to be regulated by microRNA172, and perturbing this interaction resulted in phenotypic differences in the barley spike morphology (Houston et al., 2013). Map comparison showed *Zeo* was located on the distal end of 2H (Johnson et al., 2008), not overlapping the *C* region close to the centromere. Functional work also showed no miRNA172 binding-site mutations analogous to *Zeo*-like control in *compactum* wheat mutants (Houston et al., 2013).

The compact spike character of *C* provides an easily distinguishable feature which has taxonomic importance in defining subspecies. For this reason, there is interest in understanding the relationship between *C* and other genes affecting spike compactness in different wheat subspecies. A dominant gene determining compact ear in *T. aestivum* ssp. *sphaerococcum* was shown to be non-allelic to *C*, and named *C2* (Goncharov and Gaidalenok, 2005). Four induced mutant genes in Russian wheat, named *Cp*, C^{769}, C^{17648}, Cp^{m}, all conferring spike compactness, were localised to 5AL and were discounted as being allelic to *C* (Kosuge et al., 2012). Instead, the authors suggested that these were alleles of

a new locus (they named *Cp1*) which they postulated is in tandem with the *Q* locus.

Common wheat (ssp. *aestivum*) has the genotype *QcS*, shot wheat (*sphaerococcum*) is *Qcs*, spelt wheat (*spelta*) is *qcS* and club wheat (*compactum*) is *QCS*. As there is little allelic variation reported at the major loci, differences in spike morphology cannot always solely be attributed to these genes. Further, all durum wheat cultivars have the *Q* allele and no D genome, therefore lacking *C* and *S*. This suggests that homoeoalleles or that different minor genes are involved in controlling spike morphology. Indeed, research has found QTLs contributing to spike compaction on almost all 21 wheat chromosomes (Faris et al., 2014c, Jantasuriyarat et al., 2004, Sourdille et al., 2000). These studies have found that compactness QTLs usually coincide with QTLs for spike length as opposed to spikelets spike^{-1}, despite compactness being a function of both these components. This would suggest that spike length and spikelets spike^{-1} are under differing genetic control. These two components were investigated in the work in this Chapter to establish the relative contribution to the compactness phenotype.

In terms of agronomic importance, club wheat is not grown widely. Club varieties are grown commercially (at diminishing levels) as a class of soft white wheat in the Pacific Northwest of the US and areas of Australia, Europe (e.g. Russia) and Turkey. In regions such as the US Pacific Northwest, club wheat has desirable grain quality characteristics: better stability in milling performance over growing seasons compared to other wheat (Jones and Cadle, 1997, Lin and Czuchajowska, 1997). In the hot, dry and windy summer conditions of these regions, stiff culms resistant to lodging and shatter-resistant spikes provide club wheat with superior adaptation to common wheat (Zwer et al., 1995). In the most comprehensive study to date on the agronomic performance of club wheats, the main components differentiating club from common wheat in the US Pacific Northwest environment were more grains per spike with a lower grain weight, whilst harvest index, spikes per unit area and overall yield were unchanged relative to common wheat (Zwer et al., 1995). It has been established that the size of the floret cavity (reduced in a densely packed club spike) is associated with grain weight (Millet, 1986) and that lower grain weight and greater grain number per spike are associated with *C* (Gul and Allan, 1972). Zwer et al. (1995) suggested that club wheats are better adapted to dryland areas where marginal

moisture at emergence provides the smaller grains of club wheat with superior establishment compared to common wheat (Gul and Allan, 1972, Zwer et al., 1995). No study to date has assessed the effects of different treatments such as N or water on compactness and related traits. This was achieved in the work presented here, in the same N and water regimes as described in Chapter 3.

The main aim of the work in this Chapter was to quantify spike compactness in glasshouse- and field-grown plants with *Rht8*. This was achieved first by measuring the compactness on the plot level of the *Rht8* x Paragon NILs described in Chapter 3. Second, tiller samples from Norwich and Reading were measured to dissect further the compact spike phenotype. A secondary aim was to determine whether the compact spike phenotype could be attributed to particular conditions, since anecdotal reports varied with environment, implying G x E interaction. To do this, spike compactness was measured across the water regimes and N treatments outlined in Chapter 3. Finally, the timing of the onset of spike compactness was measured by assessing spikes at GS 30 – 39 in the fine-mapping *Rht8* population (described in detail in Chapter 5). The spike compactness measured here is also put into the context of current research on the genetics of spike morphology.

4.2 QTL for compact-spike overlaps with *Rht8* introgression

Johnson and co-workers (2008) mapped *C* to two bins either side of the 2D centromere. In their map, the 2DS bin (designated a position of 0.33 cM from the centromere) was two bins away from and 0.7 cM more proximal to the centromere than *Xgwm261* and *Xcfd53*. *Xgwm261* and *Xcfd53* were mapped to a 2DS bin designated a position of 0.47 – 1.00 cM. These markers were the same two used for marker-assisted selection during the *Rht8* NIL development in the Paragon background by Gasperini (2010) (described in Chapter 3). Additionally, a stable spike compactness QTL was found to overlap the *Xgwm261-Xcfd53* region by Manickavelu et al., 2011. Further, a recent paper stated that from personal communication *C* was believed to be on 2DS (Faris et al., 2014c). Since a more precise location of *C* is unknown (or at least unpublished) and given the relatively poor marker density on the resolution of the spike compactness QTL compared

to Johnson's map, it is possible that the spike compactness QTL and *C* are the same locus. Given this, it is probable that the *Rht8* introgression includes this ambiguous genetic region. In order to rationalise the spike compactness QTL with the *Rht8* work in this thesis, *Xgwm261* and *Xcfd53* were integrated in the most recent wheat bioinformatics and comparative genomics resources, which were used in Chapters 5 and 6 to further fine-map *Rht8*.

	Wheat					
Assembly/ source	Manickavelu et al. 2011		IWGSC		Chapman assembly	
	cM	QTL	contig	bin (cM)	scaffold	bin (cM)
Xgwm261	33.1	QCpt-07/08	2DS_5318891	17.34	518430	13.642
Xcfd53	37.2	QCpt-07/08	2DS_5378845	16.95	6258899	14.7795

centromere 2D ⬇

	Barley				**Brachypodium**		**Rice**	
Assembly	IBSC-1.0 (082214v1)		WGS Morex		v1.0		IRGSP-1.0	
	Gene	pos	contig	bin (chr:cM)	Gene	pos	Gene	pos
Xgwm261	MLOC_66589	Chr 5	51801	5: 139	BRADI1G626040	21142086	-	-
Xcfd53	MLOC_76709	20513991	72474	2: 19	BRADI1G69730	68206285	OS10G0399700	13467597

Table 4.1: Integrating existing knowledge on compact-spike QTL reported by Manickavelu et al. 2011 with wheat bioinformatics and comparative genomics used in fine-mapping Rht8 in this project. The arrow indicates the downstream position of the centromere. Further explanation of the resources used is in Chapter 5.

The two markers were placed in adjacent bins in the International Wheat Genome Sequencing Consortium (IWGSC)-2 data; estimated to be 0.39 cM apart (based on population sequencing (POPSEQ) data) and over 1 cM apart in the Chapman scaffolds (Table 4.1). The order of the IWGSC-2 bins is incorrect based on the Chapman assembly and already well-established genetic maps which place *Xgwm261* more distal to the centromere than *Xcfd53*. However, this discrepancy is due to the limitation of current wheat resources and is only a good approximation. Crucially, the position of *Xgwm261* and *Xcfd53* in the IWGSC-2 bins was identical to Gasperini's flanking markers to *Rht8*, *DG279* and *DG371* (shown in Table 5.4). Comparative genomics resources indicated poor synteny with the *Rht8* region (compared with Table 5.4), a problem already associated with work in *Tg1* also on 2DS (Faris et al., 2014b). Taken together, the region harbouring *Rht8* introgressed to make NILs for the *Rht8* locus in Paragon (reported in Chapter 3) overlaps with a spike compactness QTL and possibly *C*. Given the observations of spike compaction in the *Rht8* NIL, further investigation was highly warranted.

4.3 Assessing compactness in *Rht8* x Paragon NILs

4.3.1 Spike morphology on the plot level

Figure 4.1: Hexaploid and tetraploid wheat Cp and C mutants demonstrating a range of compactness phenotypes. The arrow indicates the 'semi-compact' morphology which was also observed in the Rht8 NIL in the Paragon background. From Kosuge et al., 2012.

Despite the common reference to club or common wheat based on the binary taxonomic distinction of a compact or lax spike, the trait itself is quantitative and different degrees of compactness are observed (Figure 4.1). Compared to other reports (e.g. Kosuge et al., 2012), the spike compaction observed in the *Rht8* NIL in the field (Figure 4.2) was of the 'semi-compact' type, and not the extreme compactness used by Johnson et al. 2008 in their *C* fine-mapping population. Indeed, more recent studies of spike compaction have quantified the trait by dividing spike length by spikelet number spike^{-1} (Faris et al., 2014c).

Figure 4.2: The compact spike observed consistently in the Rht8 NIL across a range of spikelet numbers. Spikes from the Rht8 NIL (right) exhibiting semi-compact morphology as classified in Figure 4.1, compared to the more lax spike in Paragon (left). The scale bar is in centimetres.

Spike compactness had been observed sporadically in the glasshouse (in the fine-mapping *Rht8* population) and field (fine-mapping population and *Rht8* NILs). In the 2013-2014 season, spike compactness in field trials at Church Farm and Reading (described in full in Chapter 3) was also observed (Figure 4.3), and measured quantitatively on the whole-plot level for the first time.

Figure 4.3: Compact spike visible on the plot level.Contrasting compactness in Paragon (left) and the Rht8 NIL (right).Taken at Church Farm in 2013-2014 growing season.

Rht8 NIL plots were scored for compactness at the plot-level by visually assessing the percentage of compact spikes in the plot (Figure 4.3 right) relative to a more lax ear (Figure 4.3 left). Compactness was only observed in the *Rht8* NIL (not in the tall NIL or Paragon) and only at the low N (40kg N ha^{-1}; N1) treatment at Church Farm. A considerable proportion (75%) of the spikes in the low N plots were compacted but visually no difference was discerned between water regimes (overlapping error bars, Figure 4.4). At Reading, every *Rht8* NIL plot in the experiment (n=15) showed compaction to some degree across all N treatments. Overall, the percentage of compact spikes was estimated to be lower than at Church Farm (~50%) and there was no significant difference between the N treatments (overlapping error bars, Figure 4.4).

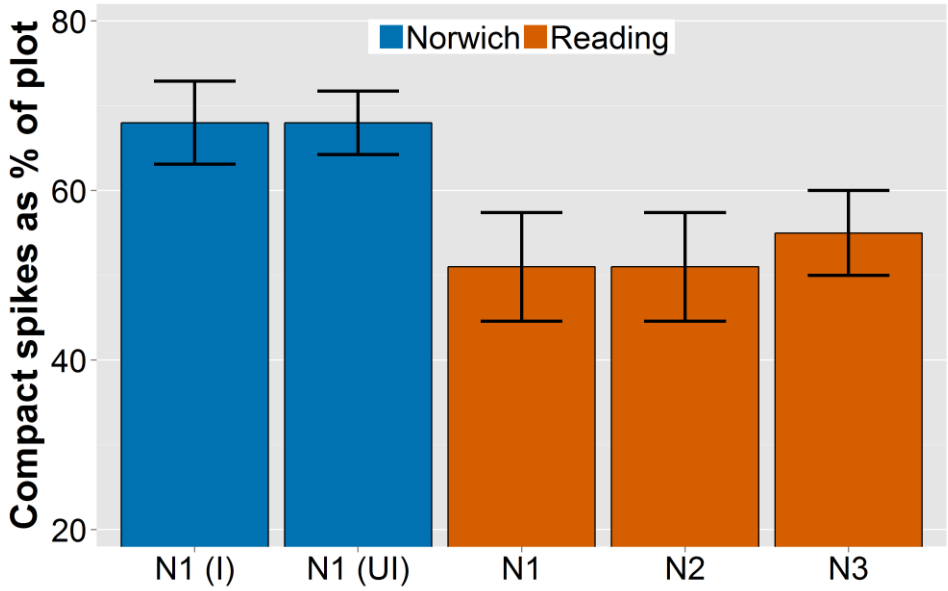


Figure 4.4: Compact spikes measured in the Rht8 NIL as a proportion of the whole plot from visual inspection N1=40kg N ha^{-1}, N2=200kg N ha^{-1}, N3=200kg N ha^{-1}, I=irrigated, UI=unirrigated (rainfed). Measurements made in 2013-2014 growing season. Error bars represent standard error. N=5 per treatment across both sites.

4.3.2 Spike morphology in tiller samples

The general observation on the plot-level that the *Rht8* NIL had more compact spikes than the tall NIL was quantified on the plant-level by analysing tiller samples taken at Church Farm from 2012 - 2014 and at Reading in 2014. Spike compactness was measured by dividing the spike length by the spikelets spike^{-1}, to achieve a unit of cm spikelet^{-1}. A smaller value is a smaller area per spikelet and hence greater compaction. Spike compactness was significantly different ($P<0.05$ and $P<0.001$) between the genotypes in every environment (Table 4.2A) with the *Rht8* NIL being consistently significantly lower than the tall NIL and Paragon (L.S.D. test) (smaller value is increased compactness). The mean decrease in cm spikelet^{-1} (increase in compactness) across environments of the *Rht8* NIL compared to the tall NIL was 15%. There was similar overall variation in compactness between genotypes, with a 0.1 cm spikelet^{-1} difference between minimum and maximum values (Figure 4.5A). As reported in Chapter 3.3 and Figure 3.4, there were some background genetic effects influencing the compactness, since the measure of compactness in the tall NIL was sometimes lower than in Paragon (Figure 4.6). However, the background effects were considerably less prominent than in the plant height data. When the data was considered overall, the distribution of values was very similar between the tall NIL

and Paragon (Figure 4.5). Across sites, contrary to the observations on the plot level outlined in 4.3.1, there was no marked difference in the spike compactness found across sites, with a -17% change in *Rht8* NIL compactness compared to the tall NIL in Church Farm 2014 data (mean compactness 0.4 cm spikelet^{-1}), whilst at Reading this value was -16% (0.39 cm spikelet^{-1}) (Table 4.2A).

lowest highest

A

Spike Compactness (cm spikelet^{-1})	Church Farm										Reading		
	2012	2013				2014					2014		
		N3	N2	UI	I	N3 I	N3 UI	N1 I	N1 UI	N2	N1	N2	N3
par	0.46	0.46	0.51	0.47	0.47	0.46	0.49	0.46	0.44	0.52	0.42	0.45	0.48
Rht8	0.41	0.38	0.42	0.39	0.39	0.41	0.42	0.41	0.37	0.41	0.37	0.36	0.43
tall	0.48	0.47	0.51	0.51	0.46	0.50	0.50	0.47	0.45	0.50	0.43	0.45	0.50
P-value	*	***		***		***				***	***		
L.S.D.	0.05	0.03		0.04		0.02				0.05	0.03		
Rht8 (% of tall)	85	82	83	77	83	82	84	87	81	82	86	80	86
difference (%)	**-15**	**-18**	**-17**	**-23**	**-17**	**-18**	**-16**	**-13**	**-19**	**-18**	**-14**	**-20**	**-14**

B

Spike length (cm)	Church Farm										Reading		
	2012	2013				2014					2014		
		N3	N2	UI	I	N3 I	N3 UI	N1 I	N1 UI	N2	N1	N2	N3
par	11.20	10.63	10.71	10.73	10.65	10.54	11.89	10.59	10.04	11.96	9.90	10.40	10.90
Rht8	10.28	8.83	8.81	9.09	8.78	10.41	9.84	9.29	8.10	9.16	9.10	8.30	10.20
tall	11.48	10.43	10.72	11.23	10.81	11.42	11.50	10.44	9.89	11.21	9.90	10.40	11.30
P-value	NS	***		***		***				***	***		
L.S.D.	1.70	0.70		0.80		1.50				0.90	0.80		
Rht8 (% of tall)	90	85	82	81	81	91	86	89	82	82	92	80	90
difference (%)	**-10**	**-15**	**-18**	**-19**	**-19**	**-9**	**-14**	**-11**	**-18**	**-18**	**-8**	**-20**	**-10**

C

Spikelets spike^{-1}	Church Farm										Reading		
	2012	2013				2014					2014		
		N3	N2	UI	I	N3 I	N3 UI	N1 I	N1 UI	N2	N1	N2	N3
par	24.11	23.11	20.94	22.83	22.61	23.22	24.33	23.00	23.11	23.11	23.56	23.22	22.89
Rht8	25.00	23.00	20.89	23.22	22.78	25.44	23.44	22.11	22.56	22.56	24.89	23.22	23.56
tall	23.89	22.22	21.00	22.22	23.33	22.89	23.11	22.00	22.00	22.67	22.89	23.33	22.67
P-value	NS	NS		NS		NS				NS	NS		
L.S.D.	1.02	1.46		1.57		2.10				2.00	1.69		
Rht8 (% of tall)	105	104	99	105	98	111	101	101	103	100	109	100	104
difference (%)	**5**	**4**	**-1**	**5**	**-2**	**11**	**1**	**0**	**3**	**0**	**9**	**0**	**4**

*Table 4.2: Spike compactness and its derivative components in the Rht8 NIL, tall NIL and Paragon. (A) Spike compactness, (B) spike length and (C) spikelets spike^{-1}. Data shown as mean values. The p-value refers to significant differences between genotypes determined by the least significant difference (L.S.D.) test. NS=means not significantly different at P<0.05, *P<0.05, ***P<0.001.*

Spike length and spikelets spike^{-1} was already reported on in Chapter 3 and Appendices, but is included here for ease of comparison. Spike length closely mirrored the pattern observed in the spike compactness. The *Rht8* NIL had a significantly (P<0.001) shorter (L.S.D test) spike than the tall NIL, with a mean

15% decrease across all environments (Table 4.2B). This mean decrease was consistent across sites, with a 14% (9.36 cm mean length) shorter spike at Church Farm in 2014 compared with a 13% reduction at Reading (9.20 cm mean length). Overall, there was a 3 cm difference between the minimum and maximum spike length measured across all three genotypes, with the median spike length much lower (9.2 cm) in the *Rht8* NIL compared to 10.75 cm in the tall NIL/Paragon (Figure 4.5B).

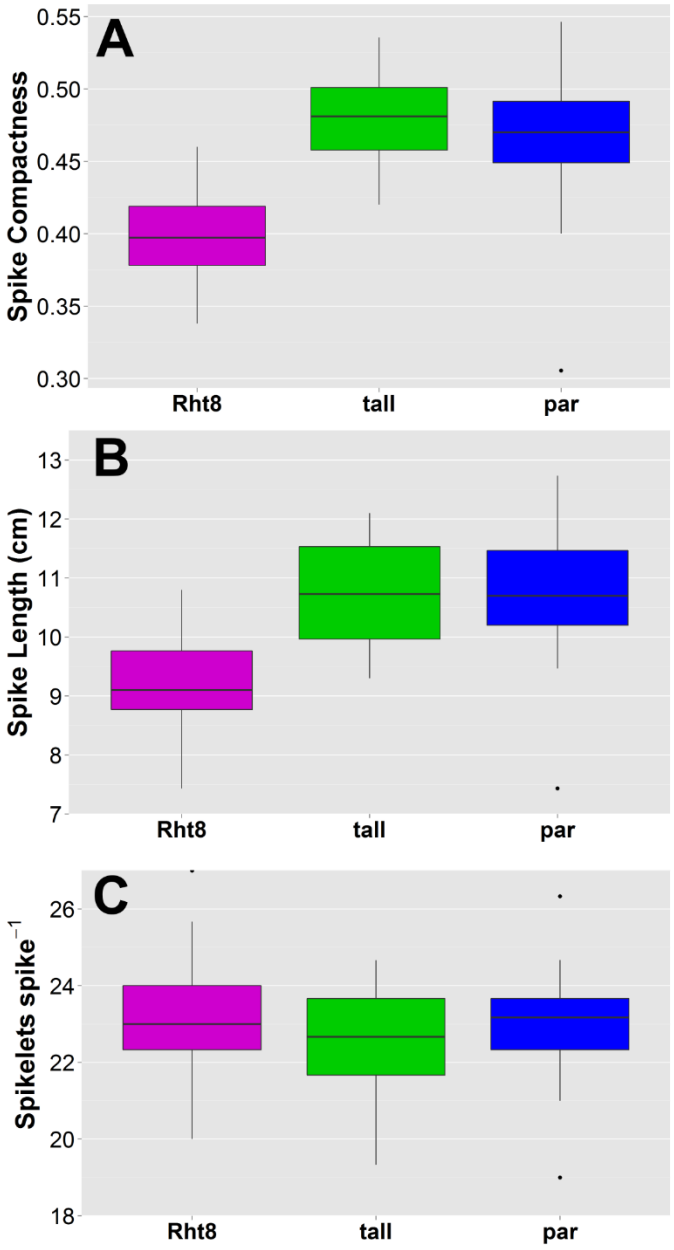


Figure 4.5: Boxplots of (A) spike compactness (units in cm spikelet^{-1}), (B) spike length and (C) spikelets spike^{-1} in the Rht8 NIL, tall NIL and Paragon. Data pooled across all sites (Norwich and Reading) where tillers were sampled. Lines represent ranges of the data, with extreme values as points. The box represents top and lower quartiles, with the median as the central line.

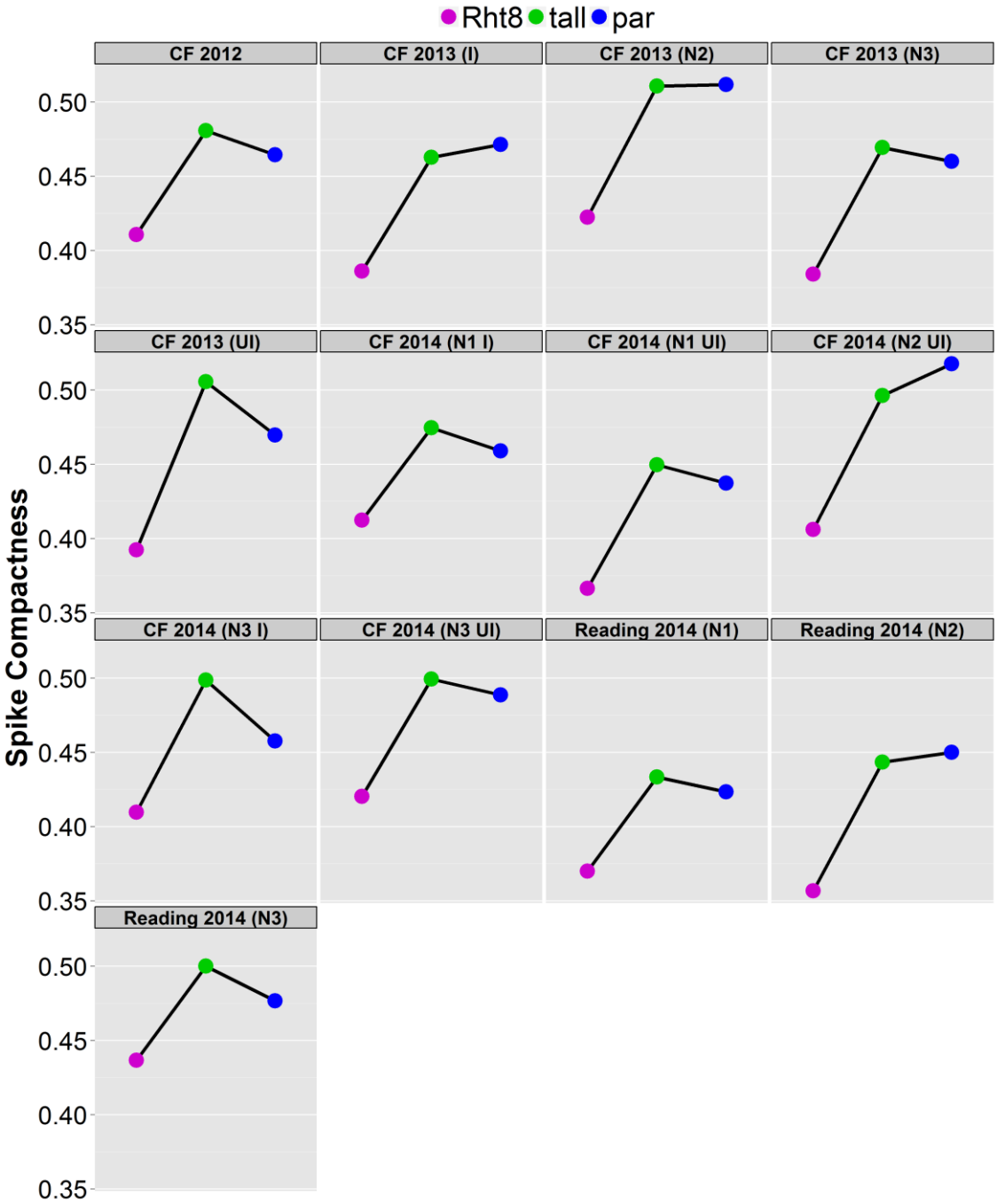


Figure 4.6: Mean spike compactness (units in cm spikelet^{-1}) of the Rht8 NIL, tall NIL and Paragon in all sites and all conditions where tillers were sampled. N1=40kg N ha^{-1}, N2=200kg N ha^{-1}, N3=200kg N ha^{-1}, I=irrigated, UI=unirrigated (rainfed).

The number of spikelets spike^{-1} was not significantly different between genotypes at the 95% confidence level in any environments (Table 4.2C), although the *Rht8* NIL did have a higher median value than the tall NIL (23 versus 22.5 spikelets spike^{-1}) (Figure 4.5C). In two environments, the *Rht8* NIL had a ~10% increase in the number of spikelets spike^{-1}, which equates to 2-2.5 more spikelets (Table 4.2C). Taken together, this suggests that despite spike compactness being a function of spike length and spikelets spike^{-1}, the difference in compactness in the *Rht8* NIL was driven by reduction in the spike length rather than an increased number of spikelets on the rachis. This was confirmed by anecdotal observation in the field, where compact spikes were collected with different numbers of spikelets (i.e. the traits were largely independent) (Figure 4.7).

To test this, correlation analysis was carried out (data pooled across environments) between spike compactness and its components. Spike compactness displayed highly positive and significant correlation with the length of the spike (r = 0.7 – 0.8, P<0.001) across genotypes. There was a small negative correlation between spike compactness and the number of spikelets spike^{-1}, but this was not significant at the 95% confidence interval. This corroborated that spike length was driving the differential spike compactness observed between the NILs.

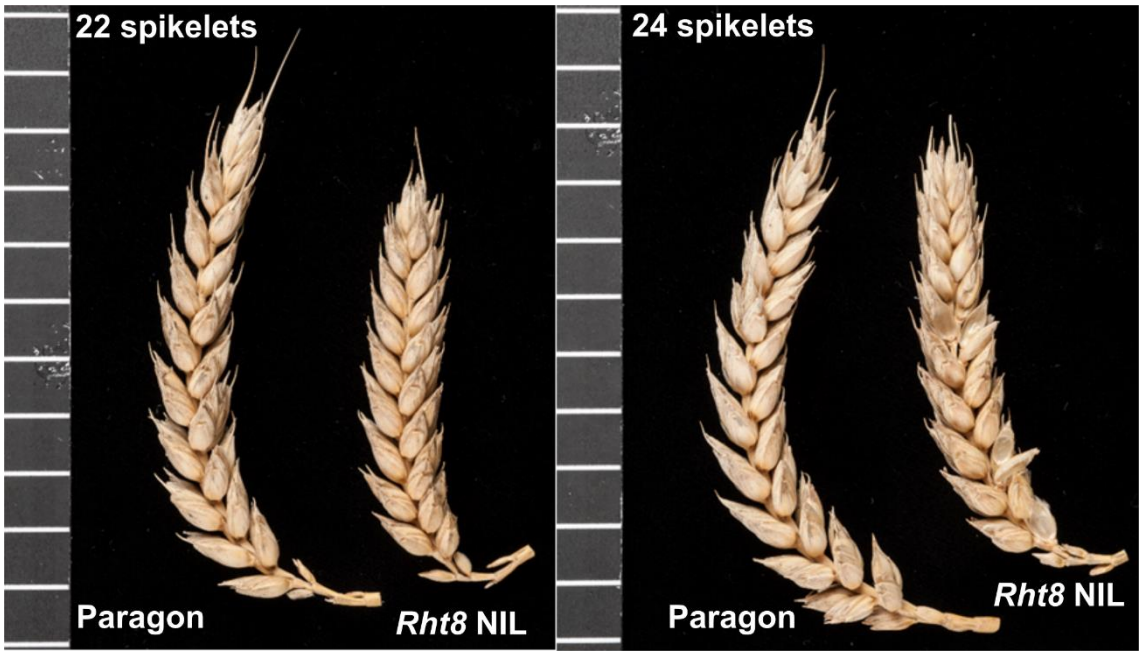


Figure 4.7: Compact spike morphology in the Rht8 NIL contrasted with Paragon, with different spikelet numbers. Scale bar is in centimetres.

lowest highest	Rht8		par		tall	
	r	p-val	r	p-val	r	p-val
Spike length	0.726	***	0.802	***	0.679	***
Spike:height	-0.010	NS	0.015	NS	0.071	NS
Spikelets spike^{-1}	-0.003	NS	-0.056	NS	-0.111	NS
Plant height	0.275	0.07	-0.018	NS	0.057	NS

*Table 4.3: Simple correlation coefficients (r) between spike compactness, spike components and total height across all environments. Spike:height is the spike length normalised for the total height (Plant height). NS=not significant at P<0.05, ***P<0.001.*

Since spike compactness was significantly reduced in the *Rht8* NIL compared with the tall, it was hypothesised that the trait could be used as a binary score to type at the *Rht8* locus, in a similar way to plant height at maturity. Testing this hypothesis was approached in two ways. First, the compactness was measured in the BC_3 NILs typed tall/short for *Rht8* on the basis of final height in 2012, from which one short and tall were selected for further field trails (described in Chapter 3). The NILs typed '*Rht8'* had significantly (P<0.001, one-way ANOVA) greater spike compaction than the tall NILs. The mean spike compaction was 0.40 cm spikelet^{-1} for short NILs compared to 0.46 cm spikelet^{-1} for tall NILs. When the NILs were ordered by spike compaction, and the plant heights plotted, a clear binary step appeared between short and tall (Figure 4.8), allowing the NILs to be typed at a high level of confidence for the *Rht8* locus.

Second, correlation analysis of the *Rht8* NIL spike compactness (pooled data) with plant height length showed a moderately weak positive correlation (r = 0.275, P=0.07) which was only significant in the *Rht8* NIL at 93% confidence interval. Correlation between compactness and height is not unexpected if the genes are linked. No significant correlation was observed for the other genotypes (P>0.1). When the spike length was normalised for the total plant height (by dividing the spike length by the plant height), the significant correlations found in raw spike length were obliterated. Taken together, this suggests first, that typing for spike compactness might indeed be easier where the plant height has a more continuous distribution less amenable to assigning a qualitative score. In this situation the spike compactness could be used in lieu of the height to type the *Rht8* locus. Second, compactness was due to spike length and only weakly correlated with overall height, and it was the raw spike length which was key, rather than the ratio of the spike:height.

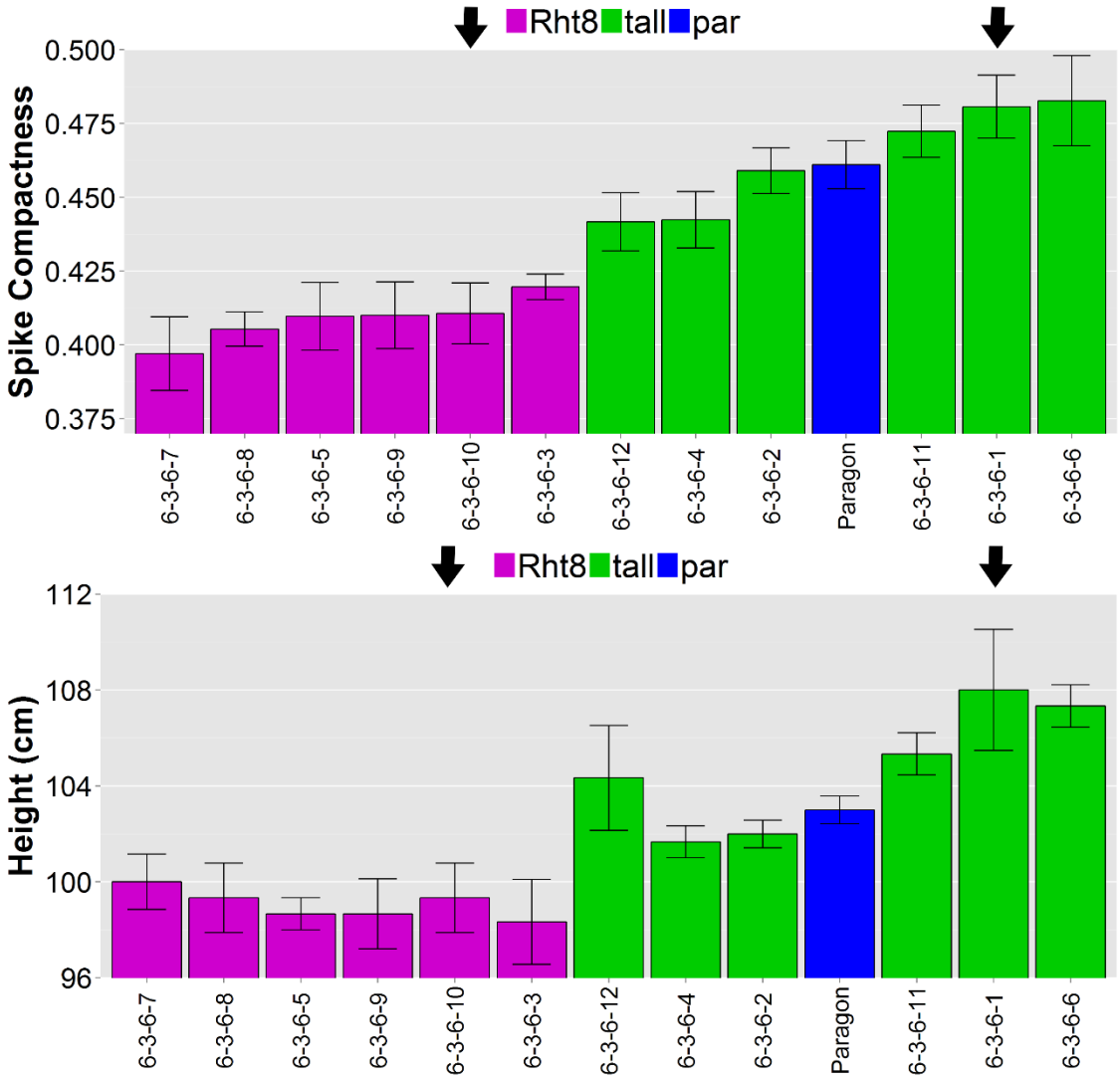


Figure 4.8: Spike compactness (units in cm spikelet^{-1}) (A) and height at maturity (B) of all Rht8 and tall NILs developed to BC_3 in the Paragon background. Data represent means, with error bars representing range of three values. Data taken from 2011-2012 growing season at Church Farm. Arrows indicate the Rht8 and tall NIL selected for further field trials. Height data is ordered in ascending order according to spike compactness.

4.4 Spike compactness in contrasting water regimes and N treatments

Using ANOVA, the effect of N and irrigation treatment on spike compactness was measured. Overall, irrigation had no effect on spike compactness in any of the genotypes in 2013 or 2014 (P=0.13; P=0.57) (Figure 4.9 and Appendix 4.1.2). The *Rht8* NIL maintained increased compactness compared to the tall NIL across all water regimes. Furthermore, there was no interaction between the genotype, water regime and N treatment (Appendix 4.1).

N treatment had a significant effect on spike compactness at both Church Farm (P<0.05) and Reading (P<0.01) in 2014, however, this affected both NILs equally since there was no N*genotype interaction (P=0.6; P=0.1) (Figure 4.10 and Appendix 4.2.3).

In sum, the data showed that the greater spike compactness in the *Rht8 NIL* remained unchanged (in the case of irrigating) and where increasing N did have an effect, the differential between the *Rht8* NIL and tall NIL was maintained since the NILs responded in the same way.

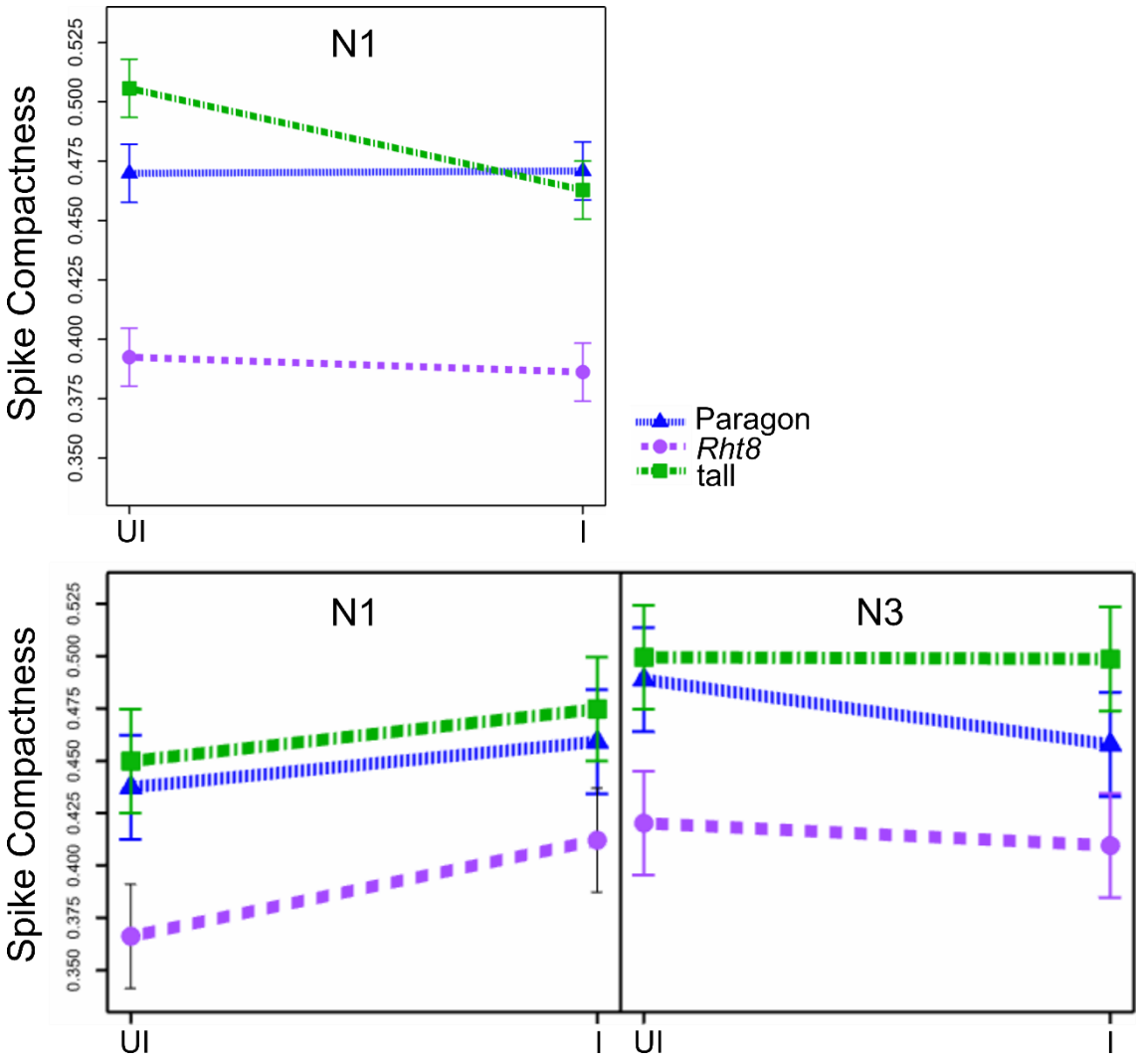


Figure 4.9: Spike compactness (units are cm spikelet^{-1}) of the Rht8 NIL, tall NIL and Paragon in contrasting irrigation regimes at Church Farm. 2012-2013 season (top) and 2013-2014 (bottom). N1=40kg N ha^{-1}, N3=200kg ha^{-1}, I=irrigated, UI=unirrigated (rainfed). Data points are means, error bars represent standard error.

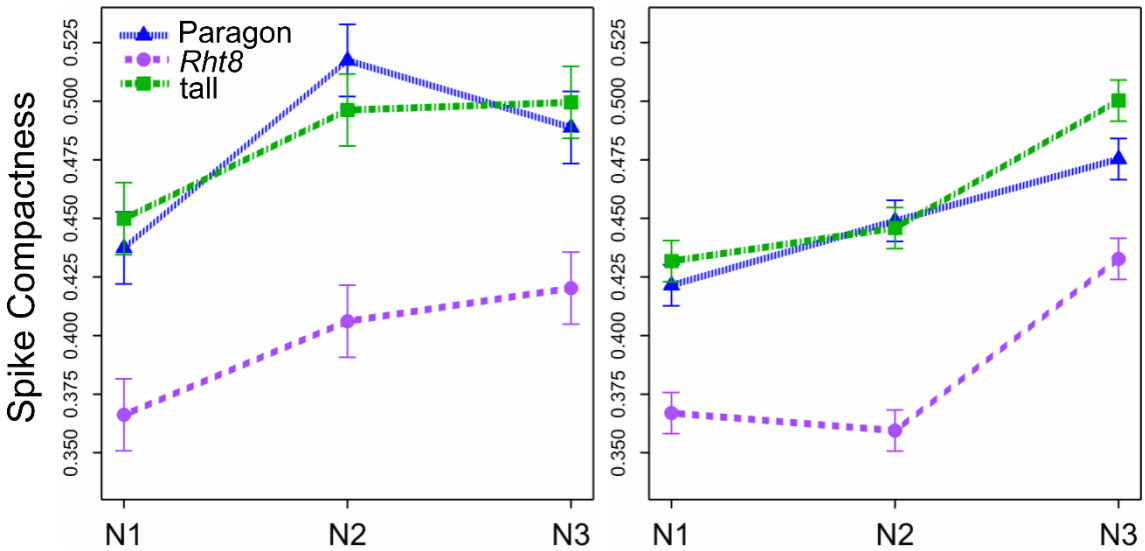


Figure 4.10: Spike compactness (units are cm spikelet^{-1}) of the Rht8 NIL, tall NIL and Paragon in different N treatments at Church Farm (left) and Reading (right) in 2014. N1=40kg N ha^{-1}, N2=100kg ha^{-1}, N3=200kg ha^{-1}. Data points are means, error bars represent standard error.

4.5 Spike compactness in the *Rht8* x Cappelle-Desprez fine-mapping population

The spike compaction reported in the Paragon NILs in this Chapter was taken from mature spike measurements. There have been no reports of spike compaction dynamics across developmental stages, yet there is a growing awareness that height and yield components have plasticity to varying degrees during the wheat growing season (Slafer et al., 2014, Slafer, 2003). The fine-mapping *Rht8* population was grown in the glasshouse and the spikes harvested for RNA-Seq (Figure 5.2) early in the reproductive phase (GS 30 – 39). A subset of 20 short and 20 tall recombinants were selected (typed on the basis of plant height at maturity) and compactness measured as before, however software was used to measure length due to the small size of the spike (2 – 5 cm, Figure 4.11B). By ANOVA (Appendix 4.4), there was no significant difference in spike compactness between the short and tall recombinants (P=0.3), although the short recombinants had decreased mean spike compaction compared to the tall (1.55 cm spikelet^{-1} versus 1.71 cm spikelet^{-1}). The components of spike compactness were also not significantly different between the short and tall groups (P>0.05) (Figure 4.11B & C).

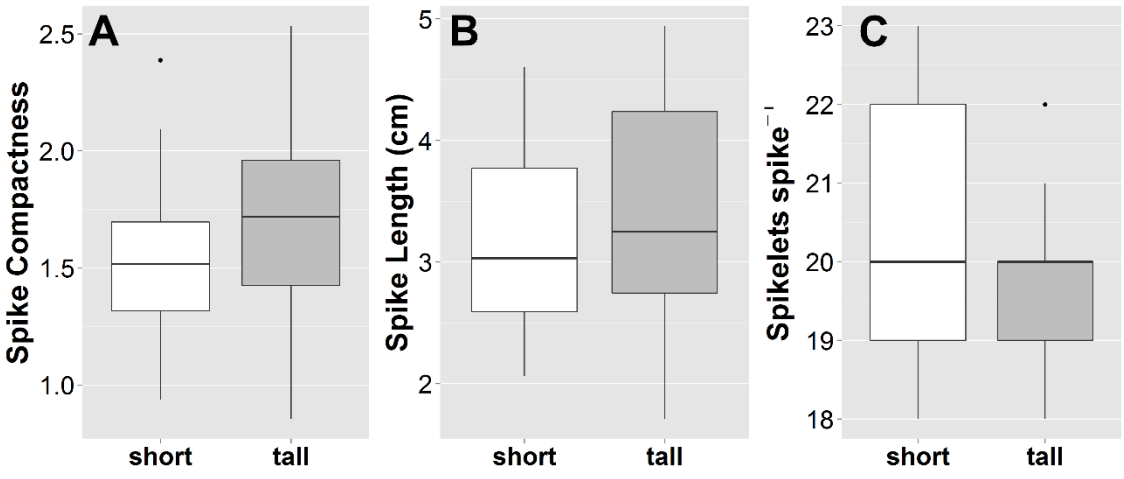


Figure 4.11: Boxplots of spike compactness (units are cm spikelet^{-1}) (A), spike length (B) and spikelets spike^{-1} (C) in a subset of fine-mapping Rht8 x Cappelle-Desprez recombinants. Recombinants were retrospectively typed short/tall at Rht8 based on final height. N=20 for short and N=20 for tall. Lines represent ranges of the data, with extreme values as points. Box represents top and lower quartiles, with median as central line.

Despite no statistically-significant difference in compactness between the genotypes, an observation was made during the measurement taking that at a similar magnification, the short NIL had a markedly compact spikelets at the tip of the spike (Figure 4.12).

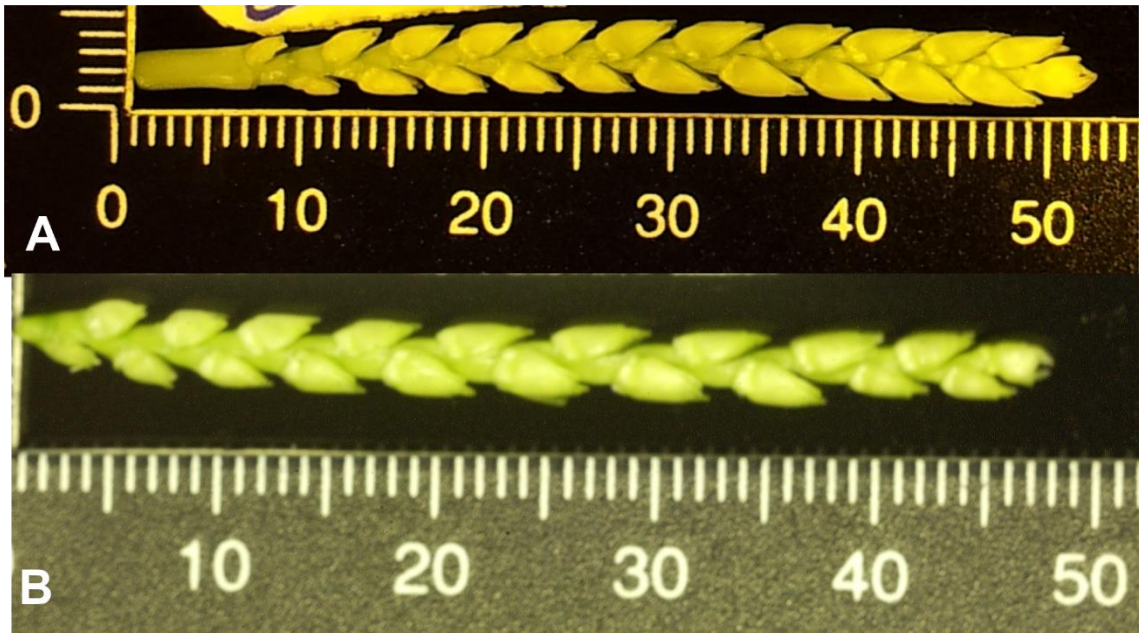


Figure 4.12: Spikes of a short and tall recombinant from the fine-mapping Rht8 x Cappelle-Desprez population. Recombinants were retrospectively typed short/tall at Rht8 locus. Recombinant 2-1-8-4 = short (A), recombinant 1-7-7-1= tall (B). Spikes were harvested within an 8-hour period on the same day at approximately GS 30 – 39. Note that scales are not equivalent.

4.6 Discussion

Johnson and co-workers (2008) mapped *C* to two centromeric bins, one on 2DS and the other on 2DL. The bin on 2DS mapped 0.7 cM proximal to the bin containing *Xcfd53* and *Xgwm261*, and therefore outside the introgression which was selected in the *Rht8* x Paragon population, though it is likely that linkage drag might include the adjacent regions. Therefore, the spike compactness observed in the *Rht8* NIL might be due to the *C* locus. Alternatively, the spike compactness observed here might not be due to *C* but instead due to another distinct QTL, such as the one reported by Manickavelu et al., 2011. It should be noted that neither Manickavelu et al., (2011) nor Johnson et al., (2008) have *Ppd-D1* (estimated at 10 cM distal to *Xcfd53* by Gapserini et al., (2012)) in their genetic maps and *Ppd-D1* could feasibly influence this trait.

Manickavelu et al., (2011) detected a QTL for spike compactness across both years of their study, indicating a stable locus. This QTL spanned *Xgwm261-Xcfd53*, with *Xcfd53* as the closest marker. It should be noted, however, that the marker density in this more recent study was much lower than that of Johnson's map. Therefore, a third possibility is that *C* and the spike compactness QTL are the same locus, and that the locus coincides with the introgressed segment harbouring *Rht8* in the *Rht8* NIL. A fourth possibility is that the spike compactness QTL found by Manickavelu et al., (2011) and the spike compactness documented in this Chapter are pleiotropic effects of Mara-derived *Rht8* only, and not due to variation at or surrounding the *C* locus. A recent paper commented on *C* being firmly located on 2DS, but cited personal communication with a researcher. Therefore, imminent but as yet unpublished work might resolve this issue. In anticipation of that, the work here contributes to the relative dearth of research underlying spike compactness, especially spike compactness and environmental interaction, which has not been reported in detail before.

The most comprehensive agronomic study on the effect of spike compactness was in the US Pacific Northwest on commercial club cultivars, which also had the *Rht-B1b* and *Rht1-Db* semi-dwarfing alleles. Zwer et al. (1995) found that club wheat had lower TGW compared to common wheat, but grain number per spike compensated to produce an overall yield equivalent to common wheat. A further finding was that the club wheat had an increased number of fertile spikelets, presumably due to the compressed nature of spikelet development in the club wheat. It was established early in this Chapter that the spike compactness observed in the *Rht8* NIL was not equivalent to the compaction of club wheats, which typically have spike length reduced by almost half compared to common wheat. Further, results in Chapter 3 already showed that the *Rht8* NIL did not have a consistent significant reduction in TGW, improved spikelet fertility or differences in grain size. Taken together, this suggested that the semi-compact phenotype had a more subtle effect on yield components than the more extreme club spike. Interestingly, results in Chapter 3 did corroborate the compensatory effect of increased grain per area on yield, described by Zwer et al., 1995 in the high-temperature conditions in Lleida. However, it is difficult to ascribe these effects to the compact spike or to *Rht8*, since no tiller data was collected from Lleida.

Work in this Chapter highlighted a discrepancy in consistency of spike compaction reported from tiller samples versus visual assessment of plots in the field. Visual assessment in 2014 only identified compaction in low N at Church Farm, whereas tiller samples showed a consistent compact spike at all N levels. One reason for this might be sampling methodology. Assessing agronomic characteristics on a plot-level has proved reliable (e.g. for lodging and height in Chapter 3) since an average is taken by eye. A disadvantage is that the assessment is qualitative, and tends to be binary (compact/lax spike). Additionally, it is only reasonably simple to assess the club spike coverage in the plot, and much harder to make judgements as to the degree of compactness (apart from to a trained breeder). The compactness conferred by the *Rht8* introgression is semi-compactness, as opposed to the 50% reduction in spike length reported in the extreme 'compact' phenotype (Kosuge et al., 2012). Therefore, this more subtle effect is difficult to assess visually. The tiller samples were based on three to five tillers selected from a 'representative' sample from the plot and are by their nature a representation of only a small subset. However, the consistent quantification of spike compaction observed across all sites and environments in the *Rht8* NIL suggests a robust effect. An alternative explanation might be that there is a pleiotropic canopy effect which is seen on a whole plot-basis, but is not necessarily gleaned from individual tillers. For example, the spike compaction might be more visible in the field when there is a reduction in tillers per unit area, since individual plants are more spaced. It was already shown in Chapter 3 that the tillering of the *Rht8* NIL was affected by N and water treatment. It would seem that several confounding factors might be involved.

Spike compaction in the *Rht8* NIL was quantitatively measured here for the first time. A robust, significant increase in spike compaction of 15% was observed across sites, water regimes and N treatments. The degree of compaction was unaffected by irrigation. The spike was more lax (decreased compaction) at the higher N treatment, but this was matched by the tall NIL response, therefore the proportional difference in compaction remained constant. For this reason, the treatment effect was due to environment rather than a differential genotypic response. It can be speculated that the differences in spike compaction between the NILs might in reality be greater than reported here, since extremely compacted spikes were not likely collected in tiller samples.

Spike compactness also has an effect on diseases associated with the spacing between florets on the rachis. Fusarium head blight (FHB) is a devastating fungal disease of hexaploid and durum wheat. The fungus infects the spike, causing production of mycotoxins and shrivelled grains, resulting in yield losses. A narrow flower opening width, likely associated with spike compactness, has been found to be associated with FHB resistance (Gilsinger et al., 2005). Therefore, the spike compactness in the *Rht8* NILs might be a breeding target to improve resistance to the disease.

The results presented in this Chapter indicate that spike length and spikelet number per spike are largely controlled by different genes. Additionally, in agreement with other studies (Faris et al., 2014c, Jantasuriyarat et al., 2004, Sourdille et al., 2000), despite spike compaction being a function of spike length and spikelets per spike, the differences in compaction were driven only by changes in spike length.

Interestingly, the background effects on height which made the tall NIL taller than Paragon (reported in Chapter 3), presumed to be due genotypic variances outside the 2DS introgression, were also found in the spike compaction data, but to a smaller degree. This might indicate that the traits are under different genetic control, but this is highly speculative.

Since spike compaction was significantly greater in the *Rht8* NIL compared to the tall NIL across environments, and was also less responsive to environmental effects than height, along with having seemingly smaller background effects, it was hypothesised that spike compactness was a useful score with which to type at the *Rht8* locus. The NILs at the BC_3F_3 stage could be easily distributed in a bimodal fashion according to the compact spike data. Thus spike compaction is a viable score with which to type recombinants at the *Rht8* locus and could be taken in conjunction with the more traditional mature plant height. However, from a practical standpoint, this would only be useful where height at maturity was unreliable and unavailable, and this must be balanced against the greater effort and delay to gather tiller samples and measure them. Furthermore, the genetic region encompassing *Rht8* and *C* requires clarification to determine which of the loci is influencing the compactness phenotype.

An attempt was made to discern the onset of the spike compaction much earlier in wheat development, at the early reproductive stage, using the recombinants to the fine-mapping *Rht8* population in the Cappelle-Desprez background. There was a large range of spike length that likely caused noise in the data and masked any possible compaction effect. This is likely first due to greater technical error with measuring small spikes, but second and more compellingly, due to the spikes being harvested according to chronological date rather than developmental time. The observation of compaction in the tip of the spike in the short recombinants suggests that measuring the spikes at the same developmental stage might yield significant results. Further, the observation also indicates that compaction is visible and measurable early on in development. This has not been reported before, but is valuable in determining the mode of action and timing of expression of *Rht8* and the surrounding genetic region. Further work in establishing spike compaction along a developmental time-series would be interesting.

The findings presented in this Chapter increase our understanding of the effect of spike compactness on agronomic traits. Even a relatively subtle compact phenotype produced quantitative, stable phenotypic differences. The background effects (albeit smaller ones) reported in spike compaction between the NILs corroborate findings in plant height in Chapter 3, and add importance to genotyping the NILs with a high-density array. The findings here are likely to become more significant with finer genetic dissection of *Rht8* and *C*, since currently, it has not been possible to discern unambiguously the precise contribution of *Rht8*, *C* and linkage blocks around those loci (assuming they are distinct) to the overall phenotype. Interestingly, a recently characterised *Rht* gene, *Rht23*, has been reported to control both spike compactness and dwarfing at a single locus (Chen et al., 2015). The markers for *Rht8* developed in Chapter 5 could greatly improve the density of the current mapping efforts of *C*, and, assuming they are polymorphic in the *C* mapping populations, provide a rapid way of resolving the location of *C* relative to *Rht8*.

Chapter 5 : Development of molecular markers within the *Rht8* interval

5.1 Introduction

Saturating the *Rht8* interval with molecular markers prior to fine-mapping, with the aim of identifying a marker co-localising with the *Rht8* phenotype, is an underlying tenet of map-based cloning (Scheible et al., 2005). Previous efforts which mapped *Rht8* to a 1.29 cM interval utilised a comparative genetics approach relying on the synteny of sequenced cereal genomes (Brachypodium, rice and sorghum) to develop single-strand conformation polymorphism (SSCP) markers (Gasperini et al., 2012). Developments in technologies and wheat resources offer an exciting opportunity to expedite traditional map-based cloning efforts in wheat. However, evaluation of approaches to filter the vast amount of data from new bioinformatic approaches and resources is not commensurate with our ability to generate data and discover variation. In this Chapter, different strategies were used and evaluated to extract biological relevance in identifying genetically-linked variation to fine-map *Rht8*. The aim of the work presented in this Chapter was to develop markers to further fine-map *Rht8*, which could be released to breeders for validation, with the ultimate goal of deploying *Rht8* into wheat breeding programs. The utility of cutting-edge technologies and wheat resources will be evaluated and the challenges of data-filtering discussed.

Rht8 was delimited to a 1.29 cM genetic interval on wheat chromosome 2DS (Gasperini et al., 2012). In Gasperini's work, a fine-mapping population was developed from recombinant-inbred lines (RILs) (Korzun et al., 1998), which originated from crosses between Cappelle-Desprez (CD) and the *Rht8* donor, RIL4. This fine-mapping population (henceforth called FM recs) was further developed to a fourth generation of self-fertilised F4 recombinants. The *Rht8* target region was saturated with gene-based markers using syntenic intervals in Brachypodium and rice and *Rht8* was mapped to a 1.29 cM interval on

chromosome 2DS. Monomorphism between markers prevented further map-based cloning and showed that polymorphism between the parent near-isogenic lines (NILs) to the original mapping population (RIL4 and CD) is low (Gasperini et al., 2012).

Emerging new technologies have revolutionised molecular breeding (Bernardo et al., 2008). Next-generation sequencing (NGS) approaches, including NGS on mRNA samples (RNA-Seq), are accelerating gene discovery (Schneeberger, 2014, Schneeberger and Weigel, 2011). Several strategies have been published in Arabidopsis (James et al., 2013). The success of these relies on a completed genome sequence in a (model) diploid organism. NGS technologies are currently underexploited in wheat due to the challenge of the large, ~17 Gb genome-size (Shewry, 2009) and the highly-related (96-98%) (Krasileva et al., 2013) A, B and D homoeologous genomes which comprise the 42 chromosomes of hexaploid wheat (6*n* = AABBDD).

Bulked segregant analysis (BSA) (Michelmore et al., 1991) is a technique that can be combined with RNA-Seq to target single nucleotide polymorphisms (SNPs) within a particular genetic interval. Two pools of individuals from a population segregating for a specific phenotype are compared, allowing identification of allelic variation from one of the parents to the population which is enriched in the appropriate bulk. The pools are 'bulked' since a number of individuals (and hence recombination events) comprise each pool. Excitingly, BSA has been combined with RNA-Seq to identify SNPs in targeted genetic intervals in tetraploid (Trick et al., 2012) and hexaploid wheat (Ramirez-Gonzalez et al., 2014). The SNPs generated were then used to fine-map to a 12.2 cM and 0.77 cM interval, respectively. This approach was also used in this Chapter, using pipelines developed by Martin Trick and Ricardo Ramirez-Gonzalez to detect SNPs between the parent NILs and then between the short and tall bulks.

Identification of SNPs between wheat varieties has recently been expedited with large-scale capture of allelic variation on high-density SNP arrays, such as the iSelect array with 90,000 (90K) SNPs (Wang et al., 2014a), and Affymetrix Axiom® 820K feature SNP array (www.cerealsdb.uk.net/cerealgenomics). These arrays have a predefined set of allelic variants against which the probes are designed. Despite relative under-representation of allelic variation on the D-

genome in both these arrays, markers in the *Rht8* interval were developed and successfully validated, demonstrating excellent potential for fine-mapping.

Developments in cereal genomics during the course of this project have provided an exciting opportunity to utilise sequence information from related plant species as well as wheat itself, to further fine-map *Rht8*. Sequence information of barley (IBGSC, 2012) has been published and the wheat D and A progenitors have been sequenced (Jia et al., 2013, Ling et al., 2013) as well as the hexaploid wheat 3B chromosome (Choulet et al., 2014). Sequence from flow-sorted hexaploid wheat chromosome arms was released in version 1.0 of the International Wheat Genome Sequencing Chromosome Survey Sequence (IWGSC CSS) (IWGSC, 2014). This sequence information was used to predict gene models in tetraploid and hexaploid wheat homoeologues (Krasileva et al., 2013). The first IWGSC version has been very recently improved with more variation data from various sources and population sequencing (POPSEQ) (Mascher et al., 2013) and released as IWGSC v2.0 (IWGSC-2) (plants.ensembl.org). In addition, a whole-genome shotgun (WGS) approach has yielded scaffolds of each of the three homoeologous genomes, covering new sequence space not completely overlapping with the IWGSC CSS contigs (Chapman et al., 2015). Very recently, the same POPSEQ map was used to genetically anchor these scaffolds (Mascher et al., 2013, CerealsDB, 2015a). These developments were integrated into marker development presented in this Chapter.

Integrating existing cereal genome sequences such as Brachypodium and rice, with new wheat sequence resources is challenging, given that many of the genome assemblies vary in sequence contiguity and annotation. Utilising the most recent cereal sequence information has also been difficult since data has been deposited in different servers with varying levels of access and user-friendliness. In the last couple of years, *Ensembl*Plants has integrated Triticeae resources into a genome browser which allows for visualisation and download of genomic information (Bolser et al., 2015). The BioMart toolkit (Kasprzyk, 2011) in *Ensembl*Plants also enables retrieval of sequence information across related species. This Chapter presents how wheat 2DS sequence, including intergenic sequence, was mined using BioMart in several iterations for marker development in the *Rht8* interval as IWGSC v 1.0 and 2.0 were released.

Wheat 2DS sequence from other sources was also exploited to discover polymorphism between the parents to the *Rht8* fine-mapping population. The gene-based reference used in the RNA-Seq approach (v3.3 cDNAs) was utilised to extract wheat sequence which could be mined for single-sequence repeats (SSRs). Sequence from a commercial partner (Limagrain) was used in a similar way. Combined marker discovery using SSRs and SNPs has been demonstrated in wheat (Lu et al., 2015) and a similar strategy was used in work presented here.

The design of genome-specific genetic markers in wheat is essential due to the hexaploid nature of the genome. Most primer-design tools are designed on diploid species thus a common approach is to align manually A, B and D sequence and identify a SNP to enable genome-specific primer design. This is time-consuming and manual sequence-scanning is prone to human-error. PolyMarker is a fast polyploid primer-design pipeline, recently developed, which was implemented in this Chapter (Ramirez-Gonzalez et al., 2015).

Physical mapping in wheat was first used to construct a flow-sorted 3B-specific Bacterial Artificial Chromosome (BAC) library (Safar et al., 2004). BACs carry large DNA fragments and are relatively immune to chimerism and insert rearrangement, hence BAC libraries are widely used for gene isolation. Given the recent advances in chromosome flow-sorting in plants, (reviewed in Dolezel et al., 2014) during this project, the process constructing a 2D-specific BAC library was initiated with Jaroslav Doležel and colleagues from the short parent NIL to the *Rht8* fine-mapping population (RIL4), with the aim of allowing precise gene isolation of *Rht8*. During this process, DNA was sorted from the single chromosome arm of 2D (Vrana et al., 2012). In a method first demonstrated in barley, DNA amplified from recovered flow-sorted chromosome fractions can be used in marker development and fine-mapping (Simkova et al., 2008). A small amount of 2D DNA from RIL4 was obtained before BAC library construction was completed. In this Chapter this DNA was used in concert with the PolyMarker tool in SNP assays to validate D-genome specificity.

This Chapter describes efforts to comprehensively exploit the most recent resources in wheat for marker discovery with the aim of further fine-mapping *Rht8*. The workflow for marker discovery in this Chapter is shown in Figure 5.1. The identification of allelic variants in the parent NILs and from the BSA approach

is described, using manifold analysis of RNA-Seq data, SNP platforms and the mining of wheat 2DS sequence as it became available for SSRs. The different strategies used to filter the large datasets for high-confidence variants are discussed and evaluated. It is shown here how the *Rht8* interval was targeted using genetic and physical data. Finally, marker validation prior to fine-mapping is discussed. These markers are the basis of the fine-mapping described in Chapter 6.

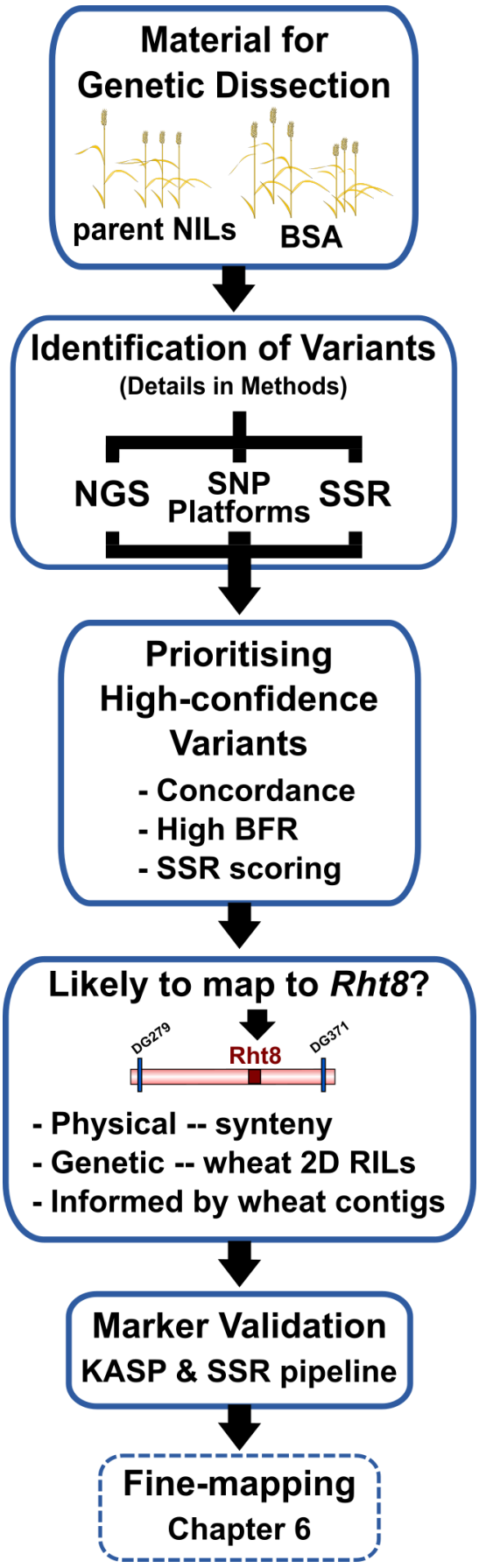


Figure 5.1: Schematic diagram of the workflow presented in Chapter 5 to generate markers targeted to the Rht8 interval.

5.2 Material for Genetic Dissection

The short and tall parent NILs (RIL4 and CD) were grown with the 73 recombinants to the fine-mapping population in the glasshouse in 2012. RIL4 was originally selected from a population of single-chromosome recombinant inbred lines (RILs) developed on wheat 2D (Korzun et al., 1998). The 2D chromosome of Mara, the *Rht8* donor, was substituted into a CD background via a series of back-crossing, to create a population of 89 RILs carrying 2D chromosome recombination events in an otherwise isogenic CD background (Korzun et al., 1998). RIL4 was selected by Gasperini as the short parent from this 2D RIL population to create a fine-mapping population (Gasperini et al., 2012) on the basis of carrying the (diagnostic for *Rht8*) 192 bp allele of *Xgwm261*.

Developing spike and elongating peduncle tissue was harvested destructively during stem elongation (GS 30 – 39) (Figure 5.2A), on multiple plants of the same genotype. *Rht8* acts to reduce cell size in the peduncle and uppermost internodes (Gasperini et al., 2012) thus it was hypothesised that *Rht8* would be expressed during this growth stage in the selected tissue. In order to maintain sampling consistency for the RNA extracted for RNA-Seq, samples were selected from the middle of the distributions of length of spike (mean = 2.5 mm) and peduncle (mean = 15 cm) (Figure 5.2B).

RNA-Seq was performed in two sequencing runs (described in Chapter 2, summarised in Table 5.1). In the first stage, samples from two biological replicates for each parent NIL and tissue were sequenced. In the second stage, the mRNA from spike tissue in the middle of the spike-length distribution (Figure 5.2A) of nine short and nine tall recombinants was pooled (Figure 5.2D). These short and tall recombinants were initially identified in the extremes of the height distribution of the 73 glasshouse-grown recombinants (Figure 5.2C). A subset of 32 recombinants from the extremes was selected for a further highly-replicated (N=24) glasshouse experiment in order to confirm the phenotype (Figure 5.2D, Appendix to Chapter 2). The strategy was to achieve higher coverage of the bulks as opposed to independent biological replicates of the recombinants within the bulks themselves. The nine recombinants within each bulk were subdivided into three libraries due to the concern of phenotyping accuracy in suboptimal glasshouse conditions, as reported in Chapter 6. The recombinants were

selected in confidence intervals of phenotype, such that the three most extreme shorts/talls were in one pool and subsequent pools consisted of recombinants with heights closer to the middle of the distribution (Appendix to Chapter 2). The six samples were randomised (Table 5.1) to avoid lane bias.

iSelect samples were prepared in two stages in a similar manner to the RNA-Seq samples. First, DNA from the parent NILs along with Mara, the original *Rht8* donor was sent for analysis. Second, in a separate array analysis, three extreme short and three extreme tall individuals from the fine-mapping population were sent for genotyping, along with pooled short and tall samples comprised of the three individuals of each phenotype.

For the Affymetrix Axiom® SNP array, DNA from the short *Rht8* x Paragon NIL and Paragon as control (detailed in Chapter 2) was sent to be analysed.

Sequencing stage	Lane	Genotype	Tissue	Sample	Raw reads
1	1	CD	Spike	P1	120915587
		RIL4	Spike	P2	94493958
	2	CD	Spike	P3	120731082
		RIL4	Spike	P4	83981466
	3	CD	Peduncle	P5	118647743
		RIL4	Peduncle	P6	87078263
	4	CD	Peduncle	P7	112118008
		RIL4	Peduncle	P8	94694351
2	5	Short bulk	Spike	B1	62007005
		Tall bulk	Spike	B2	76466586
		Short bulk	Spike	B3	75119011
	6	Tall bulk	Spike	B4	62430624
		Short bulk	Spike	B5	74208678
		Tall bulk	Spike	B6	75472124

Table 5.1: Details of the samples used for RNA-Seq in two stages: experimental design of multiplexing and randomisation across lanes, and the number of reads achieved per sample. The lane count used is 1-4 for the first sequencing stage and continues 5-6 for the second stage for clarity.

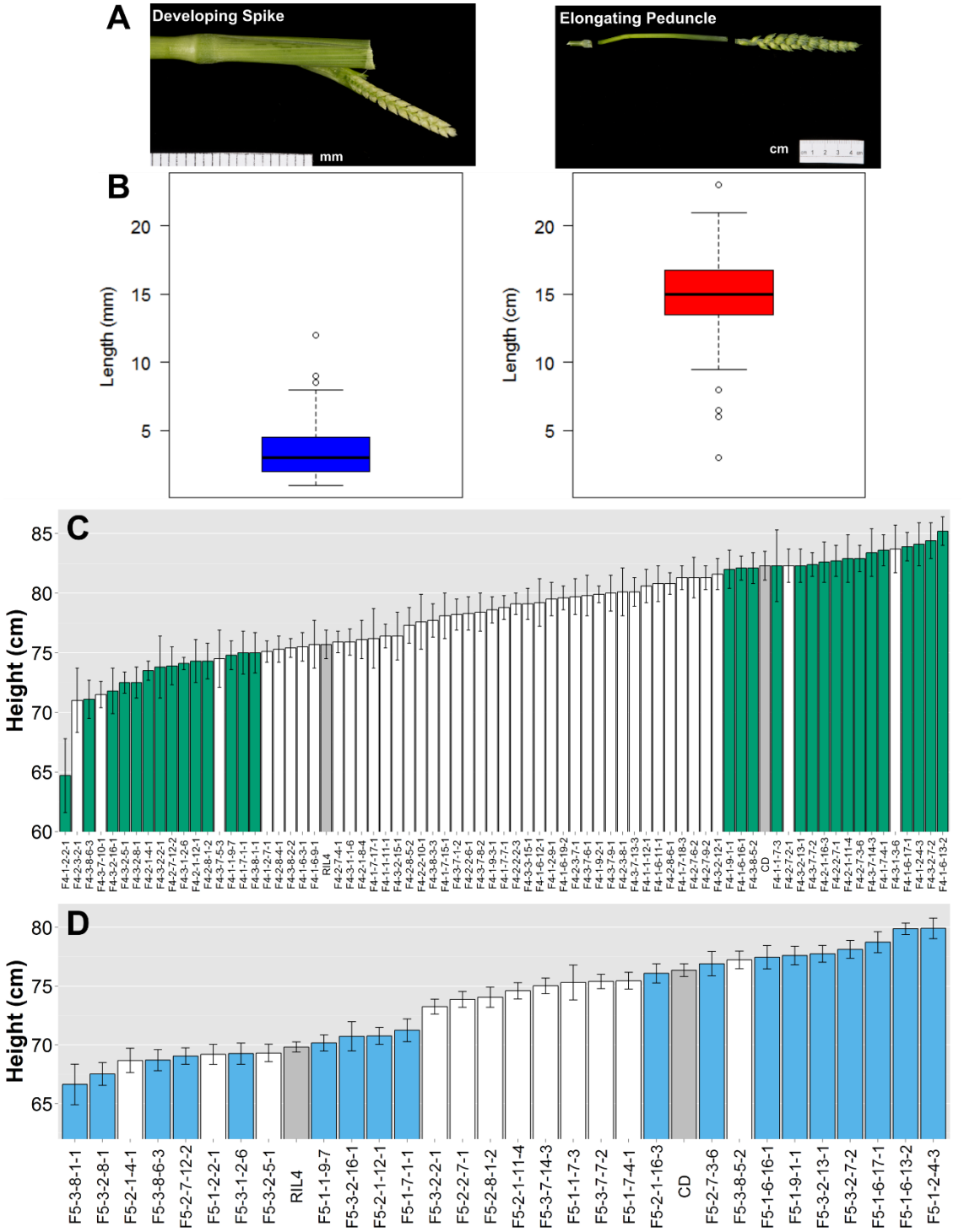


Figure 5.2: Sampling of tissue from spike and peduncle (A) Tissue sampled from the two organs indicated for each genotype during stem elongation (B) Multiple biological replicates were collected for each genotype and measured. Tissue from the middle of the distributions of the lengths all the biological replicates from the developing spike (left, N=85) and elongating peduncle (right, N=80) were selected for RNA-Seq (C) The heights of the 73 fine-mapping recombinants in the glasshouse, 2012-13, in height order (N=8). The recombinants coloured in green were extreme short and talls (a total of 30: 14 short and 16 tall) selected for a further glasshouse experiment in spring 2013. The grey bars indicate the parent NILs (N=16) (D) The subset of recombinants (N=24) from the 2012-13 glasshouse experiment (ordered by height) grown to validate heights for bulks for RNA-Seq. The individuals in blue were selected for short and tall bulks. The grey bars are the short and tall parent NILs (N=63). Details of the individuals allocated to each library within the short/tall bulks is in the Appendix to Chapter 2.

5.3 Identification of Variants

5.3.1 Combining SNP and microsatellite variation

Molecular genetic markers exploit nucleotide variation to study heritable traits and can be applied to diversity analysis, association studies, marker-assisted selection and genetic mapping (Duran et al., 2009). Molecular marker technology in wheat has advanced rapidly over recent times, as outlined in Chapter 1. SSRs and SNPs predominate among markers currently applied in genetic analysis. The molecular basis of SSR and SNP polymorphisms is quite different. SSRs are caused by replication slippage and can be multi-allelic in nature whereas SNPs are generated via point mutation and are bi-allelic (Duran et al., 2009). Therefore, the different marker systems capture different genetic variation, as was confirmed in a recent study comparing an SSR panel with the 90K SNP array used in this Chapter (Jiang et al., 2015b).

Combining the two marker systems for novel marker discovery and fine-mapping has been reported, for example in mapping a powdery mildew resistance gene in *Agropyron cristatum*, a perennial Triticeae species (Lu et al., 2015). In other work, when *de novo* transcriptome data in *T. monococcum* was mined for SNP and microsatellite sites, the overlap when measured by counting the sampled barley genes was relatively low, despite SNP discovery outnumbering SSR sites over 20 times (Fox et al., 2014). Taken together, it was decided to identify both SNPs and microsatellites in this project to capture different variation which might be underexploited if relying only on the SNP marker system.

5.3.2 Targeting genome-specific allelic variation

Identification of allelic variation in wheat is not trivial, given the highly-related (96 – 98%) (Krasileva et al., 2013) A, B and D genomes. Correct identification of variants on 2DS as opposed to the homoeologues was important for marker development in this project. Different strategies and resources were used to attempt to do this *in silico* and *in vivo*.

In silico genome specificity was attempted wherever possible prior to marker validation. The release of IWGSC-1 (IWGSC, 2014) generated genome-specific

contigs by flow-sorting individual chromosome arms prior to sequencing. In identifying SSR variants in this project, only sequence from IWGSC CSS contigs mapping to 2DS, as opposed to 2AS or 2BS, were used.

The limitation of the IWGSC-1 scaffold information (and the gene models based upon them) is that robust genome assignation, for example by a BLASTN homology search, is limited by the availability of all three homoeologues. For example, in most cases SNPs mapping to a 2DS CSS contig with ≥99% nucleotide identity were prioritised, since those with lower identity matches (97 – 99%) might imply one of the homoeologous genomes is missing (in the cases where such a hit is returned as the best hit by homology in a BLASTN search). Therefore, in filtering variation, group 2S SNPs were considered wherever possible. However, the concern of excluding SNPs mapping to 2AS or 2BS where the 2DS CSS contig was absent had to be balanced with the need to prioritise SNPs more discriminately.

The bioinformatic tool, PolyMarker, was used to increase the likelihood and efficiency of designing genome-specific markers (Ramirez-Gonzalez et al., 2015). PolyMarker aligns the three homoeologous wheat genomes around the target sequence SNP sequence, using IWGSC-1 scaffolds. An output identifies a SNP as homoeologous (which could be discarded) or varietal. PolyMarker then generates KASP primers by incorporating the varietal SNP (in this case between CD and RIL4) into the 3’ end of the VIC and FAM primers, whereas the common primer is designed to be genome-specific based on a SNP which can discriminate between all three genomes (on the 3’ end of the sequence). The success of designing genome-specific markers again relies on sequence from all three genomes being present in the alignments. This limitation is explored in depth in 5.6.1. Where SNP availability was low, there were cases where semi-specific (amplifying the D-genome and one other homoeologue) or non-specific markers were used (amplifying all three homoeologues). Despite these limitations, pursuing allelic SNPs as a priority based on the (limited) sequence information available was a cost- and time-effective strategy.

In vivo, targeted genome analysis was facilitated initially by validating markers with nulli-tetrasomic Chinese Spring DNA (in the case of SSRs). The complete set of nulli-tetrasomic DNA on chromosome 2 was used. The null 2D DNA (2DA and 2DB) showed no amplification (data not shown). Later, specificity to the 2DS genome for both SSR and KASP markers was tested with amplified 2D flow-sorted DNA from RIL4 (Simkova et al., 2008). The purity of the sorted fraction was reported as 94.44%, with contamination mainly from chromosome 7D. The DNA was tested further by using KASP markers amplifying DNA from different chromosomes (including 2A, 2B and 7D), which were known to be polymorphic between Mara and CD. A total of 15 markers were used to genotype the 2D DNA (details in Chapter 2). Most of the markers tested (12 out of 15) failed to amplify the 2D DNA, with the 2D DNA clustering with the no-template control (Figure 5.3A). However, three out of 15 markers, comprising two markers on 7D and one marker on 5A (Figure 5.3B), amplified the 2D DNA. This suggested that the contamination in the estimated 5% of the sample (which mainly consisted of 7D) was sensitive to amplification using the KASP system. The amplification from a small proportion of markers outside 2D indicated that assessing genome specificity from the flow-sorted DNA had to be done with caution. However, crucially, the markers on 2A and 2B did not amplify the flow-sorted DNA, indicating that the DNA was a useful resource to assess whether markers designed on 2DS were genome-specific.

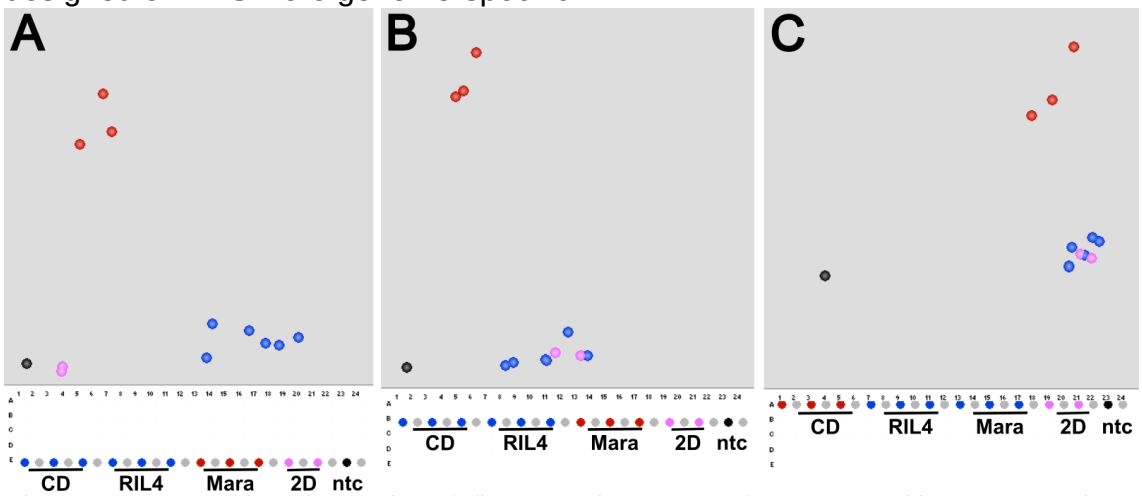


Figure 5.3: Assessing the purity of flow-sorted 2D DNA from RIL4 with KASP markers. Representative results are shown (A) 12 out of 15 markers previously mapped to chromosomes outside group 2 showed no amplification of 2D DNA (B) Three out of 15 markers tested amplified the 2D DNA (C) A positive control of a 2DS-specific marker designed on a RIL4/CD SNP. 'CD' = Cappelle-Desprez, tall parent NIL to the fine-mapping population; '2D' = flow-sorted DNA from chromosome 2D of RIL4, the short parent NIL to the fine-mapping population; 'ntc' = no template control. Coloured red on y-axis is the FAM-labelled adapter, coloured blue on the x-axis is VIC, coloured pink is the 2D DNA.

5.3.3 Identifying SNP variation in NGS data

SNPs were identified in two stages: first in the parent NILs using the customised UniGene reference (Harper et al., 2015) and second in a BSA approach, using the best gene-based in-house reference available at the time, the v3.3 cDNAs.

The reads generated from the parent NILs were aligned to a customised wheat reference as described in Chapter 2. The objective following the first RNA-Seq stage of the parent NILs was to identify putative SNPs between the short (RIL4) and tall (CD) parent NILs (Table 5.1) and proceed along the workflow in Figure 5.1. SNPs between RIL4 and CD and representing allelic variation (as opposed to inter-genome SNPs between homoeologous genomes or varietal SNPs between CD and the reference), were called between each sample (P1 – P8, Table 5.1) and the UniGene reference by Martin Trick (Chapter 2).

The SNP-calling process identified a total of 60,454 putative SNPs between any of the eight samples and the reference across 32,663 unique UniGenes (80% of the reference set) (Table 5.2). Most of the SNP-containing UniGenes could be aligned to the CSS contigs. A proportion of the UniGenes (68 – 82%) could also be annotated with information from at least one of barley, Brachypodium or rice (Figure 5.4).

RNA-Seq reads from a CD sample used by another research group were used in the initial SNP calling as a control, but the SNPs found between this sample and those from the CD samples (Table 5.1) were incongruent with each other. The extraneous CD sample had 14,283 SNPs not found in any of the CD samples. The DNA from the extraneous sample could not be obtained for typing with markers which would enable unambiguous comparison of genetic background. Therefore, this sample and all SNPs detected using those reads were discarded from further analysis. This finding highlights the importance of genotypic screening to identify errors in germplasm selection prior to sequencing (as was carried out for all samples in this project) and maintenance of pure genetic stocks between research groups, to avoid the costly error of sequencing incorrect material.

In the second stage of SNP identification in a BSA approach, the short and tall bulks (Table 5.1) as well as the previously generated parent NIL reads were

aligned to an in-house v3.3 cDNA reference. The v3.3 cDNA reference was compiled by Martin Trick and described in Chapter 2. In essence, the v3.3 cDNA reference comprised a non-redundant set of 75,419 gene models anchored to IWGSC CSS contigs. The v3.3 cDNAs were partitioned into an ordered and unordered section based on whether gene models could be anchored on the CS x Paragon map. Ordered gene models were given a chromosome position, those which could not be anchored due to lack of polymorphism in the mapping population were assigned to the 'unordered' bin (remaining assigned to chromosome arm only according to the CSS).

In the BSA approach, the objective following the RNA-Seq of the short and tall bulks was to identify SNPs that were enriched for the parental allele in the corresponding bulk i.e. SNPs found in the short parent also present in the short bulks and vice versa for the tall parent/bulk combination. To achieve this, first the parent NIL and bulk samples were aligned to the v3.3 cDNA reference (Chapter 2). Due to concern over typing the plants as short or tall in suboptimal glasshouse conditions (explained in 6.2.2), the reads from each of the bulks were separated into three libraries (Table 5.1). However, by the time the samples were sequenced, experimental data had validated the phenotyping (6.2.3). Consequently, all short reads were merged *in silico* into one bulk and all tall reads merged into a tall bulk. The short and tall parent NIL reads from biological replicates and tissues were also merged. This strategy had been demonstrated previously as an effective SNP-calling strategy to score SNPs in genes with relatively lower expression (Ramirez-Gonzalez et al., 2014).

The aligned files were passed through a pipeline for SNP discovery using bulk frequency ratios by Ricardo Ramirez-Gonzalez (described in full in Ramirez-Gonzalez et al., 2014). Briefly, first varietal SNPs were identified between the parents. Then, for each bulk, the frequency of the base at each SNP position was calculated and the bulk-frequency ratio (BFR) between bulks determined. The BFR provides a relative measure of SNP enrichment in both bulks, which is normalised for coverage. A high BFR indicates that the allelic variation is contributed from one bulk and absent in the other bulk.

SNP-calling in NGS data				%		
Reference	**Dataset**	**Total putative SNPs**	**No. of genes with SNPs**	**A**	**B**	**D**
UniGenes	parent NILs	60,454	32,663	33	47	20
v3.3 cDNAs, BFR>6	parent NILs	90	50	90	3	7
	BSA	31,350	7251	32	40	28
2D v3.3 cDNAs, 59 genes	parent NILs	401	51	-	-	100
	BSA	388	47	-	-	100

SNP-calling in SNP-platform data

Platform	Dataset	Total	Without NC	%
iSelect 90K	parent NILs	1557	412	26
	BSA	970	314	32
Axiom® 820K	*Rht8* NIL vs Paragon	56114	6089	11

Table 5.2: SNP-calling results.

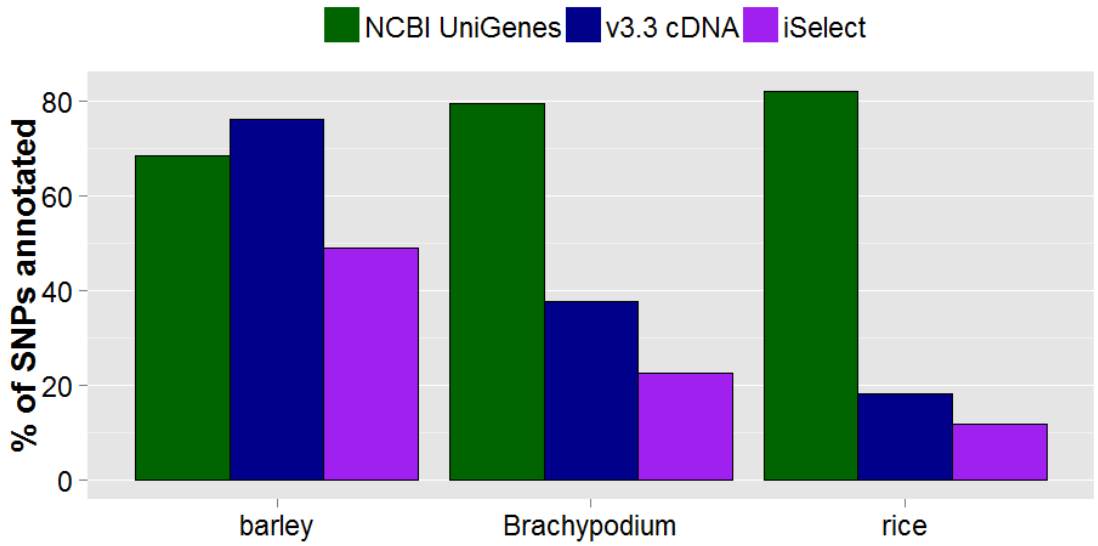


Figure 5.4: Proportion of putative SNPs which could be annotated with syntenic information from the different references and SNP platform. SNPs with an annotation from any of the syntenic species were counted as a percentage of the total SNP set.

When the highest stringency of 100% was applied to identify varietal SNPs (i.e. 100% of the bases in the reads of one parent differed to the reference at the SNP position – explained further in 2.3.7.2), total of 90 putative SNPs over 50 unique genes was identified (Table 5.2 and Appendix 2.8). Most of these mapped to chromosome 1A. Crucially, none of these stringent SNPs were located on chromosomes of group 2S, where the *Rht8* introgression is located.

Due to concern of discarding potentially informative SNPs from downstream analysis resulting from homoeologue miss-assignment in the IWGSC CSS scaffolds, as well as no group 2S SNPs in the most stringent varietal SNPs, the

threshold to identify varietal SNPs was lowered to 20%. This threshold had proved successful in SNP discovery at another locus in hexaploid wheat (Ramirez-Gonzalez et al., 2014). The SNPs identified in this way were used for all further downstream analysis. The BFR SNP-calling with the 20%-threshold varietal SNPs identified 31,350 putative SNPs over 7,251 unique genes (Table 5.2). The SNP-containing genes were aligned to the IWGSC CSS contigs to map them to a chromosome arm. Most SNPs could be annotated with a barley physical position, though only a small proportion of genes could be annotated with information from Brachypodium and rice (Figure 5.4).

To specifically target the Rht8 interval, a second iteration of BSA was performed using a narrowed 2D interval on the ordered section of the v3.3 cDNAs. The 2D interval was narrowed using the pre-existing information from the fine-mapping of *Rht8*, using the flanking markers DG279 and DG371 (Gasperini et al., 2012) (described in Chapter 2 and 5.3.5). The unordered 2D contigs were not considered, sincthe position was unknown. SNPs were called in this narrowed interval (herein termed 2D v3.3 cDNAs) using VarScan 2.0 (Koboldt et al., 2012) and additional customised steps (described in Chapter 2). SNP-calling identified 401 putative varietal SNPs between the parent NILs and the reference and 388 putative varietal SNPs between the bulks (Table 5.2).

5.3.4 Identifying SNP variation in SNP platform data

SNP discovery between wheat varieties has recently been propelled by advances in high-density arrays to capture pre-defined allelic variation. Two such arrays used in this project are the iSelect array with 90,000 (90K) SNPs (Wang et al., 2014a) and the Affymetrix Axiom® 820K SNP array (www.cerealsdb.uk.net/cerealgenomics). These arrays are used widely for genome-wide association (GWAS) and diversity studies (among others Ishikawa et al., 2014 and Zanke et al., 2014) since they incorporate a large number of wheat varieties with genome-wide marker coverage. However the arrays also offer opportunity for fine-mapping and have increasingly been exploited to this end (Babiker et al., 2015, Knight et al., 2015). Only one analysis has been published very recently which applied the BSA approach to the iSelect 90K array (Lu et al., 2015).

In this project, the iSelect 90K array was used in two analyses to genotype first the parent NILs and second, the short and tall bulks in a BSA approach. The Axiom® 820K array was used to genotype the short *Rht8* x Paragon NIL and Paragon (the provenance of this is described in Chapter 3).

Data from the arrays is interpreted by a polyploid version of GenomeStudio (Wang et al., 2014a). The data is displayed as 'AA', 'AB' or 'BB' calls. Where data is missing due to low signal, an 'NC' is returned. This does not allow for discrimination between missing data due to a deletion or technical limitation. Heterozygous 'AB' data is likely caused by inter-genome hybridisation due to a non-specific assay for that marker (since the array should capture allelic variation only). The filter in GenomeStudio identifies data in spatial clouds, with datapoints distributed within these clouds. Thus it cannot be unambiguously ascertained whether an 'AB' call is a true polymorphism relative to 'AA' or 'BB'. In published analyses, typically markers which return missing values or show ambiguous SNP calls are discarded (Babiker et al., 2015, Lu et al., 2015). In this project, a conservative approach was adopted throughout, retaining putative SNPs based on 'AB' calls.

From the first iSelect analysis of the parent NILs, SNPs were found between RIL4 and CD genome-wide, including 'AB' calls. Most of this number was due to missing data ('NC') from one of the genotypes being called with respect to the other, leaving 412 SNP variants (Table 5.2). From the total probes on the iSelect array (~81K), only 11% were mapped to an Avalon x Cadenza map. From this subset, there was a dearth of SNPs between the parent NILs genome-wide (Appendix 5.1). Therefore, it was not possible to simply use the pre-existing genetic map to target SNPs on 2D. In the second iSelect analysis of the BSA, there were 970 SNPs between the pooled DNA of the short and tall bulks, and a third were retained when eliminating SNPs based on missing data (Table 5.2). The SNPs on the iSelect array were annotated with orthologues (Figure 5.4) and also designated a position on the in-house wheat pseudomolecules, as well as annotated with the corresponding CSS contig. Later, the array data was improved with the chromosome, arm and cM position on the wheat chromosome following the publication of Wang et al., 2014.

Genotyping Paragon and the *Rht8* x Paragon NIL on the Axiom® 820K array identified 6089 variants once the SNPs due to missing data were discarded (Table 5.2).

5.3.5 Mining for SSRs in wheat sequence

5.3.5.1 Identifying microsatellites

SSRs are short stretches of tandem repeats of mono-, di-, tri-, tetra-, penta- and hexa-nucleotides, which can be interrupted by non-repeat nucleotides, or found adjacent to each other. Several computational tools are available for the identification of SSRs in sequence data, reviewed by Duran et al., (2009). Most of these tools require a pre-defined repeat-length. Here, WebSat (Martins et al., 2009) was used to find microsatellites with parameters set to identify motif length from mono- to hexa-nucleotide (full details in Chapter 2). SSR markers were designed on wheat sequence from multiple sources, outlined below. Crucially, intergenic sequence could also be considered in this way, circumventing the limitation of using genic references in SNP discovery. Different types of variation between the parent NILs were identified: presence/absence of a peak (Figure 5.5A) or variation in peak size (Figure 5.5B – D).

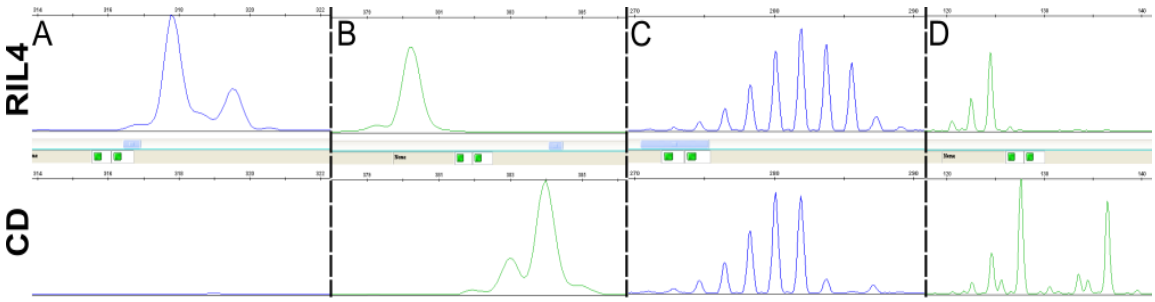


Figure 5.5: Identifying SSR variation as polymorphism between parent NILs. A: presence/absence of a peak; B: polymorphism in peak size; C: polymorphism in peak shape; D: polymorphism in number of peaks.

5.3.5.2 Utilising IWGSC data with syntenic *Rht8* intervals

Rht8 had been previously mapped to a 1.29cM interval between SSCP markers *DG279* and *DG371*. Good overall conservation of gene content (synteny) and order (collinearity) was reported in the syntenic intervals in Brachypodium and rice (Gasperini et al., 2012). Synteny has been used extensively in fine-mapping genes in wheat (Krattinger et al., 2009a). This approach was made more powerful

during the course of this project with new resources in barley and wheat being published. At the beginning of this project, a WGS assembly of the barley genome was released with gene models (IBGSC, 2012) followed by a genetic map (Mascher et al., 2013). At the time of performing the work, these resources were fragmented. The genes were first made available on the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) server (Deng et al., 2007) as 'high-confidence' (HC) and 'low-confidence' (LC) Munich Information Center for Protein Sequences (MIPS) gene models, as well as the WGS assembly of the cultivar Morex (herein referred to as Morex assembly). The IWGSC coordinated the flow-sorting (Safar et al., 2004) of chromosome arms from the cultivar Chinese Spring, followed by chromosome-by-chromosome shotgun sequencing and assembly into contigs of average size 2.5kb (IWGSC, 2014). The genome-specific wheat chromosome arm assemblies were made BLASTable (Altschul et al., 1997) on the Unité de Recherche Génomique Info (URGI) server (URGI, 2013). Later, *Ensembl*Plants integrated these Triticeae resources into a genome browser in release IWGSC v1 (IWGSC-1), which allowed for visualisation and download of genomic information (Bolser et al., 2015).

SSRs are highly transferrable between wheat varieties and even species (Duran et al., 2009). Therefore, unlike SNP markers, which might be expected to differ between Chinese Spring and the parents to the fine-mapping *Rht8* population, it was posited that mining Chinese Spring sequence for SSRs would transfer well to Cappelle-Desprez and RIL4.

	EST name	Brachypodium	Rice
DG279	TA44444_4565	Bradi5g03460	Os04g0132100
DG371	BJ307036	Bradi5g04710	Os04g0191400

Table 5.3: The previously anchored syntenic Rht8 intervals in Gasperini's work based on Brachypodium and rice. This information was the starting point of the marker development work in this project. Originally, the Brachypodium assembly used was Bd21 Genome Annotation v1.0; the Rice assembly was MSU 6.

The first step to mine wheat sequence in the syntenic intervals was to anchor the previously delimited *Rht8* interval (Table 5.3) using the most current Triticeae resources. This was updated as resources become available during this project. To achieve this, iterations of the BioMart toolkit (Kasprzyk, 2011) in *Ensembl*Plants were used in conjunction with other servers hosting barley (IPK) and wheat sequence information (URGI) (full details in Chapter 2) as well as in-house JIC resources. The *Rht8* interval was anchored in barley and the

Brachypodium and rice positions consolidated with the unified assembly in *Ensembl*Plants (Table 5.4). Table 5.4 was consolidated when new resources became available, is described in the appropriate place in the text. Crucially, the syntenic intervals were combined with IWGSC-CSS data in order to consider only those genes with evidence of 2DS localisation. This information was used to create genome zippers for each of the syntenic species (the three zippers are shown in Appendix 5.2). The genes in the zippers were annotated with the highest nucleotide identity hit to wheat 2DS contigs by BLASTN. This gave the best wheat 2DS sequence based on gene prediction from barley, Brachypodium and rice.

A total of 61 unique wheat 2DS contigs (some <200bp long) were found to correspond to the syntenic intervals across the three species. Within this sequence, 78 SSRs were identified (Table 5.5).

Resource	URGI	Ensembl Plants						barley			UC Davis		In-house					URGI
Assembly	IWGSC-1	IBSC-1.0 (082214v1)			v1.0		IRGSP-1.0	MIPS	WGS Morex		*Ae. tauschii*		NCBI UniGenes		v3.3 cDNA	Chapman		IWGSC-2
	Contig	% Id	barley	Pos	Bd		rice	HC	Identifier	bin (cM)	bin (cM)	BAC scaffold	UniGene	Pos	Best 2D identifier	Scaffold	bin (cM)	bin (cM)
DG279	2DS_5377037	96	MLOC_5957	2: 15601547	Bradi5g03460		Os04g0209200	MLOC_5957	contig_136914	12.11	30.994	ctg494	D_comp3873_c0	4337321	mrna093230 2D: 929999	1111368	13.642	17.34
	2DS_5343528	96																16.95

Resource	URGI	Ensembl Plants						barley			UC Davis		In-house					URGI
Assembly	IWGSC-1	IBSC-1.0 (082214v1)			v1.0		IRGSP-1.0	MIPS	WGS Morex		*Ae. tauschii*		NCBI UniGenes		v3.3 cDNA	Chapman		IWGSC-2
	Contig	% Id	barley	Pos	Bd		rice	HC	Identifier	bin (cM)	bin (cM)	BAC scaffold	UniGene	Pos	Best 2D identifier	Scaffold	bin (cM)	bin (cM)
DG371	2BS_5196588	99	MLOC_58453	2: 18522971	Bradi5g04710		Os04g0261400	AK375036	contig_1562744	15.44	37.978	ctg4363	D_comp225442_c0	5895457	mrna079612 2D: 1295398	2409583	15.917	17.34
	2DS_5389857	90	MLOC_12182	2: 18489177	Bradi5g04673													

Table 5.4: Anchoring of the Rht8 interval in the most current Triticeae resources and NGS references. The flanking markers DG279 and DG371 were anchored in the physical data available for barley, Brachypodium and rice, as well as wheat. Full details are in Chapter 2.

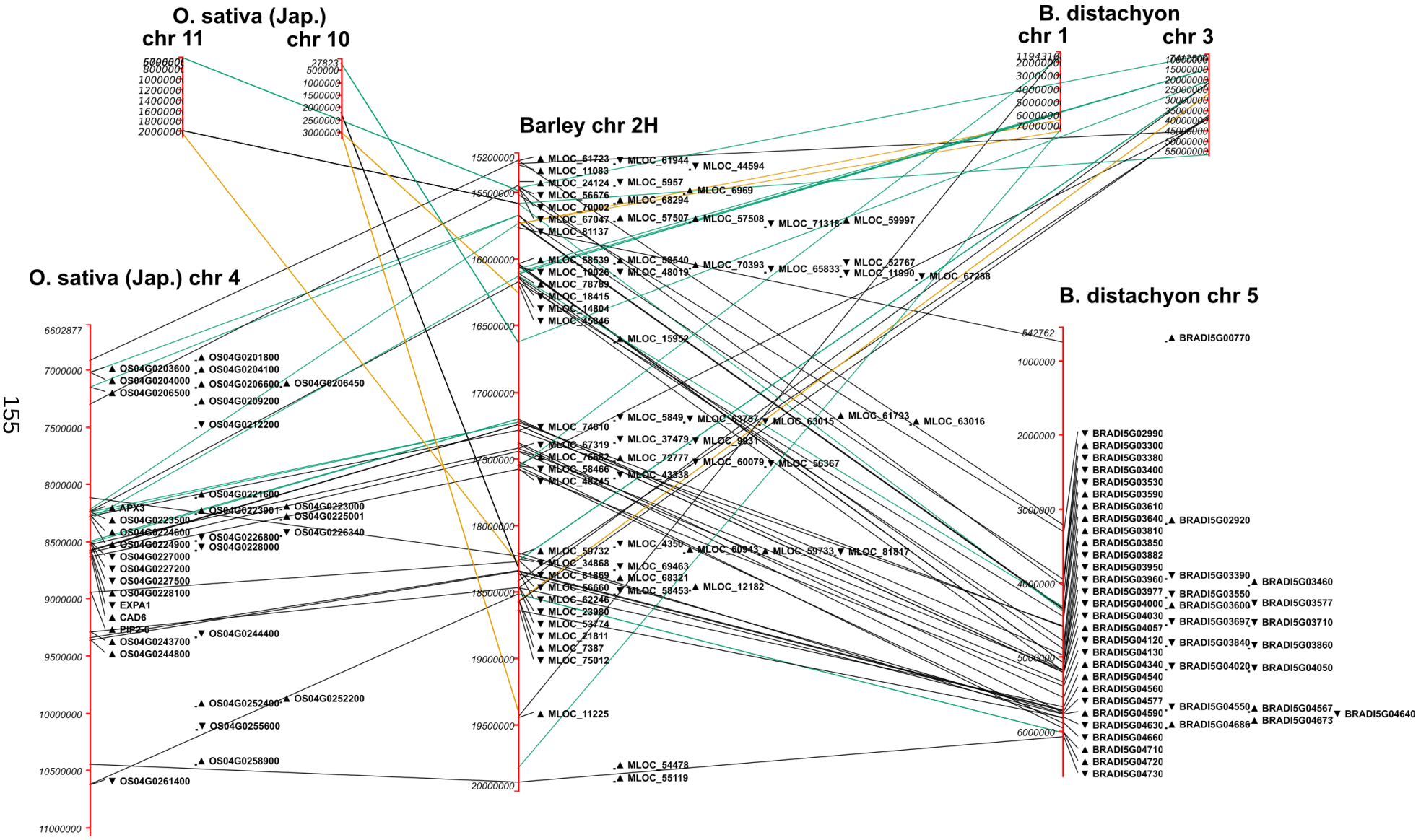
In order to examine synteny between comparative species, ArkMap (Paterson and Law, 2013) was used to show the orthologous relationships between the barley interval (Table 5.4), Brachypodium and rice (Figure 5.6).

The first observation was that the barley interval (Appendix 5.2.1) had no annotated HC genes from 2HS:16200000 – 17200000, 17800000 – 18600000 and 18700000 – 9400000, a total of ~2.5 Mb. The second observation was that the syntenic relationship between Brachypodium and rice, when considering those genes with orthology to barley, was not as good as previously described (Gasperini, 2010). The barley-to-Brachypodium interval contained orthologous genes on Brachypodium chromosomes 1, 3 and 5. The barley-to-rice interval had orthologues on rice chromosomes 4, 10 and 11. Overall, the barley-to-Brachypodium gene order was superior to barley-to-rice. There were considerably more one-to-one orthologues between barley and Brachypodium chromosome 5, compared with barley and rice chromosome 4 (these are both the syntenic chromosomes identified previously (Gasperini, 2010)).

Taken together, the data suggested that first, it was not prudent to consider only barley (the closest sequenced syntenic relative to wheat) to mine sequence space, since only wheat sequence space corresponding to barley HC genes was reported. There were extensive annotation gaps in the interval. To circumvent this, the ~ 300 WGS Morex assemblies in the 3cM space from 12.11-15.44cM, corresponding to the best-anchored *Rht8* interval (Table 5.4), were used as query sequences in BLASTN homology searches of CSS 2DS contigs. In this way, wheat 2DS sequence which could be anchored to the lower confidence barley genes was also considered for microsatellite variation. Second, the idiosyncratic nature of the break down in synteny across the different species suggested that considering wheat sequence independently from each of the comparative species was best to extract the maximum possible 2DS sequence.

Figure legend precedes the figure to facilitate maximum figure enlargement.

Figure 5.6: Synteny between the barley Rht8 interval and Brachypodium and rice. The barley interval on 2H (15200000 – 20000000) was used, as defined in Table 5.4 along with generous flanking margins with side. Conserved synteny was shown in Brachypodium and rice, based on orthologous relationship defined by a 125/200 combined similarity threshold. Orthologous relationships between genes are shown with lines as follows: black = one-to-one, green = one-to-many, brown = many-to-many. Genes on the non-syntenic chromosomes in Brachypodium and rice are not shown for clarity, instead physical position alone is indicated. ArkMap used EnsemblGenomes release 25. The assembly details are: barley IBSC-1.0, Brachypodium v1.0, rice Oryza Sativa Japonica IRGSP-1.0.


O. sativa (Jap.)
chr 11
chr 10
Barley chr 2H
B. distachyon
chr 1
chr 3
O. sativa (Jap.) chr 4
B. distachyon chr 5
APX3
EXPA1
CAD6
PIP2-6
OS04G0203600
OS04G0204000
OS04G0206500
OS04G0201800
OS04G0204100
OS04G0206600
OS04G0206450
OS04G0209200
OS04G0212200
OS04G0221600
OS04G0223901
OS04G0223000
OS04G0225001
OS04G0223500
OS04G0224600
OS04G0224900
OS04G0227000
OS04G0227200
OS04G0227500
OS04G0228100
OS04G0226800
OS04G0228000
OS04G0226340
OS04G0244400
OS04G0243700
OS04G0244800
OS04G0252400
OS04G0252200
OS04G0255600
OS04G0258900
OS04G0261400
MLOC_61723
MLOC_11083
MLOC_24124
MLOC_56676
MLOC_70002
MLOC_67047
MLOC_81137
MLOC_61944
MLOC_44594
MLOC_5957
MLOC_6969
MLOC_68294
MLOC_57507
MLOC_57508
MLOC_71318
MLOC_59997
MLOC_58539
MLOC_58540
MLOC_70393
MLOC_65833
MLOC_52767
MLOC_10026
MLOC_48019
MLOC_11990
MLOC_67288
MLOC_78789
MLOC_18415
MLOC_14804
MLOC_45846
MLOC_15952
MLOC_74610
MLOC_5849
MLOC_63787
MLOC_63015
MLOC_61793
MLOC_63016
MLOC_67319
MLOC_37479
MLOC_9931
MLOC_76682
MLOC_72777
MLOC_60079
MLOC_56367
MLOC_58466
MLOC_43338
MLOC_48245
MLOC_59732
MLOC_4350
MLOC_60943
MLOC_59733
MLOC_81817
MLOC_34868
MLOC_69463
MLOC_61869
MLOC_68321
MLOC_56660
MLOC_58453
MLOC_12182
MLOC_62246
MLOC_23980
MLOC_53774
MLOC_21811
MLOC_7387
MLOC_75012
MLOC_11225
MLOC_54478
MLOC_55119
BRADI5G00770
BRADI5G02990
BRADI5G03300
BRADI5G03380
BRADI5G03400
BRADI5G03530
BRADI5G03590
BRADI5G03610
BRADI5G03640
BRADI5G02920
BRADI5G03810
BRADI5G03850
BRADI5G03882
BRADI5G03950
BRADI5G03960
BRADI5G03390
BRADI5G03460
BRADI5G03977
BRADI5G03550
BRADI5G04000
BRADI5G03600
BRADI5G03577
BRADI5G04030
BRADI5G03697
BRADI5G03710
BRADI5G04057
BRADI5G04120
BRADI5G03840
BRADI5G03860
BRADI5G04130
BRADI5G04340
BRADI5G04020
BRADI5G04050
BRADI5G04540
BRADI5G04560
BRADI5G04577
BRADI5G04590
BRADI5G04550
BRADI5G04567
BRADI5G04640
BRADI5G04630
BRADI5G04686
BRADI5G04673
BRADI5G04660
BRADI5G04710
BRADI5G04720
BRADI5G04730

5.3.5.3 Extending the sequence space searched with new wheat resources

In 2015, at the end of this project, *Ensembl*Plants (in release 26, March 2015) incorporated updated wheat sequence as IWGSC v2 (referred to herein as IWGSC-2). At the time of writing, the most recent release (June 2015) re-named this to IWGSC-1+POPSEQ, but the former name will be used here, in line with when the information was accessed. The release refined the CSS contigs by mapping them into chromosome genetic bins using wheat POPSEQ data (Mascher et al., 2013) and aligned additional datasets such as barley, Brachypodium and rice to the CSS. Additionally, a genome zipper based on IWGSC-2 and 90K iSelect array data (Wang et al., 2014a) (shown in Appendix 5.2) was compiled by MIPS and was downloaded from URGI in March 2015 (URGI, 2015a).

Further, in 2015, a WGS *de novo* assembly of a synthetic hexaploid wheat (cultivar W7984) was published (herein named Chapman assembly) (Chapman et al., 2015). The Chapman assembly compared similarly to the CSS in terms of total genome assembled (9.1Gbp and 10.1 Gbp, respectively). However, of particular interest in this project, the Chapman assembly had much better contiguity, with average contig size more than double that of the IWGSC assembly. This made it possible to anchor almost double the fraction of the genome to chromosome locations using the same POPSEQ information that was used to anchor the IWGSC assembly (Chapman et al., 2015). Furthermore, the authors estimated that the gene space sampled by their assembly did not completely overlap with the IWGSC. Taken together, excitingly the Chapman assembly provided novel sequence space which had a higher likelihood of capturing intact genes than the previously considered IWGSC data.

In light of these developments, a second iteration of the process described in 5.3.5.2 with IWGSC-1 was performed, with some modifications. Using the now genetically-mapped CSS contigs, *DG279* and *DG371* were mapped to the 17.3 cM POPSEQ bin (Table 5.4). There were ~300 CSS contigs in this bin, which it was not possible to consider in the limited time available. For this reason, two strategies were used to prioritise marker discovery. First, new 2DS CSS contigs now anchored into the syntenic intervals were interrogated for variation. Second,

marker discovery was prioritised around genes which might be involved in plant growth and development. As marker development progressed, markers were mapped to the genetic bins. This identified that most of the markers mapped to the 17.3 cM and 33.1 cM bins (Appendix 5). The nucleotide sequence within the four genetic bins from 17.3 cM – 33.1cM totalled 4.62 Mb. The gene models in *Ensembl*Plants from this sequence were extracted and annotated as described in Chapter 2 using syntenic orthologues and functional annotations. Sequence space around some of the resulting 115 genes was mined for SSR variation (this is annotated in Appendix 6.9 and 6.10 and outlined in Chapter 6). One notable gene for which a polymorphic marker was developed is *BRU1*, which encodes a brassinosteroid-regulated protein in *Ae. tauschii*.

Interrogating the IWGSC data in this way contributed to the SSR marker tally shown in Table 5.5.

The Chapman assembly was utilised to extend the synteny approach once it was hosted on the CerealsDB website (CerealsDB, 2015a) in April 2015. The flanking markers *DG279* and *DG371* were mapped to different genetic bins (Table 5.4). A total of 253 Chapman scaffolds were anchored between these cM bins. It was not possible within the time restrictions of this project to examine these without aligning the NGS data to these scaffolds. As a priority, SSR identification was extended to novel wheat sequence that anchored to the *Rht8* interval by synteny (Figure 5.7). To achieve this, the 2DS CSS contigs in the syntenic intervals (Appendix 5.1) were used as queries against the Chapman assembly, to retrieve Chapman scaffolds which extended beyond the sequence of the queries (Figure 5.7). A total of 56 new SSRs was identified in the 1.5 Mb of sequence space and 23 of these were prioritised (Table 5.5) based on location around the centre of the syntenic *Rht8* intervals, since a polymorphic marker here would halve the interval.

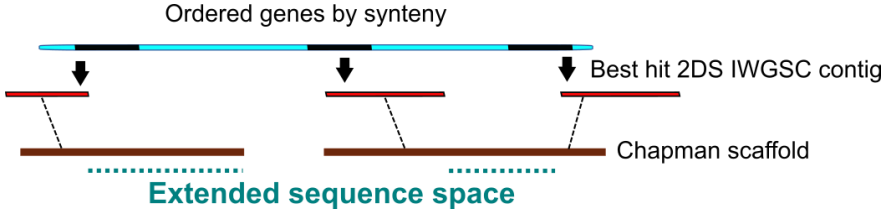


Figure 5.7: Strategy behind mining the Chapman scaffolds for variation within sequence that could be anchored to the Rht8 syntenic regions.

Sequence space searched	no. of SSRs identified
IWGSC-1	139
IWGSC-2	72
Chapman	23
NCBI UniGenes	6
v3.3 cDNAs	27
Axiom data	42
Limagrain	44

Table 5.5: Summary of wheat sequence space searched for SSRs and the number of markers identified.

5.3.5.4 Informed searching: mining IWGSC wheat sequence in the NGS references and SNP arrays

The *Rht8* interval was anchored in the NGS references described in 5.3.3 as shown in Table 5.4. The 2DS CSS corresponding to SNPs within those intervals (not already found in the IWGSC/Chapman assembly) were mined for SSRs. In the case of the v3.3 cDNAs, the 2DS CSS contigs corresponding to the genes within the 2D interval were targeted.

A collaboration on an unrelated project investigating a QTL on 2D provided confidential wheat 2D sequence from a commercial partner (Limagrain). It was unclear how this 2D sequence had been assembled and curated. The sequence was unannotated. The markers *DG279* and *DG371* were anchored in the sequence by BLAST and 44 SSRs were identified in the intervening sequence space (Table 5.5).

Since the Axiom® SNP array was used to genotype Paragon and the *Rht8* x Paragon short NIL, the SNP variants identified in probe sequences were not certain to be present between the parent NILs to the fine-mapping *Rht8* population. Additionally, the array was not annotated, apart from containing the CSS contigs the probe sequences mapped to. To prioritise marker discovery, SNPs mapping to 2DS contigs were targeted and from these 120 unique 2DS contigs, only the SNPs without heterozygous (AB) calls were retained. A total of 42 new SSRs was identified in the corresponding 2DS sequence (Table 5.5).

5.4 Synteny – how good is it?

It was important to evaluate how good the synteny was as a whole between wheat 2D and the comparative species, as well as establishing the collinearity along the syntenic intervals. In 5.3.5.2, genome zippers (Appendix 5.2) were constructed using the IWGSC CSS contigs to retain genes within the syntenic intervals which showed evidence of 2DS localisation (Table 5.4). These results showed that barley had the highest number of genes with 2DS localisation, followed by Brachypodium and finally rice (Appendix 5.2). By directly comparing the zipper for each species with orthologous genes in the other two species, it was evident that synteny between the species was not as good as had been reported previously (Gasperini et al., 2012). This was examined further (Figure 5.6) to reveal that even in barley, which had the highest number of 2DS-localised genes, there were large assembly/annotation gaps in the *Rht8* interval. As resources developed during this project, it became possible to establish a 2D interval a wheat genetic bin in wheat (5.3.5.3, Table 5.4). This made it possible to study the wheat-to-barley, wheat-to-Brachypodium and wheat-to-rice synteny directly, rather than using barley. Using the wheat 2D interval, ArkMap (Paterson and Law, 2013) was used to show orthologous relationships between the wheat 2D genetic bin and barley, Brachypodium and rice (Figure 5.8).

Of the 162 genes in the wheat sequence, 71 had orthologues in barley, 57 in Brachypodium and fewest (39) in rice. The synteny between wheat 2D and barley 2H in this region was good and very few genes (two) mapped to another barley chromosome (3H). The micro-collinearity progression along barley broadly mirrored that of wheat, however large annotation gaps remained in barley, as found previously (Figure 5.6 and 5.3.5.2). However, significant numbers of orthologous relationships were found with Brachypodium chromosomes 4 and 3, as well as rice chromosomes 7 and 11 (Figure 5.8).

Figure 5.8: Synteny between the wheat 17.3 cM bin, barley, Brachypodium and rice.The 17.3 cM bin was identified as most likely to contain the Rht8 interval as defined in Table 5.4. The coordinates used were wheat 2D 6478405 – 10885088. Conserved synteny was shown in barley, Brachypodium and rice, based on orthologous relationship defined by a 125/200 combined similarity threshold. Orthologous relationships between genes are shown with lines as follows: black = one-to-one, green = one-to-many, brown = many-to-many. Genes on the non-syntenic chromosomes in Brachypodium and rice are not shown for clarity, instead physical position alone is indicated. ArkMap used EnsemblGenomes release 25. The assembly details are: wheat IWGSC-2, barley IBSC-1.0, Brachypodium v1.0, rice Oryza Sativa Japonica IRGSP-1.0. Not all relationships are shown for clarity (described in Chapter 2). The density of annotation was reduced to 1 in 10 for barley, 1 in 10 for Brachypodium and linked genes in rice.

5.5 Prioritising High-confidence Variants

From the large SNP platform and NGS datasets, as well as extensive wheat sequence space available, strategies were required to prioritise the most reliable variants. In order to attempt to sift the high-confidence variants 'noise', three main criteria were considered.

5.5.1 Concordance

5.5.1.1 Parent NILs

Concordance within the parent NILs and BSA was considered differently. In the NGS data of the parent NILs, there was considerable discordance (Appendix 5.4) between the biological replicates and tissues (P1-8, Table 5.1) within the 60,454 putative SNPs identified. Only 1% of the total putative SNPs (a count of 638) were completely concordant whereby the base call within all four samples of the parent NIL was consistent and this base was also distinct between the other parent (Appendix 5.4). These SNPs were targeted as being high-confidence variants.

The distribution of completely concordant putative SNPs between the parent NILs was plotted with the mean SNPs found per chromosome arm, which could be considered as 'background noise'. The SNPs were normalised as described in Chapter 2 to account for the relatively under-represented D genome (Table 5.2) and also to mitigate the concern that 3B could have more variation reported since it has been fully sequenced (Choulet et al., 2014). There was enrichment on chromosome group 2 well above this mean line of 7%, particularly on 2AS and 2DS. There were also high densities of SNPs found on chromosomes 3B, 6AL, 7BS and 7DS (Figure 5.9A). Unfortunately, of the highly concordant SNPs on 2DS, there was a dearth on SNPs in the middle of the *Rht8* interval (Appendix 5.5.1).

The SNPs between the parent NILs in the dataset without missing calls on the iSelect array numbered 412 (Table 5.2). From this, a total of 85 SNPs mapped to chromosome group 2S. Since the total variation captured between the parent NILs would be inherently limited by the location of pre-defined probes on the array, the SNPs were normalised according to the number of probes mapping to

each chromosome arm (Figure 5.9B). Similarly to the UniGene data, 2AS and 2DS had high SNP densities, but notably 2BS was much lower than the other homoeologues, this time below the threshold set for background noise. The SNP hotspots on 3B, 6AL, 7BS and 7DS found between the parent NILs in the NGS dataset were once again prominent between the parent NILs in the iSelect data, though 7DS was reduced compared to the UniGenes (Figure 5.9A versus Figure 5.9B). The enriched SNP densities outside group 2S found between the parent NILs in two independent datasets were important because translocations into the *Rht8* genetic interval would complicate the fine-mapping and would also considerably limit the comparative genomics strategy. Some background SNPs outside 2DS were expected since the short and tall parent NILs were estimated as 97% similar on the basis of DArT markers (Gasperini, 2010). However, any such background noise that did not correlate with the height phenotype should have been minimised by the BSA unless it was genetically linked to or originated from a translocation into the *Rht8* genetic interval on 2DS.

5.5.1.2 BSA

In the BSA analysis of the iSelect array, three short individuals and three tall individuals, as well as bulks combining the DNA from short/tall individuals were genotyped and added to the data from the parent NILs. Surprisingly, there was considerable discordance between the variants identified. Some discordance was expected due to the BSA approach, since in close proximity to *Rht8*, recombination in individuals of the same bulk might lead to contrasting 'AA'/'BB' calls. However, the bulk samples often had opposite SNP calls to the individuals (all concordant) comprising that bulk. Additionally, the parent NILs frequently had the opposite SNP call to the individuals in the bulk of the same phenotype (Appendix 5.6). An alternative scenario was that a SNP variant identified between the parent NILs was unanimously monomorphic in the short and tall individuals comprising the bulks (Appendix 5.6). A conservative strategy was adopted and (conflicted) SNP variants identified were taken forward in the workflow (Figure 5.1). However, this exemplifies that with the great efforts to increase bioinformatic resources for wheat, the discourse on the data processing and analysis, including reliability and unsuccessful strategies, is not commensurate with the volume of data produced.

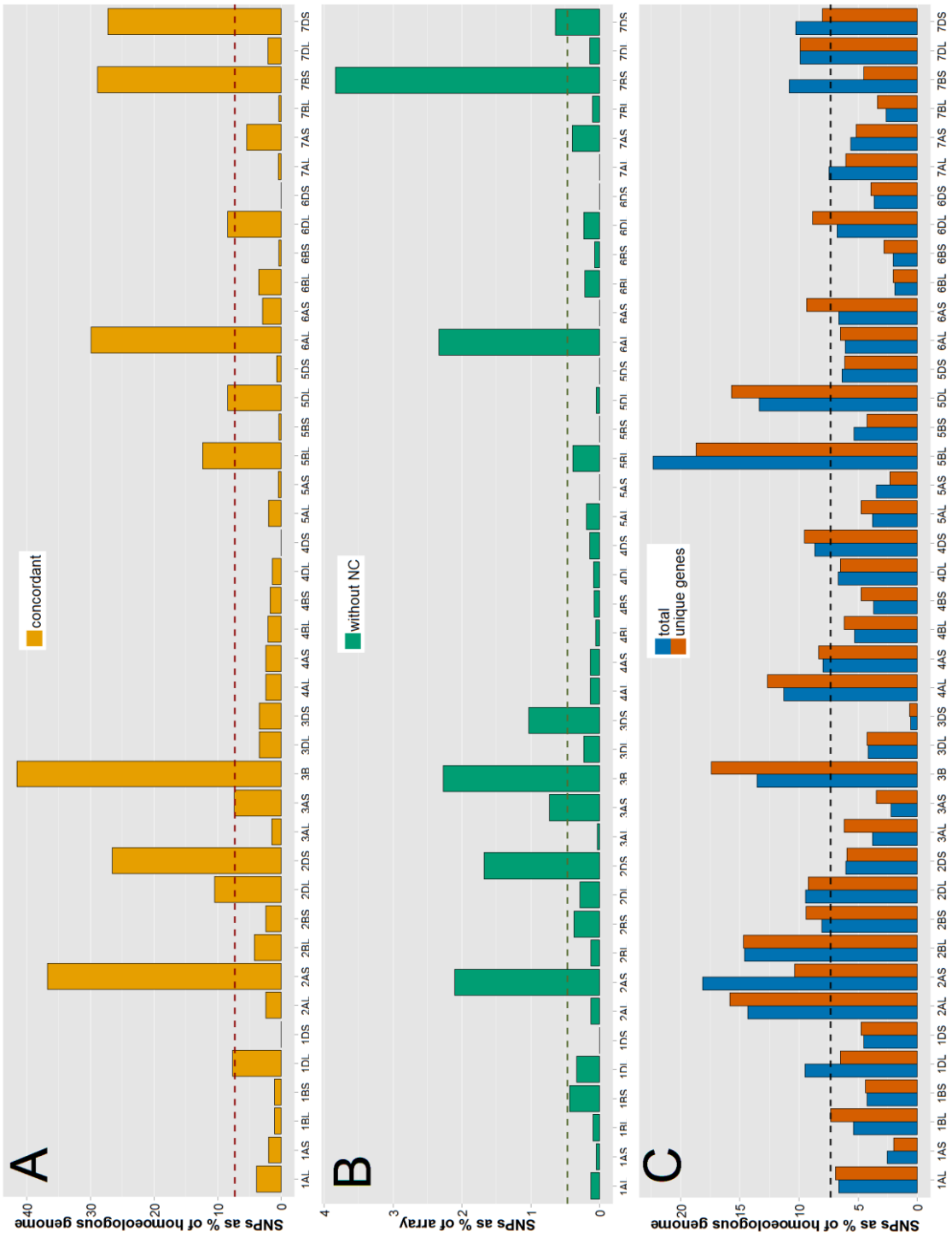


Figure 5.9: The distribution of putative SNPs over the genome in parent NILs and BSA as mapped to the CSS contigs. (A) Concordant SNPs (638) between the parent NILs from the UniGene reference (B) SNPs between the parent NILs on the iSelect 90K array, without missing data (412) (C) SNPs from BSA with BFR>6 on the v3.3 cDNAs, with the total number (7666) and the unique genes from this set (2055). The data was normalised in the following way: (A) & (C) were normalised to calculate SNP distribution as a percentage of the homoeologous genome; (B) SNPs were normalised to account for the total number of pre-defined probes on each chromosome arm in the 90K array. The dashed lines show the mean SNP percentage (per chromosome arm): 7.317% for (A) & (C); 0.472% for (B).

5.5.2 High BFR

To prioritise the high-confidence variants from the BSA analysis of the NGS data and ensure SNPs on chromosome group 2S were adequately sampled, a minimum BFR threshold was determined as BFR>6. The majority of putative SNPs (75%) were reported with a BFR of infinity, indicating 100% enrichment from one parent and absence in the other parent. However, the vast majority of SNPs with BFR = infinity were being called due to low coverage (<20 x) in the parent with the SNP present. To eliminate this large number of potentially low-confidence SNPs, the ratio threshold for the informative parent of SNPs with BFR = infinity was set to ratio ≥ 0.2. Setting both the BFR and ratio thresholds retained a total of 7666 putative SNPs genome-wide on 2055 unique genes (Figure 5.9C). The duplicated genes (total vs unique) did not overly influence the SNP distribution over the whole genome.

The high-confidence SNPs found by BSA were examined to see if they mapped to similar locations as in the parent NILs. The BSA data had high SNP densities (SNP hotspots) on 3B, 5BL, 7BS and 7DS, in common with the parent NILs. The SNP density on 5BL and 5DL was much enriched whereas the 7BS and 7DS hotspots were reduced in prominence in the BSA relative to the parent NILs (Figure 5.9C versus Figure 5.9A). High SNP densities were also found on group 2L in the BSA which had not been found in the parent NILs (Figure 5.9C). 2AS had the highest SNP density of group 2S, which was also reported in the parent NILs, but 2DS fell below the mean SNP density per chromosome arm.

5.5.3 Putative chromosome rearrangements

Interchromosomal translocations involving chromosomes 4A, 5A and 7B in wheat have been well characterised (Devos et al., 1995, Liu et al., 1992, Nelson et al., 1995). Most recently, in a study of 720 genes representing putative interchromosomal rearrangements in wheat, 40% were reported outside of these well-documented locations, scattered across chromosomes. Of particular interest to this project, one or two genes from 12 locations (1BS, 1BL, 3B, 4BS, 4DL, 5AL, 5DS, 5DL, 6AS, 6BS, 7DS, 7DL) were found to be translocated to 2DS (Ma et al., 2015a). In another study utilising the chromosome arm locations from IWGSC-1, intrachromosomal rearrangements, where sequences found on homoeologous

chromosomes were located on a different arm in one of the homoeologues were reported. One of the conclusions was that there was strong evidence of intrachromosomal rearrangement on 2D short arm to long arm (Ma et al., 2014).

Taken together, data from the parent NILs and BSA indicated a relatively high number of putative SNPs outside 2DS that could not be ruled out were translocations to 2DS or potential intrachromosomal SNPs on 2DL actually originating from 2DS. Of particular interest were putative SNPs reported on 3B, 6AL and 7BS, since these had high SNP densities reported consistently in both the independent datasets comparing the parent NILs (Figure 5.9A & B). To investigate this further, SNPs with a high degree of concordance in calls between the parent and bulk iSelect array data (Appendix 5.6.2) or with a high BFR (Appendix 5.7.2) were prioritised for marker discovery. Markers mapping to 5B and 7BS were polymorphic between RIL4 and CD and these were taken through the workflow (Figure 5.1) to assess whether the markers were likely to map to *Rht8.*

5.5.4 Prioritising SSR variants

Often, multiple SSRs were located in close proximity. In these cases, SSRs were prioritised in two main ways. First, SSRs based on mono-nucleotides were less desirable than the longer repeat lengths, due to the incomplete nature of the sequence of the CSS contigs (and hence possibility of unresolved sequencing error or missing sequence leading to stretches of mononucleotide repeats which might be artefactual). For this reason, where multiple SSRs were identified in the same CSS contig, those based on two to six repeats were prioritised. However, markers based on mono-nucleotides had been found to work and were still considered if in a prime physical location. Second, where multiple SSR markers mapped to the same CSS contig, those which were based on size polymorphism were used over presence/absence markers (Figure 5.5). This was to mitigate for the uncertainty associated with scoring false negatives. The incidence of this last case was low.

5.6 Likely to Map to *Rht8*?

The high-confidence variants identified in the previous step in the workflow were assessed prior to marker validation on the fine-mapping *Rht8* population for whether they were likely to map to the *Rht8* interval using genetic and physical information.

5.6.1 Physical location on wheat chromosome 2D

As described in 5.3.3, the physical ordering of the v3.3 cDNAs was used to delimit an interval most likely to contain *Rht8* based on pre-existing markers anchored onto the ordered section of the reference (Table 5.4). The genes were ordered using a Chinese Spring x Paragon map and the limitation of the reference was that lack of polymorphism in the mapping population assigned a large proportion of genes on 2DS into the unordered bin. Nevertheless, this was the best resource available at the time. To specifically target SNPs likely to map to *Rht8*, a second iteration of BSA was performed using the narrowed 2D interval comprising 59 genes in the ordered section of the v3.3 cDNAs.

The individual SNP distribution per gene in the narrowed 2D interval is shown in Appendix 5.8. The distribution was normalised (details in Chapter 2) to account for within-sample bias (gene length) and between sample bias (sequencing depth, Appendix 5.9). This resulted in a unit of SNPs read^{-1} base^{-1} for each of the 59 gene models (Figure 5.10). The normalised distribution was markedly different from the raw SNP distribution (Appendix 5.8). The gene model *mrna057019*, outside the strict interval, had almost ten-fold the SNP density at 30 SNPs read^{-1} base^{-1} of the next highest genes, *mrna096393* and *mrna023290* (within the strict interval) at approximately 5 SNPs read^{-1} base^{-1} (Figure 5.10). The high SNP density of *mrna057019* was due to the extremely low coverage (x 0.19) rather than a high raw SNP count (Appendix 5.9).

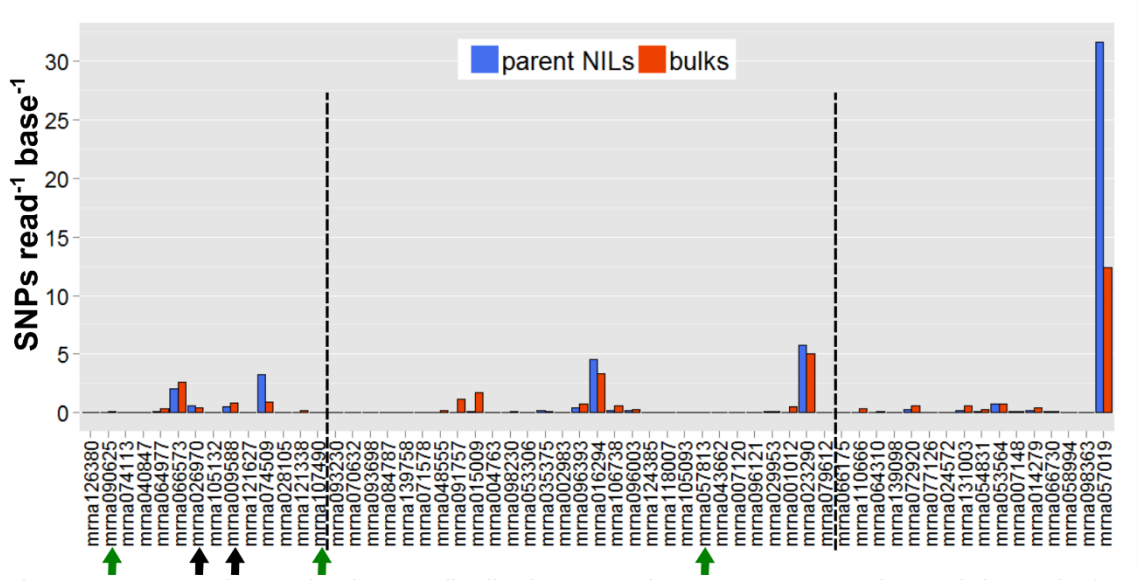


Figure 5.10: Putative varietal SNP distribution over the v3.3 cDNA 2D interval (a total of 59 cDNAs), with the genes ordered from left to right as they are anchored in the ordered v3.3 cDNA reference. The varietal SNPs between the parent NILs/ bulks and the reference are shown. SNPs were normalised from the raw number in Appendix 5.8 by the formula: SNP count/coverage where coverage was calculated as illustrated in Appendix 5.9. The dashed lines indicate the anchoring of DG279 (left) and DG371 (right) as outlined in Table 5.4. The black arrows indicate successful SNP-based KASP markers developed as described in 5.3. The green arrows indicate SSR markers developed based on the corresponding IWGSC contig to the gene highlighted.

The SNPs in Figure 5.10 were further analysed to select the most confident SNPs to validate by developing markers. Two different approaches were used, as shown in Appendix 5.10 and described in full in Chapter 2. The first approach mirrored a BSA approach and identified SNPs which overlapped in the corresponding parent NIL and bulk (Appendix 5.10). This approach found that in both short/tall datasets, there was more overlap in SNPs between the converse parent NIL/bulk (Appendix 5.11). The SNPs found in this way (Appendix 5.12) were aligned to the CSS contigs and the 18 identified as varietal were monomorphic between the parents NILs (Appendix 5.13).

Trouble-shooting this first approach found two main issues which were attributed to the poor validation rate. The first issue pertained to the SNP-calling quality threshold with the software used. Due to the high-depth of coverage (Appendix 5.9), the SNP-calling software (Koboldt et al., 2012) passed SNPs with high-quality scores, despite having a relatively low frequency (Appendix 5.13) (where frequency is used to describe the reads at the base supporting the variant call as opposed to reference, a measure of the SNP 'confidence'). In fact, the majority of putative SNPs (80% of the total) had a frequency <50% (Appendix 5.13 and Appendix 5.14), meaning that these SNPs were called despite most of the reads supporting the reference base. Personal communication with the developer

confirmed this bias, and highlights the need for software and pipelines which are developed for the intricacies of the wheat genome, rather than adapting existing pipelines which are designed for analysis of sequenced and annotated genomes.

The second issue concerned the identification of varietal SNPs. Case-studies of a number of putative SNPS (described in Chapter 2) revealed that there were instances of sequence differences between the v3.3 cDNA reference and the IWGSC CSS contigs. The difference in a single base was sufficient to result in a homoeologous SNP being called using contig alignments from one source versus a varietal SNP being called using the other source. This was unexpected, given that the v3.3 cDNA originated from gene models predicted using the IWGSC CSS contigs. The discrepancy was attributed to sequencing error and further, from redundancy between overlapping IWGSC CSS contigs (described in detail in 2.3.7.4) which complicated the situation since it was not possible to judge which redundant contig was more reliable where sequence differences arose. This shows that there are still formidable challenges when working with a hexaploid organism with an incompletely-assembled genome.

In an attempt to resolve both these issues, in the second marker validation approach, SNPs were filtered first by high frequency and second by two alignments (steps shown in Appendix 5.10.2). From the SNPs validated by markers using this process, two were found to be polymorphic between the parent NILs (Appendix 5.15). These are indicated in Figure 5.10.

5.6.2 Synteny

Work in this Chapter showed that the synteny in the *Rht8* intervals identified in comparative species (Table 5.4) was disrupted (5.4). Nevertheless, synteny (albeit limited) could still be used to target variation to *Rht8* region.

Physical information from syntenic intervals in barley, Brachypodium and rice was used as shown in Table 5.4. The comparative genomics approach relied on good annotation of the SNP platforms and NGS references. Barley and Brachypodium physical information was more prevalent than rice but the iSelect array as a whole was more poorly annotated than the other references (Figure 5.4). Notably, using syntenic information to consider high-confidence variants recovered many SNPs mapping to 2AS or 2BS CSS contigs that would have been discarded by only

considering SNPs mapping to 2DS. This complemented the strategy outlined in 5.5.2. Synteny was used to identify high-confidence variants (from 5.5) in the NGS data and SNP-platform data: between parent NILs (Appendices 5.3, 5.6.1, 5.5.2, 5.5.3) and BSA (Appendix 5.7.1).

5.6.3 Informed by wheat contigs from IWGSC

5.6.3.1 2DS provenance

To avoid relying only on comparative genomics since not all variants which mapped to 2DS had such annotation, high-confidence variants which mapped to 2DS were targeted using the IWGSC chromosome arm assignments. In the variants between parent NILs, 34 highly-concordant allelic SNPs mapping to 2DS were identified (Appendix 5.5.1). From the BSA data, 10 new allelic SNPs were identified in the iSelect array that mapped to 2DS (Appendix 5.6.2). A total of 132 putative SNPs that mapped to group 2S were identified in the v3.3 cDNAs, and PolyMarker was used to select SNPs on which markers could be designed that included all three homoeologues in the alignments to enable primer design which was D-genome specific or semi-specific (distinguishing between 2DS and another genome). Of the 48 putative allelic SNPs tested, 11 were polymorphic (Appendix 5.7.1).

5.6.4 2D RIL Population

The 2D RIL coarse-mapping population was used to identify markers from the workflow in this Chapter that were likely to map to *Rht8* (Figure 5.11). The population had been developed in the first genetic analysis of *Rht8* on 2DS (Korzun et al., 1998) and was used in further fine-mapping of *Rht8* (Gasperini et al., 2012). The population had been genotyped using markers described by Gasperini (2010) which are coloured black in Figure 5.11. A core set of markers mapping close to the *Rht8* interval was identified in group B (coloured red in Figure 5.11B), which was then fine-mapped in Chapter 6. Group B surrounded the *DG279*/*DG371* markers on 2D used in mapping by Gasperini et al., 2012. Group B contained 62 markers which were arranged into 22 marker classes (Appendix 5.16), according to their graphical genotypes (Appendix 5.17).

The outgroups (Figure 5.11A & C) contained markers which did not map to 2D and were not passed along the workflow (Figure 5.1) to Chapter 6. Group A contained markers from SNPs with a high BFR, mapping to 2S but not the *Rht8* region (Appendix 5.7.1). Group C contained markers from SNPs mapping outside chromosome 2S: 5BS and 7BS (Appendix 5.6.2, Appendix 5.7.2 and Chapter 2). All the markers were tested with the flow-sorted 2D DNA from the short parent NIL (RIL4). All markers in group A amplified the 2D DNA in the KASP assay distinctly from the tall NIL (Figure 5.11A). This validated that the SNPs were on 2D but did not map genetically to the *Rht8* region. For the SSR assays, all markers had a peak identical to the short parent NIL (Appendix to Chapter 2). Markers in group C showed mixed results. Most of the markers did not amplify 2D DNA (C.1 and C.2). Since these markers were polymorphic between the parent NILs but were developed on SNPs outside of 2D, this indicated that the SNPs were not interchromosomal rearrangements (translocations into 2DS), but instead SNPs on chromosome 7BS. One marker (labelled D), did amplify the 2D DNA. This either represents an interchromosomal rearrangement – a translocation from 7BS to 2D, or contamination in the 2D DNA. In 5.3.2, which describes the testing of the 2D DNA, two markers on 7D and one marker on 5A amplified the 2D 2DNA, and a small amount of contamination was reported (6%). Therefore, it is not possible to unambiguously ascertain which of the two alternatives ‘D’ represents.

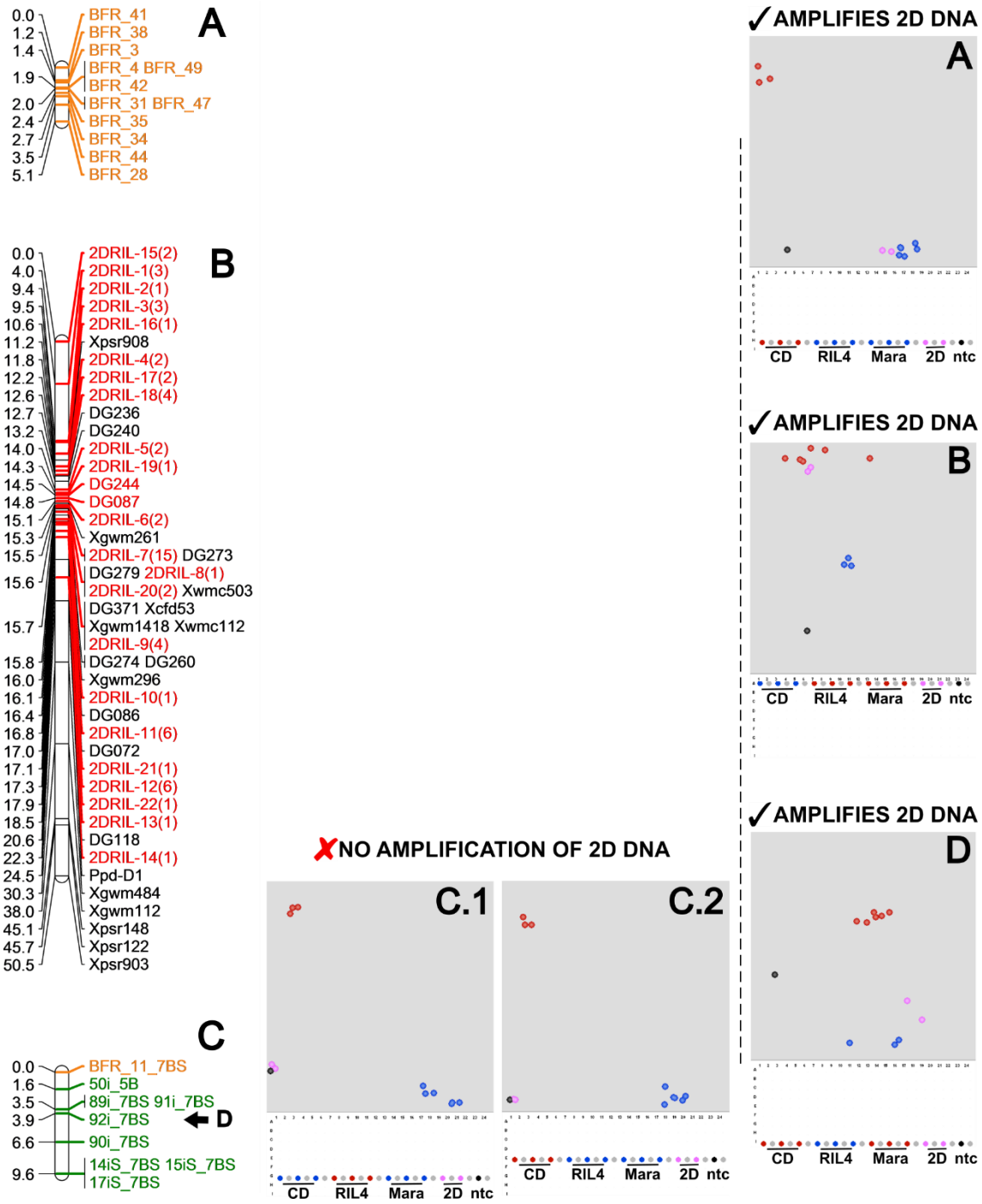


Figure 5.11: Coarse-mapping of markers developed in Chapter 5 with the 2D RIL population. Marker classes 1-22 in group B are highlighted in red. Markers developed from the high BFR BSA data are shown in orange. Markers from the iSelect BSA are shown in green. Distances in the mapping (left, in cM) are relative measures only for ordering marker classes. The specificity of the KASP assays as ascertained by amplification of the flow-sorted 2D DNA from RIL4 (highlighted pink), CD (Cappelle-Desprez, tall parent NIL), RIL4 (short parent NIL) and Mara (Rht8 donor to RIL4) is shown and labelled according to the genetic map groupings. 'ntc' (coloured black) is the no template control, which had assay mix but no DNA as a control. Coloured red on y-axis is the FAM-labelled adapter, coloured blue on the x-axis is VIC.

5.7 Validating variants with markers

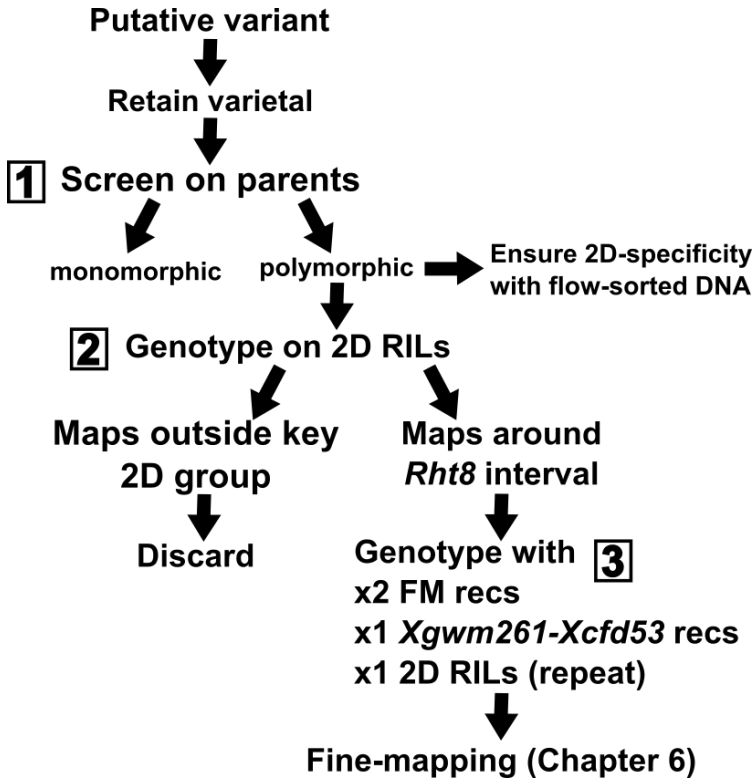


Figure 5.12: Schematic diagram to show the marker validation pipeline of testing putative variants. Three main steps were involved in the workflow. The populations used were: FM = fine-mapping Rht8 population; 2D RILs = 89 Recombinant Inbred Lines on 2D used as a coarse-mapping population; Xgwm261-Xcfd53 = Recombinants between the original two flanking markers from which the FM population was selected. The populations are described in more detail in Chapter 6.

Markers were developed to validate the variants likely to map to *Rht8,* identified by the workflow in Figure 5.1. Markers were tested in the pipeline showed in Figure 5.12. Putative variants were either SSRs or SNPs. Both types of marker were validated in the same way. SNPs were tested using the KASP assay (He et al., 2014). First, markers were tested on the parent NILs to check if they were polymorphic. If the marker was polymorphic, the 2D flow-sorted DNA was used to ensure specificity to the D genome to target genome-specific allelic variation (outlined in 5.3.2). The proportion of markers tested which were polymorphic ranged from 14 – 38% (Figure 5.13). Second, the coarse-mapping population, described in 5.6.4, was used to identify the markers which genetically mapped close to *Rht8.* A total of 62 (Table 5.6) markers were taken forward to Chapter 6, since they fitted into group A in Figure 5.11. In the final step, the markers were genotyped on all the mapping populations (Appendix 5.18), including reproducing results for robustness (Figure 5.12, part 3). The markers were then used to further fine-map *Rht8*, explored in Chapter 6.

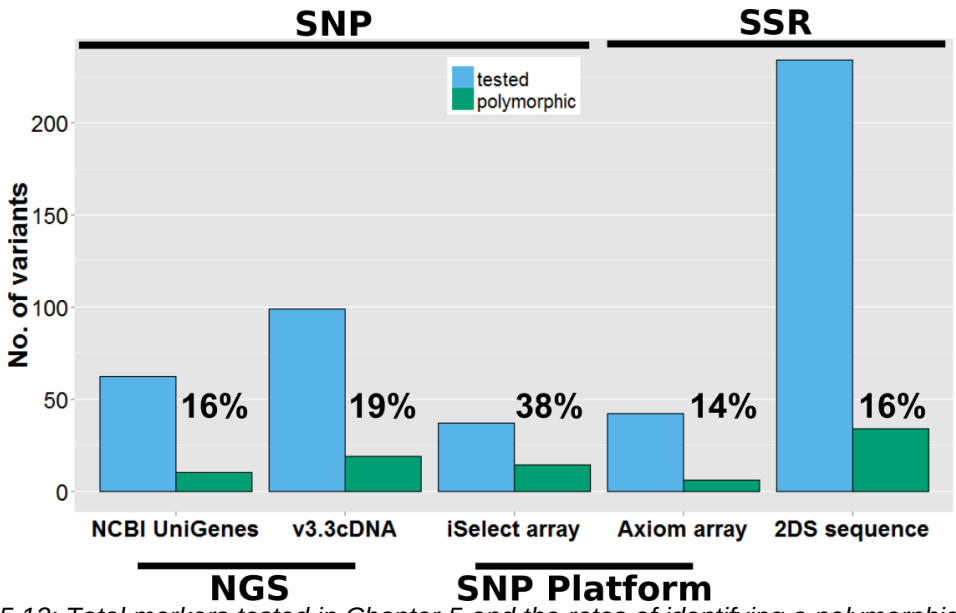


Figure 5.13: Total markers tested in Chapter 5 and the rates of identifying a polymorphic marker across the different approaches.

	Source	Initial cost	Cost/marker	Marker count taken forward for fine-mapping
SNP	**NCBI UniGenes**	£7,580	£24.50	10
	v3.3cDNA	£11,550	£24.50	3
	iSelect array	£500	£24.50	6
SSR	**Axiom array**	£600	£65	6
	2DS sequence	0	£65	37
				Total: 62

Table 5.6: Cost break-down for developing markers in Chapter 5. A more detailed description of how the costings were calculated is shown in Chapter 2.

5.8 Discussion

The aim of this Chapter was to saturate the *Rht8* region with markers prior to further fine-mapping. Wide-ranging strategies were used to identify variants and then prioritise the variants most likely to map to *Rht8*. In total, 62 markers were generated during the workflow presented at the beginning of the Chapter. Other markers which were developed were discarded because they did not map to the same linkage group as *Rht8*.

5.8.1 Identification of variants – cost and efficiency

Among different marker systems SSRs and SNPs are the markers favoured in wheat breeding (Gupta et al., 1999). SSR assays are more time-consuming and in the work in this Chapter were found to be over double the expense of the SNP KASP assay, on an individual-assay level (Table 5.6). Anecdotally, breeders prefer SNP-based high-throughput assays for large populations. Additionally, SNPs are bi-allelic and the most abundant genetic variations, with higher frequencies found distributed evenly throughout the genome (Allen et al., 2011). Despite this, the rate of marker validation between the two variant types was broadly similar (Table 5.6). SNP assays allowed for a lower cost per marker once high-density arrays or NGS sequencing were used to identify the variants (Table 5.6). Nevertheless, SSRs and SNPs target different variation and have been used in concert previously for marker development (Jiang et al., 2015a), and in this Chapter, both types of marker were developed. The alternative approaches will be considered in turn.

5.8.2 SSR variation in wheat sequence

Different sources of wheat sequence were targeted for identifying microsatellite variation. Comparative genomics is widely used in map-based cloning of Triticeae (Krattinger et al., 2009a). The advantage of the IWGSC CSS contigs was that they allowed for genes in the syntenic intervals to be considered only if they had evidence of 2DS-localisation. By anchoring existing *Rht8* markers in physical sequence, work in this Chapter established three zippers for barley, Brachypodium and rice syntenic intervals on chromosomes 2H, 5 and 4

respectively. The 2DS CSS contigs across the zippers were mined for SSR variation. The NGS data was also searched using these contigs to return SNPs which mapped to the syntenic intervals. Low polymorphism between the parent NILs, which was reported previously when fine-mapping *Rht8* (Gasperini et al., 2012) was also found to hinder marker development here and is in line with extensive reports of low polymorphism on the D-genome in wheat research. Developments in the last six months of this project saw the release of a WGS wheat assembly. The Chapman assembly is an excellent resource because the sequence space sampled is not completely overlapping with the IWGSC CSS contigs and the scaffolds have much better contiguity, on average double that of the CSS contigs (Chapman et al., 2015). The IWGSC CSS contigs corresponding to the zippers were used to query the Chapman scaffolds and this returned an additional 1.5 Mb of novel sequence space. Due to time limitations, only half of the SSRs identified were validated by markers and only one was found to be polymorphic. Therefore the Chapman assembly did provide novel sequence which was not fully utilised here, but the low polymorphism on the D-genome would likely remain problematic even with more time to examine remaining sequence space.

5.8.3 Limitation of synteny in the Rht8 *region*

Compiling the zippers of barley, Brachypodium and rice revealed that when only genes with evidence of 2DS-localisation (using the IWGSC CSS) were considered within the delimited intervals, micro-collinearity was disrupted. Using the POPSEQ genetic map of wheat (discussed directly below), the synteny directly between wheat and the three comparative species corroborated the initial findings. Barley had the highest density of genes collinear with the wheat 2D interval. The medium resolution synteny between the demarcated wheat 2D region and barley 2HS was good, however, there were large annotation gaps in the barley data within the *Rht8* interval. The synteny with Brachypodium and rice was not as good as previously reported (Gasperini et al., 2012) and inferior to barley. In particular, there were clear breakpoints in synteny where orthologous genes were found on Brachypodium chromosomes 3 and 4, and rice chromosomes 7 and 11. The work in this Chapter strongly pointed towards barley

being the best comparative species to use out of the three considered here, but the quality of the draft whole-genome shotgun assembly was a limitation.

The disrupted micro-collinearity does not exclude the possibility that a marker mapping closest to *Rht8*, or the *Rht8* gene itself, could be a gene found in some or all of the candidate-gene intervals across the species used here. This is investigated further in Chapter 6. Different strategies used in marker development described in this Chapter and taken further into Chapter 6 relied on synteny to varying extents. However, the imperfect micro-collinearity indicates that it is important to be aware that strategies relying on this might be limited.

5.8.4 Low-resolution wheat genetic map

Late in the project, an improved release of the IWGSC CSS data used POPSEQ to order the CSS contigs into chromosomal pseudomolecules in genetic bins (Mascher, 2014, IWGSC, 2014). Previously, the IWGSC CSS contigs were unanchored. This development meant that a coarse genetic map could be used to target sequence space for variants likely to map to *Rht8*. During the process of marker development in this Chapter, ordering the successful polymorphic markers revealed approximately ~5 Mb of space from the POPSEQ-anchored contigs. Due to the time limitations, a targeted approach around some genes identified in this sequence space was used to identify microsatellite variation. The genetic anchoring of wheat sequence to bins was an improvement to the unordered IWGSC CSS contigs. However, POPSEQ uses several individuals from a doubled-haploid or recombinant inbred line population, sequenced to low coverage (~1.5 x). SNPs identified between individuals are then used to map the sequenced contigs (Mascher et al., 2013). Since the wheat POPSEQ for both the Chapman scaffolds and IWGSC-2 was based on analysis of a small doubled-haploid population of 80-90 individuals (Sorrells et al., 2011), this map only provided low-resolution. This low-resolution genetic map was used for fine-mapping in Chapter 6. However, the large bin size in terms of contig number and low-resolution of the map meant that a more efficient strategy was required to target variation around candidate genes. This was considered once fine-mapping had narrowed the sequence space, which is described further in Chapter 6.

5.8.5 SNP variation in NGS data

Bulked segregant analysis (BSA) was used to target variation which was enriched for the corresponding parental allele in the relevant bulk. Importantly, the BSA strategy did not rely on synteny, which as discovered in this Chapter, had its limitations. The UniGenes had the greatest proportion of syntenic relationships (60 – 80%), which is slightly higher than previous genome-wide studies have identified (60 – 70%) (Massa et al., 2011, Luo et al., 2013). This figure in the v3.3 genes was lower at 20 – 70%. As with previous studies in tetraploid and hexaploid wheat (Trick et al., 2012, Ramirez-Gonzalez et al., 2014), the tens of thousands of variants generated required filtering to a manageable shortlist. Variants were prioritised which showed six-fold depletion or enrichment (BFR>6) in the corresponding bulk. This value was empirically determined as being suitable to capture SNPs on 2S. Of the genes with the highest (non-infinity) BFRs (BFR>20), only two out of 47 were localised to 2S (Appendix 5.19). One of the SNP assays to validate these was monomorphic, whilst the marker on the other SNP mapped outside the *Rht8* linkage group, to group A (Figure 5.11).

5.8.6 SNP variation in SNP-platform data

The iSelect array was used to identify SNPs between the parent NILs and between the bulks. The iSelect array provided the highest proportion of polymorphic markers at 38% (Figure 5.13), and was also cost-effective compared to the cost of sequencing (Table 5.6). The iSelect array was limited since only 15% of the pre-defined allelic variation on the array mapped to the D genome (Wang et al., 2014a). The markers which mapped to syntenic intervals in the zippers was low, and most of these were monomorphic (Appendix 5.1). However, even normalising for the pre-defined markers on each chromosome arm, the data showed high SNP densities outside of 2DS, which is explored below.

The BSA data showed high discordance between bulks and parent NILs in contrasting ways, explored in this Chapter, some of which was due to heterozygous calls. There is also the possibility that the difference in iSelect runs confounded analysis. The parent NILs and bulks were genotyped on separate arrays and by the second array, the SNP-calling had improved, since the number

of SNPs called due to missing data (NC) decreased across the whole array (down by 5%).

The heterozygous call assignment was made using the software GenomeStudio, which cannot always assign calls in overlapping clusters (Wang et al., 2014a). However, this issue was representative of the array genome-wide, rather than this particular analysis. This could be due to deletions or missing data, variation arising due to copy number rather than presence/absence and the limitations of using wheat sequence from the reference research wheat, Chinese Spring, in alignments.

The Axiom® 820K array could not be fully capitalised on because the parent NIL material or BSA was not used, instead, the *Rht8* NIL in the Paragon background was genotyped. Only the 2DS contigs from these SNPs could be used for marker development, and 120 of these were identified. The contigs from SNPs without heterozygous calls were prioritised. Since the markers on the Axiom® array at the time of writing were not mapped to any genetic maps or annotated with syntenic information, specific targeting of SNPs was hindered. This was not pursued further since the SNPs might not transfer to the parent NILs to the fine-mapping population. However, the array shows considerable potential for developing markers mapping close to *Rht8*, since a large number of SNPs mapped to the syntenic *Rht8* intervals in the genome zippers. These are highlighted in Appendix 5.1. Even though the SSRs on the corresponding CSS contigs were mostly monomorphic, the SSRs were designed on microsatellites identified in Chinese Spring whereas SNP assays from the array would target variation specific to the genotyped varieties.

5.8.7 Ensuring genome specificity

Designing D-genome-specific markers was important for cost and efficiency. An *in silico* approach was used in this Chapter, and then verified by 2D flow-sorted DNA from the short parent NIL. PolyMarker was used to prioritise SNPs for marker validation by using homoeologous IWGSC CSS contig alignments (Ramirez-Gonzalez et al., 2015). At times, this was limited by the absence of some homoeologue sequence, however the tool was useful for at least narrowing

down large numbers of SNPs for consideration to SNP assays which could be confirmed to be semi-specific or specific to the D genome.

One major technical problem that was found in this Chapter was the discrepancy between the identification of varietal and homoeologous SNPs caused by the redundancy of IWGSC CSS contigs. Manual alignment and visual assessment of a number of case studies identified two main problems. First, due to redundancy in the CSS contigs, often at least two highly similar (by nucleotide identity in BLAST) contigs were returned. Due to the much shorter gene models in the v3.3 cDNA reference compared with the CSS contigs, often the alignment in PolyMarker did not match the contig that the gene model was anchored to. For an unknown reason (assumed to be sequencing error), the CSS contigs, otherwise aligning with high-identity around the SNP position, would have a different base-call at the SNP. Therefore, there were cases where there was a difference in calling a SNP as varietal or homoeologous depending on which contig was used. A systematic review of how frequently this occurred in the IWGSC CSS contigs was beyond the scope of this project, but estimates report that that 3 – 5% of the IWGSC CSS contigs are duplicated (IWGSC, 2014).

The v3.3 cDNAs were a non-redundant set of gene models with the longest transcript retained in the case of multiple splice variants. It is plausible that given slightly different splice variant lengths, the IWGSC CSS contig returned as a best match (by nucleotide identity in BLAST) to a gene model varied depending on which splice variant was used. The implications of this could extend to other projects using the v3.3 cDNA reference. To my knowledge, the v3.3 cDNAs have been used in one other project (Borrill, 2014) where this was not reported. However, it is likely that this was not identified due to the greater emphasis on gene expression in that work (unpublished). To circumvent this problem, manual alignments were carried out in a subset of high-confidence variants, and doing this did improve the marker validation rate. However, time constraints did not permit for manual validation of all the SNPs already filtered *in silico*. For this to be an effective strategy, a smaller stretch of sequence would have to be considered, based on the fine-mapping of *Rht8* and physical anchoring of sequence. This will be reviewed further in Chapter 8 in light of the fine-mapping findings described in Chapter 6.

The flow-sorted 2D DNA was used to validate genome-specificity. The flow-sorted DNA was found to amplify some non-2D-localising markers used in genetic maps in the Griffiths' group. However, the DNA was found to discriminate well between 2A, 2B and 2D. This was not unexpected, given the 95% purity of the flow-sorted fraction (the remaining 5% was identified mostly as 7D). This finding has consequences for 2D BAC library construction based on the flow-sorted DNA. Crucially, the purity also limited the confidence with which SNPs apparently mapping outside 2DS could be validated.

5.8.8 Low marker validation rate

It is difficult to find one causal factor to explain the abysmally low marker-validation rate in both the SNPs identified between the parent NILs, and in the SNPs identified in the BSA strategy to target variation associated with *Rht8*. In this project, these figures were 16 – 19% (Figure 5.13). Previous studies in wheat reported ~55% success rate of marker validation (Trick et al., 2012, Ramirez-Gonzalez et al., 2014). Different angles can be considered and some of these are done so here. However, further evaluation of the BSA approach and the v3.3 cDNA reference is carried out in Chapter 8 in light of the results to fine-mapping.

5.8.9 Technical – sequencing and mapping

The strategy and SNP-calling pipelines were identical to those used before in two projects involving SNP discovery in wheat (Trick et al., 2012, Ramirez-Gonzalez et al., 2014). There is no evidence to suggest that sequencing or mapping was inferior in this project; in fact the analysis in this work built on recommendations by both sets of authors.

One study in tetraploid wheat (Trick et al., 2012) mapped the grain protein content gene *GPC-B1* to a 0.4 cM interval, with a 58% success rate in terms of validated SNPs. That work used the same UniGene reference and SNP-calling procedure that was used in the first approach of SNP discovery between the parent NILs. The experimental details of the sequencing in this Chapter were similar to that study: Trick et al., (2012) sequenced one parent per lane, and here, four samples of each parent were sequenced over two lanes. Effectively, this results as one parent per lane, stratified into four smaller samples, which were then merged.

Mapping details were better here (average of 59% across parent NIL samples, Appendix 2) than that reported in the tetraploid wheat study (48%).

The authors of that work suggested a minimum of eight-fold coverage across genes for BSA analysis. They found that increasing coverage from 8 x to 16 x reduced the number of putative SNPs identified by 60% and increased the validation rate from 57% to 83% (Trick et al., 2012). The coverage across genes used here was a minimum of 20 x, identical to the BSA analysis in a project in hexaploid wheat by Ramirez-Gonzalez et al., (2014).

5.8.10 Experimental design

Within the parent NIL samples, spike and peduncle tissue was harvested and the mRNA sequenced. Since the bulk samples were only from spike tissue, there was the possibility that the different tissues might result in different SNP datasets. This was investigated using *in silico* mixes (outlined in Chapter 2). It was decided to consider all SNPs identified within the different *in silico* mixes together, since there was no consistent bias identified when excluding spike and peduncle in turn from the parent NIL data (shown in Chapter 2). Also, it was desirable to maximise coverage in order to increase marker validation rate, since this was so low in this project.

5.8.11 SNP discovery and filtering

The specific numbers of SNPs reported between the studies in wheat and in this project vary in detail due to experimental design, which is expected. The biggest difference is the extremely low enrichment of SNPs on 2DS found here, when compared with enrichment of the target chromosome in the other studies.

In the parent NILs, 60,454 SNPs were identified between any one of the eight samples and the reference. Of these, 1% were highly concordant where all the samples within each parent NIL had the same base call. Nevertheless, even in the highly concordant dataset, there were high SNP densities outside of 2DS. Of the SNPs that mapped to 2DS and were prioritised for validation based on synteny and concordance, only 16% were validated as polymorphic and taken forward for fine-mapping.

It was expected that the BSA approach would enrich putative SNPs in the *Rht8* region on 2DS. The chromosome arms 2AS and 2BS were also considered since the absence of a 2DS CSS homoeologue would result in a gene aligning to a 2AS or 2BS contig instead. Of the genes with the highest (non-infinity) BFRs (BFR>20), only two out of 47 were localised to 2S (Appendix 5.19). Only one of these could be validated but did not map to the *Rht8* linkage group (*BFR_4* in group A, Figure 5.11). This contrasted with the study in tetraploid wheat, where the highest BFR identified was of a SNP which mapped closely to the candidate gene (Trick et al., 2012). In that study, the approach used was to validate all SNPs with a BFR>3 on unique genes (regardless of location), which was ~100 markers. Of those, 58% were polymorphic and 60% of polymorphic markers mapped close to the target gene.

Here, more stringent filtering was necessitated by the greater number of SNPs identified (SNPs with BFR>6 were found across >7000 unique genes). Of the putative SNPs with infinity BFR ratios, indicating that they were completely absent in one bulk, only 7% were on 2S (data not shown). The high-density of SNPs outside of 2DS even in the most highly enriched (therefore perhaps most reliable) putative SNPs, was indicative of the dataset as a whole. When all SNPs on unique genes with BFR>6 were considered (~2500 unique genes), 9% were on 2S which was above the average per chromosome arm (7%), but there were higher SNP densities outside of 2DS (Figure 5.9).

5.8.12 Variation outside 2DS

Some variation in the genetic background between the two parent NILs was expected, even though the two parent NILs were near-isogenic outside of chromosome 2D (Korzun et al., 1998). Around 50% of the total SNPs on the iSelect array were assigned a genetic position at the time of writing. Analysing the variation genome-wide on the array identified some SNPs outside of 2DS, e.g. on 3B but only the genetically-mapped variants could be resolved (Appendix 5.1). When considering the IWGSC CSS SNPs mapped to and not genetic position, for both the parent NIL NGS and SNP-platform data there were high SNP densities on chromosomes 3B, 6AL and 7BS.

The extent of intra- and interchromosomal rearrangements genome-wide in hexaploid wheat is only beginning to be documented (Ma et al., 2015a, Ma et al., 2014). A recent study found that 40% of translocated genes were outside of the well-known translocations involving 4A, 5A and 7B and instead were scattered across the remaining wheat chromosomes and sub-genomes.

SNPs between parent NILs in locations that were also enriched for SNP density in the BSA were focused on. The BSA data showed enrichment of SNPs on 3B and 7S in common with the parent NILs. Of these, SNPs on 7BS and on 5B could be validated and these were found to map outside the *Rht8* linkage group. None of these markers, with the exception of one, amplified the 2D DNA. This would suggest that these are not interchromosomal rearrangements but SNPs enriched in the BSA that are outside the *Rht8* region. The one exception, taken with the incomplete purity of the 2D DNA, means that there is a possibility that some interchromosomal translocations are present in the *Rht8* region.

Since the resolution of the BSA approach is a combination of marker density and the recombinations sampled in each bulk, this strategy will be evaluated along with the points discussed here in Chapter 8, in light of the fine-mapping.

Chapter 6: Fine-mapping and further characterisation of the *Rht8* interval

6.1: Introduction

Map-based (positional) cloning is a technique for characterising a gene with a particular altered phenotype usually caused by sequence polymorphism (Huang et al., 2003). The basis of this technique is the linking of molecular markers to a phenotype in order to identify a small interval harbouring the gene, which is defined by the two most closely-linked flanking markers. The ultimate goal is to find a set of makers that co-segregate (no recombination) with the gene of interest. These flanking markers can then be used to screen clone-based physical maps, such as Bacterial Artificial Chromosome (BAC) libraries. From these, physical information can be extracted and further steps such as sequencing, sequence annotation and the identification of expressed regions can take place to identify a candidate gene. Finally, there is the verification step of the identified gene by sequencing of isolated alleles, the introduction of a wild-type copy of the gene into a mutant, or overexpression/knock-out studies (Scheible et al., 2005). The prerequisites for map-based cloning of *Rht8* form the basis of much of this Chapter.

The first step for map-based cloning is the creation of an appropriate mapping population. This population should enable initial low-resolution mapping subsequently followed by mapping to enrich for recombinants around the gene. This is directed towards the definition of a minimal interval of molecular markers that includes the locus of interest. Establishing these populations can cause considerable delays in projects. The work in this Chapter was supported by previous work which developed these populations. First, a recombinant-inbred line (RIL) population based on 2D substitution lines (Korzun et al., 1998) was developed. The 2D RIL population was used in this Chapter in the first step of low-resolution mapping. Next, a fine-mapping population was developed by Gasperini et al., (2012), which originated from crosses between Cappelle-Desprez (CD) and the *Rht8* donor, RIL4. From 3104 F_2 individuals, recombinants

with respect to the parents were selected and developed to further generations to map *Rht8* between *Xgwm261* (distal) and *Xcd53* (proximal). The interval was flanked by the single-strand conformation polymorphism (SSCP) markers *DG279* and *DG371*. This population was used in the second fine-mapping step presented here. Finally, a subset of the fine-mapping population was developed to a fourth generation of self-fertilised F4 recombinants, which were fixed (recombinant with respect to the parents) homozygotes at the *Rht8* locus. These recombinants were used to order markers around *Rht8* between *DG279* and *DG371*. The populations established from both these previous efforts are excellent resources for further map-based cloning of *Rht8*, described in this Chapter.

The second requirement for map-based cloning is markers which saturate the region of interest. In the absence of a physical map for wheat chromosome 2DS, comparative genomics was used to fine-map *Rht8.* Previously, the target region was saturated with gene-based markers using syntenic intervals in Brachypodium and rice, but low polymorphism between the parent NILs to the original mapping population hampered further marker development (Gasperini et al., 2012). In Chapter 5, markers were developed from SNP arrays and RNA-Seq-enabled bulked segregant analysis. As well as this, the IWGSC CSS contigs were used to target marker design to genes with the greatest probability of lying in the syntenic gene intervals based on 2DS localisation. In addition to the CSS contigs, a WGS assembly (named Chapman scaffolds) was used to extend the sampled sequence space (Chapman et al., 2015). However, comparative genomics approaches rely on a good degree of micro-collinearity between candidate gene intervals in related species. In certain situations collinearity has been sufficient to enable cloning of wheat genes e.g. *Ph1* (Griffiths et al., 2006). In other cases, too many local rearrangements have hindered gene isolation, therefore working on a case-by-case basis is key (Krattinger et al., 2009a). The draft sequence of barley, more closely related to wheat than Brachypodium and rice (Chapter 1), showed superior synteny to the wheat *Rht8* interval than the other two species. This was explored in Chapter 5.

Resources made available at the end of this project offered a way to explore the micro-collinearity further and even circumvent the limitations. A WGS draft sequence of *Ae. tauschii*, the D-genome progenitor to hexaploid wheat, was published in 2013 (Jia et al., 2013). In the same year, a 4.03 Gb physical map

was released (Luo et al., 2013). In 2015, this data was compiled into one location (http://aegilops.wheat.ucdavis.edu/ATGSP/, 2015). As part of the work, BAC clones were compiled into minimal tiling paths to assemble BAC contigs. These BAC contigs were further combined with the WGS reads to assemble scaffolds. The BAC contigs were anchored using a genetic map containing over 7000 SNPs from an iSelect SNP array. The contigs, scaffolds and SNP-marker sequences were deposited in the database (Luo et al., 2013). The physical map was also made comparative between *Ae. tauschii,* Brachypodium and rice (Luo et al., 2013). This provided a powerful resource to identify an *Rht8* interval which had higher genetic resolution than the *T. aestivum* population sequencing (POPSEQ) resources, described in Chapter 5, and more closely related to wheat than the previously considered syntenic species. Additionally, the comparative genomics map enabled a thorough evaluation of the micro-collinearity in the *Rht8* region between resources.

Finally, for successful map-based cloning of *Rht8*, there must be accurate phenotyping. This is crucial to establish the order of recombinants and subsequently markers around the genetic interval. This Chapter presents the work which was used to score the graphical genotype at the *Rht8* locus based on plant-height data.

In this Chapter, the markers from Chapter 5 were fine-mapped and syntenic intervals harbouring *Rht8* established. The fine-mapping strategy was made possible by extensive phenotyping at the *Rht8* locus in order to establish the graphical genotype for accurate ordering of markers. Further, genetic intervals were demarcated in the most current wheat genomic resources: the IWGSC-2 CSS contigs anchored to genetic bins by POPSEQ, and the Chapman scaffolds also anchored by the same doubled haploid population. The major recent advances in *Ae. tauschii* genetic and physical maps enabled a more systematic assessment of the collinearity of the *Rht8* interval. The highly-saturated maps were also used to evaluate the performance of the wheat reference used in the bulked segregant analysis (BSA) approach outlined in the previous Chapter. Finally, the gene content in *Ae. tauschii* and *T. aestivum* is explored here, focusing particularly on differentially expressed genes and genes related to biological processes which could explain the *Rht8* phenotype.

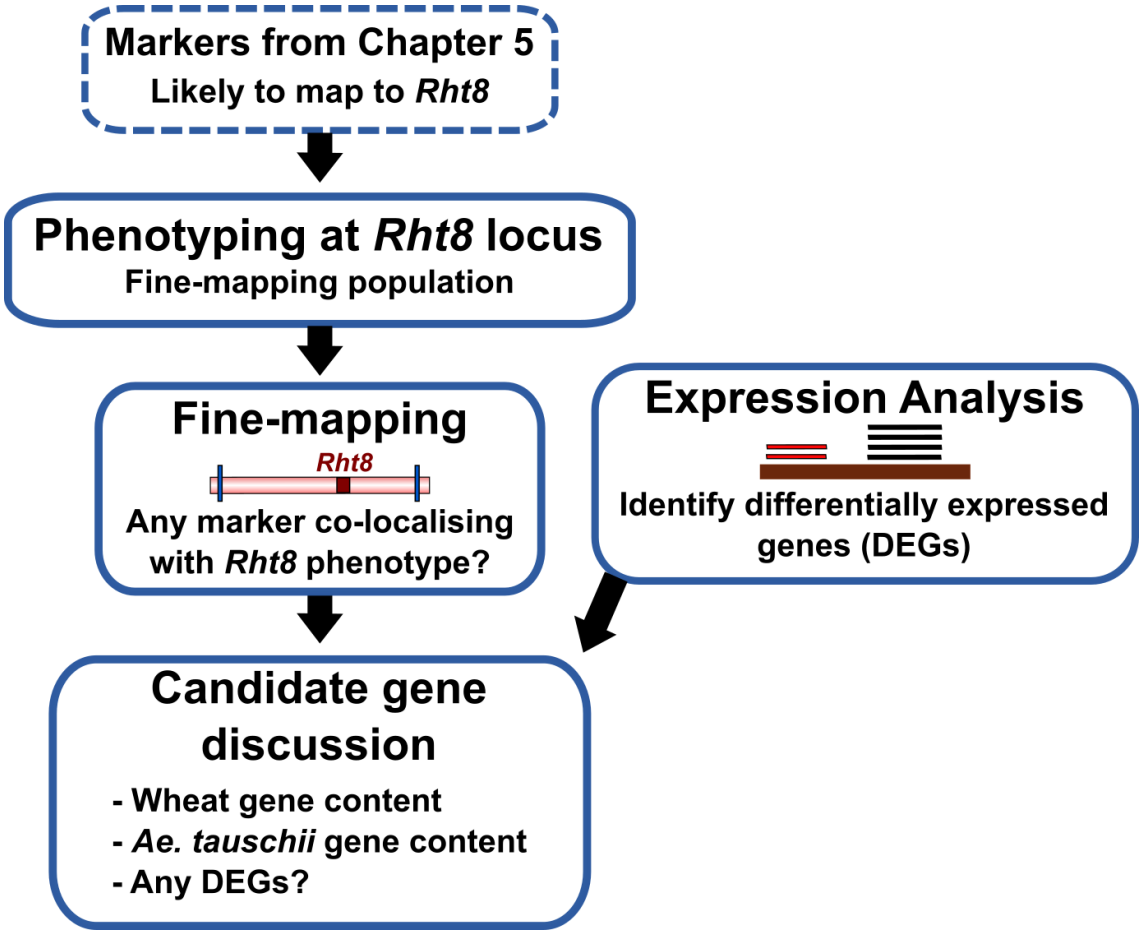


Figure 6.1: Schematic diagram of the workflow in Chapter 6 to fine-map and further characterise Rht8.

6.2 Phenotyping the fine-mapping population at the *Rht8* locus

6.2.1 Measuring height in glasshouse-grown plants

A graphical genotype score of short/tall at the *Rht8* locus of each recombinant in the fine-mapping population was a pre-requisite to mapping markers developed in Chapter 5. Accurate assignment of recombinants as short/tall was also critical for the RNA-Seq strategy outlined in Chapter 5.

In order to ascertain the phenotype, the fine-mapping F4 recombinants were grown alongside the short and tall parent NILs, initially in a glasshouse trial in eight randomised blocks in the winter of 2012. Plant height was measured at maturity. The short parent NIL RIL4, homozygous for *Rht8*, had a mean height of 75.7±1.2 cm (N=14). The tall parent Cappelle-Desprez (CD), homozygous for *rht8*, had a mean height of 82.3±1.2 cm (N=16). The difference in height between the two parents was 6.6 cm, somewhat less than the 8 – 10 cm reported by Gasperini (2010). The reduced differential in height was due to a decreased height of CD, since the mean height of RIL4 was comparable to that found previously (77 cm). The mean heights of the F4 recombinants (73 individuals) and parents are available in Appendix 6.1. The height frequencies were plotted in a histogram and a bimodal distribution was observed, with a split at 76.75 cm which enabled recombinants to be assigned to two distinct groups based on phenotype: recombinants below 76.75 cm were typed 'short', those above typed 'tall' (Figure 6.2). The mean height of the 'shorts' was 74.8±0.5 cm (N=32) (shown by a dashed line) and the mean height of the 'talls' was 81.0±0.4 cm (N=42). The mean height of each recombinant group short/tall was below the mean of the corresponding short/tall parent (shown by a continuous line and arrow in Figure 6.2A).

A total of 14 short and 16 tall recombinants from the extreme ends of the distribution were selected as candidates for the short and tall bulks for RNA-Seq, described in Chapter 5. The short outlier at 65 cm was not included in the bulk.

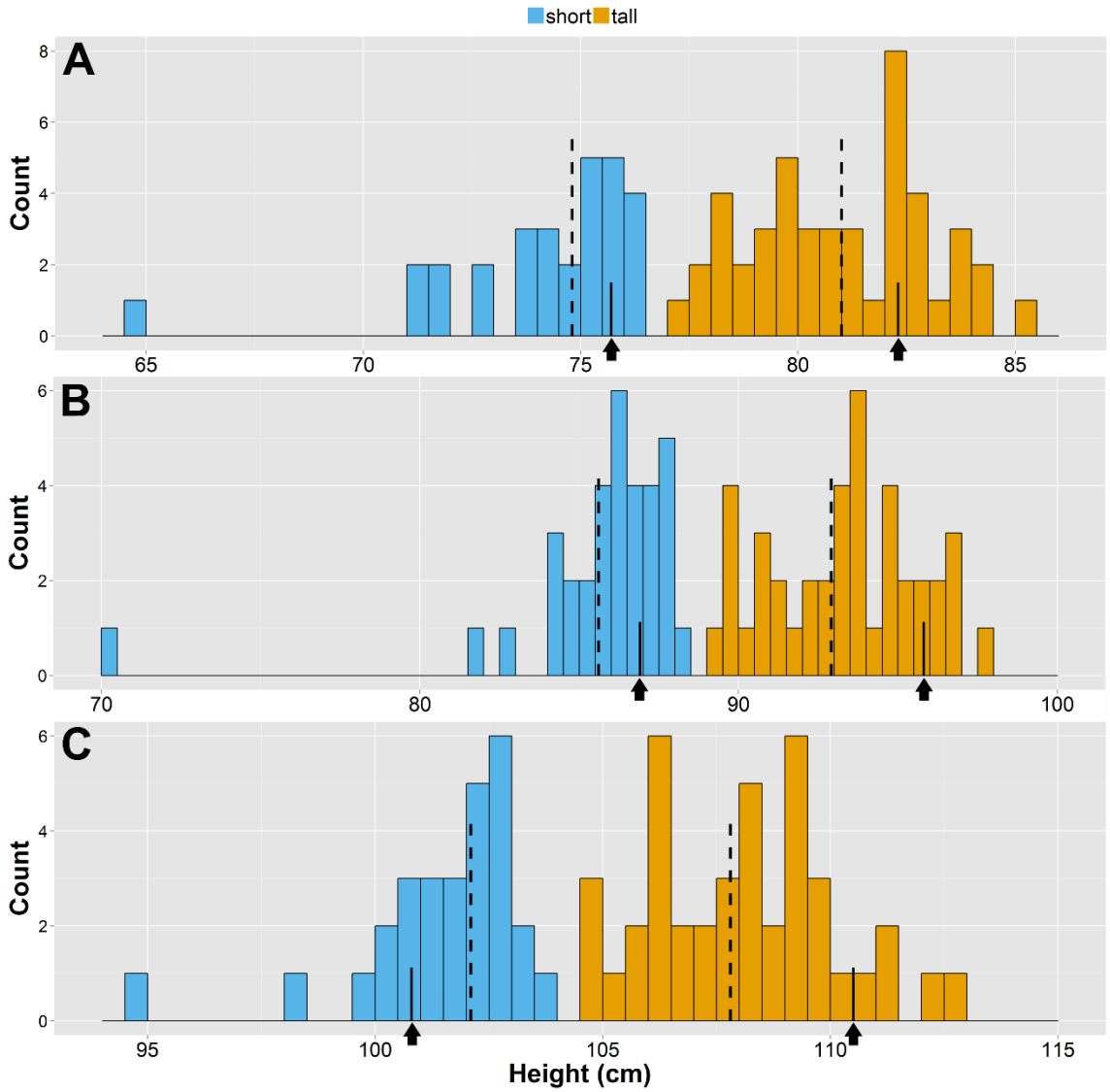


Figure 6.2: Height frequencies of the recombinants and parents to the fine-mapping Rht8 population across three locations. Mean height data was used. The y-axis represents the frequency count. The dashed lines represent the mean of the short (left) and tall (right) recombinants. The solid lines highlighted with the arrows represent the height of the corresponding parent of the short/tall group (RIL4 or Cappelle-Desprez). The recombinants were typed short or tall based on the bimodal split at each location. (A) Glasshouse 2012. Bimodal split at 76.75 cm. (B) Church Farm 2014. Bimodal split at 88.75 cm. (C) Morley 2014. Bimodal split at 104.25 cm. Full data shown in Appendix 6.1. The extreme short outlier (F4-1-2-2-1) was not selected for RNA-seq (Appendix 6.2) but was used for fine-mapping.

6.2.2 Sterility induced in glasshouse conditions

Sterility was observed in approximately half of the plants in the glasshouse experiment. There was visible absence of grain in florets of the spike, and it was hypothesised that this also reduced the overall height. The subset of extreme short and tall recombinants (30 in total) were measured for spike length and sterility was scored on a scale of 0 to 5 according to severity (5 being totally sterile) as described in Chapter 2. A linear model was fitted and sterility had a significant effect on spike length ($P<0.05$) (shown in Appendix 6.2), however

there was no recombinant*sterility interaction, suggesting that all recombinants were affected similarly. It was critical to have a reliable height phenotype for fine-mapping and for the selection of short/tall bulks for RNA-Seq. Since the latter was of primary concern earlier on in the project (see 5.3.3), the subset of 30 recombinants selected as candidates for the short and tall bulks were grown to obtain more phenotypic data. This was achieved in an experiment in the spring of 2013 with 24 randomised blocks. All plants were measured for sterility and 93% of the 880 plants had some degree of sterility (Appendix 6.3). Sterility had a highly significant effect on total height ($P<0.001$) and there was also a significant genotype*sterility interaction ($P<0.05$) (Appendix 6.2). A datalogger which measured glasshouse conditions at canopy level recorded temperatures of +50°C during grain filling (Appendix 6.4) which likely caused the severe sterility. Due to the concern of mis-typing recombinants, coupled with lack of confidence in appropriate glasshouse conditions to eliminate sterility, the recombinants (now F5 but herein called F4 for consistency with the glasshouse results) were sown in the field in winter 2013 in two Norwich sites.

6.2.3 Measuring height in the field and final typing at the Rht8 locus

The 73 recombinants were grown with the short and tall parents in the 2013-14 season in randomised blocks with five replicates in two locations. At Morley, six recombinants were not sown due to lack of seed (specified in Appendix 6.1). Seed was drilled 12 cm apart using the spaced drill, which meant that in actuality the replication was much higher than N=5, since plants within each row could be measured as individual replicates if necessary and an ‘average by eye’ was conducted when selecting plants within a row. A representative plant from each row was measured at maturity.

Overall, the plants were taller in both field experiments than the glasshouse (Figure 6.2B&C), and the differential between short and tall parent was similar to the 8-10 cm reported previously by Gasperini (2010) and Korzun et al. (1998). The short parent, RIL4 had a mean height of 86.9$\pm$0.5 cm (N=40) at Church Farm, which was 8.9 cm shorter than CD (95.8$\pm$0.7 cm, N=35). Morley had the tallest plants overall and the greatest difference in height between the parents. RIL4 had

a mean height of 100.8±0.3 cm (N=53) and CD was 9.7 cm taller (110.5±0.4 cm, N=49). The mean height frequencies of the recombinants were bimodally distributed as before (Figure 6.2) and independently typed as short or tall at the two field locations.

Using the independently-assigned phenotype score the three locations, each recombinant was assigned a consensus score at the *Rht8* locus. Mostly, the same score was assigned at each location. For 13 recombinants, there was a contrasting score in one of the locations, and the consensus score was based on the majority score (confirmed by two out of three locations). These are marked by a darker colour and with an asterisk in Appendix 6.1. Of these contrasting scores, the majority (8 out of 13) were from the glasshouse. This highlighted first, the importance of the field experiments to provide more reliable height data, and second, the ambiguity in scoring a trait seemingly trivial to phenotype.

6.3 Fine-mapping

Rht8 was fine-mapped in stages of increasing resolution since mapping directly with the fine-mapping (FM) recombinants (Figure 6.3) would give spurious results with markers which were outside the *Rht8* interval when considered at lower resolution.

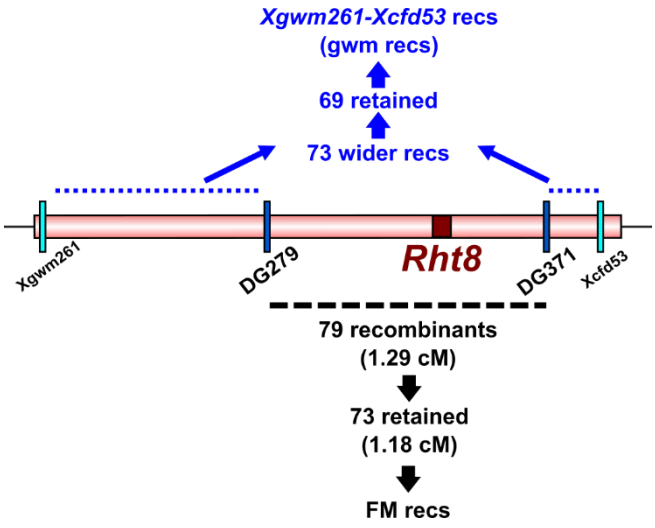


Figure 6.3: Defining the FM and gwm recombinants used in step 2 and 3 of the fine-mapping of Rht8.

6.3.1 Step 1: Coarse mapping with 2D RILs

First, the 2D RIL population (described in Chapter 5) was used to coarsely map the markers. The markers were arranged into 22 2D RIL classes according to markers which shared the same genotype (full details in Appendix 5). The main group of interest surrounded the *DG279*/*DG371* markers on 2D used in mapping by Gasperini (2010), since this group was likely to map to *Rht8*. This group is marked 'B' in Figure 6.4. The outgroups to group B contained markers on chromosomes outside 2D ('C' in Figure 6.4) or on 2D but outside group B (labelled 'A'). These markers were discarded (blue dotted line in Figure 6.4). Group B contained 62 markers across 2D RIL classes 1-22. These markers were retained (marked with the black dashed line in Figure 6.4). This first step was used to filter markers in Chapter 5 which were likely to map to *Rht8*.

6.3.2 Step 2: Medium-resolution mapping with *Xgwm261-Xcfd53* recombinants

In the second step, the fine-resolution mapping population for *Rht8* was considered to discriminate which markers were within the *DG279-DG371* interval from those in the wider region on 2DS. The fine-mapping population was developed from crossing selected 2D RILs containing the *Mara*-derived *Rht8* allele into the *Cappelle-Desprez* background, as outlined by Gasperini (2010). From the original F_2 population of 3104 F2 individuals, a total of 152 recombinant families with genotypes different to the crossing parents were identified. These were used by Gasperini et al., (2012) to fine-map markers between *Xgwm261* and *Xcfd53.* From this F2 recombinant subset, 79 recombinant individuals were used to resolve *Rht8* between *DG279* and *DG371*. Of these 79 recombinants, some were found to be scoring errors and others sterile, so 73 were retained in this work (taken to F4 and phenotyped in 6.2), which decreased the interval size (Figure 6.3). The wider recombinants outside *DG279-DG371* which could be retained were used in the second step (abbreviated to gwm recs). The 62 markers from the first step (Figure 6.4) were grouped into 30 distinct genotypic marker classes, based on markers which shared the same graphical genotype when typed with the gwm recs (full details in Appendix 6.5). There was a clear separation between marker classes which mapped between *Xgwm261* and

DG279 (discarded, shown in blue, Step 2 in Figure 6.4) and the marker classes which grouped within the *Rht8* interval (shown in red, Step 2 in Figure 6.4). These marker classes (13-29), containing 33 markers, were retained for the final fine-mapping step.

6.3.3 Step 3: Fine-mapping with FM recombinants

In the third and final step of fine-mapping, the remaining 33 markers could be grouped into seven classes, labelled A-G, according to their genotypes when typed with the 73 FM recombinants (Figure 6.4 and in full in Appendix 6.6). The recombinants were grouped into 12 recombinant classes (labelled I – XII in Table 6.1) according to their graphical genotypes and by scoring each of the recombinants as short (b)/tall (a) at *Rht8*. As described previously in 6.2, some recombinants had conflicting short/tall scores and where these conflicts arose, the majority score (according to two out of the three locations) was used to establish a consensus. These cases are marked with an asterisk. Where a recombinant with a less confident *Rht8* score was in a recombinant class with individuals of a confident score (classes VI, VIII and XII), both scores were included for transparency (Table 6.1). Given the ambiguity in height score in recombinant classes with either a single (VII) or small number (three, in VIII) of recombinants, there was reduced confidence in defining the exact short/tall boundary at the *Rht8* locus.

Nevertheless, the recombinants were ordered using the height data and arranged to minimise double recombination events. A total of 75 recombination events were counted between the seven marker classes. The recombination frequency was used to identify eight genetic intervals, shown in Step 3 of Figure 6.4. *Rht8* was mapped to a 1.015 cM interval between marker classes D and E, with the largest interval of 0.95 cM between D and *Rht8*. This placement of *Rht8* was different to the more central location of *Rht8* between *DG279* and *DG371* mapped previously (Gasperini et al., 2012), which is likely due to the differences in phenotype scoring at the *Rht8* locus. Crucially, no marker class co-segregated with *Rht8*.

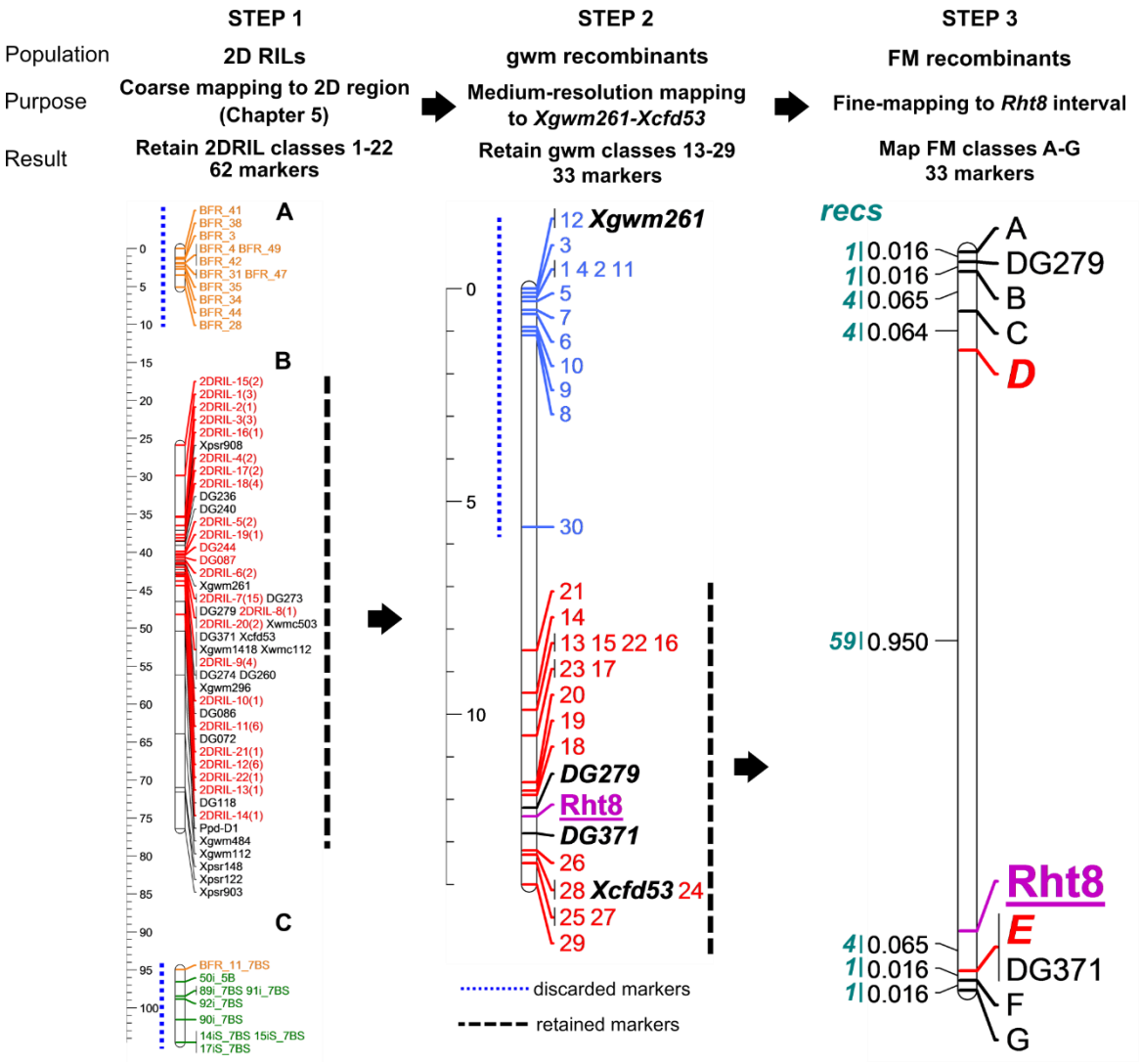


Figure 6.4: Fine-mapping Rht8. Rht8 was fine-mapped in steps from left to right, described in the top panel. Markers retained from each step are marked in black dashed lines and used in the next step as shown by the arrows. Markers discarded since they did not map in close proximity to Rht8 are indicated with blue dashed lines. Marker classes retained in the 2D RIL population and Xgwm261-Xcfd53 recombinants are coloured red. In the fine-mapping using the FM population, 73 recombinants were used and a total of 75 recombination events were counted. These are indicated in the 'recs' column (far right), along with the cM distance of the intervals between the marker classes. Rht8 was fine-mapped to a 1.015 cM interval between marker classes D and E. Distances in the mapping using the 2D RIL population (left) and Xgwm261-Xcfd53 recombinants (centre) are relative measures only for ordering marker classes. The difference in interval size between classes B-C and C-D in Step 3 (right) is due to rounding in the mapping software (the value is 0.0644 to 4 d.p.).

															Total
				No. markers in class	17	-	1	3	1	-	5	-	5	1	**33**
	Mean height of FM class			Marker Class	A	DG279	B	C	D	**Rht8**	E	DG371	F	G	
No. recs in class	GH	CF	MOR	FM class											
1	64.7	70.4	94.8	I	b	b	b	b	b	b	a	a	a	a	
-	75.7	86.9	100.8	RIL4	b	b	b	b	b	b	b	b	b	b	
1	71.5	85.4	-	II	a	b	b	b	b	b	b	b	b	b	
1	78.1	86.3	102.7	III	a	a	b	b	b	b*	b	b	b	b	
2	75.3	86.0	100.9	IV	a	a	a	b	b	b	b	b	b	b	
1	76.2	87.6	102.5	V	a	a	a	a	b	b	b	b	b	b	
25	75.0	86.1	102.4	VI	a	a	a	a	a	b/b*	b	b	b	b	
1	79.2	86.9	102.9	VII	a	a	a	a	a	b*	b	b	b	a	
3	78.1	86.9	105.8	VIII	a	a	a	a	a	a/a*	b	b	b	b	
2	80.7	91.5	108.2	IX	b	b	b	a	a	a	a	a	a	a	
3	81.8	92.3	109.1	X	b	b	b	b	a	a	a	a	a	a	
1	80.0	93.3	108.6	XI	b	b	b	b	b	a	a	a	b	b	
32	81.2	93.6	107.9	XII	b	b	b	b	b	a/a*	a	a	a	a	
-	82.3	95.8	110.5	CD	a	a	a	a	a	a	a	a	a	a	
Total	**73**														

Table 6.1: Graphical genotypes of the fine-mapping population grouped in recombinant classes according to their genotype and phenotype at the Rht8 locus. Markers were grouped into classes and shown along with the short (RIL4) and tall (CD) parent NILs. The mean height of each recombinant class from the three locations are shown, where 'GH' = Glasshouse 2012, 'CF' = Church Farm 2014, 'MOR' = Morley 2014.

Marker class A was estimated to be 0.016 cM away from *DG279* due to one recombination event. It is likely that *DG279* was previously mis-scored at this recombinant, since the genotype in the individual could not be reproduced. Therefore, *DG279* could be combined into marker class A. However, *DG279* was presented in a separate class to marker class A (Table 6.1) for consistency with the previously published data (and not correcting this likely error).

A single marker comprised marker class D (*2DS_5375260*, Table 6.2). This SSR marker was designed to amplify a microsatellite on 2DS_5375260: 4772-5019. The marker *55_uni*, in marker class C, four recombination events apart from marker class D (0.064 cM, Figure 6.4) was a KASP marker designed on a SNP at the beginning of the same CSS contig. Both these markers mapped to the same 2D RIL and gwm rec marker class (Table 6.2). Since the genotypic data

was consistent and reproducible, these markers were left in distinct marker classes.

Marker class E co-segregates with *DG371*, but was positioned separately for clarity and because one marker in the class (*2DS_208*) was ordered by barley physical position to define a smaller region/closer to *Rht8* than *DG371* (*2DS_208*, barley gene *MLOC_58453*, chromosome 2H:18521524, versus *DG371*, chromosome 2H:18522971, Table 5.4). Marker classes F and G were placed distal to marker class E to minimise gene conversion events with recombinant class XI, however, this class comprises a single recombinant. Therefore, it is possible that a genuine gene conversion event could place classes F and G closer to *Rht8*. The ordering of these classes based on physical data from *Ae. tauschii* and *T. aestivum* later confirmed the existing genotypic order (6.3.5).

6.3.4 Syntenic relationship of the Rht8 *locus with barley, Brachypodium and rice*

It was already established in Chapter 5 that previous comparative analysis between rice and Brachypodium in the *Rht8* region (Gasperini et al., 2012) gave an incomplete picture of the gene content of the locus. It was shown in Chapter 5 that when only 2DS CSS contigs in the region were considered, barley had the closest synteny to wheat, and that the synteny between the barley candidate-gene interval and Brachypodium and rice was poor. The poor micro-collinearity in the *Rht8* region was also seen when fine-mapping *Rht8* using the marker classes A-G. The marker information from Table 6.2 (with the exception of asterisked markers which had no comparative information or were mapped to non-syntenic chromosomes) was used to anchor markers onto comparative maps of barley, Brachypodium and rice and homologues joined to show the relationship between the species (Figure 6.5).

The first observation was that barley had the most anchored markers, followed by Brachypodium and then rice. This is in line with the findings in Chapter 5, where barley had the best synteny of these three in the region. Due to this attrition across the three comparative species, not all markers could be anchored across the three maps and therefore, obtaining the smallest region in one species was not necessarily confirmed with the same markers in all three maps. Further, due

to the sparseness of the rice map compared to barley, for example, analysing the rice map in isolation by only retaining the markers which map to chromosome 4 might lead to a false sense of high collinearity (Figure 6.5D).

The second observation was that macro-synteny was good in places. The marker classes from A-G which were best annotated in barley did show a general progression along barley chromosome 2. Further, in marker class A, physical positions from barley synteny could be used to order markers (Figure 6.5A) which otherwise co-localised to the same genetic marker class, and this translated well across the three comparative maps. In other marker classes, such as D and E, the synteny was much more disrupted. It should also be noted that a significant number of markers within classes A – G could not be anchored onto the homologous chromosomes, since they mapped outside of chromosomes 2H, 5 and 4 in the corresponding species. This is shown in Table 6.2. Furthermore, many markers which were developed following the reasoning that they first mapped to a wheat 2DS contig and second were also anchored physically to either one of the syntenic intervals in Figure 6.5 (as described in Chapter 5) were discarded in the first two fine-mapping steps, since the genotypes did not localise close to *Rht8*.

Perhaps most crucially, combining the observations made above, it was not possible to define the *Rht8* region using the tightest (genetically) flanking marker classes D and E in all cases, or indeed using the same markers across all three species. Nevertheless, it was possible to define physical intervals for *Rht8* in all three comparative species, summarised in Figure 6.5. The smallest barley region that could be demarcated was between *2DS_6* in marker class A (*MLOC_62798*, chromosome 2H:15618954) and, mapping the other side of *Rht8*, *2DS_208* in marker class E (*MLOC_58453*, chromosome 2H:18521524), defining a 2.9 Mb region on barley chromosome 2H. On Brachypodium chromosome 5, *52i* (*Bradi5g03460*, 4042855) in marker class B and *2DS_138* (*Bradi5g04130*, 5393012) in marker class F defined a 1.36 Mb interval and the same markers also defined a 1.34 Mb interval between *Os04g0209200* and *Os04g0229100* on rice chromosome 4.

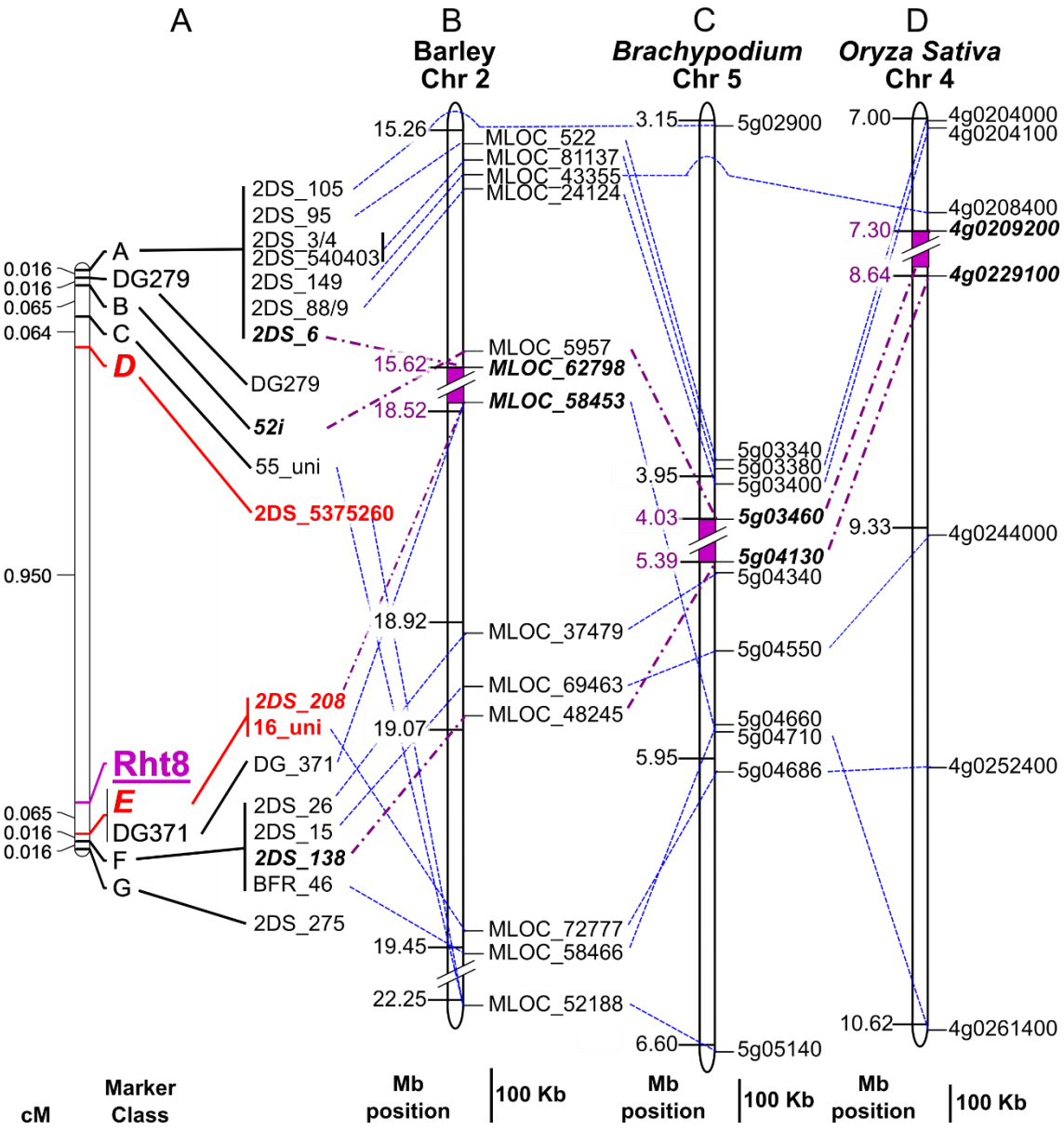


Figure 6.5: Fine-mapping of the Rht8 locus and alignment with the homologues of barley, rice and Brachypodium on physical maps. (A) Fine-mapping marker classes A-G are sub-divided into individual markers since markers within the same class mapped to different comparative map positions. The closest flanking markers to Rht8 in classes D and E are shown in red. Markers that mapped to genes used to define the smallest syntenic Rht8 intervals described in the text are in bold italics. The bottom of the map is towards the centromere, and the top towards the telomere. (B) The physical map of barley chromosome 2 (C) The physical map of Brachypodium chromosome 5 and (D) The physical map of rice chromosome 4. Homologues between the genomes are joined with blue dashed lines. Where no homology was found between all three species, the line loops over the absent map, or is discontinued where no further map position was found. Genes used to define the smallest syntenic Rht8 intervals are shown in purple dot-dash lines. The genes defining these intervals are highlighted in bold italics and the intervals themselves are filled in purple. The sizes of the syntenic Rht8 intervals were calculated from flanking Mb positions, which are coloured purple. Each comparative map is drawn to its own scale indicated at the bars along the bottom. Breaks in the scale which were necessitated to contract the maps in order to fit the page are indicated by the double diagonal bars and the break in scale is indicated by the Mb positions.

							Ae. tauschii												Chapman	
Marker		**Source**	**2D RIL class**	**gwm class**	**FM class**	**SNP marker**	**Score**	**cM**	**BAC contig**	**BAC scaffold**	**IWGSC contig**	**Barley**	**B. distachyon**	**Rice**	**v3.3 cDNAs**	**IWGSC2 bin**	**Scaff**	**Bin**
63_uni	*	UniGenes	20	14	A	-	-	-	-	-	2DS_5359909	-	15266465	-	Os10g21560	mrna008489	-	17.34	860255	13.64
2DS_1	*	IWGSC-1	7	14	A	-	-	-	-	-	2DS_5359909	-	15266465	-	Os10g21560	mrna008489	-	17.34	860255	13.64
2DS_95		IWGSC-1	7	17	A	-	-	-	-	-	2DS_5341322	MLOC_522	15276497	Bradi5g03340	-	-	-	17.34	1822283	13.64
2DS_3		IWGSC-1	7	13	A	AT2D1060	8128	30.45	ctg494	4046.5	2DS_5337443	MLOC_81137	15280265	Bradi5g03380	Os04g0204000	mrna103570	-	-	472854	13.64
2DS_4		IWGSC-1	7	13	A	AT2D1060	8128	30.45	ctg494	4046.5	2DS_5337443	MLOC_81137	15280265	Bradi5g03380	Os04g0204000	mrna103570	-	-	472854	13.64
2DS_540403		IWGSC-2	7	22	A	AT2D1060	1751	30.45	ctg494	4046.5	2DS_540403	MLOC_81137	15280265	Bradi5g03380	Os04g0204000	-	-	17.34	472854	13.64
2DS_149		IWGSC-1	12	15	A	-	-	-	-	-	2DS_5319467	MLOC_43355	15293278	-	Os04g0208400	-	-	17.34	1194015	13.64
2DS_88		IWGSC-1	7	17	A	AT2D1061	2586	30.54	ctg494	4046.5	2DS_5321770	MLOC_24124	15296824	Bradi5g03400	OS04G0204100	-	-	17.34	1111368	13.64
2DS_89		IWGSC-1	7	17	A	AT2D1061	2586	30.54	ctg494	4046.5	2DS_5321770	MLOC_24124	15296824	Bradi5g03400	OS04G0204100	-	-	17.34	1111368	13.64
2DS_6		IWGSC-1	10	14	A	AT2D1057	5293	30.22	ctg5606	3836.1	2DS_5321865	MLOC_62798	15618954	Bradi4g21260	Os11g0215100	mrna052919	-	17.34	1648489	13.64
2DS_210	*	Limagrain	12	16	A	-	-	-	-	4046.3	2DS_5337059	MLOC_16024	475472562	-	-	-	-	17.34	1822283	13.64
2DS_211	*	Limagrain	12	16	A	-	-	-	-	4046.3	2DS_5337059	MLOC_16024	475472562	-	-	-	-	17.34	1822283	13.64
2DS_212	*	Limagrain	12	16	A	-	-	-	-	4046.3	2DS_5337059	MLOC_16024	475472562	-	-	-	-	17.34	1822283	13.64
2DS_217	*	Limagrain	12	16	A	-	-	-	-	-	2DS_5319965	-	-	-	-	-	-	-	-	
2DS_215	*	Limagrain	12	22	A	-	-	-	-	-	2DS_5379317	-	-	-	-	-	-	-	-	-
2DS_105		IWGSC-1	7	23	A	-	-	-	-	-	2DS_5390456	-	-	Bradi5g02900	-	-	-	17.34	5108430	13.64
8_uni	*	UniGenes	7	21	A	AT2D1056	1759	30.17	ctg7056	4414.4	2DS_5388088	-	-	Bradi4g02250	Os04g012560	-	881779	-	1425718	13.64
52i		iSelect	19	20	B	AT2D1062	4364	30.99	ctg494	4046.11	2BS_4748675	MLOC_5957	15601547	Bradi5g03460	Os04g0209200	-	-	-	1069842	2B
2DS_222	*	Limagrain	7	19	C	-	-	-	-	-	2DS_5342594	MLOC_70393	22198957	Bradi3g16570	Os04g0652600	-	-	17.34	824877	13.64
2DS_223	*	Limagrain	7	19	C	-	-	-	-	-	2DS_5342594	MLOC_70393	22198957	Bradi3g16570	Os04g0652600	-	-	17.34	824877	13.64
55_uni		UniGenes	7	18	C	AT2D1065	4313	31.41	ctg494	3820.1	2DS_5375260	MLOC_52188	22256327	Bradi5g05140	Os03g0804300	mrna102641	7779113	-	2461570	13.64
2DS_5375260		IWGSC-2	7	18	D	AT2D1065	4313	31.41	ctg494	3820.1	2DS_5375260	MLOC_52188	22256327	Bradi5g05140	Os03g0804300	mrna102641	7779113	-	2461570	13.64
2DS_208		Limagrain	9	28	E	-	-	-	-	-	2DS_5389857	MLOC_58453	18521524	Bradi5g04710	Os04g0261400	mrna074899	-	17.34	2409583	15.92
16_uni		UniGenes	7	28	E	-	-	-	-	-	2DS_5364728	MLOC_72777	19455074	Bradi5g04686	Os04g0252400	mrna103168	7646108	17.34	909875	15.92
2DS_192	*	Limagrain	9	28	E	-	-	-	-	-	2DS_5378845	MLOC_76709	39057764	Bradi1g69730	Os10g0399100	-	-	16.95	6258899	14.78
2DS_187	*	Limagrain	9	28	E	-	-	-	-	-	2DS_5354335	-	-	Bradi5g15565	-	-	-	17.34	5459982	15.92
2DS_201	*	Limagrain	9	28	E	-	-	-	-	-	2DS_5389660	MLOC_5590	Chr_1	-	-	-	-	17.34	944725	15.92
2DS_26		IWGSC-1	6	26	F	AT2D1081	2350	37.39	ctg2121	6629.19	2DS_5390977	MLOC_37479	18924902	Bradi5g04340	Os03g0856000	mrna057813	1178953	33.06	6525784	19.33
2DS_15		IWGSC-1	6	25	F	-	-	-	-	-	2DS_5390752	MLOC_69463	19016685	Bradi5g04550	Os04g0244400	-	1189621	33.06	2457722	19.33
2DS_138		IWGSC-1	5	27	F	AT2D1083	3917	37.98	ctg4363	3809.1	2DS_5318296	MLOC_48245	19071820	Bradi5g04130	Os04g0229100	-	-	33.06	1488113	19.33
BFR_46		v3.3 cDNAs	7	25	F	AT2D1072	12180	35.66	ctg1775	4452.3	2DS_5371750	MLOC_58466	19491674	Bradi5g04660	Os02g0319800	-	1264216	17.34	3953562	15.92
2DS_66	*	IWGSC-1	5	25	F	-	-	-	-	-	2DS_5357871	MLOC_26534	608666509	-	-	mrna102226	-	33.06	1978876	20.69
2DS_275		IWGSC-2	14	29	G	-	-	-	-	-	2DS_5344159	-	-	-	-	-	-	33.06	771234	24.77
DG279						AT2D1062	10080	30.99	ctg494	4046.11	2DS_5377037	MLOC_5957	15601547	Bradi5g03460	Os04g0209200	-	-	17.34	1111368	13.64
DG371						AT2D1083	84.2	37.98	ctg4363	4243.1	2BS_5196588	MLOC_58453	18522971	Bradi5g04710	Os04g0261400	-	-	17.34	2409583	15.92

Table 6.2: Markers used in the final step to fine-map Rht8. Markers are annotated by marker class at each mapping step and with comparative genomic data to syntenic species. Wheat annotation is shown in terms of the POPSEQ bin the 2DS CSS contig was mapped to in the IWGSC-2 and Chapman datasets. No information indicates the contig is not in the POPSEQ data. Each contig was also anchored on the v3.3 cDNAs where possible. Anchoring of DG279 and DG371 is shown fully in Chapter 5 but included here for ease of comparison between all markers.

6.3.5 Identification of Rht8*-equivalent region in* Ae. tauschii *and integration with* T. aestivum *resources*

Towards the end of this project, major advances in gene discovery and building of a highly-saturated genetic *Ae. tauschii* map were achieved using survey sequencing (Jia et al., 2013). Additionally, an *Ae. tauschii* physical map was reported, which spanned 672 Mb of chromosome 2D defined by 385 anchored BAC contigs carrying 1282 markers (Luo et al., 2013). The completion of these resources is still ongoing, but they were compiled in one downloadable location in 2015 (http://aegilops.wheat.ucdavis.edu/ATGSP/, 2015).

To identify the orthologous *Rht8* region in *Ae. tauschii*, the linkage map of the *Rht8* region (Figure 6.5A) was anchored to BAC contigs in the *Ae. tauschii* chromosome 2D physical map (Figure 6.6; Luo et al. 2013). This was achieved by BLASTN homology searches of the 2DS CSS contigs to the markers in classes A-G as queries against the *Ae. tauschii* SNP marker extend contig database (http://probes.pw.usda.gov/WheatDMarker/phpblast/blast.php). The 2D gene list was then consulted to retrieve the BAC contigs anchored by the SNP markers. The same 2DS CSS contigs were queried against the IWGSC-2 data, where contigs had been anchored to POPSEQ bins (Mascher, 2014). The contigs were also used as queries against the Chapman scaffolds which were ordered by the same high-density genetic map and hosted on CerealsDB (Mascher et al., 2013, CerealsDB, 2015a).

The marker order in the *Rht8* linkage map was maintained between the *Ae. tauschii* physical map (Figure 6.6A&B) and *T. aestivum* (Figure 6.6C&D), though only 13 markers out of 33 across all classes A – G could be mapped to an *Ae. tauschii* SNP marker. Most markers could be mapped to the POPSEQ bins (Table 6.2).

In the POPSEQ data, marker classes A – E anchored to the 17.3 cM bin in IWGSC-2, with marker classes F and G anchoring to the 33.1 cM bin. This placed *Rht8* in the 17.3 cM bin. The Chapman scaffolds showed a higher resolution, with marker classes A – D mapping to 13.64 cM and E to 15.9 cM. This resolved *Rht8* to a 2.3 cM interval between classes D and E, compared with 1.015 cM in the

Rht8 fine-map. Classes F and G then progressed down in a linear fashion to 24.8 cM in class G (Table 6.2 and (Figure 6.6C & D).

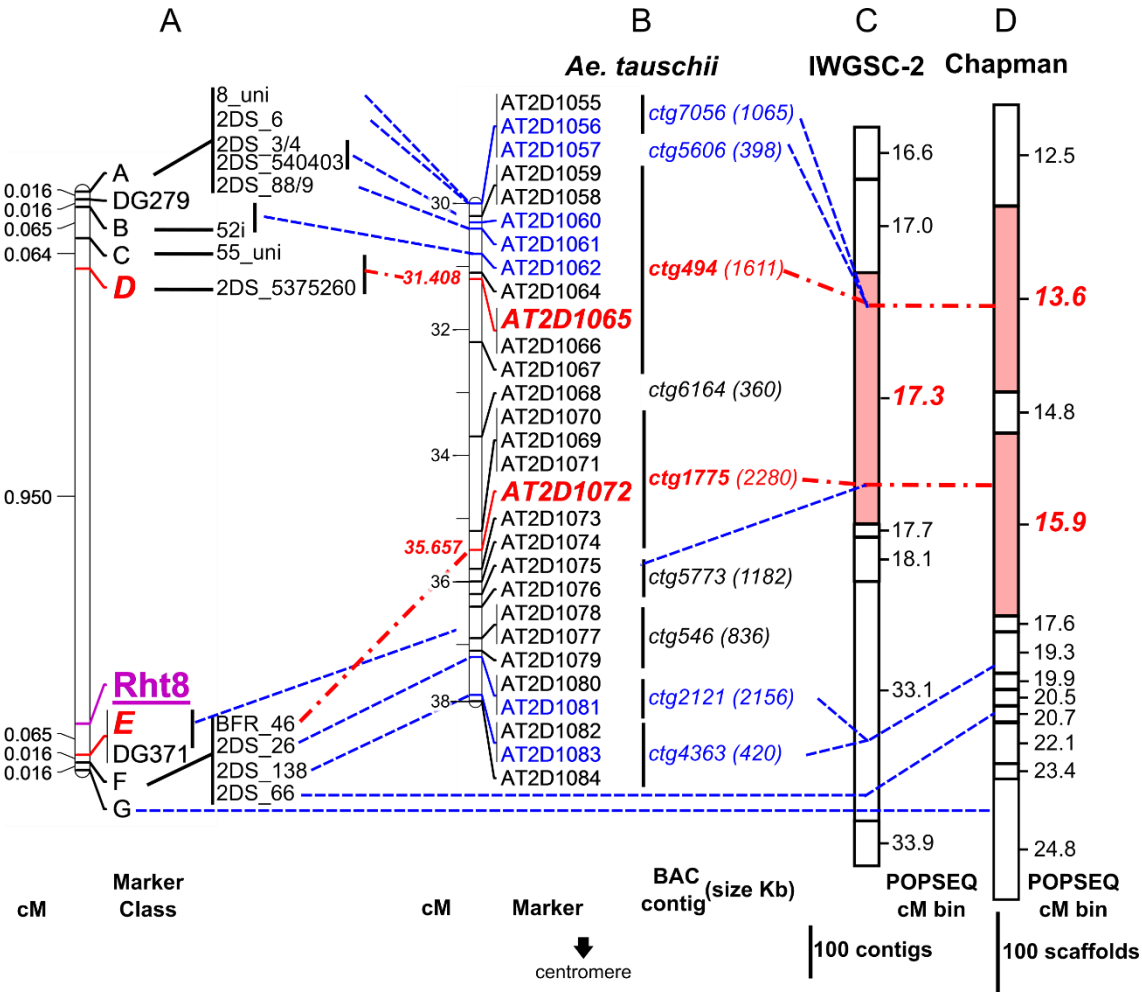


Figure 6.6: Linkage map of Rht8 and anchoring to Ae. tauschii BAC contigs and wheat genetic maps from POPSEQ data (A) The linkage map of Rht8 and details of which markers from the marker classes A-G were anchored onto BAC contigs and wheat POPSEQ genetic bins. The orthologous relationships are shown in blue dashed lines. The markers used to define the smallest intervals defined in Ae. tauschii and wheat are shown in red dot-dash lines. Markers which did not anchor in Ae. tauschii but anchored in wheat pass under (B). (B) The genetic position of the Ae. tauschii SNP markers and the corresponding BAC contigs along chromosome 2D. The smallest genetic interval defined in Ae. tauschii is highlighted in red along with the corresponding markers and BAC contigs. (C) The wheat CSS contigs and (D) Chapman scaffolds anchored to genetic POPSEQ bins. The bins were delimited using the 2DS CSS contig corresponding to the markers in (A). The smallest interval is highlighted in red. The POPSEQ bins are scaled to represent the relative size in terms of contig/scaffold number.

The markers from the *Rht8* linkage map (Figure 6.6A) anchored to six BAC contigs in the *Ae. tauschii* physical map (Figure 6.6B). There was no instance of a marker from each side of the *Rht8* interval mapping to the same BAC contig or scaffold (Table 6.2). Therefore, it was not possible to anchor *Rht8* to a single BAC or to overlapping BACs (since the assembly is incomplete). The *Ae. tauschii* SNP markers *AT2D1065* and *AT2D1072* defined a 4.25 cM genetic interval on 2D.

Marker class E was used to define one end of the *Ae. tauschii* interval. *BFR_46* mapped to 35.657 cM and BAC contig 1775. The other end of the *Ae. tauschii* interval was anchored using *2DS_5375260* in marker class D, which was assigned to the 31.408 cM bin and BAC contig 494.

The 4.25 cM interval between marker classes D and F in *Ae. tauschii* compares with the 1.031 cM between the same marker classes found using the fine-mapping *Rht8* population (Figure 6.6A). The *Ae. tauschii* interval contained six SNP marker sequences, with sizes from 5 – 13 Kb.

The exact physical space of this interval cannot be ascertained since the gaps between the BAC contigs have not been closed. However, anchored genetically in the interval between ctg494 and ctg1775 is ctg6164, which is 360 Kb (Figure 6.6B). Taking the three BAC contigs together and using the extreme genetic position at either end (AT2D1059 at 30.401 cM and AT2D1074 at 36.203 cM) defines a 5.8 cM genetic distance which contains 4.251 Mb (sum of BAC contig size). This results in a recombination rate of 1.36 cM/Mb, compared with the average on *Ae. tauschii* 2D of 0.32 cM/Mb (Luo et al., 2013) and wheat 3B ranging from 0 to 2.3 cM/Mb with an average of 0.16 cM/Mb (Choulet et al., 2014).

Using the physical map positions, a gene list, gene zipper file and comparative map were downloaded (http://probes.pw.usda.gov/WheatDMarker/downloads). The data was annotated by using the Brachypodium homology to highlight which 2DS CSS contigs had been tested for SSRs or SNPs from the parent NIL and BSA approaches previously, and which of the tested markers were polymorphic (Appendix 6.7). Combining the Brachypodium genes predicted from both the gene list and zipper, 23 genes were identified within the 4.251 Mb interval, with a density of 5.4 genes/Mb (Appendix 6.7.1 and 6.7.2).

Further attempts to decrease this interval in *Ae. tauschii* were unsuccessful since markers which were tested (highlighted in red, Appendix 6.7) were monomorphic between the parent NILs to the fine-mapping population. There was not enough time to consider sequence from the non-syntenic Brachypodium and rice genes which punctuated the *Ae. tauschii* interval – for example, no sequence on BAC ctg6164 was tested in this project. There was also insufficient time to directly mine the extended SNP markers within the delimited interval.

The comparative map of the *Ae. tauschii* interval (Appendix 6.7.3) showed the comparison with the Brachypodium and rice orthologous pseudomolecules. The delimited interval in *Ae. tauschii* was extended either side on chromosome 2D to better establish the context of gene rearrangements and collinearity. The first observation was that most of the genes with the highest homology for each *Ae. tauschii* marker in the demarcated region were on Brachypodium chromosome 5 and rice chromosome 4, which confirmed the previous macro-synteny identified by Gasperini et al. (2012). However, the second observation is that the collinear progression along the Brachypodium and rice chromosomes disintegrated upon approach to the 4.25 cM *Rht8* interval. Between 30 – 31 cM, there is a collinear progression from *Bradi5g02990 - Bradi5g03460* and *Os04g12560 - Os04g13210*, though still with punctuations from Brachypodium chromosome 1, 3 and 4 and rice chromosomes 1 and 7. Closer inspection of the gene list (Appendix 6.7.1) reveals that this region includes a tandem eight-copy gene duplication of *Bradi5g03380/Os04g12980* which encodes the UDP-glucosyltransferase 74F2.

In the 4.25 cM interval from ~31 – 35 cM, the shaded dark green cells in the comparative map (Appendix 6.7.3) highlight gene translocations. The intrachromosomal rearrangements are shown in detail in Appendix 6.7.1 and Appendix 6.7.2, including *Bradi5g05140/Os03g58960* and *Bradi5g09000/Os04g20880.* The zipper (Appendix 6.7.2) reveals that *Bradi5g0900 – Bradi5g09090* is a cluster of seven genes with limited homology to rice (only one out of seven of the cluster). Further to these translocations from within the same chromosome, interchromosomal rearrangements were also present from Brachypodium chromosomes 4 and 2/rice 2 and 5. As the progression on *Ae. tauschii* 2D continues towards the centromere, outside the delimited interval, the synteny is also punctuated with both intra- and interchromosomal rearrangements.

Taken together, this confirmed the emerging picture first observed in barley, that the synteny surrounding the *Rht8* interval is good on a medium resolution between rice and Brachypodium, but this does not extend to the fine-detail. The *Ae. tauschii* data was best (most densely) annotated with Brachypodium genes and this comparison highlighted gene duplications which occurred in *Ae. tauschii* after divergence from its common ancestor with Brachypodium. There were substantial numbers of genes in *Ae. tauschii* which had no homology to

Brachypodium or rice in the genome zipper data. Also, there was insertion of transposed chromosomal blocks from both intra- and interchromosomal regions.

In summary, the analysis of the *Ae. tauschii Rht8* interval indicates genomic divergence across the grass genomes which was not observed when looking at the model grasses individually (Chapter 5) or the *Rht8* linkage map between barley, Brachypodium and rice. The inversions and rearrangements, taken together with genes with no homology to rice or Brachypodium, indicate that the *Ae. tauschii* genome in this region expanded upon divergence from Brachypodium and rice. This is corroborated by the physical data, since the *Rht8* interval was narrowed to 1.36 Mb in Brachypodium and 1.34 Mb in rice, whereas the three BAC contigs in the *Ae. tauschii* anchored to the *Rht8* interval define a distance close to ~4 Mb. This distance could be larger still since the BAC contigs are not overlapping.

6.4 Gene content of the *Rht8* interval

6.4.1 Expression analysis

To prioritise genes of interest within the intervals found in *Ae. tauschii* and the POPSEQ-anchored IWGSC-2 and Chapman bins, differentially expressed genes in spike tissue between the parent NILs (UniGenes) and BSA (v3.3 cDNAs) were identified. The relative expression level of genes was estimated by calculating the transcript abundance expressed as reads per kilobase per million mapped reads (RPKM) (Mortazavi et al., 2008). Martin Trick analysed the datasets and calculated RPKM values as described in Trick et al., 2012 and Harper et al., 2015.

Ascertaining criteria for differentially expressed genes (DEGs) from RNA-Seq data is an ongoing area of research (Tarazona et al., 2011, Yendrek et al., 2012). For this reason, conservative criteria were adopted both for filtering out lowly expressed genes and then for establishing differential expression by 1.5 fold difference (Warden et al., 2013). The DEG data is displayed in full in Appendix 6.8.

Globally, 1735 DEGs were identified between the parent NILs. Only 4% of these genes localised to a 2AS, 2BS or 2DS CSS contig. Most of the DEGs were relatively overexpressed in the short parent NIL (Figure 6.7). The DEGs were also annotated when the gene appeared in the SNP data, reported in Chapter 5. There was 8% overlap between DEGs and genes with putative SNPs. One notable DEG that was observed (first entry in Appendix 6.8.1) encoded a brassinosteroid (BR) insensitive 1-associated receptor kinase on wheat chromosome 4, corresponding to *Bradi4g14000*. This gene was overexpressed in the short parent by 1.67-fold, though overall expression was low (0.11 RPKM in the tall parent and 0.35 RPKM in the short parent). In the BSA data, 20 DEGs were identified (Appendix 6.8.4) using the same criteria as before, but in addition, the genes had to be overexpressed toward the same parent/bulk combination (e.g. short/short). The brassinosteroid-related gene identified in the parent NILs was not recovered in the BSA. Most of the BSA DEGs were overexpressed in the short parent/bulk. None of these DEGs were on chromosome 2D.

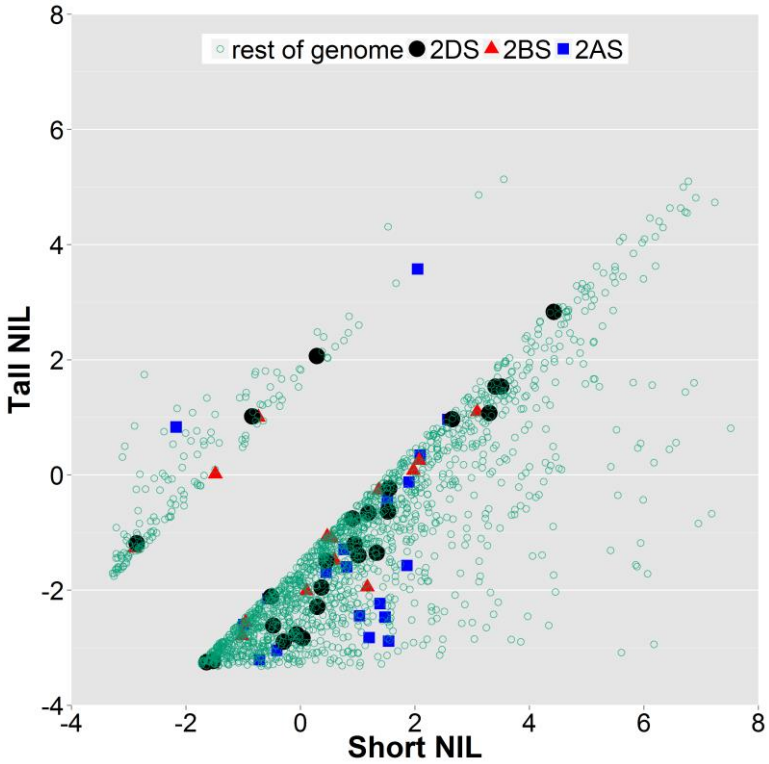


Figure 6.7: Differentially expressed genes (DEGs) between the parent NILs in the UniGene dataset. Values are Log2 RPKM values. There were 1735 genes in total. Most DEGs (>95%) were outside chromosome 2S. The full list is shown in Appendix 6.8.1. In order to be differentially expressed, a gene had to have an RPKM of >0.1 and have a 1.5-fold difference in expression between short and tall parent NIL.

Since Brachypodium offered the most comprehensive annotation of the parent NIL dataset (Chapter 5), the Brachypodium annotation was used to compare the enrichment of molecular function and biological process in the global DEGs

(Appendix 6.8.2 and Appendix 6.8.3). Relative to the background library of Brachypodium genes, the DEG dataset had enrichment for genes involved in metabolic processes. In terms of molecular function, there was double the enrichment in the DEGs for catalytic activity, comprising mainly of enrichment for catalytic oxidoreductase activity and hydrolase activity of O-glycosyl compounds (Appendix 6.8.3).

6.4.2 Ae. tauschii *and* T. aestivum

Work in this Chapter strengthened the emerging picture first explored in Chapter 5 of poor micro-collinearity in the *Rht8* region between wheat and the models Brachypodium and rice. Synteny with barley was better preserved than with these models, but barley is still relatively poorly annotated and the draft sequence has large assembly gaps in the *Rht8* region. For this reason, candidate genes were explored in the *Ae. tauschii* interval and wheat POPSEQ bins, the latter described in 6.3.5. Given the hyposenstivity to BR conferred by *Rht8* (Gasperini et al., 2012) and reduced stature explored in Chapter 3, genes which could have a biological role in processes related to BR signalling and biosynthesis, cell growth and plant development were explored with particular focus.

Despite being based on the same high-density genetic map using population sequencing (Mascher et al., 2013), the *Rht8* linkage map resolved the IWGSC-2 and Chapman data differently. An interval of 2.3 cM was demarcated in the Chapman scaffolds between 13.6 cM and 15.9 cM bins (Figure 6.6D). Taking these bins together, there were over 250 scaffolds. There was insufficient time to design markers on this sequence space beyond the approach outlined in Chapter 5. Apart from being made available for BLAST homology searches on CerealsDB, there are no gene models or annotations available for these scaffolds, therefore the Chapman assembly was not used directly to predict gene content.

The CSS contigs from marker classes D and E (either side of *Rht8*) anchored into the same bin in the IWGSC-2 data (17.3 cM bin, Figure 6.6C). This was a strong indication that the neighbouring bins (18.1 – 33.1 cM, Figure 6.6C) were outside the *Rht8* interval. Indeed, multiple markers from classes E and F, the proximal side of *Rht8* relative to the centromere anchored into the 17.3 cM bin (Table 6.2).

Since the *Ae. tauschii* genetic map was based on SNP markers from a 10K SNP array to which BAC contigs were anchored (Luo et al., 2013), it is reasonable to assume that the resolution would be superior to that proffered by POPSEQ analysis of 80-90 doubled haploid lines (Sorrells et al., 2011) used to genetically order IWGSC-2 data (Mascher, 2014). For this reason, the *Ae. tauschii* genetic map was used to annotate the IWGSC-2 data. This was achieved by first downloading the peptides from the relevant IWGSC-2 bins from *Ensembl*Plants, along with their syntenic annotation. Second, since the peptide sequence would be better conserved than nucleotide sequence, the peptides were used as a query in BLASTP searches against the NCBI database of flowering plants. The highest hit by identity and query coverage was retrieved. From the first step, the Brachypodium genes were used to retrieve the positions of genes which overlapped with the *Ae. tauschii* genome database (UC Davis Plant Science and USDA, 2015). Annotation of the genes in the 18.1 – 33.1 cM IWGSC-2 bins with the *Ae. tauschii* genetic positions confirmed that this bin was outside the *Rht8* interval (Appendix 6.10). For this reason, the genes in these bins will not be considered in detail here. One notable gene within the 18.1 cM bin is the BR-regulated *BRU1,* which had no Brachypodium or rice homologues. A marker to this gene was designed and validated (*2DS_280*) but was discarded in Step 2 of the fine-mapping process (gwm class 12, Figure 6.4).

The 17.3 cM IWGSC-2 bin identified in the *Rht8* linkage map contained 483 2DS contigs, totalling 2.48 Mb. This sequence was annotated as described above. Additionally, the genes were annotated with the expression data (Appendix 6.8.1 and described in 6.4.1) and the genes were marked if they appeared in the parent NIL SNP dataset (Appendix 6.9). There were a total of 60 genes, including duplications, which were marked in the second column of Appendix 6.9. Of these genes, 11 could be anchored outside the *Ae. tauschii Rht8* genetic interval. These are shown separately in Appendix 6.9.2 and are not considered further here.

Of the remaining genes, a minority (11) overlapped between the *Ae. tauschii* and *T. aestivum* data, whilst others were loci only found in either one of the two species.

6.4.2.1 Loci common to *Rht8* intervals in *Ae. tauschii* and *T. aestivum*

The genes in the 17.3 cM IWGSC-2 bin and also within the 4.35 cM *Rht8* interval in *Ae. tauschii* encoded proteins involved in a range of biological functions. These proteins included Bradi5g04686, a probable Sec1a protein based on similarity to Arabidopsis Sec1 (AtSec1a), involved in vesicle trafficking during exocytosis (http://www.uniprot.org/uniprot/Q9C5P7); a predicted ATP-binding protein and ABC transporter (Bradi5g03477/Os04g0209300); the nucleic-acid binding cleavage and polyadenylation specificity factor Bradi5g04673/Os04g0277400 and the membrane-bound O-acyltransferase (MBOAT) Bradi5g09000. MBOATs are involved in Golgi vesicle transport and metabolism (http://www.uniprot.org/uniprot/Q9SV07). The *Ae. tauschii* genome zipper (Appendix 6.7.2) revealed that *Bradi5g09000* was part of a gene cluster from an interchromosomal translocation from *Bradi5g09000* – *Bradi5g09064*, comprising six genes, three of which had no ontology; the others (*Bradi5g09010* and *Bradi5g09020*) were genes predicted to have prephenate dehydratase activity, which are enzymes involved in amino acid metabolism. Markers were designed on wheat 2DS sequence to these genes, but all were monomorphic (details in Appendix 6.9.1). *Bradi5g04630* was found as a single copy gene in the 35.3 cM bin on *Ae. tauschii* 2D (Appendix 6.7.2). This gene was also found in the *T. aestivum* data, but here a large-scale duplication was found with seven copies in the *T. aestivum* gene models based on the nucleotide sequence, all orthologues of *Bradi5g04630.* This duplicated gene encodes a plant pollen protein whose biological function is unknown (http://www.ebi.ac.uk/interpro/entry/IPR006041), but structurally related to the *Olea europaea* (olive tree) pollen protein, Ole e I (Villalba et al., 1993).

6.4.2.2 *Triticum aestivum*-specific loci in IWGSC-2

6.4.2.2.1 Loci with possible roles in BR signalling and metabolism

The cytochrome P450 family (abbreviated to CYP) is one of the largest and diverse superfamilies in plants, bound to membranes and involved in lipid metabolism, hormone metabolism and defence (Bak et al., 2011). In Arabidopsis, there are 244 genes and 28 pseudogenes within this superfamily organised into

two main clades: the A-type and non-A-type (Bak et al., 2011). Though pathways have not been fully elucidated, CYP family members in the 'non-A-type' clade are involved in hormone metabolism and modulate BR precursors. For example, CYP51 is an obtusifoliol 14α-demethylase involved in BR metabolism and the gene sequence encoding it has been identified in wheat (Cabello-Hurtado et al., 1997). CYP90 and CYP85 also participate in BR biosynthesis (Bak et al., 2011). Three P450s were identified: TRAES_2DS_985CFD29C/Bradi4g09000, CYP71D11/CYP71AF3P (different nomenclature between species), and TRAES_2DS_7EDB434AD/Bradi3g15020, CYP71D8/CYP71AB5P. There was also the isoflavone hydroxylase, Bradi4g07480. The CYP database (Nelson, 2009) was searched to assess their relationship with BR biosynthetic genes and other CYPs. Bradi4g07480 is CYP81M1, identified in Brachypodium and rice, of unknown function. Bradi4g09000 and Bradi3g15020 are both encoded by pseudogenes which remain uncharacterised. In fact, the CYP81 and CYP71 subfamilies, to which the three CYPs in the interval belong, are particularly badly characterised, even in the rice/Arabidopsis data, which has the best plant annotation. The CYP71A subfamily has one member known to be involved in BR metabolism (CYP71A6, http://www-ibmp.u-strasbg.fr/~CYPedia/index.html#CYP71) but this did not show homology with either of the CYPs identified here. Therefore, none of the P450s identified in the genetic bin have a known function in BR metabolism.

One UDP-glycosyltransferase (UGT74E1) was identified in the interval. As mentioned in 6.3.5, there was a tandem eight-copy gene duplication encoding UGT74F2 just outside the *Ae. tauschii* interval (genetic position 30.4 cM, Appendix 6.7.1). UGTs are of interest in the context of BR signalling, since in Arabidopsis the UTG73 family glucosylates brassinolide, the biologically active form of BR, which renders the substrate inactive. Overexpression of two members of a *UGT73* gene cluster (*C5* and *C6*) leads to a BR deficiency in Arabidopsis, both in phenotype and in reduced endogenous BR levels (Husar et al., 2011). The UGT74E1 in the *T. aestivum* interval has limited characterisation in Arabidopsis. The precise *in vivo* function remain unknown, but UGT74F1 acts on salicylic acid (Dean and Delaney, 2008). The closely related gene *UGT74F2* (found outside the interval) is expressed at higher levels in Arabidopsis and is

better characterised in glycosylating jasmonates in a defence response (Ostrowski and Jakubowska, 2014, Lim et al., 2002).

6.4.2.2.2 Protein kinases with a possible role in cell elongation

Two protein kinases (PK) were identified in the interval, both of which could feasibly modulate cell expansion. This is of interest since *Rht8* causes internode-length reduction (Chapter 3) by decreasing cell expansion (Gasperini et al., 2012). PKs act in signalling transduction and are classified according to their primary sequence and the type of protein phosphorylation activity in Ser/Thr, His or Tyr PKs (Chevalier and Walker, 2005). Traes_2DS_CB771B9DF was annotated as receptor-like Ser/Thr PK SD1-8 and the gene in Arabidopsis, *SD18*, is involved in cell expansion (http://www.uniprot.org/uniprot/O81905). The other kinase was a cell wall-associated kinase (WAK). WAKs and WAK-like kinases are transmembrane receptor-like kinases which contain a Ser/Thr kinase domain and extracellular region. There is growing evidence that WAKs are pectin receptors, both for shorter pectins generated during pathogen exposure or wounding, and for longer pectins in cell walls which are modulated during cell expansion (Kohorn and Kohorn, 2012). WAKs are commonly found in tandem repeats, but this was not observed here.

6.4.2.2.3 Other biological functions

The LRR extensin-like protein 2 encoded by *Bradi1g18280* was of interest since it was also a DEG, with 2.32-fold higher expression in the short parent NIL relative to the tall parent NIL. There are around 20 extensins in Arabidopsis, which are highly-abundant glycoproteins involved in primary cell wall architecture (Lamport et al., 2011). Pollen-specific LRR/extension 2 is AtPEX2 in Arabidopsis and the gene is highly expressed in mature pollen and during anthesis (Noir et al., 2005).

Other proteins with catalytic functions encoded by genes in the interval included Bradi5g02900, a reticuline oxidase, Bradi1g69730, a cystathionine gamma-synthase, localised to the chloroplast, the phospholipid synthase (Bradi5g09110), a cytosolic sulfotransferase (Bradi3g09500) and a caffeic acid methyltransferase (Bradi1g14870).

There were three transporters within the interval: SWEET6a, a bidirectional sugar transporter, a vacuolar amino acid transporter (Bradi5g02920) and a chloride transporter (Bradi2g11652).

Also in the interval were *T. aestivum* orthologs of genes encoding proteins involved in plant defence, such as Bradi1g22860, a synthase of Momilactone A, a secondary metabolite and four disease resistance-response proteins. There were also a number of proteins of unknown function.

Markers were designed on microsatellite sequence around some of these genes, but all were monomorphic (Appendix 6.9.2). None of the genes mentioned in this section had putative SNPs between parent NILs.

6.4.2.3 *Ae. tauschii*-specific loci

Loci found in the *Ae. tauschii* gene list (Appendix 6.7.1) from 31.4 cM – 35.7 cM which were not represented in the 17.3 cM *T. aestivum* bin were genes encoding Bradi2g20430, of unknown function with the domain DUF594; proteins involved in nucleic acid binding (Bradi4g267670, a ribonuclease and Bradi4g34520, a nucleotide transferase) and Bradi2g16396, a 6-phosphgluconate dehydrogenase. Two genes found in a tandem repeat in *Ae. tauschii* of which only a single-copy was found in *T. aestivum* were *Bradi5g03530* and *Bradi5g03550*, which encode WAKs.

6.4.2.4 v3.3 cDNAs

The 2D interval in the v3.3 cDNAs consisted of 59 genes. In the fine-mapping described in 6.3, this interval was narrowed to 20 genes between *52i* (position 929999) and *2DS_26* (position 1178953), an interval of 0.25 Mb. The full details of this interval are shown in Appendix 6.11. To determine whether the interval identified in the v3.3 cDNAs was congruent with the *Rht8* intervals and gene content in *Ae. tauschii* and IWGSC-2 genetic data, the Brachypodium orthologues to the v3.3 cDNAs were used, as described in 6.4.2. Almost all the genes in the v3.3 cDNA interval could be assigned a position in either the *Ae. tauschii* or IWGSC-2 data, or both (Appendix 6.11).

Most genes anchored into the 18.3 – 33.1 cM POPSEQ bins, outside the 17.3 cM previously identified as most likely to contain *Rht8*. The *Ae. tauschii* data confirmed that most of the genes were likely not in candidate-gene *Rht8* regions. Only one gene in the v3.3 cDNA interval (*mrna070632*, with a Brachypodium orthologue of *Bradi5g03510*) was identified to be inside the 4.25 cM interval in *Ae. tauschii*. The gene directly preceding this was *mrna020368*, which was anchored to the 31.0 cM POPSEQ bin. The proceeding genes were anchored to POPSEQ bins in the vicinity of 38 – 42 cM. Taken together, the data indicated that there was a very poor representation of the *Rht8* interval in the v3.3 cDNAs. This was because the resolution was not great enough to resolve the region apart from the one identified gene, which was already present in the other datasets. Indeed, analysis of the SNP data confirmed this. The SNP dataset (BFR>6) described in Chapter 5 contained genes from 145 2DS contigs. Of these, only 51 overlapped with the POPSEQ-anchored CSS contigs, and of this subset, only three anchored into the 17.3 cM bin. Most of the putative SNPs identified in the BSA on these three contigs had been identified as homoeologous SNPs in the PolyMarker alignments, thus markers were not developed to validate these SNPs. The one contig with a putative varietal SNP was validated with *BFR_46*, which was mapped to marker class F.

In sum, these findings showed that the v3.3 cDNA reference was a resource with limited potential for fine-mapping *Rht8*, since the gene models did not cover the *Rht8* region with sufficient density to provide enough resolution power for finer mapping.

6.4.3 Is there a candidate for Rht8*?*

The *Ae. tauschii* and IWGSC-2 *T. aestivum* data was used as the best representation of wheat gene content of the *Rht8* interval. Higher resolution of the *Ae. tauschii* genetic map facilitated the narrowing-down of the large 17.3 cM bin to which the *Rht8* interval was anchored in the IWGSC-2 data. A number of genes with functions feasibly involved in BR-signalling or metabolism were identified, however a number of genes remain uncharacterised. There is no clear candidate which co-localises with the *Rht8* phenotype (6.3.3) and only two genes in the 17.3 cM bin had differential expression. Therefore, all the genes within the interval should still be considered. In the current project, the lack of polymorphism

in sequence around genes identified in the interval hampered finer mapping and there was insufficient time to fully capitalise on the emerging genetic and physical resources in *Ae. tauschii*. The best onward strategy is discussed in the next section and in wider context in Chapter 8.

6.5 Discussion

The strategy to fine-map *Rht8* was to saturate the region with markers using a variety of approaches (Chapter 5) and then to map these markers with increasing resolution, with the aim of finding markers which co-localised with the phenotype. Ultimately, *Rht8* was mapped to a 1.015 cM region flanked by two marker classes, D and E (Figure 6.4). The markers across all fine-mapping classes, not only the closest flanking, were used to resolve syntenic intervals of 1.34 Mb, 1.36 Mb and 2.9 Mb in rice, Brachypodium and barley, respectively. Using the latest genetic and physical resources in *Ae. tauschii*, the candidate-gene interval was narrowed to 4.25 Mb, which had high micro-collinearity to *T. aestivum* genetic bins in IWGSC-2 and Chapman assembly data. This enabled the wheat gene content of the *Rht8* regions to be considered.

The map-based cloning strategy of mapping with increasing resolution was effective in progressively retaining the markers most closely segregating with the *Rht8* phenotype. From over 60 markers, 33 were fine-mapped within a 1.18 cM genetic interval. The mapping population is an excellent resource, and there are still 63 recombinants between the flanking marker classes D and E for future mapping efforts.

A discernible, confidently-scored phenotype was critical for accurate ordering of the marker classes around the *Rht8* region. In this Chapter, it was shown that highly-replicated field-sown plants were required to ameliorate the initial height data from the glasshouse. The scoring of recombinants based on bimodal distribution of plant height at maturity showed first, that the height differential between recombinants scored ‘short’ and ‘tall’ was decreased in the glasshouse which made these scores less reliable. Second, the glasshouse assignations did conflict with field scores (across both field sites) in a number of cases. In the end, there was only one recombinant (*F4-2-3-2-1*) for which, due to missing data in

one field site, there was a 1:1 opposing score. Therefore, overall, a consensus score could be reached for all the individuals to the fine-mapping population.

Anchoring the markers across classes A – G in syntenic *Rht8* intervals in rice, Brachypodium and barley provided a close study into the degree of micro-collinearity in the *Rht8* region between wheat 2DS, *Ae. tauschii* 2D, barley 2H, Brachypodium chromosome 5 and rice chromosome 4. Barley anchored the most markers by orthology to 2H. Across the barley *Rht8* interval, the collinear progression from marker class A to G was generally good. For example, the barley physical map could be used to order co-segregating markers in marker class A (Figure 6.5B). The exception in this was in the collinearity for markers in classes D and E, which was unwelcome, given that these flanked *Rht8*. The micro-collinearity between the wheat *Rht8* interval and Brachypodium and rice was poor, confirming the initial findings in Chapter 5. As shown in Figure 6.5D, taking a more sparsely annotated syntenic interval, as in the case of rice which had the fewest markers anchored, gave a false sense of the collinearity, taken at the medium-resolution level. This likely explained the previous under-reporting of the micro-collinearity disintegration by Gasperini et al., (2012). In fact, in addition to the poor micro-collinearity in the final *Rht8* linkage map, a high number of markers developed in Chapter 5 on 2DS CSS contigs anchoring to the syntenic intervals were discarded at Step 2, suggesting the possibility of intrachromosomal rearrangements around the region which mean that genes which were apparently in physical syntenic intervals were not genetically linked to *Rht8*. Taken together, this confirmed that the comparative genomics approach used in Chapter 5 was limiting.

Comparative genomics was used as a tool to prioritise marker development in Chapter 5. However, a *de novo* assembly of wheat was also performed by Martin Trick to circumvent the limitations of the incomplete wheat reference. This approach has been used in several studies in analysing wheat transcriptomes (Duan et al., 2012, Oono et al., 2013). The Trinity assembler was used since it was developed specifically for short-sequence reads as was the case here (100bp) (Grabherr et al., 2011). The same BSA approach was used as with the v3.3 cDNA reference. All genes which overlapped with genes in the original A/B/D genome progenitor leaf assemblies were removed to avoid duplicating sample space already examined and to ensure genes were spike-specific. The *de novo*

assembly only generated 29 putative SNPs mapping to 2AS, 2BS or 2DS CSS contigs which were not already represented in the v3.3 cDNA (data not shown). All these SNPs were tested with markers, but were monomorphic between the parent NILs. No SSRs were found on the eight 2DS contigs within the data, and these 2DS CSS contigs were localised to bins well outside the proximity of the 17.3 cM bin identified in this Chapter. This assembly was not used further, since it appeared that it did not offer a rich source of novel SNPs on genes likely to map to *Rht8*.

The ordered part of v3.3 cDNA reference only contained one gene on 2DS which mapped to the *Rht8* interval by comparative genomics and comparison to the wheat genetic bins (identified in Figure 6.6C). This, together with the technical difficulties reported due to redundancy in CSS alignments and a high level of background noise (Chapter 5), explains the poor return on the marker development based on these SNPs. Of course, it is possible that genes mapping to the *Rht8* interval are present in the unordered part of the v3.3 cDNAs, but these were not successfully identified in the BSA, since all SNPs with high-BFR on 2DS were considered in Chapter 5. In sum, the work in this Chapter showed that the v3.3 cDNA reference limited the resolution with which gene content of the *Rht8* interval could be determined.

The genetic and physical map resources in *Ae. tauschii* that were developed late in to this project offered the opportunity to bridge the gap between the limitations of comparative genomics already described and existing wheat genetic maps (Borrill et al., 2015). In *Ae. tauschii*, the markers from the *Rht8* linkage map were used to delimit a 4.25 cM region. The physical distance could not be exactly determined since the BAC contigs are not overlapping. However, one estimate was reached of 4.25 Mb based on three non-overlapping BAC contigs. The outer extent of the BAC contigs was wider than the 4.25 cM interval, and using this wider interval, a 1.36 cM/Mb recombination rate was calculated, with 5.4 genes/Mb. In the physical mapping of wheat chromosome 3B, an average of 9 $\pm$ 5 genes/Mb was reported across the whole chromosome, with much higher density in the distal region (Choulet et al., 2014). The highest recombination rates found in global comparison of the *Ae. tauschii* physical data were around 1.5 – 2.0 cM/Mb and were found in distal regions (Luo et al., 2013), in line with the recombination rate found in this small interval, though the number reported here

cannot be precise. Interestingly, in hexaploid wheat, estimates of 1 cM in genetic distance corresponding to an average of 4.4 Mb of physical distance (Faris and Gill, 2002) seem to be in line with the genetic interval in the *Rht8* linkage map and the physical data from *Ae. tauschii*.

The comparative map from the *Ae. tauschii* data was used to study the gene rearrangements and expansions between species. There were intra- and inter-chromosomal rearrangements in *Ae. tauschii* relative to Brachypodium and rice. Multiple genes were found on chromosomes outside of Brachypodium chromosome 5 and rice chromosome 4, previously identified to be syntenic with wheat 2DS in this region (Gasperini et al., 2012). A tandem eight-copy gene duplication of *Bradi5g03380/Os04g12980*, encoding a UDP-glucosyltransferase of unknown function was found proximal to the *Rht8* interval. The relatively high non-collinearity found here, taken together with the relatively high recombination rate described in the previous paragraph is in line with the findings across the *Ae. tauschii* genome as a whole: first, showing that non-collinear genes correlated with recombination rates along chromosomes and second, that a faster rate of genome evolution was found between *Ae. tauschii* – Brachypodium than that with *Ae. tauschii* – rice (Luo et al., 2013).

The *Ae. tauschii* genomic resources have already empowered genetic and genomics studies which could be translated to wheat, for example in the isolation of the stem rust resistance gene *Sr33* (Periyannan et al., 2013). The *Ae. tauschii* resources provide excellent scope to the further map-based cloning of *Rht8*. The individual non-syntenic genes identified here could be targeted for marker development. As a priority though, the BAC contig spanning 360 Kb in the middle of the *Rht8* interval should be mined for polymorphism, followed by the other two contigs, as well as the extended SNP-marker sequences used to anchor the BACs. If markers co-segregating with the *Rht8* locus are found, then working backwards from the BAC contig, markers can be mapped onto the minimal tiling path (MTP) BACs, the individual BACs then sequenced and physical sequence isolated. A similar strategy has been used with the same resources in projects involving the mapping of genes on wheat sub-genomes B and D in *T. dicoccoides* and *Ae. tauschii* (Liang et al., 2015, Zhang et al., 2015).

The *Rht8* linkage map was used to identify the 17.3 cM bin in the IWGSC-2 CSS contigs, ordered by POPSEQ, as most likely to contain *Rht8*. The flanking marker classes D and E both anchored into this bin. However, the Chapman scaffolds showed higher genetic resolution here than IWGSC-2, since the flanking marker classes delimited a 2.3 cM distance between bins. Although the precision of these genetic bins is limited by the low-resolution of POPSEQ, these findings confirm that the gene space sampled by the Chapman assembly does not completely overlap with the IWGSC contigs (Chapman et al., 2015).

The gene content of the 17.3 cM bin containing *Rht8* was focused on with improved resolution using the *Ae. tauschii* genetic map. Particular emphasis was placed on genes that could have a biological role in plant growth and development, based on NCBI annotation. There was no clear candidate based on the *Rht8* linkage map since no markers co-segregated with the *Rht8* locus and only two genes in the 17.3 cM bin showed differential expression. One high-copy number duplication in the *T. aestivum* data that was identified is the incompletely characterised pollen Ole1-like protein. However, the gene was not differentially expressed between the parent NILs, and the duplication might indicate genome expansion from *Ae. tauschii* to *T. aestivum* rather than *Rht8*-related function. Three cytochrome P450s were identified, as well as a UDP-glycosyltransferase, but these proteins are encoded by large gene families and the particular families found here remain poorly characterised. The P450s found in the interval belong to subfamilies CYP71 and CYP81, which are non-A-type P450s. Non-A-type P450s contain families known to be involved in BR metabolism, but the non-A-types also include more divergent sequences than the A-types. Many families show more similarity to non-plant P450s than to other plant P450s and in Arabidopsis, sequence identity among family members can be less than 20% (Bak et al., 2011). Wheat is likely to have a much more sequence divergence. For this reason, high-quality gene models and annotation will be crucial in identifying genes such as these.

Chapter 7: Germplasm development to study rare alleles at the *Xgwm261* locus

7.1 Introduction

In bread wheat, there is a general correlation between reduced height and reduced yield (Law et al., 1978). The most influential breeding strategy of the 20th century was the introduction of the major semi-dwarfing genes *Rht-B1b* and *Rht-D1b* into germplasm at the International Maize and Wheat Improvement Centre in Mexico (CIMMYT) by Norman Borlaug. These genes break the height/yield correlation. *Rht-B1b* and *Rht-D1b* are gibberellin (GA) insensitive and in optimal conditions, reduce plant height by 15-35% (Gale and Youssefian, 1985, Trethowan et al., 2001) whilst increasing yield to similar levels (Worland and Law, 1986). Originally derived from the Japanese cultivar 'Norin 10', these genes became prevalent in CIMMYT wheat varieties and are now found in the majority of modern wheat cultivars (Hedden, 2003). However, *Rht-B1b* and *Rht-D1b* are not universally beneficial. Where heat stress occurs during ear emergence, interactions between these dwarfing genes and the environment have been shown to reduce fertility resulting in a yield penalty (Worland and Law, 1986). Furthermore, poor seedling emergence due to reduced coleoptile length and maladaptation to dry environments are other problems associated with *Rht-B1b* and *Rht-D1b* (Botwright et al., 2005, Rebetzke and Richards, 1999, Trethowan et al., 2001). In these conditions, the GA-responsive semi-dwarfing gene *Rht8* produces a semi-dwarf phenotype without the undesirable effects of the Norin 10-derived genes (Ellis et al., 2004, Rebetzke and Richards, 1999). Pre-dating Borlaug, the Italian wheat breeder Strampelli introduced *Rht8* to Europe from the Japanese variety 'Akakomugi'.

Using a chromosome substitution line between Cappelle-Desprez and the Strampelli cultivar Mara, Korzun et al. (1998) reported a tight linkage of 0.6 cM between *Rht8* and a 192-bp allele at the microsatellite locus *Xgwm261.* A screen

of over 800 wheat varieties revealed that 90% of varieties carried the three most common alleles of 165-bp, 174-bp or 192-bp at this locus (Worland et al., 1998b, Worland et al., 2001). A height-reduction of 7 – 8 cm was attributed to the 192-bp allele relative to the 174-bp allele; a 3 cm reduction was found in varieties carrying the 174-bp allele relative to 165-bp, and the 165-bp allele was found to be neutral for height. It was therefore suggested that genotyping at *Xgwm261* represents a simple method to assay for variants at the *rht8* locus, and that a 192-bp allele was diagnostic for *Rht8* (Worland et al., 1998b, Worland et al., 2001).

However, it was reported by Ellis et al. (2007) that the 192-bp allele at this locus is not always diagnostic for the height-reducing *Rht8*. Instead, *Xgwm261*-192-bp is only indicative of *Rht8* in wheat cultivars that have inherited this allele from Akakomugi or a Strampelli-wheat ancestor. The authors found that Norin 10-derived material has an identical 192-bp allele at the *Xgwm261* locus which is not associated with *Rht8* and suggested that this alternative haplotype evolved prior to the *Xgwm261*-192-bp linkage with *Rht8*. Furthermore, Gasperini et al. (2012) reported that *Xgwm261* maps further away from *Rht8* (1.95 cM) than previously described (0.6 cM, Korzun et al., 1998). Taken together, the linkage between the 192-bp allele at *Xgwm*261 and *Rht8* can be broken, thus a 192-bp allele at this locus is insufficient to unequivocally determine whether a particular cultivar carries *Rht8*.

Despite the more complex relationship between the *rht8* locus and *Xgwm261* than initially believed, genotyping at *Xgwm261* is likely to remain a popular method to assess allelic variation at *rht8*, at least in conjunction with other information, such as pedigree or height-reducing effect. This is for two main reasons. First, *Xgwm261* is multi-allelic whereas *DG279* and *DG371*, previously-reported flanking markers to *Rht8*, are bi-allelic and showed very low polymorphism across a diversity wheat panel (Gasperini, 2010). The markers developed in Chapters 5 and 6 remain untested for the extent of their multi-allelism. Further, when novel KASP markers closely linked to *Rht8* have been provided to breeders during the course of this PhD, they have tested the performance of these relative to *Xgwm261*. One breeder reported 100% match with the 192-bp allele in Akakomugi-derived material (personal communication). Therefore, breeders will likely use these new high-throughput markers in addition

to the well-established *Xgwm261* screen. Second, *Xgwm261*, in addition to the SSR-markers developed in Chapters 5 and 6, will still be used by breeders in countries where SNP markers and the associated technology is not yet prevalent.

Despite the importance of understanding the variants at the *Xgwm261* locus, almost all research has focused on height-related effects only. Germplasm development to enable better understanding is needed. The 192-bp allele at *Xgwm261* is found in Bulgaria, Greece, former Yugoslavia, Ukraine, China, North America and more recently Australia (reviewed in Asplund et al., 2012). Other than the most common 165-, 174- and 192-bp alleles, genotypic screens have reported distinct and less prevalent ('rare') *Xgwm261* alleles ranging from 180- to 220-bp (Ahmad and Sorrells, 2002, Asplund et al., 2012, Bai et al., 2004, Bakshi and Bhagwat, 2012, Chebotar et al., 2001, Ganeva et al., 2005, Liu et al., 2005, Worland et al., 1998b, Worland et al., 2001, Yediay et al., 2011). The precise number of these rare alleles remains uncertain since it has been demonstrated that variations of 2 – 5 bp are 'stutter' as a result of polymerase slippage during amplification of the alleles (Schmidt et al., 2004), and the same authors speculated that many scientists did not adjust allele sizes to produce uniformity of results in line with previous investigations. For these reasons, descriptions of 'novel' *Xgwm261* alleles varying by only two bp from previous reports should be treated with caution (e.g. as in Bakshi and Bhagwat, 2012). Despite ambiguities about the precise number, 'rare' alleles at *Xgwm261* exist and their adaptive significance remains poorly understood. This is despite 'rare' alleles (in terms of global distribution) being highly prevalent in certain germplasm collections e.g. Argentinian wheat with 42% of varieties reported to contain a 210-bp allele (Worland et al., 2001), indicating non-random selection by breeders or founder effects.

To determine the agronomic significance (including height and yield components) of allelic variants at *Xgwm261* rather than only cataloguing diversity at the locus, the alleles need to be studied in a common genetic background. Work to achieve this was first mentioned in Worland et al., 2001 (p.159). A range of alleles at *Xgwm261* were selected and backcrossed into a UK-adapted winter wheat, Mercia, used by Worland and colleagues to study other genes such as *Ppd*. Since this first mention in 2001, adaptive significance of the alleles at the *Xgwm261* locus remains poorly studied. One analysis of distribution of 192- and non-192-

bp genotypes showed no advantage of the 192-bp allele to coleoptile elongation in 135 US and Chinese winter wheat cultivars (Bai et al., 2004). Further, a screen on a 19th century wheat collection revealed no correlation between genotype at *Xgwm261* and plant height, but the authors cited height measurements taken from small, non-replicated plots as a possible reason for this result (Asplund et al., 2012). Another study found that all Bulgarian cultivars carrying the rare 203-bp allele were the earliest in heading and also had increased yield due to increased spikes per area (Ganeva et al., 2005). However, these effects were not dissected away from other genes determining earlier flowering on 2D (e.g. *Ppd-D1*), since the allele was studied in different genetic backgrounds. Clearly, isogenic lines grown in yield-size plots in replicated conditions are required in order to unambiguously determine the pleiotropic effects of the *Xgwm261*-allelic variants. The germplasm first mentioned in Worland et al. (2001) was recovered during the course of this PhD, from the JIC. The development of this germplasm is described in this Chapter.

Our current knowledge of the adaptive significance of variants at *Xgwm261*, the extent to which they reveal variation at *Rht8* and the pleiotropic effects of *Xgwm261* variants is poor. The importance of studying the agronomic performance in a comprehensive way, other than just height effects, was demonstrated in Chapter 3. The *Rht8* allele from Mara has great agronomic importance in reducing lodging and has no yield penalty (and a non-significant higher mean yield) in certain agronomic conditions. Additionally, in Chapter 4, it was shown that an interesting spike morphology closely segregates with *Rht8.* The allelic diversity of the *Rht8* flanking markers developed in Chapters 5 and 6 mapping closer to *Rht8* than *Xgwm261* remains untested, but they could be used by breeders in conjunction with typing for *Xgwm261*. With the aim of filling this gap in our knowledge of variants at *Xgwm261*, the isogenic lines first described by Worland et al. (2001), were developed in this Chapter and will provide the basis for ongoing work.

7.2 Recovered germplasm and development pipeline

Rare *Xgwm261* alleles, as well as the 192-bp Mara-derived allele, were introgressed into Mercia in 2006 by Liz Sayers at JIC, in multiple streams. Mercia carries the 174-bp allele at *Xgwm261* and also *Rht-D1b* (GRIS, 2015). The alleles which were introgressed were derived from Maringa, a Brazilian wheat, Pliska, of Bulgarian origin (reported initially as 201-bp by Worland et al., (2001) and later as 203-bp (Ganeva et al., 2005)), Klein 157 and Klein 49 (Argentinian wheats, reported originally as 210-bp and 215-bp, respectively, by Worland et al., (2001)). Extant allele sizes as described in the initial screen were used here for continuity (Sayers, personal communication), even though the actual sizes detected were larger due to the tailed primer with a labelled adapter (shown in Appendix 7.1). Four streams produced fertile heterozygous seed (Figure 7.1). The F_2 seeds from each stream were planted and genotypes collected. The segregation patterns for the introgressed *Xgwm261* allele did not significantly deviate from 1:2:1 Mendelian ratios (Table 7.1). The plants homozygous for the parent and donor alleles were grown to maturity in the glasshouse and bagged in order to bulk seed.

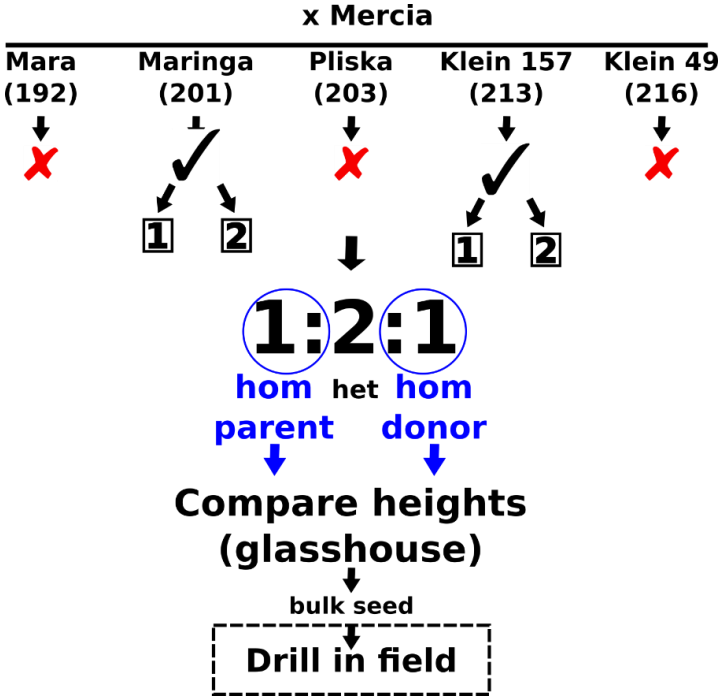


Figure 7.1: Germplasm development pipeline for rare Xgwm261 variants. Material was recovered from the introgression of rare Xgwm261 alleles (sizes indicated in brackets) into a common Mercia background. Four successful streams are highlighted from which the homozygous plants for donor and parent allele were selected to be grown in the glasshouse. This seed has been bulked and will next be drilled in replicated plots in the field in order to assess agronomic performance.

Cross and stream	Size		Frequency	%	χ^2 1:2:1	*P* 2 d.f.
Maringa x Mercia 4-2	174 and 201	het	27	56	0.792	0.673
	174	hom parent	10	21		
	201	hom donor	11	23		
Maringa x Mercia 5-6	174 and 201	het	24	50	0.000	1.000
	174	hom parent	12	25		
	201	hom donor	12	25		
Klein 157 x Mercia 1-3	174 and 213	het	28	58	2.833	0.243
	174	hom parent	13	27		
	213	hom donor	7	15		
Klein 157 x Mercia 4-4	174 and 213	het	19	40	3.125	0.210
	174	hom parent	17	35		
	213	hom donor	12	25		

Table 7.1: Segregation for Xgwm261 in the F_2 germplasm in the Mercia background. The p-value at two degrees of freedom was calculated for each Chi-square value. The test value for Chi-square tests is: H_0 = 5.99, N=48.

7.3 Preliminary height measurements

To determine the effect of the donor allele on height, the plant height and internode lengths were measured at maturity. The homozygotes for the donor versus parent (Mercia) allele were compared using the Student's T-test. Cappelle-Desprez and RIL4 were grown concurrently to compare the effects of the rare *Xgwm261* alleles to the 192-bp allele. In glasshouse conditions, the Mara-derived 192-bp allele in RIL4 had an 8 cm height-reducing effect and highly-significant reduction in spike and internode length relative to the wild-type Cappelle-Desprez (Figure 7.2). Neither streams from the Maringa introgression into Mercia showed a significant difference in height between parent and donor alleles. The Maringa allele was also neutral for reduction in the spike and internode lengths. In the Klein 157 x Mercia streams, there were two significant differences between the donor- and Mercia-allele. In one stream (4-4), the donor allele had a significant overall height-reducing effect of 3.6 cm, but no further differences in height components. The other stream (1-3), had no difference in overall height, but the donor allele had a length-promoting effect in the first internode (Table 7.2).

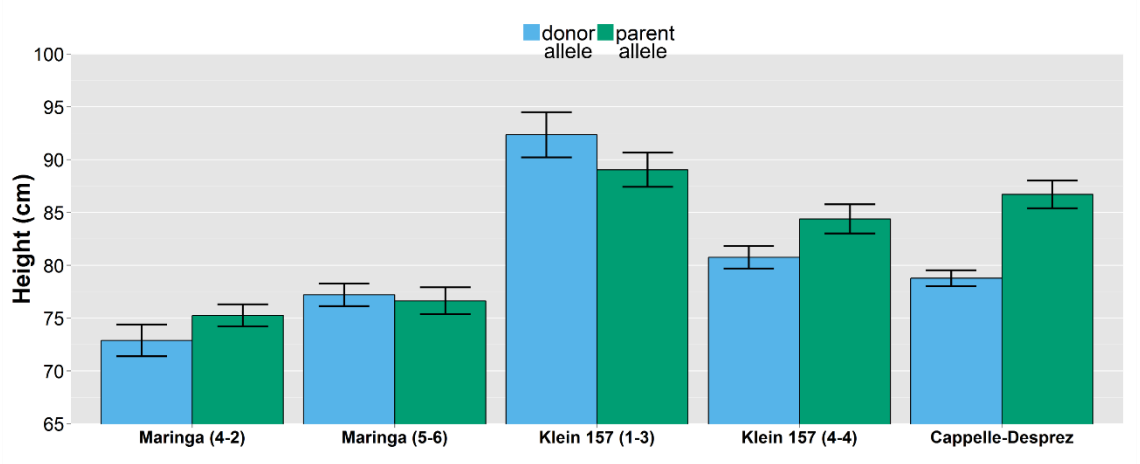


Figure 7.2: Plant height at maturity of homozygous individuals within each stream, contrasting for donor and parent allele at the Xgwm261 locus. Cappelle-Desprez parent is wild-type, donor is RIL4. Data represent means, error bars represent standard error. N shown in Table 7.1 and for Cappelle-Desprez and RIL4, N=16.

lowest highest

		Total height		Spike length		Internode-1		Internode-2	
		mean	p-val	mean	p-val	mean	p-val	mean	p-val
Cappelle-Desprez	parent	86.7	***	10.2	**	32.4	***	17.9	***
RIL4	donor	78.8		9.5		26.4		16.4	
Maringa x Mercia 4-2	parent	75.3	NS	9.2	NS	26.0	NS	16.7	NS
	donor	72.9		9.4		25.6		16.3	
Maringa x Mercia 5-6	parent	76.6	NS	9.7	NS	26.3	NS	17.0	NS
	donor	77.2		9.7		24.7		17.3	
Klein 157 x Mercia 1-3	parent	89.0	NS	8.5	NS	28.9	*	21.7	NS
	donor	92.4		8.9		33.5		22.8	
Klein 157 x Mercia 4-4	parent	84.4	*	8.5	NS	29.2	NS	22.2	NS
	donor	80.8		9.4		32.4		22.4	

NS P>0.05
* P<0.05
** P<0.01
*** P<0.001

Table 7.2: Plant height and height components of Xgwm261 allele introgressions. P-values to student's T-test comparing the donor and parent alleles within each stream are shown. Means are based on N shown in Table 7.1. All units are in centimetres.

7.4 Discussion

The work here progresses the development of germplasm which will enable the agronomic effects of rare *Xgwm261* alleles to be assessed in the field. Measurements from plants grown in the glasshouse revealed an overall height-reducing effect of the Maringa allele (201-bp) in only one of the streams by 3.6 cm. This is approximately comparable to the reduction of the 174-bp allele relative to 165-bp allele described previously (3 cm) (Worland et al., 1998b), but modest with respect to the 8 cm reduction from the 192-bp allele. The isogenic lines contrasting for the rare donor and common-background parent allele have been bulked which will enable replicated plots to be drilled in the field and other agronomic traits to be measured.

As seen in Chapter 5, it can be crucial to verify height effects observed in glasshouse-grown plants in the field. In Chapter 5, a reported 8 cm difference in heights was not observed in the glasshouse due to confounding factors. Therefore, the even more modest height effects reported here require verification in the field.

Germplasm development is a crucial resource, since advancing generations requires time and often limits projects in crops. At the same time, germplasm development can be perceived to be an unglamorous task and usually continued over time by different people. Often, continuity and records are lacking. The work here in advancing a germplasm resources was therefore included for transparency.

Cloning *Rht8* will be an assured way of understanding the adaptive significance of variants at *Xgwm261* and will also likely identify new alleles at the *Xgwm261* locus.

Chapter 8: Summary and Outlook

8.1 Summary

The aims of this PhD were broadly divided into two parts: agronomic characterisation of *Rht8* and the cloning of the gene. At the start of the project, *Rht8* had been fine-mapped to a 1.29 cM region on 2DS with single-strand conformation polymorphism (SSCP) markers, primarily using Brachypodium and rice synteny. A population had been developed of NILs contrasting for Mara-derived *Rht8* and the *rht8* allele from Cappelle-Desprez into a photoperiod sensitive background (*Ppd-D1b*). One of the aims of this PhD was to assess the performance of *Rht8* for the first time in UK-adapted germplasm in northern latitudes, without tight-linkage to *Ppd-D1a* which confounded many previous reports. This enabled the evaluation of the adaptive significance of *Rht8*, as opposed to 'escape' from terminal stresses due to earliness conferred by *Ppd-D1a* (Semenov et al., 2014). Most agronomic reports into *Rht8* have focused on height, leaving yield and yield component traits poorly understood. A thorough evaluation of these traits was conducted in this thesis. Another objective was to investigate the spike compactness following anecdotal reports of the phenotype in *Rht8*-related material. The other main aim was to further fine-map *Rht8*.

In this Chapter, I summarise the findings and how they further our knowledge of *Rht8*. I evaluate questions pertaining to some of the key strategies used in this thesis. I also outline suggestions for future work.

In the first part of this thesis, a short and tall NIL contrasting for the *Rht8* (short)/*rht8* (tall) allele were selected from a BC_3F_3 population of NILs in the UK spring wheat background, Paragon, without *Ppd-D1a* or any known major height genes. The *Rht8* and tall NIL were grown alongside the recurrent parent across two sites in the UK and a high-temperature site in Spain, which had irrigation to field capacity. Field experiments included contrasting N treatments and irrigation regimes.

There was no significant difference in the developmental traits examined (heading, anthesis or senescence) therefore it was confirmed that the height

reduction was due to *Rht8* and not premature development. Plant height at maturity was reduced by ~11% consistently across sites and different N and water treatments and was principally due to the shortening of the spike, peduncle and top internode. This reduction in stature is within the 7 – 18% range reported in varying genetic backgrounds and conditions (Wang et al., 2015b, Rebetzke et al., 2012b, Lanning et al., 2012, Ellis et al., 2004).

In high-yield potential UK sites under standard agronomic conditions of high N and no irrigation, the *Rht8* NIL had a ~10% yield penalty which agrees with previous reports of the negative impact of *Rht8* on yield (Wang et al., 2015b, Lanning et al., 2012). The key findings of Chapter 3 were that the yield penalty was abolished at low N input, in irrigated conditions and at the high-temperature, lowest yield-potential site in Spain. This is a novel finding and indicates that *Rht8* might be usefully deployed in low-input agriculture including, but not limited to, organic systems where yield stability is often more important than maximal yield in an individual season (Wolfe et al., 2008). To add to this, Chapter 3 showed very preliminary indicators that *Rht8* had improved N uptake efficiency at low N, which further strengthens the case for future trials in organic systems. There is growing demand for organic wheat in the EU, increasing 15% year-on-year from 2000 to 2010. Currently the production of organic bread wheat in the EU is well below demand resulting in high prices and reliance on imports (David et al., 2012). Therefore, there is a clear agronomic niche which *Rht8* could fill.

Dissecting the yield components revealed that the yield penalty was driven by a concomitant reduction in grains m^{-2} and spikes m^{-2}, and not grain weight or harvest index, where there was no significant difference between the *Rht8* and tall NIL. There have been mixed reports previously, with significant differences reported in grain weight but not grain number in different environments and much smaller trials (either pots or smaller field plots) (Wang et al., 2015b, Rebetzke et al., 2012b). The interplay in yield components suggest that studying tiller dynamics and floret generation, both established early in wheat development, is key to further explaining these findings.

A further key finding of Chapter 3 was that the *Rht8* NIL conferred root-lodging resistance at agronomically-relevant N levels. On average, every three or four years, widespread lodging occurs on up to 20% of the UK wheat-growing area

which reduces grain yield and quality (Berry et al., 2004). *Rht8* could be deployed where lodging risk outweighs the 10% yield penalty. This work forms a basis from which to extend our understanding of the mechanism of improved anchorage in *Rht8* by investigating the spread and depth of the root plate. The role that root traits might play in increasing wheat yield has not been fully explored even though genetic diversity in root traits has been reported in bread wheat (Narayanan et al., 2014, Herrera et al., 2012). The effects of the *Rht* alleles on root systems are unclear with contradictory results reported in different growing conditions. In the most recent investigation in near-isogenic lines in the wheat variety Mercia, the presence of *Rht8, Rht-B1b* or *Rht-D1b* had no effect on total root length in seedlings and the authors ascribed contradictory reports in other studies to differences between methodologies such as laboratory gel experiments versus field experiments (Wojciechowski et al., 2009). Recent developments in image-based high-throughput analysis of crop roots could ameliorate future investigations into *Rht8* roots in the field (Downie et al., 2015, Bucksch et al., 2014).

In Chapter 4, I set out to investigate the spike compactness which had been reported anecdotally. A semi-compact spike phenotype was quantified as a significant ~15% increase in compaction in the *Rht8* NIL compared to the tall. The compaction was due to a decreased spike length and not a change in the spikelet number per spike. Spike compactness was unaffected by the different N treatments and water regimes. The findings from this Chapter provide a novel way with which to type for *Rht8*, in addition to the well-established plant height at maturity. The methodology in Chapter 4 showed that visual assessment of compaction is unreliable and further work needs to be quantitative. In light of this, further work is being undertaken within our group to measure spike compactness in *Rht8*-related material and a landrace collection with more extreme club spikes. The findings in Chapter 4 pave the way for future investigation into how spike compactness in *Rht8* might affect disease resistance, for example Fusarium Head Blight (FHB), which infects the spike and is considered one of the most devastating diseases of wheat worldwide. Further work is being done at the JIC with the *Rht8* NIL to assess the potential in reducing spike infection in FHB.

Chapters 5 and 6 presented the results to efforts to isolate *Rht8* by map-based cloning. Work in Chapter 5 used manifold approaches to generate markers likely

to map to *Rht8*. The strategy to fine-map *Rht8* was to saturate the region with markers and then to map these markers with increasing resolution, with the aim of finding markers which co-localised with the phenotype, or at least flanked *Rht8* and aligned to a contiguous region of a wheat physical map. Ultimately, no co-segregating marker or contiguous sequence was identified so we could not clone *Rht8*. However a total of 33 markers from Chapter 5 were used to map *Rht8* to a 1.015 cM region and resolve syntenic intervals of 1.34 Mb, 1.36 Mb, and 2.9 Mb in rice, Brachypodium and barley, respectively. Using the latest genetic and physical resources in *Ae. tauschii*, the candidate-gene interval was narrowed to 4.25 Mb, which had high collinearity to *T. aestivum* genetic bins in IWGSC-2 and Chapman assembly data. This enabled the wheat gene content of the *Rht8* region to be considered.

The gene models in the IWGSC-2 genetic bin identified as most likely to contain *Rht8* were examined in Chapter 6, but attempts at further marker development mainly focused on genes on barley chromosome 2, Brachypodium chromosome 5 and rice chromosome 4, due to the late stage the *Ae. tauschii* data was obtained. However, there is scope to examine the other genes mapping to the wheat population sequencing (POPSEQ) bin which were not represented in syntenic intervals. The coarse resolution of the POPSEQ bins still makes mining this sequence a formidable challenge in its current state.

Of the genes within the narrowed wheat genetic bin, there were feasible candidates which could be involved with brassinosteroid signalling. Three cytochrome P450s (CYPs) were identified, as well as a UDP-glycosyltransferase (UGT), which both belong to large gene families. The particular families found in the *Rht8* region remain poorly characterised. The P450s found in the interval belong to the non-A-type P450s subfamily and some members of this sub-family are known to be involved in BR metabolism. Since the particular families found here (CYP71 and CYP81) are poorly characterised, this could not be further investigated. UGT74F1 acts on salicylic acid in Arabidopsis but the mechanism is not fully understood. Future improvements in gene annotation will be useful, but all the genes in the interval should be considered.

Many aspects of the fine-mapping strategy were evaluated. At the highest level, fine-mapping was hampered by both low polymorphism and low recombination.

Low polymorphism was reported in the previous fine-mapping attempt of *Rht8*, where only 4% of the tested markers were polymorphic. In this project the figure was ~15%. The low polymorphism is a product of the relatively recent addition of the D-genome and bottlenecks during hexaploid wheat evolution as outlined in Chapter 1. Historical limitations which make cloning genes on the wheat D-genome more challenging cannot be reversed. However, contributing variables can be broken down into poor micro-collinearity in the *Rht8* region, low success of marker validation, limitation of the wheat reference/arrays and the limitation of the bulked segregant analysis approach. The latter two are discussed in the next sections.

Marker development in Chapter 5 identified technical limitations which limited marker validation. These included the redundancy in the IWGSC CSS contigs and the scoring of variants on SNP arrays. For example, *52i* was completely monomorphic within the bulks on the iSelect SNP array, but polymorphic between the parent NILs. This SNP would not be found following the filtering of variant markers with a high number of absent or heterozygous calls, as some studies have done (Lu et al., 2015, Jiang et al., 2015a). More transparent reporting is needed within the research community to explore limitations in bioinformatics, since going into the future, data analysis will be a bottleneck faced by researchers rather than data generation.

Development in genomic resources allowed closer study of synteny around the *Rht8* interval. By focusing on genes on 2DS CSS contigs and by comparing orthologous relationships between genes within the IWGSC-2 genetic bins on wheat 2D, barley was identified as having the highest density of genes collinear with the wheat 2D interval, as might be expected given its more recent divergence from wheat. Unfortunately, there are large annotation gaps in the barley data in the *Rht8* interval. The synteny with Brachypodium and rice was worse than previously reported (Gasperini et al., 2012). In particular, there were clear synteny breakpoints where orthologous genes were found on Brachypodium chromosomes 3 and 4, and rice chromosomes 7 and 11. The poor micro-collinearity does not exclude the possibility that a marker mapping closest to *Rht8*, or the *Rht8* gene itself, could be a gene found in some or all of the syntenic gene intervals across the different species considered here. However, there is clearly a need to consider the non-syntenic gene content.

At the end of this project, genetic and physical map resources in *Ae. tauschii*, including genome zippers were made available. Given the low polymorphism experienced during this project and coarse resolution of extant wheat genetic maps, these resources are an excellent marker source for cloning *Rht8.* One analysis comparing *Ae. tauschii* 5D with barley 5H and wheat 5D found that the relative positioning of genes had fewer perturbations between *Ae. tauschii* and wheat than *Ae. tauschii* and barley (Akpinar et al., 2015). This fits with evolutionary distances between the species and adds weight to the usefulness of *Ae. tauschii* as a tool to examine hexaploid wheat gene content, in particular for genes on the D-genome. With this in mind, the comparative map from the *Ae. tauschii* data was used in Chapter 6 to study the local gene order around the *Rht8* interval. There were intra- and inter-chromosomal rearrangements evident in *Ae. tauschii* relative to the model grasses. This is reflective of genome-wide findings that the gene space in the smaller grass genomes is more stable than the larger *Aegilops* genome (Massa et al., 2011, Luo et al., 2013). The finer genetic map of *Ae. tauschii* was used to narrow down the large IWGSC-2 wheat genetic bin, but many non-syntenic genes identified in the *Ae. tauschii* data along with ~4 Mb of physical space could not be investigated fully due to the lack of time available to complete the research part of the PhD.

Map-based cloning strategies and the bulked segregant analysis (BSA) strategy adopted in this PhD require confident phenotyping for the trait of interest. Scoring the heights of the fine-mapping recombinants in the field was found to be more robust than the glasshouse data, where sterility and high temperatures were encountered and contracted the height differential – for example, the short parent NIL was 25 cm taller in the Morley field location than in the glasshouse. For the first time in this work, the *Rht8* score for each of the individual fine-mapping recombinants could be ascertained, based on multiple replicates across three locations. Previously, markers had been ordered around *Rht8* based on mean scores of recombinant classes. However, despite the high replication of the measurement of a single trait across multiple environments, there were still 13 instances of recombinants where conflicts in short/tall typing were encountered. Phenotyping technologies based on 3D lasers are emerging. These platforms can measure multiple traits, including height, under field conditions in a non-destructive manner and without affecting photosynthesis (Busemeyer et al.,

2013, Kjaer and Ottosen, 2015). The Phenospex PlantEye high-resolution 3D laser scanner is being trialled at the JIC and might be able to offer dynamic height measurements during the growing season and provide canopy heights to ensure more robust typing. Directly for fine-mapping *Rht8*, the spike compaction might offer an alternative approach independent of plant height at maturity, as outlined in Chapter 4.

The markers developed in Chapter 5 fine-mapped with some redundancy across seven marker classes. This redundancy was useful for defining syntenic *Rht8* intervals, since not all the markers could be mapped across all species. The redundancy is also convenient since it gives generous scope to identifying markers which could be polymorphic in a wide range of populations, and hence beneficial from a breeders' perspective. Though the SSR markers developed here are less applicable to breeding in Western Europe, where SNP technologies are more cost-effective, they are of particular importance in countries and research institutes where SNP technology is lacking. To this end, the markers are being disseminated to breeders and colleagues of EWAC, the European Wheat Aneuploidy Consortium.

As bioinformatic and genomic resources develop in wheat, the development and maintenance of mapping populations is likely to be the rate-determining step in cloning genes. Work in Chapter 7 continued the germplasm and resource development to generate an allelic series of *Xgwm261* in a common background. *Xgwm261* is still used in many breeding programs as a proxy of the genotype at the *Rht8* interval and we do not yet understand how variation at *Xgwm261* reflects variation at *Rht8*. For this reason, the work in Chapter 7 is useful in developing a future resource which will be tested in field trials.

Overall, the RNA-Seq strategy used in this project had a sound rationale based on the limited previous work in tetraploid and hexaploid wheat which used similar approaches (Ramirez-Gonzalez et al., 2014, Trick et al., 2012). The practical implementation of the strategy over the course of Chapters 5 and 6 brings up two key elements: evaluation of the wheat reference used and the BSA approach.

8.1.1 Was the v3.3 cDNA reference fit for purpose?

The v3.3 cDNA was the best ordered representation of 75,419 gene models which became available during the PhD. At the time, the alternative resources were the customised UniGene reference which relied on ordering from Brachypodium synteny together with a low-resolution Chinese Spring x Paragon map, or the raw CSS contigs released in IWGSC-1, which were not ordered. Work in Chapter 6 showed that gene models on the ordered section of the v3.3 cDNA did were of insufficient density around *Rht8* to facilitate fine-mapping.

Assessing the gene content in the unordered section of the v3.3 cDNAs is difficult since the cDNAs have no position or order relative to each other. However, the overlap between the BFR>6 SNPs on genes anchored to IWGSC CSS contigs across the v3.3 cDNA reference as a whole and the IWGSC-2 POPSEQ-anchored contigs was low. The SNPs with BFR>6 covered 145 2DS CSS contigs and only three of these were in the 17.3 cM POPSEQ bin most likely to contain *Rht8*. One of these was validated with the marker *BFR_46* and the other two mapped outside the *Rht8* linkage group. Most of the 2D markers developed on SNPs from BSA did not map to the *Rht8* linkage group. There were also technical limitations of IWGSC CSS redundancy which complicated marker validation. Taken together, there is strong evidence that the v3.3 cDNA reference was not fit for purpose. Better gene model representations now exist, which are considered below.

8.1.2 Did the BSA methodology work?

The BSA strategy was built on the rationale that variation captured in the parent NILs to the fine-mapping *Rh8* population would be enriched in the bulks in regions which mapped to *Rht8*.

8.1.2.1 Background noise

The first significant finding was a high level of background noise across several chromosomes outside of 2DS, the chromosome which *Rht8* maps to. High numbers of SNPs between the parent NILs were identified on chromosomes 3B, 6AL, 7BS and 7DS. Some background noise in the parent NILs was expected but

through BSA we anticipated that only variation that was genetically linked to *Rht8* would be enriched in the bulk data.

Instead, BSA identified high SNP densities on chromosomes 3B, 5BL, 7BS and 7DS. Background noise across chromosomes has been reported in previous BSA approaches in wheat, but not to the extent encountered here. Trick et al., (2012) found one significant peak on a different wheat chromosome to that with the candidate gene, and one SNP from that chromosome mapped close to the candidate region, implying a small insertion. The authors in the study found that increasing gene coverage and considering the highest BFR-SNPs reduced background noise and was successful at identifying SNPs in the collinear regions. Both these recommendations were followed here but were relatively unsuccessful. In the study upon which the BSA analysis was developed (Ramirez-Gonzalez et al., 2014), using an F_2 segregating population for the yellow rust gene *Yr15,* 60% of the mapped SNPs with BFR>6 aligned to group 1 chromosomes (the gene is close to the centromere on 1BS, but the *Yr15* introgression included regions from 1BS and 1BL). In the v3.3 cDNA dataset, only 20% of the mapped SNPs with BFR>6 aligned to group 2 chromosomes.

The second significant finding was that for the markers which could be developed on the SNPs enriched in the BSA, on 7BS and 5B, most did not amplify the 2D flow-sorted DNA and did not map close to *Rht8*. This implies that SNPs were enriched in the BSA which were not linked to the 2D *Rht8* interval, although one of the markers did amplify the 2D DNA so might be a translocation to 2D. However, some SNPs with a high BFR mapping to 2S which were validated by polymorphic SNPs assays did not map to the *Rht8* linkage group. This implied that the resolution of the BSA might be limited to identify markers mapping close to *Rht8*. This will be considered next.

8.1.2.2 Genetic resolution from BSA limited by the reference and SNP array

One viewpoint is that the BSA strategy worked as well as it could have, but the genetic resolution was instead limited by other factors. In light of the evaluation of the v3.3 cDNAs above, it is possible that the limitations of the gene models captured in the reference meant that more tightly-linked SNPs with *Rht8* could

simply not be identified because the genes in the *Rht8* interval were not adequately represented. This is also true of BSA approach with the pre-defined variants on the iSelect array.

A systematic analysis of the pre-defined SNP-probes on the iSelect array mapping to 2D either genetically (the Akhunov map captured half of the probes, Appendix 5.1) or in the syntenic intervals, found low polymorphism between parent NILs and bulks. Even in the barley *Rht8* interval, the best-annotated of the syntenic species, the density of SNP-probes was low, with seven within the strict barley *Rht8* interval (Appendix 5.6.3). Only one marker designed on those pre-defined probes was found to be polymorphic between parent NILs and mapped to the *Rht8* interval (*52i*). Therefore, a valid view is that in this case, both the v3.3 cDNA reference and the iSelect array were not complete enough for the precision required for mapping *Rht8*.

In terms of using a capture array, there is potential to use the Axiom® 820K array for developing markers mapping close to *Rht8*, since a large number of variants on the array mapped to the syntenic zippers constructed in Chapter 5. This is a function of the greater (~ten-fold) number of pre-defined variants on the array, and the wider germplasm sampled (including landraces) to construct that array (CerealsDB, 2015b).

The aim of the *de novo* assembly was to circumvent the limitations of an incomplete reference or limited variant capture on an array. A *de novo* assembly would allow SNPs to be detected on novel transcripts as long as they were expressed to levels compatible with the depth of coverage (Wang et al., 2009).

The *de novo* assembly also had considerable background noise across non-2S chromosomes. The assembly only generated 29 putative SNPs mapping to 2S which were on novel genes not sampled in the v3.3 cDNA reference. The putative varietal SNPs identified from IWGSC CSS alignments did not generate any polymorphic markers. Since the SNPs on 2DS CSS contigs did not map to the POPSEQ bin identified as most likely to contain *Rht8* (or close by), these were not pursued further.

8.1.2.3 Did we capture *Rht8* in the material sampled?

The resolution of the BSA approach is a combination of marker density and the number of recombinations sampled in each bulk. The question of marker density and the possibility that this was limited by technical issues in the marker validation and limitations inherent to the references used have already been explored. It is also crucial to address the question as to whether it was realistic for the sampling strategy to identify the causal *Rht8* SNP. So rather than exploring the question: how did the BSA fail? We could consider: could the BSA ever have worked?

The criticisms of our BSA approach pertain to first, the recombinants sampled within the bulks, and second, the tissue selected for RNA-Seq.

The first point was of great concern at the start of the project. Accurate phenotyping was a major determinant in the success of the BSA approach and this was not underestimated. The work in Chapter 6 has shown that assigning recombinants to short and tall phenotypes was not trivial even with highly-replicated trials. The suboptimal conditions in the initial glasshouse experiments did delay the sequencing and as was shown in Chapter 6, field data changed the phenotype of a small number of recombinants. Phenotyping for a BSA in a quantitative trait such as height is more difficult than previous work in cereals, almost all of which has studied disease resistance where a qualitative score of resistant/susceptible between bulks is more certain (Ramirez-Gonzalez et al., 2014, Quarrie et al., 1999, Michelmore et al., 1991). Measures were taken to account for the potential phenotyping ambiguity, such as stratifying the bulks and sampling at the extremes of the height distributions. Looking back at the end of the project, it seems that only recombinants with confident phenotypes were included in the BSA, but there is a small possibility this was not the case.

A criticism of our BSA approach could be the small number of individuals (nine) which comprised each bulk. These individuals originated from 3104 F_2 plants and were selected from a fine-mapping *Rht8* population which mapped to a 1.29 cM interval harbouring *Rht8*. The number of individuals comprising bulks for BSA is typically much larger in Arabidopsis, ranging from 50 – 500 (Schneeberger et al., 2009, Austin et al., 2011). Other studies in cereals have used bulks comprising 14 – 20 plants (Trick et al., 2012, Ramirez-Gonzalez et al., 2014, Quarrie et al.,

1999). Although the absolute number of individuals is low in this case, since they were selected from a 1.29 cM interval, the mapping resolution was high – theoretically 0.07 cM. This supposes that the individuals comprising the samples captured enough recombination events to allow for this resolution. Given the large co-segregating block of 59 recombinants between marker class D and *Rht8,* which could not be resolved, there might be a lack of further recombination within the *Rht8* interval in the fine-mapping population used here. An alternative is that the relatively low number of individuals in the bulks did not capture enough recombination events to allow for adequate mapping. By sampling at the extreme of the height distributions, the recombinants with potential for ambiguity in phenotype were discounted. It might be that these recombination events were closest to *Rht8* and were not sampled. Given the concerns of incorrect typing of recombinants at the start of this project, it was decided imprudent to increase the size of the bulks. The lack of recombination within the interval will be explored in the future strategies below.

In our sampling strategy at the start of this project, we had to take a best estimate on the time-point and organ where *Rht8* would be actively expressed at high levels. It was reasonable to assume that this might be in the developing spike, given the reduction of spike length by *Rht8* reported previously by Gasperini (2010). If *Rht8* is linked to or causes the spike compaction reported in Chapter 4, it is feasible that the window of expression in terms of developmental time could be narrow and also at a different time point to our sampling at GS 30 – 39. Undetectably low *Rht8* expression would render our RNA-Seq approach ineffective. To resolve this, further work in terms of a developmental time-course of the action of *Rht8* looking at both wheat culms and spike compaction would be invaluable.

Finally, the very question of whether *Rht8* transcripts were captured in our strategy should be examined. The RNA-Seq strategy used here assumes that *Rht8* is a gene generating mRNA which we could detect. Some reasons why this might not have been the case have been explored. It is also possible that the mode of action of *Rht8* is more complex and related to gene dosage/copy number (for example, the duplicated gene cluster in Chapter 6) or epigenetic variation, which the RNA-Seq approach would not detect.

Epigenetic variation from variable methylation has been identified via bisulfite sequencing to be involved in the modulation of certain genes in wheat such as *Ppd-B1* (Sun et al., 2014) and the expansin gene *TaEXP1* (Hu et al., 2013). Further, transcriptional silencing has been found to be involved in a third of single-copy gene homoeoalleles with organ-specific and temporal control found to be common (Bottley et al., 2006). In addition, small non-coding RNAs have been shown to regulate expression in plants at the transcriptional and post-transcriptional level by binding to gene targets (Vazquez et al., 2010). The small RNA transcriptome has been analysed, revealing dynamic homoeologue regulation mediated by small RNAs (Li et al., 2014). One example of this is a putative *miRNA172* binding site in an exon of the domestication gene, *Q*, pointing to a possible role of miRNA regulation (Zhang et al., 2011). Since the sequencing of the wheat genome and its progenitors has identified large numbers of small non-coding RNAs (IWGSC, 2014, Ling et al., 2013, Jia et al., 2013), we are only at the beginning of understanding how this might alter expression of agronomically-important genes such as *Rht8*.

8.2 Future directions

Deployment of *Rht8* in agriculture

Further trialling of the *Rht8* NIL in low-input agricultural systems is endorsed in light of the findings of Chapter 3. Favourable results would broaden the potential agronomic application of *Rht8*. The emphasis on low-input agricultural systems is only going to increase into the future as resources become more limited. Further investigation into the mechanism of root-lodging is also warranted, given that variation for anchorage in the UK remains largely untapped (Berry et al., 2007).

Developmental time-course of *Rht8*

Work in Chapter 3 implicated the involvement of tiller dynamics in modulating the yield of *Rht8.* This, together with the spike compaction which was visually observed in the developing spike in Chapter 4, suggests that studying the effect of *Rht8* only at full maturity is insufficient. Taken together with the evaluation of the BSA strategy, it is clear that a developmental time-course to study the spatial and temporal effects of *Rht8* would increase our understanding both on the

molecular and agronomic level. New phenomics technology being trialled at the JIC could be used to measure canopy dynamics on a plot-level in the field. This could be complemented by experiments in controlled-growth conditions, by destructive assays measuring internode and spike dynamics during development, and also investigating the development of the floret to observe how early compaction takes place.

Interplay between *Rht8* and *compactum*

The intriguing findings of Chapter 4 highlight the dearth of knowledge in spike compactness and niche interest in club wheats in scientific literature. It is particularly important to dissect the genetic effects of *Rht8* and *compactum* (*C*), if they are separate loci, for fine-mapping purposes. It is also important to determine the extent to which, if any, *C* contributes to some of the undesirable agronomic effects found in Chapter 3 such as yield penalty. The markers developed in Chapter 5, including the markers mapping to 2D but outside of the *Rht8* linkage map provide a fast way of determining genetic linkage between *Rht8* and *C* if they are polymorphic in *compactum* mapping populations. A collaboration with the small *compactum* research community could provide novel links between the well-documented semi-dwarf effect of *Rht8* and the relatively unknown semi-compact spike phenotype.

Short-term strategies for targeted marker development

The *Ae. tauschii* resources could not be fully capitalised on due to time limitations. An immediate strategy would be to utilise the ~4 Mb of sequence identified in the genetic interval. The highest priority should be mining the BAC contig spanning 360 Kb, in the middle of the *Rht8* interval and corresponding SNP marker sequences for polymorphism. Additionally, the Axiom® SNP array showed potential for marker discovery since it captured SNPs within the syntenic *Rht8* intervals. Genotyping the parent NILs with this array would be a relatively fast and inexpensive method to probe these variants.

Physical information to improve the low recombination in the *Rht8* interval

Usually a target interval of <0.5 cM is required to establish a physical contig (Krattinger et al., 2009a), which was not achieved in this case. However, in some situations, recombination is limited which does not make this a feasible target

(Adamski et al., 2013). Similarly in this case, there is low recombination within the *Rht8* interval since of the 63 fine-mapping recombinants mapping between the flanking marker classes; 59 are in a co-segregating block. These recombinants are only useful for fine-mapping if they can be resolved further. In light of this uncertainty, and with the evaluation of the flow-sorted 2D DNA in Chapter 5, a 2D BAC library from RIL4 would be an invaluable resource for the future. This could be screened with the markers developed in this project, followed by chromosome walking. A recent development is the fluorescence in situ hybridization in suspension (FISHIS) method (Giorgi et al., 2013). Using this method, chromosomes are flow-sorted using a fluorescent label. This method is currently run as a service for any wheat variety. This provides the opportunity to obtain a BAC library from flow-sorted 2D DNA of the short parent NIL in a matter of weeks.

Developing another fine-mapping population to capture more recombination events around *Rht8*

Success of map-based cloning of *Rht8* requires high polymorphism and a high recombination rate. Recombination in the interval was limited. In particular, there is a recombination dead-lock in the co-segregating block of 59 recombination events between marker class D and *Rht8.* Using another mapping population which might capture more recombination in this region should be constructed to increase mapping resolution. The fastest way of doing this would be to use some of the BC_5-material in Paragon generated previously (Gasperini, 2010). The first step would be to self-pollinate the heterozygotes from a further back-cross (say BC_6) to generate a BC_6F_2 population which could then be used for fine-mapping. An alternative would be to construct a new population with greater polymorphism by hybridisation of D-genome donors with tetraploid wheat to create a synthetic hexaploid-wheat mapping population.

Future NGS strategies should use an improved wheat reference

The v3.3 cDNA wheat reference used in this project was the best representation of gene models at the time. By the end of this thesis, it can be seen that the reference was limiting for the purposes of fine-mapping *Rht8*. Currently, there are more complete sets of gene models on *Ensembl*Plants, comprising over 100,000 genes (for example, the PGSB gene models, based on the flow-sorted CSS assemblies) (EnsemblPlants, 2015). These are improved references which could

be used to align to with the sequenced reads and for SNP-calling. However, a limitation is that truncated genes that are split across different CSS contigs are not represented in those gene models. Given the prediction of obtaining greater contiguity across coding regions from the Chapman assembly (Chapman et al., 2015), developments in the near future incorporating genes from that dataset will improve the gene models available to the wheat community.

Future strategies to clone *Rht8*

One strategy to further define the *Rht8* interval would be the use of overlapping deletions in γ-irradiated mutants. This approach has been used successfully for the *Ph1* locus, which was also suffering from a lack of recombination (Griffiths et al., 2006). Deleted segments of the *Rht8* interval could be identified and the plants harbouring these deletions grown and phenotyped. Subsequently, candidate genes mapping to the deletions essential for *Rht8* function could be further tested, but subtle height effects would be difficult to detect robustly.

One way in which candidate genes identified in this project and in the future could be tested is through the use of Targeting Induced Local Lesions IN Genomes (TILLING). TILLING is a reverse genetics approach which introduces SNPs from chemical mutagenesis to induce deleterious mutations and then uses high-throughput screening to identify the mutations (Uauy et al., 2009). Sequenced tetraploid and hexaploid TILLING populations are being made available at the JIC later in 2015, whereby it will be possible to identify a mutation in a gene of interest *in silico* and order the seed online. The database which will host these resources will be searchable using the IWGSC CSS scaffold or *Ensembl*Plants gene model as a query. This will be a valuable resource for functional characterisation of candidate genes. A candidate gene would be further validated by transformation studies such as stable transformation or virus-induced gene silencing and by complementation of null or knock-down mutants by transgenesis (Krattinger et al., 2009a).

Appendices

The appendices are presented grouped into chapters, with the relevant chapter number preceding the figures and tables.

Appendix to Chapter 2

```
##R Script to calculate day length over the growing season 2013-14
##Ania Kowalski
##June 2015

##loading packages
library("ggplot2")
library("maptools")
library("scales")
#Norwich data
##using site co-ordinates
x_seq_N <- seq(from = as.POSIXct("2013-09-01", tz = "GMT"), length.out = 365, by = "days")
coord_N <- matrix(c(-1.29,52.63), nrow = 1)
sunrise<-sunriset(coord_N, x_seq_N, direction="sunrise", POSIXct.out=TRUE)
sunset<-sunriset(coord_N, x_seq_N, direction="sunset", POSIXct.out=TRUE)
day_length=as.numeric(sunset$time-sunrise$time)
n<-data.frame(date=as.Date(sunrise$time),day_length)

##Lleida data
##using site co-ordinates
x_seq_L <- seq(from = as.POSIXct("2013-09-01", tz = "CET"), length.out = 365, by = "days")
coord_L <- matrix(c(0.63, 41.62), nrow = 1)
sunrise_L<-sunriset(coord_L, x_seq_L, direction="sunrise", POSIXct.out=TRUE)
sunset_L<-sunriset(coord_L, x_seq_L, direction="sunset", POSIXct.out=TRUE)
day_length_L=as.numeric(sunset_L$time-sunrise_L$time)
L<-data.frame(date=as.Date(sunrise_L$time),day_length_L)

setwd("E:/PhD/Paragon x Rht8/R/data")
write.csv(L, file="Lleida_daylength.csv")
write.csv(n, file="Norwich_daylength.csv")

##now combine these into one csv and load it in
days<-read.csv("daylength_combined.csv",as.is=T)
days$date <- as.Date(days$date, format="%d/%m/%Y")
days$location <- factor(days$location, levels=c("Norwich","Lleida"))
class(days$date)

dayl<-ggplot(days, aes(x=date, y=day_length, group=location)) +
geom_line(size=2,aes(colour=location)) +
scale_colour_manual(values = c("#0072B2", "#CC0000")) +
scale_x_date(breaks=date_breaks("1 month"), labels=date_format("%b"), expand=c(0,0)) +
scale_y_continuous(limits=c(0,20), breaks=seq(0,20, 4), expand=c(0,0)) +
ylab("Day length (hours)") +
  xlab("") +
  theme(panel.grid.minor.x=element_blank(), panel.grid.major.x=element_blank(),
        plot.title = element_text(lineheight=.4, face="bold"),
        axis.title = element_text(size=25, face="bold", colour="black"),
        axis.text.y = element_text(size=20, colour="black"),
        axis.text.x = element_text(hjust=-1, size=20,face="bold", colour="black"),
        strip.text.x =element_text(size=22, face="bold"),
        strip.background=element_rect(colour = "black"),
        legend.text= element_text(size = 22),
        legend.title=element_blank(),
        legend.position="top") +
  guides(fill=guide_legend(title=NULL))
```

A2.1: R script to calculate day length over the 2013-14 growing season in Norwich and Lleida.

```
//Image procesing
//run("Threshold...");
// Color Thresholder 1.48i
min=newArray(3);
max=newArray(3);
filter=newArray(3);
a=getTitle();
run("HSB Stack");
run("Convert Stack to Images");
selectWindow("Hue");
rename("0");
selectWindow("Saturation");
rename("1");
selectWindow("Brightness");
rename("2");
min[0]=34;
max[0]=117;
filter[0]="pass";
min[1]=0;
max[1]=255;
filter[1]="pass";
min[2]=0;
max[2]=255;
filter[2]="pass";
for (i=0;i<3;i++){
  selectWindow(""+i);
  setThreshold(min[i], max[i]);
  run("Convert to Mask");
  if (filter[i]=="stop")  run("Invert");
}
imageCalculator("AND create", "0","1");
imageCalculator("AND create", "Result of 0","2");
for (i=0;i<3;i++){
  selectWindow(""+i);
  close();
}
selectWindow("Result of 0");
close();
selectWindow("Result of Result of 0");
rename(a);

// Pixel counting
blackPixels = 0;
totalPixels = getHeight * getWidth;
for (j = 0; j < getHeight(); j+=1){
        for(i = 0; i< getWidth(); i+=1){
                val=getPixel(i,j);
                if(val==255){
                        blackPixels +=1;
                }
        }
}
percent=(blackPixels)*100/totalPixels;
//Results
n=nResults;
e= getTitle();
setResult("File", n,a);
setResult("% Ground cover", n, percent);
```

A2.2: ImageJ (.ijm) script to calculate ground cover written by Oscar Gonalez.

	Glasshouse 2013				BSA	
F4 rec	Score	Height (cm)	N	St. error	RNA-Seq	iSelect
F4-1-2-4-3	a	79.9	24	0.9	B2	
F4-1-6-13-2	a	79.9	22	0.5	B4	T1
F4-1-6-17-1	a	78.7	25	0.9	B4	T2
F4-3-2-7-2	a	78.1	24	0.8	B4	T3
F4-3-2-13-1	a	77.8	24	0.7	B2	
F4-1-9-1-1	a	77.6	24	0.8	B2	
F4-1-6-16-1	a	77.5	24	1.0	B6	
F4-3-8-5-2	a	77.2	24	0.7	-	
F4-2-7-3-6	a	76.9	22	1.0	B6	
Cappelle	a	76.2	63	0.6	-	
F4-2-1-16-3	a	76.1	24	0.8	B6	
F4-1-7-4-1	a	75.5	24	0.7	-	
F4-3-7-7-2	a	75.4	24	0.6	-	
F4-1-1-7-3	a	75.3	24	1.5	-	
F4-3-7-14-3	a	75.0	24	0.7	-	
F4-2-1-11-4	a	74.6	25	0.7	-	
F4-2-8-1-2	b	74.1	24	0.9	-	
F4-2-2-7-1	b	73.9	23	0.7	-	
F4-3-2-2-1	b	73.3	24	0.6	-	
F4-1-7-1-1	b	71.2	25	1.0	B3	
F4-2-1-12-1	b	70.8	23	0.7	B3	
F4-3-2-16-1	b	70.7	23	1.2	B3	
F4-1-1-2-9	b	70.2	24	1.1	-	
F4-1-1-9-7	b	70.2	24	0.7	B5	S3
RIL4	b	69.8	63	0.4	-	
F4-3-2-5-1	b	69.3	24	0.7	-	
F4-3-1-2-6	b	69.3	20	0.9	B5	S2
F4-1-2-2-1	b	69.2	24	0.9	-	
F4-2-7-12-2	b	69.0	22	0.7	B1	
F4-3-8-6-3	b	68.7	22	0.9	B1	
F4-2-1-4-1	b	68.7	23	1.0	-	
F4-1-1-10-5	b	68.5	24	0.7	-	
F4-3-2-8-1	b	67.5	24	1.0	B5	S1
F4-3-8-1-1	b	66.6	23	1.7	B1	

BSA	Height (cm)	N	St. error
Short (B1, B3, B5)	**69.3**	**204**	**0.4**
Tall (B2, B4, B6)	**78.0**	**213**	**0.3**

A2.3: Mean heights of the subset of short and tall recombinants identified following the first glasshouse experiment. The selection of recombinants for the short and tall bulks for RNA-Seq and also for BSA using the iSelect SNP array is indicated.

Marker	Gene	Pos	Barley			Brachypodium			Rice	
2DS_AX4	mrna105701	573682..576170	MLOC_5618	2	9707617	Bradi5g02040	-		Os04g01740	heat shock protein 82
	mrna105285	581614..582602	MLOC_74324	4	207672522	Bradi1g63430	-		Os06g03770	ATP-binding cassette sub-family B member 7, mitochondrial
	mrna094242	585492..594064	MLOC_53502	2	9665488	Bradi1g63430	-		Os06g03770	ATP-binding cassette sub-family B member 7, mitochondrial
	mrna006241	-	-	3	170328523	Bradi4g00290	-		Os12g44360	-
	mrna091443	617087..617504	MLOC_65860	2	9586895	Bradi2g11550	-		Os01g19740	CP12-1
	mrna037735	626889..630688	MLOC_65968	2	7996695	Bradi5g02160	-		Os04g01590	arginase
	mrna026681	633698..638867	MLOC_15515	5	347207284	Bradi5g02170	acetyl-CoA C-acetyltransferase activity		Os09g07830	acetyl-CoA acetyltransferase, cytosolic 1
	mrna063205	647427..647522	MLOC_57107	2	9598506	Bradi5g02200	uridylate kinase activity		-	-
	mrna102282	677583..675726	MLOC_54900	2	9627250	Bradi5g02300	rRNA (uridine-C5-)-methyltransferase activity		Os04g01480	nucleic acid binding protein
	mrna047620	682344..676292	MLOC_54900	-	-	Bradi5g02300	-		Os04g01480	-
	mrna038321	686600..689111	MLOC_5776	4	350822582	Bradi5g02340	-		Os04g01290	PCI domain containing protein
	mrna126380	716490..723135	MLOC_18035	2	10827107	Bradi5g02400	-		Os04g01230	phosphoglycerate mutase-like protein
2DS_242	mrna090625	724876..729232	MLOC_13122	2	12807171	Bradi5g02400	-		Os04g01230	phosphoglycerate mutase-like protein
	mrna074113	737560..741664	MLOC_62712	-	-	Bradi5g02450	-		Os04g01150	-
66_uni	mrna040847	764510..755080	MLOC_65493	2	12631261	Bradi5g02490	cyclopropane-fatty-acyl-phospholipid synthase		Os07g29220	cyclopropane-fatty-acyl-phospholipid synthase
27_uni	mrna064977	773718..772828	MLOC_62750	2	11080613	Bradi5g02510	-		Os09g09320	-
	mrna066573	-	-	1	421952842	-	-		-	-
Freq_2	mrna026970	779625..784077	MLOC_62749	2	11088281	Bradi4g39940	serine-type carboxypeptidase activity		Os11g24374	circumsporozoite protein precursor, putative
	mrna105132	789621..784897	MLOC_38821	2	9986033	Bradi5g02520	D-glucose transmembrane transporter activity		-	-
1_al	mrna009588	809811..811325	MLOC_56811	2	12512964	Bradi5g02860	NBS LRR		Os04g11430	disease resistance RPP13-like protein 1, putative
	mrna121627	820329..817278	MLOC_56812	2	12519925	-	-		-	-
72_uni	2DS_5343763									
	mrna074509	865258..860151	MLOC_61723	2	14128009	Bradi5g02920	L-tyrosine transmembrane transporter activity		Os04g12499	amino acid permease, putative
	mrna028105	868800..875398	MLOC_13573	2	23130708	Bradi5g02940	NAD-dependent histone deacetylase activity		Os04g20270	mono-ADP-ribosyltransferase sirtuin-6
	mrna121338	895032..896599	MLOC_38009	2	27646647	-	-		-	-
2DS_235	mrna107490	905879..906241	MLOC_24124	2	15297328	Bradi5g03300	UDP-galactose:glucosylceramide beta-1,4-galactosyltransferase activity		-	-
DG_279	mrna093230	929999..929511	MLOC_5957	2	15606953	Bradi5g03460	xenobiotic-transporting ATPase activity		Os04g13210	multidrug resistance-associated protein 4
52i	mrna020368		MLOC_5957							
	mrna070632	934335..937458	MLOC_37835	2	283333993	Bradi5g03510	single-stranded DNA specific 3'-5' exodeoxyribonuclease activity		Os04g23830	oligoribonuclease
	mrna093698	942301..940006	MLOC_54824	2	283296081	Bradi1g55690	tyrosine-tRNA ligase activity		Os04g23820	multisynthetase complex auxiliary component p43
	mrna084787	952651..950993	MLOC_48019	2	17529225	Bradi5g03580	-		-	-
	mrna139758	967144..975027	MLOC_57508	2	15598089	Bradi5g03600	-		Os04g13470	expressed protein

A2.4: Annotated 2D v3.3 cDNA interval, delimited from the ordered section of the v3.3 cDNA reference. Physical position is indicated. Coloured red: markers developed in Chapter 5 which could be anchored via IWGSC 2DS contig; coloured blue: new markers developed in Chapter 6; shaded green: limits of the 59 genes used as the 2D interval; shaded grey: delimited interval by Gasperini (2010).

A2.4 continued

Marker	Gene	Pos	Barley			Brachypodium		Rice	
	mrna071578	980992..982656	MLOC_14804	2	17593210	Bradi5g03640	cytochrome-c peroxidase activity	Os04g14680	OsAPx3 - Peroxisomal Ascorbate Peroxidase
	mrna048555	987628..992477	MLOC_72300	-	-	-	-	-	-
	mrna091757	996752..998090	MLOC_78870	2	433617556	Bradi5g11360	-	Os04g36062	expressed protein
	mrna015009	1008566..1009383	MLOC_16798	2	17491605	Bradi3g20960	YUCCA-like flavin monooxygenase	Os04g14690	monooxygenase/ oxidoreductase
	mrna004763	1014979..1022844	MLOC_57069	2	293513169	Bradi5g03740	FAD binding	Os04g20990	tRNA-dihydrouridine synthase A
	mrna098230	1034572..1032265	MLOC_11990	2	15662153	Bradi5g03810	-	Os04g14790	mitochondrial precursor
	mrna053306	-	-	2	17554661	Bradi5g03830	-	Os05g22970	-
	mrna035375	1052274..1048977	MLOC_45846	2	17496478	Bradi5g03850	-	Os04g20590	nucleus protein
	mrna002983	1067514..1055341	MLOC_52767	2	17439859	Bradi5g03860	ATP binding	Os05g22940	acetyl-coenzyme A carboxylase
	mrna096393	1078530..1075332	MLOC_58539	2	17418895	Bradi5g03960	leucine-rich repeat protein kinase, subfamily LRR-XII	Os04g15560	receptor kinase-like protein, putative
	mrna016294	1097699..1098161	MLOC_10026	2	17639762	-	-	-	-
	mrna106738	1109640..1110322	MLOC_61794	2	17692687	Bradi5g03980	-	-	-
	mrna096003	1117860..1116717	MLOC_61793	2	17399188	Bradi5g04000	leucine-rich repeat protein kinase, subfamily LRR-XII	Os04g15660	receptor-like protein kinase 5 precursor, putative
	mrna124385	1128085..1128538	MLOC_63016	2	17704245	Bradi5g04060	-	Os04g15800	expressed protein
	mrna118007	1133627..1133074	MLOC_21811	2	17600858	Bradi4g08500	ribulose-bisphosphate carboxylase activity	Os12g17600	ribulose bisphosphate carboxylase small chain C
	mrna105093	1141782..1137361	MLOC_64679	5	308971823	Bradi4g08800	ribulose-bisphosphate carboxylase activity	Os12g17600	ribulose bisphosphate carboxylase small chain C
2DS_26	mrna057813	1178953..1175902	MLOC_10084	2	19080339	Bradi4g40600	protein tyrosine/serine/threonine phosphatase activity	Os12g09120	mRNA capping enzyme
	mrna043662	1187529..1177350	MLOC_10084	2	19203419	Bradi4g40600	protein tyrosine/serine/threonine phosphatase activity	Os12g09120	mRNA capping enzyme
	mrna007120		-						
	mrna096121	1213914..1221010	MLOC_81817	-	-	Bradi5g04580	-	Os04g14510	-
	mrna029953	1228749..1227278	MLOC_62246	2	19049035	Bradi5g04590	RING finger protein 13	-	-
	mrna001012	1247138..1246392	MLOC_71561	2	19532719	Bradi5g04630	-	Os10g05970	proline-rich protein
	mrna023290	1261929..1263458	MLOC_60079	2	18991726	Bradi3g59380	-	Os08g33910	NAC domain-containing protein 9
DG371	mrna079612	1295398..1277393	MLOC_4350	2	19442462	Bradi5g04670	-	Os04g18010	160 kDasubunit
	mrna066175	1304350..1308155	MLOC_8932	2	557067051	Bradi1g15750	-	Os03g36439	F-box domain containing protein
	mrna110666		-						
	mrna064310	1319922..1325751	MLOC_36970	2	145179275	Bradi1g54980	-	Os03g58600	argonaute-like protein
	mrna139098	1320119..1326260	MLOC_36970	2	145179275	Bradi1g54980	-	Os07g09020	argonaute-like protein
	mrna072920	1345527..1345991	MLOC_43713	-	-	Bradi1g16060	-	Os03g33650	-
	mrna077126	1347080..1350291	MLOC_59733	2	19470273	Bradi1g16100	-	Os03g39740	expressed protein
	mrna024572	1352649..1353401	MLOC_58913	2	454551540	Bradi1g16240	F-Box	Os03g30920	F-box domain containing protein
	mrna131003	1354903..1352804	MLOC_58913	2	454551973	Bradi1g16240	F-Box	Os03g30920	F-box domain containing protein
	mrna054831	1357377..1359480	MLOC_58913	2	454556219	Bradi5g12610	-	Os04g38980	pentatricopeptide repeat protein PPR1106-17
	mrna053564	1362385..1362937	MLOC_60251	6	524296736	Bradi1g22150	-	Os01g62000	pectate lyase 4 precursor, putative
	mrna007148	1366856..1364433	MLOC_17685	-	-	Bradi1g16460	-	-	-
	mrna014279	1369240..1366797	MLOC_17685	-	-	Bradi1g16460	-	-	-
	mrna066730	1384586..1383619	MLOC_72613	2	24814834	Bradi1g16480	subfamily RLCK-VIIa	Os07g49470	protein kinase APK1B, chloroplast precursor
	mrna058994	1386885..1386044	MLOC_81154	2	24919963	Bradi1g16490	-	-	-
	mrna098363		-						
	mrna057019		-						
	mrna057573	-	-	2	358279548	Bradi1g16520	-	Os07g49390	-

A2.5 continued over 12 pages: UNIX and BASH commands to align parent NIL and bulk reads to v3.3 cDNA reference, 2D v3.3 cDNAs and call SNPs in the 2D v3.3 cDNA interval using VarScan.

```
###BASH commands to align parent NIL and bulk reads to v3.3 cDNA reference
# Ania Kowalski
# April-June 2014
#files were received as gzip forward (R1) and reverse (R2) reads
#in fastq format
##files split into separate folder per sample and reads within each folder were concatenated

###############1####################
##concatenate the files into R1 and R2
##This merges all forward read together and reverse reads together, for each sample.
##Parent reads shown here, repeated in the same way for bulks samples B1-B6

#S1_CD_Spike
bsub -q NBI-Test128 cat *R1*.fastq.gz >P_S1_CD_Spike_R_1.fastq.gz
bsub -q NBI-Test128 cat *R2*.fastq.gz >P_S1_CD_Spike_R_2.fastq.gz
#S2_RIL4_Spike
bsub -q NBI-Test128 cat *R1*.fastq.gz >P_S2_RIL4_Spike_R_1.fastq.gz
bsub -q NBI-Test128 cat *R2*.fastq.gz >P_S2_RIL4_Spike_R_2.fastq.gz
#S3_CD_Spike
bsub -q NBI-Test128 cat *R1*.fastq.gz >P_S3_CD_Spike_R_1.fastq.gz
bsub -q NBI-Test128 cat *R2*.fastq.gz >P_S3_CD_Spike_R_2.fastq.gz
#S4_RIL4_Spikecd ../S4
bsub -q NBI-Test128 cat *R1*.fastq.gz >P_S4_RIL4_Spike_R_1.fastq.gz
bsub -q NBI-Test128 cat *R2*.fastq.gz >P_S4_RIL4_Spike_R_2.fastq.gz
#S5_CD_Peduncle
bsub -q NBI-Test128 cat *R1*.fastq.gz >P_S5_CD_Peduncle_R_1.fastq.gz
bsub -q NBI-Test128 cat *R2*.fastq.gz >P_S5_CD_Peduncle_R_2.fastq.gz
#S6_RIL4_Peduncle
bsub -q NBI-Test128 cat *R1*.fastq.gz >P_S6_RIL4_Peduncle_R_1.fastq.gz
bsub -q NBI-Test128 cat *R2*.fastq.gz >P_S6_RIL4_Peduncle_R_2.fastq.gz
#S7_CD_Peduncle
bsub -q NBI-Test128 cat *R1*.fastq.gz >P_S7_CD_Peduncle_R_1.fastq.gz
bsub -q NBI-Test128 cat *R2*.fastq.gz >P_S7_CD_Peduncle_R_2.fastq.gz
#S8_RIL4_Peduncle
bsub -q NBI-Test128 cat *R1*.fastq.gz >P_S8_RIL4_Peduncle_R_1.fastq.gz
bsub -q NBI-Test128 cat *R2*.fastq.gz >P_S8_RIL4_Peduncle_R_2.fastq.gz

##Copying all reads into reads_fastq file
bsub -q NBI-Test128 cp *.fastq.gz /net/group-data/ifs/NBI/Research-Groups/Simon-
Griffiths/Ania_seq/reads_fastq/

###############2####################
##indexing genome for alignment
first source bowtie2-2.1.0
bsub -q NBI-Test128 bowtie2-build v3.3_cdna.fasta bowtie2_v3.3_cDNA  #cDNA reference
from Martin Trick

###############3####################
##aligning with bowtie
##navigate to Reads_input folder where .fastq files are
source bowtie2-2.1.0

##run alignments in batches:
##for bulks, run one alignment with the shorts and second alignment with the talls
##example bowtie2 -p 12 -x bowtie2/NC_012967.1 -1 SRR030257_1.fastq -2
SRR030257_2.fastq -S SRR030257.sam
```

##note: considered relaxing parameters to allow one mismatch (default is 0) by using -N 1,
but in the end used default parameters
the more relaxed mapping gave hardly any difference,
therefore do with more stringent parameters (i.e. no need for N -1)

##single alignments for the shorts
##S1/S3/S5
bsub -q NBI-Test256 "bowtie2 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 B_S1_short_R_1.fastq.gz -2 B_S1_short_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/short_S1.sam"
bsub -q NBI-Test256 "bowtie2 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 B_S3_short_R_1.fastq.gz -2 B_S3_short_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/short_S3.sam"
bsub -q NBI-Test256 "bowtie2 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 B_S5_short_R_1.fastq.gz -2 B_S5_short_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/short_S5.sam"

##single alignments for the talls
#talls S2/S4/S6
#save tall_S2_S4_S6 to Ania_seq/alignments/bowtie2

bsub -q NBI-Test128 "bowtie2 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 B_S2_tall_R_1.fastq.gz -2 B_S2_tall_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/tall_S2.sam"
bsub -q NBI-Test128 "bowtie2 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 B_S4_tall_R_1.fastq.gz -2 B_S4_tall_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/tall_S4.sam"
bsub -q NBI-Test128 "bowtie2 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 B_S6_tall_R_1.fastq.gz -2 B_S6_tall_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/tall_S6.sam"

##combining bulks
#shorts S1, S3, S5
#save short_S1_S3_S5 to Ania_seq/alignments/bowtie2

bsub -q NBI-Test256 bowtie2 -p 10 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 B_S1_short_R_1.fastq.gz,B_S3_short_R_1.fastq.gz,B_S5_short_R_1.fastq.gz -2 B_S1_short_R_2.fastq.gz,B_S3_short_R_2.fastq.gz,B_S5_short_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/short_S1_S3_S5.sam

#talls S2, S4, S6
#save tall_S2_S4_S6 to Ania_seq/alignments/bowtie2

bsub -q NBI-Test256 bowtie2 -p 10 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 B_S2_tall_R_1.fastq.gz,B_S4_tall_R_1.fastq.gz,B_S6_tall_R_1.fastq.gz -2 B_S2_tall_R_2.fastq.gz,B_S4_tall_R_2.fastq.gz,B_S6_tall_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/tall_S2_S4_S6.sam

##for parents, run one alignment per biological replicate
##save CD_Spike_S1_S3 to Ania_seq/alignments/bowtie2

```
bsub -q NBI-Test256 bowtie2 -p 10 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 P_S1_CD_Spike_R_1.fastq.gz,P_S3_CD_Spike_R_1.fastq.gz -2 P_S1_CD_Spike_R_2.fastq.gz,P_S3_CD_Spike_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/CD_Spike_S1_S3.sam

## SAVE RIL4_Spike_S2_S4 to Ania_seq/alignments/bowtie2
bsub -q NBI-Test256 bowtie2 -p 10 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 P_S2_RIL4_Spike_R_1.fastq.gz,P_S4_RIL4_Spike_R_1.fastq.gz  -2 P_S2_RIL4_Spike_R_2.fastq.gz,P_S4_RIL4_Spike_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/RIL4_Spike_S2_S4.sam

##save CD_Peduncle_S5_S7 to Ania_seq/alignments/bowtie2
bsub -q NBI-Test256 bowtie2 -p 10 -x /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/reference/genes/bowtie2_v3.3_cDNA -1 P_S5_CD_Peduncle_R_1.fastq.gz,P_S7_CD_Peduncle_R_1.fastq.gz  -2 P_S5_CD_Peduncle_R_2.fastq.gz,P_S7_CD_Peduncle_R_2.fastq.gz -S /net/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/alignments/bowtie2/CD_Peduncle_S5_S7.sam

################4####################
##index genome for SAMTOOLS
##cd into samtools_bowtie2

bsub -q NBI-Test128 samtools faidx v3.3_cdna.fasta

############5##########
## Convert .sam to .bam using SAMTOOLS
## first check header is present in the sam files

head filename.sam

#  if header info is there, use -bS option
##header is there if there is @ visible
#navigate into alignments/spike for bowtie sam files

bsub -q NBI-Test256 "samtools view -bS CD_Peduncle_S5_S7.sam > CD_Peduncle_S5_S7.bam"
##Note correcting for incorrect SAM file naming - now corrected with RIL S6 S8 (same file called CD mistakenly)
bsub -q NBI-Test256 "samtools view -bS CD_Peduncle_S6_S8.sam > RIL4_Peduncle_S6_S8.bam"
bsub -q NBI-Test256 "samtools view -bS CD_Spike_S1_S3.sam > CD_Spike_S1_S3.bam"
bsub -q NBI-Test256 "samtools view -bS RIL4_Spike_S2_S4.sam > RIL4_Spike_S2_S4.bam"
bsub -q NBI-Test256 "samtools view -bS short_S1_S3_S5.sam > short_S1_S3_S5.bam"
bsub -q NBI-Test256 "samtools view -bS tall_S2_S4_S6.sam > tall_S2_S4_S6.bam"

#####6#################
##QC using samstat on sam files
##and also samtools flasgstat of bam files to get statistics
##note this requires indexing and sorting
##navigate to alignments/bowtie2
bsub -q NBI-Test128 "samtools flagstat CD_Peduncle_S5_S7.bam"
bsub -q NBI-Test128 "samtools flagstat RIL4_Peduncle_S6_S8.bam"
bsub -q NBI-Test128 "samtools flagstat CD_Spike_S1_S3.bam"
bsub -q NBI-Test128 "samtools flagstat RIL4_Spike_S2_S4.bam"
bsub -q NBI-Test128 "samtools flagstat short_S1_S3_S5.bam"
bsub -q NBI-Test128 "samtools flagstat tall_S2_S4_S6.bam"
```

```
#####7########
##removing duplicates
##Need to first sort by read names -n
#sorting by name rather than the default which is chromosome location

##navigate to alignments/bowtie2

bsub -q NBI-Test256 "samtools sort -n CD_Peduncle_S5_S7.bam
CD_Peduncle_S5_S7_nsorted"
bsub -q NBI-Test256 "samtools sort -n RIL4_Peduncle_S6_S8.bam
RIL4_Peduncle_S6_S8_nsorted"
bsub -q NBI-Test256 "samtools sort -n CD_Spike_S1_S3.bam CD_Spike_S1_S3_nsorted"
bsub -q NBI-Test256 "samtools sort -n RIL4_Spike_S2_S4.bam RIL4_Spike_S2_S4_nsorted"
bsub -q NBI-Test256 "samtools sort -n short_S1_S3_S5.bam short_S1_S3_S5_nsorted"
bsub -q NBI-Test256 "samtools sort -n tall_S2_S4_S6.bam tall_S2_S4_S6_nsorted"

##second step: because samtools rmdup works better when the insert size is set correctly,
#samtools fixmate can be run to fill in mate coordinates, ISIZE and mate related flags from a
name-sorted alignment

bsub -q NBI-Test256 "samtools fixmate CD_Peduncle_S5_S7_nsorted.bam
CD_Peduncle_S5_S7_nsorted_fixm.bam"
bsub -q NBI-Test256 "samtools fixmate RIL4_Peduncle_S6_S8_nsorted.bam
RIL4_Peduncle_S6_S8_nsorted_fixm.bam"
bsub -q NBI-Test256 "samtools fixmate CD_Spike_S1_S3_nsorted.bam
CD_Spike_S1_S3_nsorted_fixm.bam"
bsub -q NBI-Test256 "samtools fixmate RIL4_Spike_S2_S4_nsorted.bam
RIL4_Spike_S2_S4_nsorted_fixm.bam"
bsub -q NBI-Test256 "samtools fixmate short_S1_S3_S5_nsorted.bam
short_S1_S3_S5_nsorted_fixm.bam"
bsub -q NBI-Test256 "samtools fixmate tall_S2_S4_S6_nsorted.bam
tall_S2_S4_S6_nsorted_fixm.bam"

##third step
##remove duplicates, treating paired-end reads as single reads
##use -S
##format = samtools rmdup -S input bam output.bam
#navigate to alignments/bowtie2
bsub -q NBI-Test256 "samtools rmdup -S CD_Peduncle_S5_S7_nsorted_fixm.bam
CD_Peduncle_S5_S7_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test256 "samtools rmdup -S RIL4_Peduncle_S6_S8_nsorted_fixm.bam
RIL4_Peduncle_S6_S8_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test256 "samtools rmdup -S CD_Spike_S1_S3_nsorted_fixm.bam
CD_Spike_S1_S3_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test128 "samtools rmdup -S RIL4_Spike_S2_S4_nsorted_fixm.bam
RIL4_Spike_S2_S4_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test128 "samtools rmdup -S short_S1_S3_S5_nsorted_fixm.bam
short_S1_S3_S5_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test128 "samtools rmdup -S tall_S2_S4_S6_nsorted_fixm.bam
tall_S2_S4_S6_nsorted_fixm_rmdup.bam"

#####8##########
##Sorting and indexing by genomic position
##since this is the genomic position, use default

bsub -q NBI-Test256 "samtools sort CD_Peduncle_S5_S7_nsorted_fixm_rmdup.bam
CD_Peduncle_S5_S7_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test256 "samtools sort RIL4_Peduncle_S6_S8_nsorted_fixm_rmdup.bam
RIL4_Peduncle_S6_S8_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test256 "samtools sort CD_Spike_S1_S3_nsorted_fixm_rmdup.bam
CD_Spike_S1_S3_nsorted_fixm_rmdup_gsort"
```

```
bsub -q NBI-Test128 "samtools sort RIL4_Spike_S2_S4_nsorted_fixm_rmdup.bam
RIL4_Spike_S2_S4_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test128 "samtools sort short_S1_S3_S5_nsorted_fixm_rmdup.bam
short_S1_S3_S5_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test128 "samtools sort tall_S2_S4_S6_nsorted_fixm_rmdup.bam
tall_S2_S4_S6_nsorted_fixm_rmdup_gsort"

#samtools index input file (automatically will create a .bai file)

bsub -q NBI-Test256 "samtools index CD_Peduncle_S5_S7_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test256 "samtools index RIL4_Peduncle_S6_S8_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test256 "samtools index CD_Spike_S1_S3_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index RIL4_Spike_S2_S4_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index short_S1_S3_S5_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index tall_S2_S4_S6_nsorted_fixm_rmdup_gsort.bam"

##Aligned and indexed files given to Ricardo Ramirez-Gonzalez
##to run on BFR pipeline June 2015
```

##BASH commands to align parent NIL and bulk reads to 2D v3.3 cDNA interval
##Ania Kowalski
##January 2015

#using 2D interval fasta file from extracted mRNA sequences on 2D within interval as reference and aligning CD and RIL4 reads

##following this pipeline:
#https://wikis.utexas.edu/display/bioiteam/Removing+duplicates+from+alignment+output
###1#####
##INDEX THE GENOME
#Indexing the genome, which is the mrnas extracted from 2D
##indexing genome for alignment
first source bowtie2-2.1.0
bsub -q NBI-Test128 bowtie2-build interval.fasta 2Dinterval_bowtie2 #mRNAs extracted from Martin

index the reference sequence in the FASTA format
bsub -q NBI-Test128 samtools faidx interval.fasta

####2###############
##ALIGN THE READS
####################
###Reads to align

##use the fastq.gz files

##copy from Reads_input folder into interval_2D folder:
bsub cp P_S1_CD_Spike_R_1.fastq.gz ../interval_2D
bsub cp P_S1_CD_Spike_R_2.fastq.gz ../interval_2D

repeat for S2-S8

bsub cp P_S2_RIL4_Spike_R_1.fastq.gz ../interval_2D
bsub cp P_S2_RIL4_Spike_R_2.fastq.gz ../interval_2D

bsub cp P_S3_CD_Spike_R_1.fastq.gz ../interval_2D
bsub cp P_S3_CD_Spike_R_2.fastq.gz ../interval_2D

bsub cp P_S4_RIL4_Spike_R_1.fastq.gz ../interval_2D
bsub cp P_S4_RIL4_Spike_R_2.fastq.gz ../interval_2D

bsub cp P_S5_CD_Peduncle_R_1.fastq.gz ../interval_2D
bsub cp P_S5_CD_Peduncle_R_2.fastq.gz ../interval_2D

bsub cp P_S6_RIL4_Peduncle_R_1.fastq.gz ../interval_2D
bsub cp P_S6_RIL4_Peduncle_R_2.fastq.gz ../interval_2D

bsub cp P_S7_CD_Peduncle_R_1.fastq.gz ../interval_2D
bsub cp P_S7_CD_Peduncle_R_2.fastq.gz ../interval_2D

bsub cp P_S8_RIL4_Peduncle_R_1.fastq.gz ../interval_2D
bsub cp P_S8_RIL4_Peduncle_R_2.fastq.gz ../interval_2D

#Now have all reads, F and R from each parental sample.
#Align each sample (S1-S8) individually and then merge at the end
#Even though bowtie2 can align multiple reads at same time (bwa cannot)

#PARENTS
#single sample at a time

```
bsub -q NBI-Test256 "bowtie2 -x 2Dinterval_bowtie2 -1 P_S1_CD_Spike_R_1.fastq.gz -2
P_S1_CD_Spike_R_2.fastq.gz -S bowtie2_S1_CD_Spike_mrna.sam"
bsub -q NBI-Test256 "bowtie2 -x 2Dinterval_bowtie2 -1 P_S2_RIL4_Spike_R_1.fastq.gz -2
P_S2_RIL4_Spike_R_2.fastq.gz -S bowtie2_S2_RIL4_Spike_mrna.sam"
bsub -q NBI-Test256 "bowtie2 -x 2Dinterval_bowtie2 -1 P_S3_CD_Spike_R_1.fastq.gz -2
P_S3_CD_Spike_R_2.fastq.gz -S bowtie2_S3_CD_Spike_mrna.sam"
bsub -q NBI-Test256 "bowtie2 -x 2Dinterval_bowtie2 -1 P_S4_RIL4_Spike_R_1.fastq.gz -2
P_S4_RIL4_Spike_R_2.fastq.gz -S bowtie2_S4_RIL4_Spike_mrna.sam"
bsub -q NBI-Test256 "bowtie2 -x 2Dinterval_bowtie2 -1 P_S5_CD_Peduncle_R_1.fastq.gz -2
P_S5_CD_Peduncle_R_2.fastq.gz -S bowtie2_S5_CD_Peduncle_mrna.sam"
bsub -q NBI-Test256 "bowtie2 -x 2Dinterval_bowtie2 -1 P_S6_RIL4_Peduncle_R_1.fastq.gz -2
P_S6_RIL4_Peduncle_R_2.fastq.gz -S bowtie2_S6_RIL4_Peduncle_mrna.sam"
bsub -q NBI-Test256 "bowtie2 -x 2Dinterval_bowtie2 -1 P_S7_CD_Peduncle_R_1.fastq.gz -2
P_S7_CD_Peduncle_R_2.fastq.gz -S bowtie2_S7_CD_Peduncle_mrna.sam"
bsub -q NBI-Test256 "bowtie2 -x 2Dinterval_bowtie2 -1 P_S8_RIL4_Peduncle_R_1.fastq.gz -2
P_S8_RIL4_Peduncle_R_2.fastq.gz -S bowtie2_S8_RIL4_Peduncle_mrna.sam"

#######################################################
########3###########
#####QC on sam files
#######################################################
# samstat on sam files to get alignment statistics
#For each input file SAMStat will create a single html page named after the input file name plus
a dot html suffix.
#more info here http://samstat.sourceforge.net/
#######################################################

#######
##First need to source samstat
source samstat-1.09
#######
bsub -q NBI-Test128 samstat bowtie2_S1_CD_Spike_mrna.sam
bsub -q NBI-Test128 samstat bowtie2_S2_RIL4_Spike_mrna.sam
bsub -q NBI-Test128 samstat bowtie2_S3_CD_Spike_mrna.sam
bsub -q NBI-Test128 samstat bowtie2_S4_RIL4_Spike_mrna.sam
bsub -q NBI-Test128 samstat bowtie2_S5_CD_Peduncle_mrna.sam
bsub -q NBI-Test128 samstat bowtie2_S6_RIL4_Peduncle_mrna.sam
bsub -q NBI-Test128 samstat bowtie2_S7_CD_Peduncle_mrna.sam
bsub -q NBI-Test128 samstat bowtie2_S8_RIL4_Peduncle_mrna.sam


#########################
##4####
#### Convert .sam to .bam
#########################
## first check header is present in the sam fil
# head filename.sam
#  if header info is there, use -bS option
##header is there if there is @ visible
#SQ is present thus us -bS option

bsub -q NBI-Test256 "samtools view -bS bowtie2_S1_CD_Spike_mrna.sam >
bowtie2_S1_CD_Spike_mrna.bam"
bsub -q NBI-Test256 "samtools view -bS bowtie2_S2_RIL4_Spike_mrna.sam >
bowtie2_S2_RIL4_Spike_mrna.bam"
bsub -q NBI-Test256 "samtools view -bS bowtie2_S3_CD_Spike_mrna.sam >
bowtie2_S3_CD_Spike_mrna.bam"
bsub -q NBI-Test256 "samtools view -bS bowtie2_S4_RIL4_Spike_mrna.sam >
bowtie2_S4_RIL4_Spike_mrna.bam"
bsub -q NBI-Test256 "samtools view -bS bowtie2_S5_CD_Peduncle_mrna.sam >
bowtie2_S5_CD_Peduncle_mrna.bam"
```

```
bsub -q NBI-Test256 "samtools view -bS bowtie2_S6_RIL4_Peduncle_mrna.sam >
bowtie2_S6_RIL4_Peduncle_mrna.bam"
bsub -q NBI-Test256 "samtools view -bS bowtie2_S7_CD_Peduncle_mrna.sam >
bowtie2_S7_CD_Peduncle_mrna.bam"
bsub -q NBI-Test256 "samtools view -bS bowtie2_S8_RIL4_Peduncle_mrna.sam >
bowtie2_S8_RIL4_Peduncle_mrna.bam"

#######################################################
########5##########
#####QC on bam files
#######################################################
### samtools flagstat of the bam files to get statistics
bsub -q NBI-Test128 "samtools flagstat bowtie2_S1_CD_Spike_mrna.bam"
bsub -q NBI-Test128 "samtools flagstat bowtie2_S2_RIL4_Spike_mrna.bam"
bsub -q NBI-Test128 "samtools flagstat bowtie2_S3_CD_Spike_mrna.bam"
bsub -q NBI-Test128 "samtools flagstat bowtie2_S4_RIL4_Spike_mrna.bam"
bsub -q NBI-Test128 "samtools flagstat bowtie2_S5_CD_Peduncle_mrna.bam"
bsub -q NBI-Test128 "samtools flagstat bowtie2_S6_RIL4_Peduncle_mrna.bam"
bsub -q NBI-Test128 "samtools flagstat bowtie2_S7_CD_Peduncle_mrna.bam"
bsub -q NBI-Test128 "samtools flagstat bowtie2_S8_RIL4_Peduncle_mrna.bam"

############
#####6#####
############
# remove duplicates
#run samtools fixmate, and remove pcr duplicates.
# To do this, first sort by read names (-n)
## further explanation here http://www.htslib.org/workflow/ and in
https://wikis.utexas.edu/display/bioiteam/Removing+duplicates+from+alignment+output
##a)sort by read name
bsub -q NBI-Test256 "samtools sort -n bowtie2_S1_CD_Spike_mrna.bam
bowtie2_S1_CD_Spike_mrna_nsorted"
bsub -q NBI-Test256 "samtools sort -n bowtie2_S2_RIL4_Spike_mrna.bam
bowtie2_S2_RIL4_Spike_mrna_nsorted"
bsub -q NBI-Test256 "samtools sort -n bowtie2_S3_CD_Spike_mrna.bam
bowtie2_S3_CD_Spike_mrna_nsorted"
bsub -q NBI-Test128 "samtools sort -n bowtie2_S4_RIL4_Spike_mrna.bam
bowtie2_S4_RIL4_Spike_mrna_nsorted"
bsub -q NBI-Test128 "samtools sort -n bowtie2_S5_CD_Peduncle_mrna.bam
bowtie2_S5_CD_Peduncle_mrna_nsorted"
bsub -q NBI-Test128 "samtools sort -n bowtie2_S6_RIL4_Peduncle_mrna.bam
bowtie2_S6_RIL4_Peduncle_mrna_nsorted"
bsub -q NBI-Test128 "samtools sort -n bowtie2_S7_CD_Peduncle_mrna.bam
bowtie2_S7_CD_Peduncle_mrna_nsorted"
bsub -q NBI-Test128 "samtools sort -n bowtie2_S8_RIL4_Peduncle_mrna.bam
bowtie2_S8_RIL4_Peduncle_mrna_nsorted"

##b) run samtools fixmate to fill in mate coordinates, ISIZE and mate related flags from a name-
sorted alignment

bsub -q NBI-Test256 "samtools fixmate bowtie2_S1_CD_Spike_mrna_nsorted.bam
bowtie2_S1_CD_Spike_mrna_nsorted_fixm.bam"
bsub -q NBI-Test256 "samtools fixmate bowtie2_S2_RIL4_Spike_mrna_nsorted.bam
bowtie2_S2_RIL4_Spike_mrna_nsorted_fixm.bam"
bsub -q NBI-Test256 "samtools fixmate bowtie2_S3_CD_Spike_mrna_nsorted.bam
bowtie2_S3_CD_Spike_mrna_nsorted_fixm.bam"
bsub -q NBI-Test256 "samtools fixmate bowtie2_S4_RIL4_Spike_mrna_nsorted.bam
bowtie2_S4_RIL4_Spike_mrna_nsorted_fixm.bam"
bsub -q NBI-Test128 "samtools fixmate bowtie2_S5_CD_Peduncle_mrna_nsorted.bam
bowtie2_S5_CD_Peduncle_mrna_nsorted_fixm.bam"
bsub -q NBI-Test128 "samtools fixmate bowtie2_S6_RIL4_Peduncle_mrna_nsorted.bam
bowtie2_S6_RIL4_Peduncle_mrna_nsorted_fixm.bam"
```

```
bsub -q NBI-Test128 "samtools fixmate bowtie2_S7_CD_Peduncle_mrna_nsorted.bam
bowtie2_S7_CD_Peduncle_mrna_nsorted_fixm.bam"
bsub -q NBI-Test128 "samtools fixmate bowtie2_S8_RIL4_Peduncle_mrna_nsorted.bam
bowtie2_S8_RIL4_Peduncle_mrna_nsorted_fixm.bam"

# c) remove duplicates, treating paired-end reads as single reads

bsub -q NBI-Test256 "samtools rmdup -S bowtie2_S1_CD_Spike_mrna_nsorted_fixm.bam
bowtie2_S1_CD_Spike_mrna_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test128 "samtools rmdup -S bowtie2_S2_RIL4_Spike_mrna_nsorted_fixm.bam
bowtie2_S2_RIL4_Spike_mrna_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test256 "samtools rmdup -S bowtie2_S3_CD_Spike_mrna_nsorted_fixm.bam
bowtie2_S3_CD_Spike_mrna_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test256 "samtools rmdup -S bowtie2_S4_RIL4_Spike_mrna_nsorted_fixm.bam
bowtie2_S4_RIL4_Spike_mrna_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test128 "samtools rmdup -S bowtie2_S5_CD_Peduncle_mrna_nsorted_fixm.bam
bowtie2_S5_CD_Peduncle_mrna_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test128 "samtools rmdup -S bowtie2_S6_RIL4_Peduncle_mrna_nsorted_fixm.bam
bowtie2_S6_RIL4_Peduncle_mrna_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test128 "samtools rmdup -S bowtie2_S7_CD_Peduncle_mrna_nsorted_fixm.bam
bowtie2_S7_CD_Peduncle_mrna_nsorted_fixm_rmdup.bam"
bsub -q NBI-Test128 "samtools rmdup -S bowtie2_S8_RIL4_Peduncle_mrna_nsorted_fixm.bam
bowtie2_S8_RIL4_Peduncle_mrna_nsorted_fixm_rmdup.bam"

####################
########7############
##SORT AND INDEX####
# a) sort by genomic position so use default i.e NO NEED FOR -N

bsub -q NBI-Test256 "samtools sort bowtie2_S1_CD_Spike_mrna_nsorted_fixm_rmdup.bam
bowtie2_S1_CD_Spike_mrna_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test256 "samtools sort bowtie2_S2_RIL4_Spike_mrna_nsorted_fixm_rmdup.bam
bowtie2_S2_RIL4_Spike_mrna_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test256 "samtools sort bowtie2_S3_CD_Spike_mrna_nsorted_fixm_rmdup.bam
bowtie2_S3_CD_Spike_mrna_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test256 "samtools sort bowtie2_S4_RIL4_Spike_mrna_nsorted_fixm_rmdup.bam
bowtie2_S4_RIL4_Spike_mrna_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test128 "samtools sort
bowtie2_S5_CD_Peduncle_mrna_nsorted_fixm_rmdup.bam
bowtie2_S5_CD_Peduncle_mrna_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test128 "samtools sort
bowtie2_S6_RIL4_Peduncle_mrna_nsorted_fixm_rmdup.bam
bowtie2_S6_RIL4_Peduncle_mrna_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test128 "samtools sort
bowtie2_S7_CD_Peduncle_mrna_nsorted_fixm_rmdup.bam
bowtie2_S7_CD_Peduncle_mrna_nsorted_fixm_rmdup_gsort"
bsub -q NBI-Test128 "samtools sort
bowtie2_S8_RIL4_Peduncle_mrna_nsorted_fixm_rmdup.bam
bowtie2_S8_RIL4_Peduncle_mrna_nsorted_fixm_rmdup_gsort"

#b) merge CD and RIL4 files, and also keep as separate bam files so have merged and
individual (S1-S8 biorep and tissue provenance)

#merge all CD alignments
#S1, S3, S5, S7
bsub -q NBI-Test256 "samtools merge bowtie2_S1357_CD_mrna.bam
bowtie2_S1_CD_Spike_mrna_nsorted_fixm_rmdup_gsort.bam
bowtie2_S3_CD_Spike_mrna_nsorted_fixm_rmdup_gsort.bam
bowtie2_S5_CD_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam
bowtie2_S7_CD_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"

#merge all RIL4 alignments
```

```
#S2, S4, S6, S8
bsub -q NBI-Test256 "samtools merge bowtie2_S2468_RIL4_mrna.bam
bowtie2_S2_RIL4_Spike_mrna_nsorted_fixm_rmdup_gsort.bam
bowtie2_S4_RIL4_Spike_mrna_nsorted_fixm_rmdup_gsort.bam
bowtie2_S6_RIL4_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam
bowtie2_S8_RIL4_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"

#now sort these pooled alignments, to be sure they are sorted by genome position
bsub -q NBI-Test256 "samtools sort bowtie2_S1357_CD_mrna.bam
bowtie2_S1357_CD_mrna_gsort"
bsub -q NBI-Test128 "samtools sort bowtie2_S2468_RIL4_mrna.bam
bowtie2_S2468_RIL4_mrna_gsort"

#c)indexing to create a .bai file
#individual alignments
bsub -q NBI-Test128 "samtools index
bowtie2_S1_CD_Spike_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index
bowtie2_S2_RIL4_Spike_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index
bowtie2_S3_CD_Spike_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index
bowtie2_S4_RIL4_Spike_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index
bowtie2_S5_CD_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index
bowtie2_S6_RIL4_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index
bowtie2_S7_CD_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools index
bowtie2_S8_RIL4_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"

#pooled CD and RIL4 alignments
bsub -q NBI-Test128 "samtools index bowtie2_S1357_CD_mrna_gsort.bam"
bsub -q NBI-Test128 "samtools index bowtie2_S2468_RIL4_mrna_gsort.bam"


##############
##FINAL QC
##############

##OUTPUT format is:
#Retrieve and print stats in the index file.
#The output is TAB delimited with each line consisting of reference sequence name, sequence
length, # mapped reads and # unmapped reads
#NOTE: this requires bam files to be indexed and have a .bai file in the same directory.

bsub -q NBI-Test128 "samtools idxstats
bowtie2_S1_CD_Spike_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools idxstats
bowtie2_S2_RIL4_Spike_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools idxstats
bowtie2_S3_CD_Spike_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools idxstats
bowtie2_S4_RIL4_Spike_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools idxstats
bowtie2_S5_CD_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools idxstats
bowtie2_S6_RIL4_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"
bsub -q NBI-Test128 "samtools idxstats
bowtie2_S7_CD_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"
```

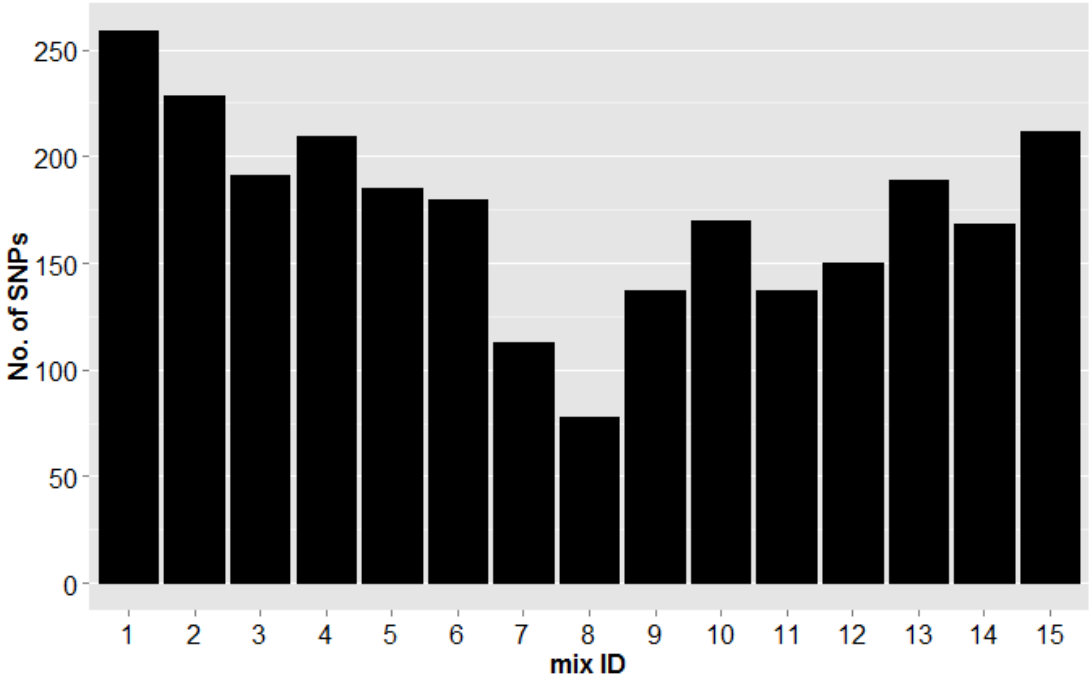
```
bsub -q NBI-Test128 "samtools idxstats
bowtie2_S8_RIL4_Peduncle_mrna_nsorted_fixm_rmdup_gsort.bam"

#pooled CD and RIL4 alignments
bsub -q NBI-Test256 "samtools idxstats bowtie2_S1357_CD_mrna_gsort.bam"
bsub -q NBI-Test256 "samtools idxstats bowtie2_S2468_RIL4_mrna_gsort.bam"

###COPYING OVER FINAL FILES TO A SUBFOLDER
###individual sample files for S1-S8
##copying bams
bsub -q NBI-Test128 "cp *rmdup_gsort.bam /net/group-data/ifs/NBI/Research-Groups/Simon-
Griffiths/Ania_seq/interval_2D/bams"
##copying bai files
bsub -q NBI-Test128 "cp *rmdup_gsort.bam.bai /net/group-data/ifs/NBI/Research-
Groups/Simon-Griffiths/Ania_seq/interval_2D/bams"
### pooled RIL4 and CD
##copying bams
bsub -q NBI-Test128 "cp bowtie2_S1357_CD_mrna_gsort.bam /net/group-
data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/interval_2D/bams"
bsub -q NBI-Test128 "cp bowtie2_S2468_RIL4_mrna_gsort.bam /net/group-
data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/interval_2D/bams"

##copying bai files
bsub -q NBI-Test128 "cp bowtie2_S1357_CD_mrna_gsort.bam.bai /net/group-
data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/interval_2D/bams"
bsub -q NBI-Test128 "cp bowtie2_S2468_RIL4_mrna_gsort.bam.bai /net/group-
data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/interval_2D/bams"

###END HERE
```

```
##BASH commands used to call SNPs using VarScan and pipe these into a .vcf file
##Using samtools to generate mpiluep files against interval.fasta 2D mrnas
##Ania Kowalski
##Run on 24-29th Dec 2014
##final edit January 2015
########################################

##input is the merged bam file alignments
##one example is shown here, the same commands were used for all bam alignments

#note, the script below was edited in vi for each file
#editing in a windows editor created ^M symbols which caused errors.
#each script was then saved as the name of the input bam file, and the .vcf file saved
accordingly.
#each script was edited with the correct .bam file and new .vcf piped output.
#insert changes with *i, then esc, then save name with :w filename.lsf, then quit vi editor with :q!
# job run as bsub < file.lsf

##6th Jan 2015 edit: add --output-vcf 1 with *i
##saved this script as vcf_
##modified output to have 'vcf'
##this is to get vcf v4.1 output, for visualising in IGV. Note, IGV accepts v4 vcf files.
##also, on 7th Jan 2015: ran pileup (not mpileup) on CD, RIL4, short and tall bulks, to see
difference.
## Note that pileup doesn't have the vcf output option.
##saved the pileups with pileup_prefix
##NOTE: PILEUP command has been removed (error message), so cannot compare this, only
use mpileup.
## Ran mpileup with all CD samples, RIL4 samples, short samples, and tall samples.
## ran above line again with vcf output, saved with usual prefix.

# LSBATCH: User input
#!/bin/bash
## LSF script to launch VarScan
BSUB -q NBI-Test128
BSUB -J VarScan
BSUB -R "rusage[mem=100000]"
source samtools-0.1.19
HOME=/usr/users/celldev/kowalska
REF=/nbi/group-data/ifs/NBI/Research-Groups/Simon-Griffiths/Ania_seq/interval_2D/bams/
cd $REF
samtools mpileup -f $REF/interval.fasta  bowtie2_S1357_CD_mrna_gsort.bam |  /software/jre-
6.0.25/x86_64/bin/java -jar $HOME/VarScan.v2.3.7.jar mpileup2snp > CD.vcf

##use this as a trick to see the progression of each job
# .bashrc
# User specific aliases and functions
# Source global definitions
if [ -f /etc/bashrc ]; then
        . /etc/bashrc
fi

export PYTHONPATH=$HOME/lib/python2.7/site-packages:$PYTHONPATH
export PYTHONPATH=$HOME/lib/python2.6/site-packages:$PYTHONPATH
export HISTTIMEFORMAT="%F %T "

alias lsrun='/nbi/common/lsf/lsf-7.0/7.0/linux2.6-glibc2.3-x86_64/bin/lsrun'
[kowalska@NBI-HPC interval_2D]$ /nbi/common/lsf/lsf-7.0/7.0/linux2.6-glibc2.3-
x86_64/bin/lsrun -SPv -m ncn-128-07 top
```

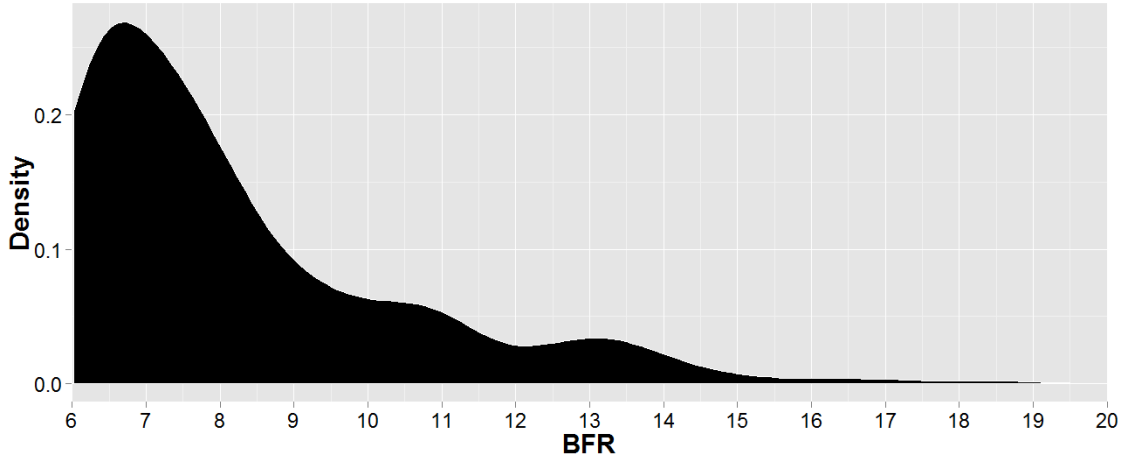
	Genotype	ID	Tissue	Total raw reads	Total reads mapped in pairs	Reads mapped in pairs of total (%)	Coverage
UniGenes	CD	P1	Spike	241,831,174	173,151,639	72	65
	RIL4	P2	Spike	188,987,916	136,685,332	72	51
	CD	P3	Spike	241,462,164	171,850,700	71	64
	RIL4	P4	Spike	173,680,844	124,603,201	72	46
	CD	P5	Peduncle	237,295,486	186,694,002	79	70
	RIL4	P6	Peduncle	174,156,526	136,755,947	79	51
	CD	P7	Peduncle	224,236,016	178,394,809	80	67
	RIL4	P8	Peduncle	189,388,702	148,341,563	78	55
v3.3 cDNAs	CD	P1	Spike	242,908,601	140,782,372	58	222
	RIL4	P2	Spike	189,833,791	111,238,568	59	175
	CD	P3	Spike	242,516,079	140,625,613	58	222
	RIL4	P4	Spike	174,429,905	101,792,495	58	161
	CD	P5	Peduncle	238,337,131	146,915,945	62	232
	RIL4	P6	Peduncle	174,924,471	109,124,171	62	172
	CD	P7	Peduncle	225,206,215	134,015,987	60	211
	RIL4	P8	Peduncle	190,256,420	117,003,831	61	185
	Short bulk	B1	Spike	124,330,880	69,318,085	56	109
	Tall Bulk	B2	Spike	153,325,474	87,709,759	57	138
	Short bulk	B3	Spike	150,650,804	88,522,097	59	140
	Tall bulk	B4	Spike	125,179,316	71,884,092	57	113
	Short bulk	B5	Spike	128,251,751	72,387,212	56	114
	Tall Bulk	B6	Spike	147,871,785	88,445,515	60	140

A2.6: Mapped reads and coverage statistics for alignments to the the v3.3 cDNAs and UniGenes.

in silico mix ID	Short parent NIL	Tall parent NIL	Short bulk	Tall bulk
1	bowtie2_S24_RIL4_Spike_short.bam	bowtie2_S13_CD_Spike_tall.bam	bowtie2_S135_bulk_Spike_short.bam	bowtie2_S246_bulk_Spike_tall.bam
2	bowtie2_S2468_RIL4_SpPed_short.bam	bowtie2_S1357_CD_SpPed_tall.bam	bowtie2_S135_bulk_Spike_short.bam	bowtie2_S246_bulk_Spike_tall.bam
3	bowtie2_S68_RIL4_Peduncle_short.bam	bowtie2_S57_CD_Peduncle_tall.bam	bowtie2_S135_bulk_Spike_short.bam	bowtie2_S246_bulk_Spike_tall.bam
4	bowtie2_S24_RIL4_Spike_short.bam	bowtie2_S13_CD_Spike_tall.bam	bowtie2_S35_bulk_Spike_short.bam	bowtie2_S246_bulk_Spike_tall.bam
5	bowtie2_S2468_RIL4_SpPed_short.bam	bowtie2_S1357_CD_SpPed_tall.bam	bowtie2_S35_bulk_Spike_short.bam	bowtie2_S246_bulk_Spike_tall.bam
6	bowtie2_S68_RIL4_Peduncle_short.bam	bowtie2_S57_CD_Peduncle_tall.bam	bowtie2_S35_bulk_Spike_short.bam	bowtie2_S246_bulk_Spike_tall.bam
7	bowtie2_S24_RIL4_Spike_short.bam	bowtie2_S13_CD_Spike_tall.bam	bowtie2_S1_bulk_Spike_short.bam	bowtie2_S2_bulk_Spike_tall.bam
8	bowtie2_S2468_RIL4_SpPed_short.bam	bowtie2_S1357_CD_SpPed_tall.bam	bowtie2_S1_bulk_Spike_short.bam	bowtie2_S2_bulk_Spike_tall.bam
9	bowtie2_S68_RIL4_Peduncle_short.bam	bowtie2_S57_CD_Peduncle_tall.bam	bowtie2_S1_bulk_Spike_short.bam	bowtie2_S2_bulk_Spike_tall.bam
10	bowtie2_S24_RIL4_Spike_short.bam	bowtie2_S13_CD_Spike_tall.bam	bowtie2_S3_bulk_Spike_short.bam	bowtie2_S6_bulk_Spike_tall.bam
11	bowtie2_S2468_RIL4_SpPed_short.bam	bowtie2_S1357_CD_SpPed_tall.bam	bowtie2_S3_bulk_Spike_short.bam	bowtie2_S6_bulk_Spike_tall.bam
12	bowtie2_S68_RIL4_Peduncle_short.bam	bowtie2_S57_CD_Peduncle_tall.bam	bowtie2_S3_bulk_Spike_short.bam	bowtie2_S6_bulk_Spike_tall.bam
13	bowtie2_S24_RIL4_Spike_short.bam	bowtie2_S13_CD_Spike_tall.bam	bowtie2_S5_bulk_Spike_short.bam	bowtie2_S4_bulk_Spike_tall.bam
14	bowtie2_S2468_RIL4_SpPed_short.bam	bowtie2_S1357_CD_SpPed_tall.bam	bowtie2_S5_bulk_Spike_short.bam	bowtie2_S4_bulk_Spike_tall.bam
15	bowtie2_S68_RIL4_Peduncle_short.bam	bowtie2_S57_CD_Peduncle_tall.bam	bowtie2_S5_bulk_Spike_short.bam	bowtie2_S4_bulk_Spike_tall.bam



A2.7.1: *In silico* mixes with different BAM-file combinations used to call SNPs in the BSA (top) and the number of SNPs identified using the different mixes (bottom).



A2.7.2: SNP density in the SNPs mapping to chromosomes 2AS, 2BS and 2DS with increasing BFR.

Gene	Base	Pos	SNP	Parent	Contig	Identity
mrna060106	C	165	G	tall	1AL_3459745	100
mrna068831	C	115	T	short	1AL_3844200	100
mrna068831	G	95	A	short	1AL_3844200	100
mrna031602	C	918	T	tall	1AL_3870255	97.95
mrna049250	A	738	G	tall	1AL_3875538	100
mrna049250	A	766	G	short	1AL_3875538	100
mrna049250	A	771	C	short	1AL_3875538	100
mrna038535	A	1608	G	tall	1AL_3888227	100
mrna038535	C	1310	T	tall	1AL_3888227	100
mrna038535	G	753	A	tall	1AL_3888227	100
mrna104149	C	1215	G	tall	1AL_3888280	99.65
mrna077919	T	174	C	tall	1AL_3900709	99.86
mrna134932	A	683	G	tall	1AL_3903825	99.69
mrna136896	G	259	C	tall	1AL_3915371	100
mrna041882	G	304	T	short	1AL_3920586	98.84
mrna115004	C	146	A	tall	1AL_3920663	100
mrna024736	C	452	T	tall	1AL_3925769	100
mrna018965	A	728	G	short	1AL_3930734	99.45
mrna056956	T	2084	C	short	1AL_3932920	100
mrna014541	T	111	C	short	1AL_3933825	100
mrna006360	G	135	C	tall	1AL_3936026	100
mrna042991	G	552	T	tall	1AL_3938636	100
mrna035206	A	352	G	short	1AL_3938726	100
mrna035206	C	298	T	short	1AL_3938726	100
mrna035206	G	289	C	short	1AL_3938726	100
mrna035206	G	322	C	short	1AL_3938726	100
mrna035206	T	1398	G	short	1AL_3938726	100
mrna035894	T	573	C	short	1AL_3963635	100
mrna132305	A	1137	G	short	1AL_3968091	100
mrna132305	A	1191	T	short	1AL_3968091	100
mrna132305	A	1429	G	short	1AL_3968091	100
mrna132305	A	960	G	short	1AL_3968091	100
mrna132305	C	1047	A	short	1AL_3968091	100
mrna132305	C	843	T	short	1AL_3968091	100
mrna132305	G	1086	A	short	1AL_3968091	100
mrna132305	G	1106	A	short	1AL_3968091	100
mrna132305	G	1164	A	short	1AL_3968091	100
mrna132305	G	624	C	short	1AL_3968091	100
mrna132305	T	1077	G	short	1AL_3968091	100
mrna132305	T	498	C	short	1AL_3968091	100
mrna132305	T	879	C	short	1AL_3968091	100
mrna132305	T	900	C	short	1AL_3968091	100
mrna132305	T	948	C	short	1AL_3968091	100
mrna077677	T	138	G	short	1AL_3968565	100
mrna019803	G	143	C	tall	1AL_3976997	100

Gene	Base	Pos	SNP	Parent	Contig	Identity
mrna039178	A	416	G	short	1AL_3977586	100
mrna039178	G	345	A	short	1AL_3977586	100
mrna059394	C	816	T	short	1AL_3977616	99.68
mrna059394	G	180	C	short	1AL_3977616	99.68
mrna082756	C	87	T	short	1AS_1644836	100
mrna026487	A	54	C	short	1AS_2196745	100
mrna026487	T	2250	C	short	1AS_2196745	100
mrna059657	C	708	T	tall	1AS_2294915	100
mrna059657	C	708	T	tall	1AS_2294915	100
mrna099156	A	135	G	tall	1AS_3252213	99.57
mrna082355	G	561	A	tall	1AS_3252288	100
mrna088602	C	150	T	tall	1AS_3263146	100
mrna088602	G	987	A	tall	1AS_3263146	100
mrna055280	T	439	C	tall	1AS_3263388	99.73
mrna107291	A	359	G	tall	1AS_3263895	99.63
mrna107291	A	763	G	tall	1AS_3263895	99.63
mrna036678	T	615	A	short	1AS_3266877	100
mrna032695	C	4	G	short	1AS_3270733	100
mrna134181	A	990	G	tall	1AS_3271165	100
mrna134181	T	1044	C	short	1AS_3271165	100
mrna123893	G	154	A	short	1AS_3272904	100
mrna013943	G	973	T	tall	1AS_3273947	100
mrna024055	A	1083	G	tall	1AS_3274088	100
mrna024055	A	1767	G	short	1AS_3274088	100
mrna024055	C	1453	T	tall	1AS_3274088	100
mrna106327	G	492	C	short	1AS_3275079	100
mrna079310	G	513	A	tall	1AS_3282639	100
mrna020301	A	123	T	tall	1AS_3287290	100
mrna031065	A	593	T	short	1AS_3295387	100
mrna037032	G	187	C	tall	1AS_3298206	100
mrna064198	T	1222	C	short	1AS_3302210	100
mrna075779	C	477	G	short	1AS_3316467	100
mrna075779	T	573	C	short	1AS_3316467	100
mrna010207	A	2362	C	tall	1AS_380247	100
mrna010207	A	2431	G	tall	1AS_380247	100
mrna010207	C	2356	T	tall	1AS_380247	100
mrna072622	A	850	G	short	1BL_3895131	95.99
mrna019962	A	1414	G	short	1DS_1875577	95.13
mrna019962	T	918	C	short	1DS_1875577	95.13
mrna019962	A	1414	G	short	1DS_1891653	95.13
mrna019962	G	787	C	short	1DS_1891653	95.13
mrna019962	T	918	C	short	1DS_1891653	95.13
mrna127780	T	90	A	tall	2DL_9876388	99.24
mrna087948	C	180	G	tall	3B_10673105	92.09
mrna072622	A	850	G	short	5BL_10869081	95.99

A2.8: Varietal SNPs in the v3.3 cDNAs with 100% variant bases at the SNP position from either short or tall parent, with a minimum-coverage threshold of 20. Columns left to right: Base = reference base at SNP position; Pos = position on gene model; SNP= variant at SNP position; Parent = short or tall progenitor from which 100% of the variant calls originated; Contig = IWGSC CSS best hit; Identity = % nucleotide identity of the gene to Contig.

Marker ID	Contig/source	FW-primer	RV-primer
2DS_1	2DS_5359909	TgtaaaacgacggccagtGTGTGGAGCCTATCCAAATGA	CCCAATGAACTGCTACATGAGT
2DS_3	2DS_5337443	TgtaaaacgacggccagtAAAAGGTAATAGAACCGGAGCC	TGTGATTGGTGAAGATGGAGAG
2DS_4	2DS_5337443	TgtaaaacgacggccagtAAAAGGTAATAGAACCGGAGCC	CATTTTCACCCCTATATGTCCG
2DS_5	2DS_5362384	TgtaaaacgacggccagtGCTTGTTGGTTTAATTGGTGG	TCTCTCTCCATAAGAAAACGCC
2DS_6	2DS_5321865	TgtaaaacgacggccagtCGACAGAAAACAAACGAGACTG	AGATTGATATGTACCTGCGCGT
2DS_7	2DS_5352598	TgtaaaacgacggccagtTCATTGTCATCCTTCTGCTGTC	CGTCTTGTAGTTGGCATCAATC
2DS_8	2DS_5352525	TgtaaaacgacggccagtTGTTTTGCTTTGATCCGTGTAG	GGAGAAGAAATGGACACACACA
2DS_9	2DS_5352525	TgtaaaacgacggccagtTGTTTTGCTTTGATCCGTGTAG	GGAGAAGAAATGGACACACACA
2DS_10	2DS_5352525	TgtaaaacgacggccagtATACGGGTCACAAATGGTCATA	AGGCTCAAGTTCTGCTGGATAG
2DS_11	2DS_5327480	TgtaaaacgacggccagtGTCACAAGGCAGCACAAACTAC	GGTTGATATGCACGATGATTTG
2DS_12	2DS_5390981	TgtaaaacgacggccagtGATGAGGAACACTCACAGCTTCT	ATAACCAGCTCCACACATTTCC
2DS_13	2DS_5379098	TgtaaaacgacggccagtACCTGCAAAACTCAAAAGTTGG	TCAGATTTATTCGCACTTGCC
2DS_14	2DS_5347513	TgtaaaacgacggccagtTTCGGTGTGCAGAAAAGTCTAA	TCTTAGTCCACCCATCTCCATC
2DS_15	2DS_5390752	TgtaaaacgacggccagtGTACCAACCTTTACGCCCCT	ACCACATCTTCACCCATATTCC
2DS_16	2DS_5390752	TgtaaaacgacggccagtATTGATGAGGAAAGGTGGAAGA	GACTCTTGAAAACGGAGCAAGT
2DS_17	2DS_5379495	TgtaaaacgacggccagtAAATCTAGGGTTGGGTTTTGGT	AAGAGGAAGAGGAAGAGGGAGA
2DS_18	2DS_5361162	TgtaaaacgacggccagtCATGAGCGAAAGGTCAGATACA	ACATCGTGTTTGTGCTTGGAC
2DS_19	2DS_5366894	TgtaaaacgacggccagtTGCTGGCTTGAGTAGTGAGAAA	ACATGACCCACACAAAACACAT
2DS_20	2DS_5389857	TgtaaaacgacggccagtCTCCTCCTCCTCAGTGTAGCAT	GACAAGCAAATTAAGACCGACC
2DS_21	2DS_5389857	TgtaaaacgacggccagtGGTCGGTCTTAATTTGCTTGTC	GTGTATGTACCAAAGCGGCATA
2DS_22	2DS_5389857	TgtaaaacgacggccagtAATATCCTCCCAGTCCTTCTCC	TAGATCCTTCCTTCCTTCCCTC
2DS_23	2DS_5389857	TgtaaaacgacggccagtAGCAGAGATGTGTGGATGGAC	CTAAACGGCCATAATAACCACC
2DS_24	2DS_5375625	TgtaaaacgacggccagtGCCATGTCTCTCTTCTTGTCCT	AGTTTCTGCTCCCCTTCCTTAC
2DS_25	2DS_5375625	TgtaaaacgacggccagtGCCATGTCTCTCTTCTTGTCCT	AGTTTCTGCTCCCCTTCCTTAC
2DS_26	2DS_5390977	TgtaaaacgacggccagtTGAGGGAAAATACAAAGAGGGA	ATGTTAAGTGGAACAGCGTGTG
2DS_27	2DS_5377094	TgtaaaacgacggccagtAAATGAAGGTTCGAGCAAAGAG	AAATGAAGGTTCGAGCAAAGAG
2DS_28	2DS_5341683	TgtaaaacgacggccagtACCATCGTTTCCTTAGCCTTCT	ACCATCGTTTCCTTAGCCTTCT
2DS_29	2DS_5347050	TgtaaaacgacggccagtGTTGTTGCGAATCAGTAATGGA	GTTGTTGCGAATCAGTAATGGA
2DS_30	2DS_5347050	TgtaaaacgacggccagtGAGCTTTGAGGTTTGATCCCTA	GAGCTTTGAGGTTTGATCCCTA
2DS_31	2DS_5383358	TgtaaaacgacggccagtATGTGGCAGTTCCTTCTTTTGT	ATGTGGCAGTTCCTTCTTTTGT
2DS_32s	2DS_5352598	TgtaaaacgacggccagtGGGACTCAATGTCTTCCTTGAC	TTGGTGTTGGATCTGCTAGTTG
2DS_33s	2DS_5352598	TgtaaaacgacggccagtCTCCGGCCACAATACCTATAAC	GTACTCTATGCAACCCTACCGC
2DS_34s	2DS_5352598	TgtaaaacgacggccagtATGAGCAGATCCACCCCTTT	AATTTACCCCTCAGCTTTTCGT
2DS_35s	2DS_5352598	TgtaaaacgacggccagtATCCACACATCAAATTCAGACG	GGCCATCTTATGCAACTTCTTT
2DS_36s	2DS_5367475	TgtaaaacgacggccagtGTCACAGCTTTCTTCCATCAAG	AGTCTTCGTTTGGGATTGGAT
2DS_37s	2DS_5367475	TgtaaaacgacggccagtAGATTCCCAACAAGCAAGGTAA	AGTGTTCACAACGAGCAAAAGA
2DS_38s	2DS_5389048	TgtaaaacgacggccagtTTCCACCACTAAATGCAGATTG	TACAATGCAGAGTGCTTAGGGA
2DS_39s	2DS_5343181	TgtaaaacgacggccagtCAACAGGCTTCATAACCAACTG	TATTGGCTGGGTCTAGTTTTGG
2DS_40s	2DS_5385061	TgtaaaacgacggccagtTCGTATCTGGTAAGTCCTCGGT	TAAACAAATCCAAACGGCCTAC
2DS_41s	2DS_5316778	TgtaaaacgacggccagtAATTGCGCTTGCTCTAACTTTC	TACTGGTATGCAACACGCCTTA
2DS_42s	2DS_5352525	TgtaaaacgacggccagtTAGCATTGGAATCTGTTGGTTG	CAAAAGGACATCTAAAACCGGA
2DS_43s	2DS_5327480	TgtaaaacgacggccagtGACCTTTAGCTTCTGCAAGGAA	TCGCCAAACTCTCTGAATAACC
2DS_44s	2DS_5327480	TgtaaaacgacggccagtCAACCAAAACCTTTATACCCCA	CCTCTCTGAAGCAAGAAGCCTA
2DS_45s	2DS_5324300	TgtaaaacgacggccagtACGGAGAATGGAAGAGGAAGAG	AGAGGGTGGGAGAGTGACCTAT
2DS_46s	2DS_5324300	TgtaaaacgacggccagtAGCCCTGCAATCCTTTATTCTA	ACTCATTTCTCTCAACGTGCAA
2DS_47s	2DS_5389716	TgtaaaacgacggccagtTCATTGGCTTGTTGCTCTCTTA	AAGGTTTCCGTGCATAGGTACA
2DS_48s	2DS_5389716	TgtaaaacgacggccagtAAAGCTCGACAACGCTCTAGTT	GTACCTCGCTACCGTCCATTTA
2DS_49s	2DS_5389716	TgtaaaacgacggccagtTGCTTGTATAGATGGTGCTTGG	CATGCCTGCTTCTTTATTACCC
2DS_50s	2DS_5389716	TgtaaaacgacggccagtTACAAAACTCGACCACGCTCTA	GATCTGTCTTTTAATTTGCCCG
2DS_51s	2DS_5341587	TgtaaaacgacggccagtACGTTCGACTTGATCCTTGATT	CCTACCCTATGTTGCCCATAAA
2DS_52s	2DS_5363870	TgtaaaacgacggccagtCAAGACATGCAACTGTGCTTTA	GCTGTTTTCTACAGACGTGGAGT
2DS_53s	2DS_5375380	TgtaaaacgacggccagtCATAGAATCATCCCAAAACCG	AAATCGAGCAAGAAAACTGGAG
2DS_54s	2DS_5339156	TgtaaaacgacggccagtCACGACTTTCGGTGAATCTTTA	AATGTGTTTGATTGGAAGGGAG
2DS_55s	2DS_5335120	TgtaaaacgacggccagtAGCAAGGTGAGATGAAGTTTCC	GGGCAAGACGCTATATTGTGAT
2DS_56s	2DS_5335120	TgtaaaacgacggccagtATGTGTCCCTTCCATTTATCGT	AATCTGTTCCCACTTGTTTTGG
2DS_57s	2DS_5343181	TgtaaaacgacggccagtCAACAGGCTTCATAACCAACTG	TATTGGCTGGGTCTAGTTTTGG
2DS_58s	2DS_5381312	TgtaaaacgacggccagtAATGCCTAACTGAATTGCCTGT	TGAGTATGTTTATTGCTTGGCG
2DS_59s	2DS_5381312	TgtaaaacgacggccagtAATGCCTAACTGAATTGCCTGT	TGAGTATGTTTATTGCTTGGCG
2DS_60s	2DS_5374739	TgtaaaacgacggccagtGGTGAATGTTGGTTGCACATAG	CAGTATTTGATGTTTGCCTTGC
2DS_61s	2DS_5379098	TgtaaaacgacggccagtGCAGTGGGAAACATGCTCTTAT	GATCAAAAGCTAGAATGGGGTG
2DS_62s	2DS_5347513	TgtaaaacgacggccagtGGATAGCCCGGAAGAGTAAAAT	GGACAAGTAAGCATGAGAACCC
2DS_63s	2DS_5347513	TgtaaaacgacggccagtACCAAACAAGCCTGAACTGAAT	CGTGTTCTGTCAATATGTGCAA
2DS_64s	2DS_5390752	TgtaaaacgacggccagtGTACCAACCTTTACGCCCCT	ACCACATCTTCACCCATATTCC
2DS_65s	2DS_5371042	TgtaaaacgacggccagtAATACGTCCTTTTCGCATCACT	ATGTGTACCAGAGAGAGGTCGG

A2.9: SSR markers. Annealing temperature with HotStar Taq was 60°C.

Marker ID	Contig/source	FW-primer	RV-primer
2DS_66s	2DS_5357871	TgtaaaacgacggccagtAATGGATGGGTAATATGGCACT	GCTAGGATGAGGTCCGAGTTTA
2DS_67s	2DS_5341487	TgtaaaacgacggccagtCCAATTCCAAACTTCCAACAAC	CAATACAATATCAACGCCCAGA
2DS_68s	2DS_5365680	TgtaaaacgacggccagtTTGTAGTTCTTTGCCTTCCGAT	AACCGAACAGACCGAAATCTTA
2DS_69s	2DS_5333000	TgtaaaacgacggccagtATCATTCCTCAACTCAAAACGG	ACTGTTCACGGGCGTAGATATT
2DS_70s	2DS_5341683	TgtaaaacgacggccagtCAAAACCAACTCAGCCTCTTCT	CCTCCATAATAAAATGCAAGGC
2DS_71s	2DS_5341683	TgtaaaacgacggccagtGCAGACTTTGTCGGTATCATCA	AGTAAGAGGAGGTTAAAGGGCG
2DS_72s	2DS_5341683	TgtaaaacgacggccagtACATTAACCCTGACATCAAGCC	CCAGTAGAGCCATCTCTCCATC
2DS_73s	2DS_5341683	TgtaaaacgacggccagtATTAAAGAGGGAGAATGGGGAA	AAACCAAGGACGATTGATGAAC
2DS_74s	2DS_5341683	TgtaaaacgacggccagtAAACGCTCAAACACAAAGGACT	CTACCCACAGGTTTAGGTGCAT
2DS_75s	2DS_5341683	TgtaaaacgacggccagtAAACGCTCAAACACAAAGGACT	CTACCCACAGGTTTAGGTGCAT
2DS_76s	2DS_5341846	TgtaaaacgacggccagtATCATTCCTCAACTCAAAACGG	ACTGTTCACGGGCGTAGATATT
2DS_77s	2DS_5375208	TgtaaaacgacggccagtCACGAAATCACAACTATGCCTC	GAGTCACGGCAATACCTCTCA
2DS_78s	2DS_5390697	TgtaaaacgacggccagtTTGAAAGACAGGGATCACAGAA	GAGAGAGACAAAAGAGTGGCAAA
2DS_79	2DS_5317970	TgtaaaacgacggccagtACTTATGGGTTTCAGGCTTTCA	GTGAACACAAATCTTCCGTCAA
2DS_80	2DS_5318660	TgtaaaacgacggccagtATTTGTACCGAGAAACACAGCA	TACGGAAACAAGTCCAGTAGCA
2DS_81	2DS_5318660	TgtaaaacgacggccagtTCTAACTTTATGACACCGCCAA	GAAGAAGGGTGAGGAAGGAACT
2DS_82	2DS_5318660	TgtaaaacgacggccagtTACTTTTGCCTTGACCAGAACC	TGCTATGTTTCTACGGCTGTGT
2DS_83	2DS_5319403	TgtaaaacgacggccagtATTTGTACCGAGAAACACAGCA	TACGGAAACAAGTCCAGTAGCA
2DS_84	2DS_5319403	TgtaaaacgacggccagtTCTAACTTTATGACACCGCCAA	GAAGAAGGGTGAGGAAGGAACT
2DS_85	2DS_5319403	TgtaaaacgacggccagtTACTTTTGCCTTGACCAGAACC	TGCTATGTTTCTACGGCTGTGT
2DS_86	2DS_5319959	TgtaaaacgacggccagtTATAGACGAGAGCGTCGTTTCA	CTTATGACTCTTTACCGCCGTC
2DS_87	2DS_5321770	TgtaaaacgacggccagtAAACCCAGTGAACAAAACAAGC	AAAGTGCCAAATAGGAATCGTC
2DS_88	2DS_5321770	TgtaaaacgacggccagtTCTAACCCTTTGGTCTTATCGC	TTCGCGGAAGATGAGTACATAA
2DS_89	2DS_5321770	TgtaaaacgacggccagtGAAGAATGAGGGAAATACGCAA	TTGACCTTCCATTGTACTGGTG
2DS_90	2DS_5330382	TgtaaaacgacggccagtAAACGCTCAAACACAAAGGACT	CTACCCACAGGTTTAGGTGCAT
2DS_91	2DS_5330382	TgtaaaacgacggccagtAAACGCTCAAACACAAAGGACT	CTACCCACAGGTTTAGGTGCAT
2DS_92	2DS_5338366	TgtaaaacgacggccagtTACTATGTTTTGGCACCCAGTT	AAGCGAGGATAAATCGGTAAGA
2DS_93	2DS_5338366	TgtaaaacgacggccagtGAGACTGGATGGGAGGGTAAAC	CGTGCATGAGAGGCAATATAGA
2DS_94	2DS_5338366	TgtaaaacgacggccagtCATGTGGTAAGCAATCCAAACT	CACGCGAGAAGCACAACTAT
2DS_95	2DS_5341322	TgtaaaacgacggccagtAACAAAGCTCGCAGTACATCAA	CCAATCCTTGAAAAGAACCCTA
2DS_96	2DS_5344563	TgtaaaacgacggccagtGTTGGACATTTGTACCCGAAAT	CTGGGTGTGCCTAGCTCTTATC
2DS_97	2DS_5349408	TgtaaaacgacggccagtCATCGGAATGAATGAAGGAAAC	AGTGCTGGTAGTGGTACGTGGT
2DS_98	2DS_5349408	TgtaaaacgacggccagtGATTGTCCGGTTCTCAGTCTTC	TGCGTGGTCCTACAGTTATGTC
2DS_99	2DS_5349408	TgtaaaacgacggccagtACTGCCGTAACCCTGAAAACT	ACGCGAAAAGAGACTTGACATT
2DS_100	2DS_5352574	TgtaaaacgacggccagtTTGATGTTAGCGAAGACATTGG	GTCATGCCACTCTTTGGTGAT
2DS_101	2DS_5352574	TgtaaaacgacggccagtGTTACAGTGCGGGAGATGTTTT	GTGTGAAAGAGCAAATGAAGACC
2DS_102	2DS_5358023	TgtaaaacgacggccagtGGTAGGTATATGTGTGCGCCTT	AGATGGGAAAAGGTAAGTGTGG
2DS_103	2DS_5368504	TgtaaaacgacggccagtCGTAGATAGTGTTCTCAGGCCG	GGCAGCTATGTTCAGTTCACAA
2DS_104	2DS_5372744	TgtaaaacgacggccagtCACAAACTTGCCTGGTATTTGA	AGTTATCATTCCTTTTGCGAGG
2DS_105	2DS_5390456	TgtaaaacgacggccagtGTACATTTTGCTTCCTTTTCCG	TAAAACCAGTTACATGCGGTGA
2DS_106	2DS_5390754	TgtaaaacgacggccagtCTACTACAACAACATTTGCCCG	ACATGAATTAACAGTCTTGCCG
2DS_107	2DS_5390754	TgtaaaacgacggccagtCAACAGAAAACCAAAGAGGGAG	TTGGGCTAGTATGGAGAGTTATGA
2DS_108	2DS_5325683	TgtaaaacgacggccagtACCTTCTATGGATCACCAGCTC	ACCTTCTATGGATCACCAGCTC
2DS_109	2DS_5325683	TgtaaaacgacggccagtAAGAAGAAGAAAGAGCGTGTGC	AAGAAGAAGAAAGAGCGTGTGC
2DS_110	2DS_5325683	TgtaaaacgacggccagtGTTCGACAACCTCTACCTCGAC	GTTCGACAACCTCTACCTCGAC
2DS_111	2DS_5325683	TgtaaaacgacggccagtAGAAATGTAGGAAAAGGTTGCG	AGAAATGTAGGAAAAGGTTGCG
2DS_112	2DS_5350706	TgtaaaacgacggccagtTGTACGATAGGACGTTTTCTGC	TGTACGATAGGACGTTTTCTGC
2DS_113	2DS_5350706	TgtaaaacgacggccagtGGTTTCTGAGTGTCGTGTGTGT	GGTTTCTGAGTGTCGTGTGTGT
2DS_114	2DS_5388494	TgtaaaacgacggccagtCTTTGTGGATGTCTTGAATGGA	CTTTGTGGATGTCTTGAATGGA
2DS_115	2DS_5390725	TgtaaaacgacggccagtGTTCTGTAACCGGCACTAAAGG	GTTCTGTAACCGGCACTAAAGG
2DS_116	2DS_5331590	TgtaaaacgacggccagtGTTTTAGTTGGAAATGGTGACG	GTTTTAGTTGGAAATGGTGACG
2DS_117	2DS_5331590	TgtaaaacgacggccagtTACTCCATTAAGTTCGCCATCA	TACTCCATTAAGTTCGCCATCA
2DS_118	2DS_5345858	TgtaaaacgacggccagtGAGCAGGAAGAAGGCGGT	GAGCAGGAAGAAGGCGGT
2DS_119	2DS_5345858	TgtaaaacgacggccagtTGGCTGTATCTCCTGCTCTGTA	TGGCTGTATCTCCTGCTCTGTA
2DS_120	2DS_5345858	TgtaaaacgacggccagtGATCTTCTTGTTCTGGTCGCTC	GATCTTCTTGTTCTGGTCGCTC
2DS_121	2DS_5340507	TgtaaaacgacggccagtCAGCACAGTAGATAGGCCAACA	CAGCACAGTAGATAGGCCAACA
2DS_122	2DS_5340507	TgtaaaacgacggccagtACTCTGGCTGTAAGGAAAGGTG	ACTCTGGCTGTAAGGAAAGGTG
2DS_123	2DS_5340507	TgtaaaacgacggccagtCCTCTCTCTCTCTCTCGACCAC	CCTCTCTCTCTCTCTCGACCAC
2DS_124	2DS_5318940	TgtaaaacgacggccagtTTCATGCAGAGTTCGTTTGTCT	TTCATGCAGAGTTCGTTTGTCT
2DS_125	2DS_5318940	TgtaaaacgacggccagtATTATACTGCCTGTTTGGCTGG	ATTATACTGCCTGTTTGGCTGG
2DS_126	2DS_5354001	TgtaaaacgacggccagtCCCGCTGAAATACTTACTTGAT	CCCGCTGAAATACTTACTTGAT
2DS_127	2DS_5390973	TgtaaaacgacggccagtGTAATGCGAGCACCAATATCAA	GTAATGCGAGCACCAATATCAA
2DS_128	2DS_5390973	TgtaaaacgacggccagtTGTTCGTGGATAAGGGAGCTAT	TGTTCGTGGATAAGGGAGCTAT
2DS_129	2DS_5390973	TgtaaaacgacggccagtAGAGAAGAAAGGAGAAAGGGGA	AGAGAAGAAAGGAGAAAGGGGA

A2.9 continued

Marker ID	Contig/source	FW-primer	RV-primer
2DS_130	2DS_5390973	TgtaaaacgacggccagtAGAACAAAGATAGCCGGTTGTG	AGAACAAAGATAGCCGGTTGTG
2DS_131	2DS_5371882	TgtaaaacgacggccagtTCGCCATAAGGTTGTGTCTATG	TCGCCATAAGGTTGTGTCTATG
2DS_132	2DS_5320325	TgtaaaacgacggccagtCCTTGGGTCTATCAAGCTGAAC	CCTTGGGTCTATCAAGCTGAAC
2DS_133	2DS_5389491	TgtaaaacgacggccagtGCATTTTCTTCCAAATCCAGTC	GCATTTTCTTCCAAATCCAGTC
2DS_134	2DS_5375359	TgtaaaacgacggccagtCAACGATGACACAGTCCGTATT	CAACGATGACACAGTCCGTATT
2DS_135	2DS_5329971	TgtaaaacgacggccagtATTTTGGTCCTTCTGAGTGCAT	ATTTTGGTCCTTCTGAGTGCAT
2DS_136	2DS_5309868	TgtaaaacgacggccagtCAGTGTAGGGGACGGATGG	CAGTGTAGGGGACGGATGG
2DS_137	2DS_5309868	TgtaaaacgacggccagtCTAAGGACCAACAGCGGTATTC	GTGACAAGCATCTATGTGGAGG
2DS_138	2DS_5318296	TgtaaaacgacggccagtCGTGCGAGAGATTACAGAGAGA	GGGGTGGTAAAACAAATCAAGA
2DS_139	2DS_5324565	TgtaaaacgacggccagtGCTACGTTGGTTGTACTTGCAC	CATCTCAGGGCATTTCACAATA
2DS_140	2DS_5324565	TgtaaaacgacggccagtCACAAATGTTGGAAAAGTGCC	GGATGTTAGTCAGGGCGATCTA
2DS_141	2DS_5361624	TgtaaaacgacggccagtATGCTGCTAGGGCAATCTAAAA	CCCATCCCTTCTAACAATATGC
2DS_142	2DS_5361624	TgtaaaacgacggccagtTGGAGTACAAGACAAGGCAGAA	CTACCTCAACCACGGCTACAC
2DS_143	2DS_5362023	TgtaaaacgacggccagtCGCGTACATACTCTCCTCTTTGA	CTCATTTGCATTGGGTTTCTTC
2DS_144	2DS_5362023	TgtaaaacgacggccagtAAATGGTTGCATGTATCACCCT	TATGCTCCTGCCCATACGATA
2DS_145	2DS_5366858	TgtaaaacgacggccagtAATAGATACCGATAACGCTCCG	ATGGACAAGCACACACACATCT
2DS_146	2DS_5381599	TgtaaaacgacggccagtCTGCCAAATTCATGTGTAGAGC	GAGTTTCCATCCCCAAATAGC
2DS_147	2DS_5388432	TgtaaaacgacggccagtGAAACAAGACCAAAGGTGGTTC	ACCCTACATGGAAATGCAAAAC
2DS_148	2DS_5319467	TgtaaaacgacggccagtGTACCAATACCAGAAAGCCTCG	GCCATTTATTCTCAACCACCAT
2DS_149	2DS_5319467	TgtaaaacgacggccagtAGATTGAGCCGACACATCATC	CGGAACGAGAAGAAGAAGCTAA
2DS_150	2DS_5379496	TgtaaaacgacggccagtCACAACAACAACTTTGAATGGC	TCACAAGATTCCCACATCTCAC
2DS_151	2DS_5354706	TgtaaaacgacggccagtGCTTGAACTCGCTGTACCTCTT	CGTCGGTCAATAGAAGGAAAAC
2DS_152	2DS_5354706	TgtaaaacgacggccagtTACCCTCACTAACAGGGCACTT	TCTCTGCAACTGTGGATTTGAT
AX1	2DS_5389432	TgtaaaacgacggccagtCAGCTAATGCAAATGTGGGTAA	CGCAGGTTTGTTGTCCAGTAG
AX2	2DS_5389432	TgtaaaacgacggccagtGTTAAATGTTATGTGCGGCCAG	TAAAAGGGTTTGATGCCAGGAG
AX3	2DS_5389432	TgtaaaacgacggccagtGTTAAATGTTATGTGCGGCCAG	TAAAAGGGTTTGATGCCAGGAG
AX4	2DS_5389432	TgtaaaacgacggccagtAAACAGCGACAACAGAAAGTGA	TCGTTAGATTTCCCATCTTCGT
AX5	2DS_5389432	TgtaaaacgacggccagtAAGAGTCGGATGAGAAGAAGGA	GCACGTAGAGCTTGATGTTGTT
AX6	2DS_5389432	TgtaaaacgacggccagtGGGGAAGCAGAACAGAAAGTTA	ACGAGAAGGTCATATTTGGGTG
AX7	2DS_5389432	TgtaaaacgacggccagtCACCCAAATATGACCTTCTCGT	GTGTCCAGCAAGTTCTTCTACAAA
AX8	1BL_3920229	TgtaaaacgacggccagtACAACTCTCCGACCATGAGG	AGCTCGCACCTACTGATAATAGAAA
AX9	1BL_3847188	TgtaaaacgacggccagtCCGGACTCTCTCATGTGTATCA	TCAAGCAAAGACAAAGGATCTG
AX10	1BL_3896023	TgtaaaacgacggccagtCCTAGATGTCCAAACACCATGA	GCTTGTTGCCAGTGTGAGATAA
AX11	1BL_3898099	TgtaaaacgacggccagtGCAGAGCGATGAACTAAAGAGA	ATACAGGGACGGATACTGAAGC
AX12	1BL_3898099	TgtaaaacgacggccagtTCCAGTAACAGAGCACATATTCC	CCTTTGAGATAGGTTTCCAAGA
AX13	1BL_3803445	TgtaaaacgacggccagtGGAGCATGGCAATTTTAGTTTC	AGGCAATGGACACAGTTGTAGA
AX14	1BL_3878973	TgtaaaacgacggccagtCCATGTCTTCCCAAAGGTTCTA	CCCACCAGATGTACGGTTCTTA
AX15	1BL_3919868	TgtaaaacgacggccagtTGAATGTGTCCTTGTTGTAGGC	ACAGAGATGGATGAGGGAAAGA
AX16	1BL_3920351	TgtaaaacgacggccagtAGGAATAAGCGACCACATCACT	TCATGGAAGGTCAAACAAGATG
AX17	1BL_3832084	TgtaaaacgacggccagtTCAGGGGTGTAGAGTGTAGTTTATGA	TGAGAATGCTTCCTTGCATCTA
AX18	2DS_5357871	TgtaaaacgacggccagtAATGGATGGGTAATATGGCACT	GCTAGGATGAGGTCCGAGTTTA
AX19	2DS_5334312	TgtaaaacgacggccagtTGAGAAAGGAACAAACGAAACC	CCATGCGTGACTGGAATTG
AX20	2DS_5334312	TgtaaaacgacggccagtCGAGTCCGATAAATAGGTTTGTG	GATCTGGCTTGTGTGTTCTGTC
AX21	2DS_5354297	TgtaaaacgacggccagtACGCCCAGATACCTAAGATTGA	TGTTTTAACGGGATCAAGTGTG
AX22	2DS_5347513	TgtaaaacgacggccagtTTCGGTGTGCAGAAAAGTCTAA	TCTTAGTCCACCCATCTCCATC
AX23	2DS_5347513	TgtaaaacgacggccagtGGATAGCCCGGAAGAGTAAAAT	GGACAAGTAAGCATGAGAACCC
AX24	2DS_5347513	TgtaaaacgacggccagtACCAAACAAGCCTGAACTGAAT	CGTGTTCTGTCAATATGTGCAA
AX25	2DS_5358861	TgtaaaacgacggccagtTCTCAATCTTCAGCACTTTCCA	GCAGCACGTCTCGTATTTACTG
AX26	2DS_5358861	TgtaaaacgacggccagtACGATGGAAGAGCGAAGAGAG	GTGACATTGAACAGGTGGTTTG
AX27	2DS_5358861	TgtaaaacgacggccagtAGCATCCAGATTTCCTCTCAAA	CTTCAGTCCCGCTTCAATATCT
AX28	2DS_5358861	TgtaaaacgacggccagtTTATCATGTGGTTGTAGGGCTG	AGAAAGAGAGGTGAGCAGTTGG
AX29	2DS_5358861	TgtaaaacgacggccagtTGCACTGAGCGTGGAAAA	ACCAACCACTATACGAGGAGGA
AX30	2DS_5388293	TgtaaaacgacggccagtAAGTACCCATACACATGCCCAG	GGTCAAAATAAACATGCCACCT
AX31	2DS_5388293	TgtaaaacgacggccagtACTACGCAGAACGGTATCAAAA	CCAGAATCCTACTTACCCCAAC
AX32	2DS_5388293	TgtaaaacgacggccagtTTGTCGAAGGAGTAGGTCTCATC	ACAATCAATTACGTCCATGCAG
AX33	2DS_5352839	TgtaaaacgacggccagtGCAAGGAGAAAGACACCAAGAT	CGACGACAATGAAAACAGAGAA
AX34	2DS_5352839	TgtaaaacgacggccagtAACAAATACTAATGGCGCACCT	CACAAAGAACGTACCAGATCCA
AX35	2DS_5352839	TgtaaaacgacggccagtTGGTCTACGGATCTGGATGTAA	AACTACACACACACACACGCTG
AX36	2DS_5352839	TgtaaaacgacggccagtGGCGGCAAGTAGAAAGACACTA	CAGCCAAACTGAACTGAAGACA
AX37	2DS_5352839	TgtaaaacgacggccagtAATGTTTGGTTACAGTTTCCGC	ATTTCACGCACCTGTCCAATAG
AX38	2DS_5352839	TgtaaaacgacggccagtTCTGAATACGAGCAAACCTGAA	GGAAATACTTGTCGGAGGAATG
AX39	2DS_5352839	TgtaaaacgacggccagtGGTGCGATTTTCTCTTTGTTTC	ACTCCTCGAAGTGCTATGCTTC
AX40	2DS_5352839	TgtaaaacgacggccagtACATACACGCACACAATCACAA	CAGAGAAAGCAAGCAAACATTC
AX41	2DS_5352839	TgtaaaacgacggccagtGCCTAGCCAACTTATGAGCTTC	AAACCATGACTCTCACGAAAGG

A2.9 continued

Marker ID	Contig/source	FW-primer	RV-primer
AX42	2DS_5317606	TgtaaaacgacggccagtGCATACATACTCTCGTGCAAGG	CGTACTAACCTTTGTCGTGCTG
AX43	2DS_5317606	TgtaaaacgacggccagtAAAAGCATCACCGAAGAGCTAC	GTCATGGATGTTGATCGAGAGA
AX44	2DS_5317606	TgtaaaacgacggccagtTCTCTCGATCAACATCCATGAC	AGACTCCATTAGGCTGAACACC
AX45	2DS_5317606	TgtaaaacgacggccagtGAGCTTCTAATATGGCGTGACC	TTTGCCCTGTATTGATGTTGAG
AX46	2DS_5359909	TgtaaaacgacggccagtTCGTCTGATTAAAGTGGTGGAA	CGACTTCTAGGTAGTGCAAGTGG
AX47	2DS_5359909	TgtaaaacgacggccagtGTGTGGAGCCTATCCAAATGA	CCCAATGAACTGCTACATGAGT
AX48	2DS_5366459	TgtaaaacgacggccagtCCCCGTCATGTGGTAGTATGTA	GGTTAATCCTGGTCAATGTGGT
AX49	2DS_5373243	TgtaaaacgacggccagtCGGCCTATTCACTTCAAAGATT	ACCGGCAAATGTAAGACAAGAT
AX50	2DS_5377660	TgtaaaacgacggccagtCTCGGGAGATCAATTAGGTTTG	AACTAGAAAGGATCGGAGGGAG
AX51	2DS_5377660	TgtaaaacgacggccagtTATGTGTGTGTGTGCGAGAGAG	AACTAGAAAGGATCGGAGGGAG
AX52	2DS_5377660	TgtaaaacgacggccagtCTAGGATTTCGCCACGTATAGC	TAGTGAGTTGAGAGCAAACCGA
AX53	5BL_1091087	TgtaaaacgacggccagtATAATGGATTTCACTCGATGCC	ACAGTATTTCAGACCCGCAAAA
AX54	5BL_1091087	TgtaaaacgacggccagtACTCCTCCATTAGCCTCCCTAC	ACACGTCCAGTGACAACAACTC
AX55	5BL_1088962	TgtaaaacgacggccagtGCACTTCGTGTACTCTCTCCCT	TTGTTGTGTGTGTGTGTGTGTG
AX56	5BL_1088962	TgtaaaacgacggccagtGCACTTCGTGTACTCTCTCCCT	TGTTGTTGTGTGTGTGTGTGTG
AX57	5BL_1088962	TgtaaaacgacggccagtCACACACACACACACAACAACA	CCCTACTCCTTCGGTGTCAGTA
AX58	5BL_1088962	TgtaaaacgacggccagtCACACTTACTGACACCGAAGGA	GGCACAGAGACACAACTAGCAC
2DS_153	v3.3 cDNA	TgtaaaacgacggccagtTGAAACAGATGGCTGATGTACG	ATCTCTCCCCTATCTCTCTGCC
2DS_154	v3.3 cDNA	TgtaaaacgacggccagtATAGGGGAGAGATTAAGTGGGC	ACTACACCACGACACTGCATTC
2DS_155	v3.3 cDNA	TgtaaaacgacggccagtATCAACCAACCAGAGACCAGAC	TTCATGTACCCAAGATGACCAG
2DS_156	v3.3 cDNA	TgtaaaacgacggccagtATGGATACAAGGGGTTCTTCCT	GACTCATCATCATCTTCGCTTG
2DS_157	v3.3 cDNA	TgtaaaacgacggccagtGAGAAGGAAGGTCAACGGTGT	CAGCAAGGACCAGCTACCC
2DS_158	v3.3 cDNA	TgtaaaacgacggccagtCTGACATAGTGACAAAAGGACCC	TCACACACACACACACACACAC
2DS_159	v3.3 cDNA	TgtaaaacgacggccagtGGATTCTCTTGATCTCTCTTGACG	TATCGTTTTAGCAAGCACTCCC
2DS_160	v3.3 cDNA	TgtaaaacgacggccagtAGAGCTTGGACCCATTTTGTAG	GGCCACTTCGAGTCAGGTTA
2DS_161	v3.3 cDNA	TgtaaaacgacggccagtGGAATGCTGGAAACTGTGTGTA	GCGAGTAGAAGTCAAGGCTCAC
2DS_162	v3.3 cDNA	TgtaaaacgacggccagtCCACTGCTTTTGCTTAATCTCA	CCATCATCTGGTTTCTTCACAA
2DS_163	v3.3 cDNA	TgtaaaacgacggccagtCAGCTCCGCCTCAACTTTT	CATTGCTTCAACTTCTCCGACT
2DS_164	v3.3 cDNA	TgtaaaacgacggccagtACACATGGCAGAAAAGCTAACA	GCCTACCGAGCACCCTTC
2DS_165	v3.3 cDNA	TgtaaaacgacggccagtATGGCTCACTTCAGTTCACCTT	CTTCGAGTTTACTGCTACGCTG
2DS_166	v3.3 cDNA	TgtaaaacgacggccagtGTGTAGATGAGACATGGCAAAA	CCTACCGAGCACCCTTCT
2DS_167	v3.3 cDNA	TgtaaaacgacggccagtAAATGACTGGCCGAACTATTTG	ATTGTATTAACCCCATGCCAAC
2DS_168	v3.3 cDNA	TgtaaaacgacggccagtACGCACAATCAGTTCCAAGATA	ATTCCAGGTTTATCGTCCCTTT
2DS_169	v3.3 cDNA	TgtaaaacgacggccagtAGAGCACCATGTTCAAAACCTT	CTGTTCATTTCAGATTGCTTGC
2DS_170	v3.3 cDNA	TgtaaaacgacggccagtTGCAACTCTTACTCTGTGCCTG	GGGTGTAGTATGTGGAAGCCAT
2DS_171	v3.3 cDNA	TgtaaaacgacggccagtGTTGAAGAACTGTGCCTACACG	TTGTTCGGACGTATGGTAGAAA
2DS_172	v3.3 cDNA	TgtaaaacgacggccagtTTGGGGTATATTTGGTTTGGAG	AAAAGAGAACTGCAAGTGGAGG
2DS_173	v3.3 cDNA	TgtaaaacgacggccagtGGTATCTCACACAGGGACAGGT	GACCTTCTTTTCAGTTTGGAGC
2DS_174	v3.3 cDNA	TgtaaaacgacggccagtGATAGCAGGGTTCTGATTTTGG	CGTGGTTTTCTCTGAATTTGG
2DS_175	v3.3 cDNA	TgtaaaacgacggccagtCACTCCCTCCATCACATTTTCT	GACAAGCTCACCCTCCTCC
2DS_176	v3.3 cDNA	TgtaaaacgacggccagtAACTTTCTCTGCAACCAAAACC	CGCGAAGAACTCGTAGAGG
2DS_177	v3.3 cDNA	TgtaaaacgacggccagtTACAAGGTCACGATTCATTTCG	AATATCCTCTCCAGCACTCCAG
2DS_178	v3.3 cDNA	TgtaaaacgacggccagtGATGATGCTCCATTACAACACC	CTACTCCCCTACTCCCAAATCC
2DS_179	v3.3 cDNA	TgtaaaacgacggccagtGTGCTAGGAAACCAAGAGGAGA	TAGTTCACTGGTGTTGATGCGT
2DS_180	v3.3 cDNA	TgtaaaacgacggccagtAAAGCACGACCTCTCTACTTGG	TTTGGAGTTTATCGGGCAATAC
2DS_181	Avalon x Cadenza interval	TgtaaaacgacggccagtGAAAGCAGAGTGGGAGAAGAAA	ATTGCGTCAACAGTCAAGGTAG
2DS_182	Avalon x Cadenza interval	TgtaaaacgacggccagtTCCAGTCTCGTTTCCCTTATGT	GTATGCACAAGCTACAGGCAAG
2DS_183	Avalon x Cadenza interval	TgtaaaacgacggccagtATGATAATTGGCGGAAGATGAG	TCAAGTAGAGGGTACATCGGGT
2DS_184	Limagrain	TgtaaaacgacggccagtCAGTTTTCCCTCGCCTATAATG	CCAGATACTAATGACGCACCAG
2DS_185	Limagrain	TgtaaaacgacggccagtCAAATCAATATGACTGCCGTGT	GTATGTGTGTGTGTGTGTGCGT
2DS_186	Limagrain	TgtaaaacgacggccagtCAAATCAATATGACTGCCGTGT	CATGTGTGTGTGTGTGTGTGTG
2DS_187	Limagrain	TgtaaaacgacggccagtCAAATCAATATGACTGCCGTGT	ACAAGGAGAGGCTGGAGTTTC
2DS_188	Limagrain	TgtaaaacgacggccagtATGGGTAACGGAAGATGTCCTA	TTCGTTGGAACTTTGACGTG
2DS_189	Limagrain	TgtaaaacgacggccagtGGTTGTGCTTTTGCGAGAGT	CCATGTTGATAGGTTTTGACGA
2DS_190	Limagrain	TgtaaaacgacggccagtCTCCAGCGACAGTACAAATACG	TCTTACCATCAGCTCAAAAGCA
2DS_191	Limagrain	TgtaaaacgacggccagtTGTGGCATAGGTCTACAAGTGG	ATCCATATTGGTCAGTTGGCTC
2DS_192	Limagrain	TgtaaaacgacggccagtTAGCAAGAGCGAGAAACACTTG	TCATCACAAATCCCTATTTCCC
2DS_193	Limagrain	TgtaaaacgacggccagtGACCCATCTTCCATCTTCTGG	AGGAGCGAAGCTAGTCAACATC
2DS_194	Limagrain	TgtaaaacgacggccagtATATACACGGTAGAACGGGTCG	TGGTTGCAGAGAAGTAGAGCAA
2DS_195	Limagrain	TgtaaaacgacggccagtTGGTTGCAGAGAAGTAGAGCAA	ATATACACGGTAGAACGGGTCG
2DS_196	Limagrain	TgtaaaacgacggccagtACAAATCGTTTAGGCTGGGTTA	GGCAAGCAACCACTATCTTGA
2DS_197	Limagrain	TgtaaaacgacggccagtAAATCCTACCACAAGATCGCAT	TGTCCTCTCTTATCTCCATGCTT
2DS_198	Limagrain	TgtaaaacgacggccagtTGTCCTCTCTTATCTCCATGCTT	AAATCCTACCACAAGATCGCAT
2DS_199	Limagrain	TgtaaaacgacggccagtTTCAGGACTTGTTGAGTTGTGC	ATAGATTATGTGGCCGGTATGG

A2.9 continued

Marker ID	Contig/source	FW-primer	RV-primer
2DS_200	Limagrain	TgtaaaacgacggccagtTTGTGCCTCTCACTAAAGGAAA	ATGCCTCTCGGAAACAGAAA
2DS_201	Limagrain	TgtaaaacgacggccagtGATCTGCACAAAATTCCATCTG	GCATTCCCCTTCTTTTCTATCC
2DS_202	Limagrain	TgtaaaacgacggccagtAAGTTCATTTGGGATCTGGTTG	GCATGTCTAGTTCTCTGACCCC
2DS_203	Limagrain	TgtaaaacgacggccagtTTTGGTTTGTCTGGAGTTTGTG	TACGTTCCCCTTTAATTTCCCT
2DS_204	Limagrain	TgtaaaacgacggccagtGAAAACATTCAAAGGCTCAAGG	ATTCAACTGGACTTCGTCGTTC
2DS_205	Limagrain	TgtaaaacgacggccagtAGGATGTTGGTTGCAGTTTTCT	AGTTGGAGTACCGCTCGTTTAG
2DS_206	Limagrain	TgtaaaacgacggccagtTGATACCTCCTTCGTTTTCTCC	GAAATCATCCACAACCACAAGA
2DS_207	Limagrain	TgtaaaacgacggccagtCTTTGTGGATGTCTTGAATGGA	CTCTGGTTTGTTTGGTTTTGGT
2DS_208	Limagrain	TgtaaaacgacggccagtAGTGTAGCATTCCATCCCATTC	TACCCAACCAAAACAAGAGGAG
2DS_210	Limagrain	TgtaaaacgacggccagtTTCCTGGTATTGTCTAGGCTCC	AGTGCCCAGTGTTAGTTCCAAT
2DS_211	Limagrain	TgtaaaacgacggccagtGTCAGTCCCTGTTGATGATGAC	TAATACACTCCGAAAAGCCACC
2DS_212	Limagrain	TgtaaaacgacggccagtTGACTGATTTGATGCGAACAC	GGAGAAGATGCAGATGTAGGGT
2DS_213	Limagrain	TgtaaaacgacggccagtGGGCCATCTTGTTTGTGAGTAT	CTTATTTTACATTTGGGTGCCG
2DS_214	Limagrain	TgtaaaacgacggccagtATTTTGGGTTGTGCTAGACGTT	GGTAGATAATTTGGTGCATGAGG
2DS_215	Limagrain	TgtaaaacgacggccagtGATGCACCATATCAAGGTCTCC	AGCCTACCAAATCGAAACACAT
2DS_216	Limagrain	TgtaaaacgacggccagtGTCCCCTATGCTCCTTCTTCTT	AGCCTACCAAATCGAAACACAT
2DS_217	Limagrain	TgtaaaacgacggccagtAGCCTACCAAATCGAAACACAT	GTCCCCTATGCTCCTTCTTCTT
2DS_218	Limagrain	TgtaaaacgacggccagtAGCCTACCAAATCGAAACACAT	GATGCACCATATCAAGGTCTCC
2DS_219	Limagrain	TgtaaaacgacggccagtTGTAGGCAAACACAAAGACACC	ACCAAAGCGACGACATTACTTAC
2DS_220	Limagrain	TgtaaaacgacggccagtTTCATGCAGAGTTCGTTTGTCT	GTCGTCTGAAATGGGATAAAGC
2DS_221	Limagrain	TgtaaaacgacggccagtATTATACTGCCTGTTTGGCTGG	GTTGCTGTTTGTGCTATGAACC
2DS_222	Limagrain	TgtaaaacgacggccagtTACATGGCAAAACACAAAGACC	TAGCAGGTTTCTACGTGATGGA
2DS_223	Limagrain	TgtaaaacgacggccagtTTAAATCACTGCTCTTCGCGT	AGCCTCATCACCAGGAAAATAA
2DS_224	Limagrain	TgtaaaacgacggccagtCTAATGTGTTTGGATGGATTCG	GAGACTTTCTTGGAGGGATGAG
2DS_225	Limagrain	TgtaaaacgacggccagtCGTTGTCACTTTCCTTTACCAA	ACGCAATAAGAAGATCCACGTT
2DS_226	Limagrain	TgtaaaacgacggccagtTTAGAGTGCATGAATGAGACGC	GCTGAAAATACAACGGTGAACA
2DS_227	Limagrain	TgtaaaacgacggccagtATGTGGACGAGTACGTGGATCT	AAGAAAGAGGATCGGGAGAAAC
2DS_228	Limagrain	TgtaaaacgacggccagtAGCTTCAACTTCCCTCACAAAA	ACGTAAACCGAAACAAATCAGG
2DS_229	2D v3.3	TgtaaaacgacggccagtCCCATCTAGGGAGAAGGTCAA	ATTCACCAATTACCGACAGGAC
2DS_230	2D v3.3	TgtaaaacgacggccagtATCAGGTCCCCACTAGAAATCA	AAAATGGGAGATACATGGGTTG
2DS_231	2D v3.3	TgtaaaacgacggccagtTACGGCGCGATATGTGGT	GAACAATCTCTACAACCATGAGCA
2DS_232	2D v3.3	TgtaaaacgacggccagtAAAAGGCTCTAATGGTTGGTGA	CAGACAGGTTTGGGACTAAAGG
2DS_233	2D v3.3	TgtaaaacgacggccagtATATTCACCAAGGAAATCGCTG	TAGTCGTTTCAGACCAAAAGCA
2DS_234	2D v3.3	TgtaaaacgacggccagtCACACTCTCCTCCTCCTCCTC	CCGGAGTATAAAACAAAGCCAC
2DS_235	2D v3.3	TgtaaaacgacggccagtGATCTATTGACGGCGATCCTAC	CTACAACCACAGTCCGAAACAA
2DS_236	2D v3.3	TgtaaaacgacggccagtCTCCCATCAAACTGAGAAAAGG	GGACTTAACTGAGGCAACCACT
2DS_237	2D v3.3	TgtaaaacgacggccagtGGATTCGGAGATAAAATGAGCA	ACTGCCGATAAAAGACAAAAGC
2DS_238	2D v3.3	TgtaaaacgacggccagtTTGGGGTTTGGTACAGCTAACTA	TCTTTGTCGTTAAATGCGACTG
2DS_239	2D v3.3	TgtaaaacgacggccagtAACCTGATCCTTGAGCTAACCA	TTGCTTAATTGTGTCTGTTCCG
2DS_240	2D v3.3	TgtaaaacgacggccagtCCATCAACTCCCTTCGTTCATA	CCCTCCCACCACATTACTTAGA
2DS_241	2D v3.3	TgtaaaacgacggccagtGGAGCTATATTTTCCCTTGCTG	GTGATGAGTGAGTACGACTGGC
2DS_242	2D v3.3	TgtaaaacgacggccagtAATGACCTCCTCGTCGTACACT	TAGTAAAAGGCGTCGATTCTCC
2DS_243	2D v3.3	TgtaaaacgacggccagtAGTTGGAGCAGGTTCATGTTTT	TAGTAAAAGGCGTCGATTCTCC
2DS_244	2D v3.3	TgtaaaacgacggccagtCATCATCTCAGGCTTTCAGCTA	CTGACTCTAACCCCAAACAAGC
2DS_245	2D v3.3	TgtaaaacgacggccagtACATACGCCTCACATCCTTGTT	GGCAGTTTCACCTTTTGCAT
2DS_246	2D v3.3	TgtaaaacgacggccagtCAGAACACAATGACACCACCTT	GGGGAAACAATAACCAACTCAA
2DS_247	2D v3.3	TgtaaaacgacggccagtGCCTGATGCTCTGCTTTTATG	GAAGATATGGGACTCTTGACCG
2DS_248	2D v3.3	TgtaaaacgacggccagtCGTCTTCTCCCAGCTCAGTATAG	TTGTACCAGTCTCATCACAGGC
2DS_249	2D v3.3	TgtaaaacgacggccagtCAATGGGAGAGAGAAACAATACG	AAGTGATGCTGATCGTTGTGAG
2DS_250	2D v3.3	TgtaaaacgacggccagtAGACACGGTTTTGCTTCTGTG	ATCGTAATTGTGCCTCTCGC
2DS_251	2D v3.3	TgtaaaacgacggccagtAGAAGGGAAGCTCTTTTATGGG	GGTGGACATGCTAATTTTCTCC
2DS_252	2D v3.3	TgtaaaacgacggccagtATGTGCAGTGTCTTACTTTCCG	TCTTGGCCTACAAAATGGTCTT
2DS_253	2D v3.3	TgtaaaacgacggccagtCAACAAGCACCCGTTATCCT	TCCCTTCTACTGATTTTCGGAG
2DS_254	2D v3.3	TgtaaaacgacggccagtGGGGTTGAATCGTAACAAAGAA	CTCAAGATGCTATGCCAACTCA
2DS_255	2D v3.3	TgtaaaacgacggccagtAAACAAAACCATGCCTCTCGT	GGTCTTTCCTAGCTCTTCGACA
2DS_540403	IWGSC-2/Ensembl	TgtaaaacgacggccagtAAACCGAGTTACCGATGAATTG	CCCGAGCAAAAGTATGTGTGTA
2DS_256	IWGSC-2/Ensembl	TgtaaaacgacggccagtATTTCTGTAACCCATTTGTCGC	TCAATCCAACCCTGTCATACAA
2DS_257	IWGSC-2/Ensembl	TgtaaaacgacggccagtCTAAGCGAGCCTTCATTTCCTA	ACCCACACATGCCCTTATACAT
2DS_258	IWGSC-2/Ensembl	TgtaaaacgacggccagtCCCGATCTTGTTTTCCTACCTA	GGCTGTCATATCTGGTCTCAAA
2DS_259	IWGSC-2/Ensembl	TgtaaaacgacggccagtTGTACGATAGGACGTTTTCTGC	CTACTCATGGCCCGTTAATTTT
2DS_260	IWGSC-2/Ensembl	TgtaaaacgacggccagtATTTCAGGCAGGCAATTAAGAG	ATGAGCACAATGAAGATGATGC
2DS_261	IWGSC-2/Ensembl	TgtaaaacgacggccagtGGCGACCTCAAACTCTACCATA	AGCCCTAGCCATATCCTCTTTC
2DS_262	IWGSC-2/Ensembl	TgtaaaacgacggccagtTTATGAACCCGCACAGTACATC	AGGCTTACATGAGAAAGGCAAG
2DS_263	IWGSC-2/Ensembl	TgtaaaacgacggccagtAACGCGCTACTGCTAGTTTGAC	AATGCCTACTGCCCTAAACAAA

A2.9 continued

Marker ID	Contig/source	FW-primer	RV-primer
2DS_264	IWGSC-2/Ensembl	TgtaaaacgacggccagtCACAATAGTAGAAGCACTCTGCG	TGATGCGCTCTCTTTGAACTAA
2DS_265	IWGSC-2/Ensembl	TgtaaaacgacggccagtATTTGAAGGGTTAATGAGCCAC	TGAAAACCAGAGAGAACACACA
2DS_266	IWGSC-2/Ensembl	TgtaaaacgacggccagtAAAAGTAGTTGGTGCAAATGGG	ACTTGACTCGGACATGCAAATA
2DS_267	IWGSC-2/Ensembl	TgtaaaacgacggccagtAGTTGCTGGTGAGAGAGTTCGT	CAGATAGAGAGAGGGAGAGGCA
2DS_268	IWGSC-2/Ensembl	TgtaaaacgacggccagtTCATCCTGGTCATCTTTATCCTG	GGTGAAAGCAAATTCATCAAGG
2DS_269	IWGSC-2/Ensembl	TgtaaaacgacggccagtCAGCACAGTAGATAGGCCAACA	TGCTTCAGTTTTGGATGGTATG
2DS_270	IWGSC-2/Ensembl	TgtaaaacgacggccagtACTCTGGCTGTAAGGAAAGGTG	TTGCACGTAAGATTGTCTGGAG
2DS_271	>IAA-amino acid hydrolase ILR1-like protein 8	TgtaaaacgacggccagtCGCGGTATAGGCAAAAGAACTA	CATGAATGGAAATGTGTGTGTG
2DS_272	>IAA-amino acid hydrolase ILR1-like protein 8	TgtaaaacgacggccagtTATCTTTCCACCATCCTTCCAC	ATCTTTGTCCACTTTGATCCCA
2DS_273	IWGSC-2/Ensembl	TgtaaaacgacggccagtCTTGATGGAGGATATGGTGGTT	TAGTGAAATTGCTAGGGCATGA
2DS_274	IWGSC-2/Ensembl	TgtaaaacgacggccagtTGATCGTGACTACATCGACTCC	AAACCTCCTCTCCTTCTCGTCT
2DS_275	IWGSC-2/Ensembl	TgtaaaacgacggccagtCAGAGATGGATGACGTGGACT	TATTGGCAGAACAGAGTGATGG
2DS_276	IWGSC-2/Ensembl	TgtaaaacgacggccagtATCAGAGTGGACAAGCAATGG	ATGCTCTCGCATGAACAGTAAA
2DS_277	IWGSC-2/Ensembl	TgtaaaacgacggccagtCACTGTCGCGGTAGAAGGTC	ACAAGATTGCTTAACGGGTCTG
2DS_278	IWGSC-2/Ensembl	TgtaaaacgacggccagtATAGCGCATGATCCATCTTTCT	ATCCTGACCCAATAACCATGAC
2DS_279	IWGSC-2/Ensembl	TgtaaaacgacggccagtGTTTCCTAGTGGACATGGTTGC	CAGAGGGTGGACCTCATTTAAC
2DS_5375260	BRADI5G05140	TgtaaaacgacggccagtAGATCCCATTGACAGAACGAAT	TTGAAAACGGTGAGCAGTTG
2DS_280	>BRU1	TgtaaaacgacggccagtCCGGACCAACTCTTTTCTTCTA	ATTCCTTTGATCTTGCATGGAG
2DS_281	>BRU1	TgtaaaacgacggccagtTTCCGGTCTCTCTGTCTAGCTC	ATCGAGTAACGCTGCATAAACA
2DS_282	>auxin_response_factor 9-like	TgtaaaacgacggccagtTGGGTAGGAGAGAGAGGGTACA	GTACACAAGCTGTTGACCTTGC
2DS_283	>UDP_glycosyltransferase 92A1	TgtaaaacgacggccagtGATCACAGGGAACAATGCCTAT	AAGCGTTTTGACTAGAAGCCC
2DS_284	>UDP_glycosyltransferase 92A1	TgtaaaacgacggccagtTGATTACACTTCTCTGCCTCCA	GATTCATGTTGCCTTGTGAAGA
2DS_285	>IAA-amino acid hydrolase ILR1-like protein 8	TgtaaaacgacggccagtATATTAACAAAGCCCGCGATAC	CTTCATAATCATCACCGGAGG
zip_1_Ft	2DS_5375260	TgtaaaacgacggccagtTTTATTGGAGAAAGCCAGCCTA	CACCTGATTTGATGTGTTCTGC
zip_2_Ft	2DS_5340329	TgtaaaacgacggccagtCTAGGCCAGGTCATCTCTCTGT	TTATGGTGCCCCTGAATCTAGT
zip_3_Ft	2DS_5340329	TgtaaaacgacggccagtGGGATTGTCTAGGTCTCAGTCG	GTTTTGCCTGGGTTCAGTTAGA
zip_4_Ft	2DS_5366894	TgtaaaacgacggccagtTGCTGGCTTGAGTAGTGAGAAA	ACATGACCCACACAAAACACAT
2DS_283b	2DS_5351773	TgtaaaacgacggccagtTCCTGATCTGTTTGTGTTCGTT	GCAATGTTTATTCAGCAAGACG
2DS_284b	2DS_5351773	TgtaaaacgacggccagtCGTCTTGCTGAATAAACATTGC	TAGTGCTTGGTATCGTTGTTGG
2DS_285b	2DS_5319489	TgtaaaacgacggccagtTAGGTAGGAAAACAAGATCGGG	GAAGCTAATGTGAAGGTGGCTC
2DS_286b	2DS_5388540	TgtaaaacgacggccagtACTAGAGAGGATGCGGAAGTGA	CGCGTCAAGTTTTCTGATCTTT
2DS_287b	2DS_5390725	TgtaaaacgacggccagtGTTCTGTAACCGGCACTAAAGG	TGTGTATGTGTGAATGGTGGTG
2DS_288	2DS_5384527	TgtaaaacgacggccagtACATGAAAACAACGAAGACACG	TGAGTACGGAAGAAGGTCAAGTTT
2DS_289	2DS_5384527	TgtaaaacgacggccagtGCCGTGAGAGATTTTGAGAGAT	GGTGTCTTCTGACTTTGTTTGC
2DS_290	2DS_5323988	TgtaaaacgacggccagtAGTTTCTGCTCCCCTTCCTTAC	GCCATGTCTCTCTTCTTGTCCT
2DS_291	2DS_5352525	TgtaaaacgacggccagtAGAGTGTTTTCGCAAGAGGAAC	TATGAAGAGATGGTTGGTGCTG
2DS_292	2DS_5390004	TgtaaaacgacggccagtGCAAATAAACACCCCACTATAACC	GCAACGGGAGAGAAAACATAAC
2DS_293	2DS_5363870	TgtaaaacgacggccagtACCCTAAGAAAGGACTGCATCA	CAAGGGGTCATAAATTGAGGTT
2DS_294	2DS_5363870	TgtaaaacgacggccagtAGACGCACAGCAGGTTAGTACA	GAATCCATCAGCCTCACTGTC
2DS_295	2DS_5363870	TgtaaaacgacggccagtGCGTACATTGGTTGGAATAACA	AGTTGGTTGTCGCACCTAGTTT
2DS_296	2DS_5366894	TgtaaaacgacggccagtCAATGTTCAGGTTGCAGTTGTT	AAAGACCATGTGAAGGAAAACG
2DS_297	2DS_5341587	TgtaaaacgacggccagtAGATGGGAAAAGGTAAGTGTGG	GGTAGGTATATGTGTGCGCCTT
2DS_298	2DS_5379098	TgtaaaacgacggccagtTCAGATTTATTCGCACTTGCC	ACCTGCAAAACTCAAAAGTTGG
2DS_299	2DS_5347513	TgtaaaacgacggccagtTCTTAGTCCACCCATCTCCATC	TTCGGTGTGCAGAAAAGTCTAA
2DS_300	2DS_5319959	TgtaaaacgacggccagtCCTCGACCCTATGTGATTTTGT	TGCATGGTACTTGTGGCTTATC
2DS_301	2DS_5319959	TgtaaaacgacggccagtTCTTCTTCTTCTTCTTCGTGGG	TGCATGGTACTTGTGGCTTATC
2DS_302	2DS_5319959	TgtaaaacgacggccagtCCCATTGAAAGAAAACCCTACA	CATCATAGGAACGGTCAGGAGT
2DS_303	2DS_5319959	TgtaaaacgacggccagtTCTGAAGAAACACACATCTCCG	TCCTCTAGTGCTCTCTCCGTTT
2DS_304	2DS_5319959	TgtaaaacgacggccagtGCCAACCTAAACAGGTGAAAAG	CCATGTGTGCTTGTTGTGTATG
2DS_305	2DS_5319959	TgtaaaacgacggccagtCCAGTAGCTTAACATCCCTTGG	TATATTTGGGGCGCTTTGTACT
2DS_306	2DS_5319959	TgtaaaacgacggccagtTTGGTACTCACACTGACATGGAC	TTCACTTTTCTTCCTCCTCTGC
2DS_307	2DS_5319959	TgtaaaacgacggccagtAAAGGTGCGAGCTATTTAGTGG	TTTGCTTGGGTAGCGATTAGTT
2DS_308	2DS_5389716	TgtaaaacgacggccagtCTCATAAGAAGACGCCACGAC	AGGGATGTGTGCAAAGGAGTAT
2DS_309	2DS_5389716	TgtaaaacgacggccagtAAACTATACCTTGGGGAGGGG	TCTGCGACACCGTATTAAAAGA
2DS_310	2DS_5390981	TgtaaaacgacggccagtCTCTTCGCCCAAAACCCTAC	GATCTCCCAATAGCAATCAAGC

A2.9 continued

		Marker class			Marker validation			Primers	
Marker	IWGSC contig	2D RIL	gwm	FM	CD (a)	RIL4 (b)	Type	Forward	Reverse
2DS_1	2DS_5359909	7	14	A	372	375	B	GTGTGGAGCCTATCCAAATGA	CCCAATGAACTGCTACATGAGT
2DS_3	2DS_5337443	7	13	A	-	261	A	AAAAGGTAATAGAACCGGAGCC	TGTGATTGGTGAAGATGGAGAG
2DS_4	2DS_5337443	7	13	A	-	292	A	AAAAGGTAATAGAACCGGAGCC	CATTTTCACCCCTATATGTCCG
2DS_6	2DS_5321865	10	14	A	283	279	B	CGACAGAAAACAAACGAGACTG	AGATTGATATGTACCTGCGCGT
2DS_66	2DS_5357871	5	25	F	411	412	B	AATGGATGGGTAATATGGCACT	GCTAGGATGAGGTCCGAGTTTA
2DS_79	2DS_5317970	11	3	-	348&349	348&350	C	ACTTATGGGTTTCAGGCTTTCA	GTGAACACAAATCTTCCGTCAA
2DS_88	2DS_5321770	7	17	A	275&285	282	D	TCTAACCCTTTGGTCTTATCGC	TTCGCGGAAGATGAGTACATAA
2DS_89	2DS_5321770	7	17	A	324 double	324 triple	C	GAAGAATGAGGGAAATACGCAA	TTGACCTTCCATTGTACTGGTG
2DS_94	2DS_5338366	11	3	-	357	362	B	CATGTGGTAAGCAATCCAAACT	CACGCGAGAAGCACAACTAT
2DS_95	2DS_5341322	7	17	A	311 double	311 single	C	AACAAAGCTCGCAGTACATCAA	CCAATCCTTGAAAAGAACCCTA
2DS_105	2DS_5390456	7	23	A	424	422	B	GTACATTTTGCTTCCTTTTCCG	TAAAACCAGTTACATGCGGTGA
2DS_137	2DS_5309868	1	7	-	280	281	B	CTAAGGACCAACAGCGGTATTC	GTGACAAGCATCTATGTGGAGG
2DS_138	2DS_5318296	5	27	F	397	395	B	CGTGCGAGAGATTACAGAGAGA	GGGGTGGTAAAACAAATCAAGA
2DS_145	2DS_5366858	11	3	-	248	249	B	AATAGATACCGATAACGCTCCG	ATGGACAAGCACACACACATCT
2DS_149	2DS_5319467	12	15	A	339	344	B	AGATTGAGCCGACACATCATC	CGGAACGAGAAGAAGAAGCTAA
2DS_15	2DS_5390752	6	25	F	411	410	B	GTACCAACCTTTACGCCCCT	ACCACATCTTCACCCATATTCC
2DS_187	2DS_5354335	9	28	E	286	282	B	CAAATCAATATGACTGCCGTGT	ACAAGGAGAGGCTGGAGTTTC
2DS_192	2DS_5378845	9	28	E	209	232	B	TAGCAAGAGCGAGAAACACTTG	TCATCACAAATCCCTATTTCCC
2DS_201	2DS_5389660	9	28	E	322&323	322&324	C	GATCTGCACAAAATTCCATCTG	GCATTCCCCTTCTTTTCTATCC
2DS_208	2DS_5389857	9	28	E	212	222	B	AGTGTAGCATTCCATCCCATTC	TACCCAACCAAAACAAGAGGAG
2DS_210	2DS_5337059	12	16	A	233	236	B	TTCCTGGTATTGTCTAGGCTCC	AGTGCCCAGTGTTAGTTCCAAT
2DS_211	2DS_5337059	12	16	A	315	321	B	GTCAGTCCCTGTTGATGATGAC	TAATACACTCCGAAAAGCCACC
2DS_212	2DS_5337059	12	16	A	253&267	267	D	TGACTGATTTGATGCGAACAC	GGAGAAGATGCAGATGTAGGGT
2DS_215	2DS_5379317	12	22	A	145&157	144&161&170	D	GATGCACCATATCAAGGTCTCC	AGCCTACCAAATCGAAACACAT
2DS_217	2DS_5319965	12	16	A	127	130&139	D	AGCCTACCAAATCGAAACACAT	GTCCCCTATGCTCCTTCTTCTT
2DS_222	2DS_5342594	7	19	C	354	353	B	TACATGGCAAAACACAAAGACC	TAGCAGGTTTCTACGTGATGGA
2DS_223	2DS_5342594	7	19	C	380	384	B	TTAAATCACTGCTCTTCGCGT	AGCCTCATCACCAGGAAAATAA
2DS_235	2DS_5365907	4	2	-	300 triple	300 double	C	GATCTATTGACGGCGATCCTAC	CTACAACCACAGTCCGAAACAA
2DS_242	2DS_5388557	3	9	-	354	355	B	AATGACCTCCTCGTCGTACACT	TAGTAAAAGGCGTCGATTCTCC
2DS_243	2DS_5388557	3	9	-	260	261	B	AGTTGGAGCAGGTTCATGTTTT	TAGTAAAAGGCGTCGATTCTCC
2DS_26	2DS_5390977	6	26	F	185	198	B	TGAGGGAAAATACAAAGAGGGA	ATGTTAAGTGGAACAGCGTGTG
2DS_275	2DS_5344159	14	29	G	401	399	B	CAGAGATGGATGACGTGGACT	TATTGGCAGAACAGAGTGATGG
2DS_278	2DS_5360680	2	10	-	350	349	B	ATAGCGCATGATCCATCTTTCT	ATCCTGACCCAATAACCATGAC
2DS_280	2DS_5323734	4	12	-	394	-	A	CCGGACCAACTCTTTTCTTCTA	ATTCCTTTGATCTTGCATGGAG
2DS_293	2DS_5363870	8	30	-	370	372	B	ACCCTAAGAAAGGACTGCATCA	CAAGGGGTCATAAATTGAGGTT
2DS_5375260	2DS_5375260	7	18	D	299	284	B	AGATCCCATTGACAGAACGAAT	TTGAAAACGGTGAGCAGTTG
2DS_540403	2DS_540403	7	22	A	226	223	B	AAACCGAGTTACCGATGAATTG	CCCGAGCAAAAGTATGTGTGTA
AX-2	2DS_5389432	1	7	-	379	376	B	GTTAAATGTTATGTGCGGCCAG	TAAAAGGGTTTGATGCCAGGAG
AX-25	2DS_5358861	11	3	-	241	239	B	TCTCAATCTTCAGCACTTTCCA	GCAGCACGTCTCGTATTTACTG
AX-28	2DS_5358861	11	3	-	398 single	398 double	C	TTATCATGTGGTTGTAGGGCTG	AGAAAGAGAGGTGAGCAGTTGG
AX-29	2DS_5358861	11	3	-	356	354	B	TGCACTGAGCGTGGAAAA	ACCAACCACTATACGAGGAGGA
AX-30	2DS_5388293	13	8	-	200	201	B	AAGTACCCATACACATGCCCAG	GGTCAAAATAAACATGCCACCT
AX-4	2DS_5389432	1	7	-	382	381	B	AAACAGCGACAACAGAAAGTGA	TCGTTAGATTTCCCATCTTCGT

A2.10: Details of polymorphic SSR markers and the allele sizes used for genotyping the mapping populations.

Marker	Source	IWGSC contig	2D RIL class	gwm class	FM class	BS code/ gene model	FAM primer (GAAGGTCGACCAAGTTCATGCT) +	VIC primer (GAAGGTCGGAGTCAACGGATT) +	Common primer	Optimal cycle number
76_uni	UniGenes	2DS_5330846	3	11	-	D_comp79633_c0_seq1:189	CGCACTCTACCCATTGAGATACTG	CGCACTCTACCCATTGAGATACTA	TCCCTCAGCTCACTATTTGCA	35
55_uni	UniGenes	2DS_5375260	7	18	C	D_comp16925_c0_seq1:2569	ACAGTACCTCATCCTCGGATTT	ACAGTACCTCATCCTCGGATTG	GTCTATGGCTACCGGTGACA	35
8_uni	UniGenes	2DS_5388088	7	21	A	D_comp211584_c0_seq2:332	AGAATTAAGCTTCCTGCCATTACG	AGAATTAAGCTTCCTGCCATTACA	AAGCCAAAGATTTCTGCGCC	35
BFR_46	v3.3 cDNAs	2DS_5371750	7	25	F	mrna042711	TGTAGTTGCCGTTGGATTGT	TGTAGTTGCCGTTGGATTGC	GGACTTGACAAAAATCCCACTT	40
16_uni	UniGenes	2DS_5364728	7	28	E	D_comp5544514_c0_seq1:82	TCCTTCTCTTTCATTGCCTTTCTT	TCCTTCTCTTTCATTGCCTTTCTC	GAGGGGAGCGAAGGAAAGG	30
21_uni	UniGenes	2DS_5353487	15	6	-	D_comp132046_c0_seq2:176	TGCAACTTATCGGCAAGGAAA	TGCAACTTATCGGCAAGGAAG	TCCATGGAGTTGTGGCTGTC	40
50_uni	UniGenes	2DS_4514573	15	6	-	D_comp132046_c0_seq1:407	TGATAAAAGGGGAGCATATCGT	TGATAAAAGGGGAGCATATCGC	CGAGCAAAGTTTGACCGTGG	35
6iS	iSelect	2DS_5354297	16	5	-	BS00185568	CGAGTTGCAAAAATTGTGGTGTT	CGAGTTGCAAAAATTGTGGTGTG	CATCGCTTTCGTAAAATTGTTCC	30
4iS	iSelect	2DS_5343763	17	2	-	BS00049514	TCAACGGGCTGTGCTTCG	TCAACGGGCTGTGCTTCA	GGTGAACCCCTCCGCACC	35
5iS	iSelect	2DS_5316382	17	2	-	BS00134231	TGGGAAATCTTCACCTTGAGATTC	TGGGAAATCTTCACCTTGAGATTT	AAACCATAGCCTTCAAGAAACAG	30
1_al	2D v3.3 cDNAs	2DS_5338366	18	2	-	mrna009588	TCTTTCCACAGATAGAACTCCAATG	TCTTTCCACAGATAGAACTCCAATC	ATTGAAAGGGGTGCTAAAAGAAG	35
66_uni	UniGenes	2DS_5362023	18	2	-	D_comp4534_c0_seq1:1233	GAGTGGGGCAGCCGGAAT	GAGTGGGGCAGCCGGAAC	GGCATTGCTCTTCTGCGCT	30
27_uni	UniGenes	2DS_5358861	18	4	-	D_comp3280_c0_seq1:727	CAAGTAAAAAGCGCAGGCATATC	CAAGTAAAAAGCGCAGGCATATT	GAAGATGATCCTATATGACCAAATTGA	35
Freq_2	2D v3.3 cDNAs	2DS_5358861	18	4	-	mrna026970	GCGAGCGTCCTCTTTGTC	GCGAGCGTCCTCTTTGTT	AGCTTCATGTTGAATTATTGCTACG	35
52i	iSelect	2BS_4748675	19	20	B	BS00164872	GCCTGAGACCTTGTCCATCG	GCCTGAGACCTTGTCCATCT	CGGCAAGTCGTCACTGCTG	35
11iS	iSelect	2DS_5343186	20	2	-	BS00181365	CGAATTGTTGTTTAGCTCGGAAG	CGAATTGTTGTTTAGCTCGGAAA	GGTTTACCATTAGCCCTTGTGT	35
63_uni	UniGenes	2DS_5359909	20	14	A	D_comp375913_c0_seq1:199	CCACCTTTTGCCTTTCTAATATCTGAC	CCACCTTTTGCCTTTCTAATATCTGAT	AGCAGCCCTATGGTAATGGC	30
72_uni	UniGenes	2DS_5343763	21	2	-	D_comp713_c0_seq1:154	GCGAGGAAGGTTTACCGTTC	GCGAGGAAGGTTTACCGTTT	AGGATTGCCATTTTTGCGAC	35
10iS	iSelect	1BS_3467454	22	1	-	BS00022234	TCTACGATAGCCCAACCCAC	TCTACGATAGCCCAACCCAT	CGAATGCCAGCCAGGTTCTA	35

A2.11: Details of KASP assays for markers within group B in coarse mapping using the 2D RIL population.

Marker	2D RIL group	Source	IWGSC contig	BS code/ gene model	FAM primer (GAAGGTGACCAAGTTCATGCT) +	VIC primer (GAAGGTCGGAGTCAACGGATT) +	Common primer	Optimal cycle number
BFR_11_7BS	C	v3.3 cDNAs	7BS_3121441	mrna084023	TACGTCGGGAGCCTGGAC	TACGTCGGGAGCCTGGAT	CGTGCAATTCGAGCGGAGA	30
91i_7BS	C	iSelect	7BS_3095060	BS00158894	ACCTCGGCATCATCTAGTAGC	ACCTCGGCATCATCTAGTAGT	CAGCATGAGGATTAGTCTGTTGG	35
92i_7BS	C	iSelect	7BS_3153065	BS00157057	TCTGAAGAAGCAGCTCATCATC	TCTGAAGAAGCAGCTCATCATT	CCGAAGTTGAAGATGGTGAACTTG	35
89i_7BS	C	iSelect	7BS_2854988	BS00182691	TGTGGCTCGAGAACTTGTCG	TGTGGCTCGAGAACTTGTCA	GTCTGTAGAAGCCGACACCT	35
90i_7BS	C	iSelect	7BS_3032904	BS00178201	TGAAACGAGTGAATTCAGATGAGC	TGAAACGAGTGAATTCAGATGAGT	CGTGGTTCAGCATCTTGTCT	35
14i_7BS	C	iSelect	7BS_196497	BS00138407	CCCTCAACCTCCGCTCCG	CCCTCAACCTCCGCTCCA	CGTCAGAACCACCGTGACAG	35
15i_7BS	C	iSelect	7BS_3032904	BS00079019	TGTTAACTTGAACTGGTCCGC	TGTTAACTTGAACTGGTCCGT	GTGCCAATTGAATTTGGGCG	35
17i_7BS	C	iSelect	7BS_3089157	BS00139230	AGCTAAACTGGTTCGGGCTG	AGCTAAACTGGTTCGGGCTC	TCACTGGCAGCAGATTTCCA	35
50i_5B	C	iSelect	5B	BS00035631	CTTCTCGCTCAAAACCACGCAAT	CTTCTCGCTCAAAACCACGCAAC	CGGCCGGAGAGGGTCGCTT	35
BFR_44	A	v3.3 cDNAs	2AS_5247752	mrna034307	GGTAACCAAACCTTTTAGATGTGCC	GGTAACCAAACCTTTTAGATGTGCT	AAATGCTGCGACCCCAGC	30
BFR_34	A	v3.3 cDNAs	2AS_2899441	mrna071136	ACATTTGCAACAGTTGATAAATCCG	ACATTTGCAACAGTTGATAAATCCA	GCAGGTTGGAGTATGTATAGACAAC	30
BFR_3	A	v3.3 cDNAs	2DS_5390412	mrna083050	GTTGTTCAGGATGCCAAGCA	GTTGTTCAGGATGCCAAGCG	CCATGCAAATTGGACACCGA	30
BFR_28	A	v3.3 cDNAs	2AS_5223115	mrna003675	GGAAGCCATCAGTTATGAAGATTTC	GGAAGCCATCAGTTATGAAGATTTG	CAGCTCGAGTGAAGGCCATG	30
BFR_31	A	v3.3 cDNAs	2AS_5188575	mrna041371	GCCGGGAAGGTGAACAGT	GCCGGGAAGGTGAACAGC	TCCTTGAGCACGAGCCGC	30
BFR_38	A	v3.3 cDNAs	2AS_5264433	mrna120229	CGCTTCAGAAAACAACACCA	CGCTTCAGAAAACAACACCG	CGCTGCCTATCACGAAGACT	30
BFR_35	A	v3.3 cDNAs	2AS_5265661	mrna083051	GTTGTTCAGGATGCCAAGCA	GTTGTTCAGGATGCCAAGCG	CCATGCAAATTGGACACCGA	30
BFR_41	A	v3.3 cDNAs	2AS_5265661	mrna035372	AGTAGGACCCGACGTAAGATAA	AGTAGGACCCGACGTAAGATAG	GGTCATGTAATGGCAAAAGGC	30
BFR_42	A	v3.3 cDNAs	2AS_5265661	mrna035372	CAGTGGGCAGTTTGTAGGATTA	CAGTGGGCAGTTTGTAGGATTG	AGTCCATGGTCTTCTTGAAATCA	30
BFR_47	A	v3.3 cDNAs	2AS_5207752	mrna070131	CAGCAGAGAAGCTTCAGCG	CAGCAGAGAAGCTTCAGCA	TCATCCTTCTTTTGGGCCAC	30
BFR_49	A	v3.3 cDNAs	2AS_5264433	mrna051701	GCACCTGTACTTGCCGTCA	GCACCTGTACTTGCCGTCG	CGAAGAGGCTGTGGAGTTT	30
BFR_4	A	v3.3 cDNAs	2AS_5237913	mrna098108	GACAAGTATGACAGGCCAAGT	GACAAGTATGACAGGCCAAGC	GGCGATATGAACTGTCATCCT	30

A2.12: Details of KASP assays for markers outside group B in coarse mapping using the 2D RIL population.

A2.13: R script used for identifying differentially expressed genes.

```
####R Script for RPKM analysis
##Two key steps
##1) rounding for RPKM cutoff of 0.1 and Log 2 transformation
##2) identify DEGs using a 1.5-fold difference
####
##Ania Kowalski
##Updated July 2015
####

rm(list=ls()) ###clears any previous workspace/data
##reading in data

##UNIGENES FIRST
setwd("E:/PhD/RNA Seq/RPKM")
uni.r<-read.csv("unigenes_raw.csv",as.is=T)

############
##searching for unigenes with 1.5-fold expression difference or higher
##first need to filter out the lowly expressed genes, where a two-fold difference
##will be occurring even at low expression - use anything under 0.1.
##This is the most conservative threshold identified by Warden et al., 2013.
##then log transform the data to avoid division by zeros
##need to keep the zero expression unigenes for the comparison,
##because zero expression in one dataset could be highly expressed in the other.
##therefore change the zeros to a small number.
##can use histograms to check the distribution of the expression of genes, which should be
normal

##selecting data, filter for expression higher than 0.1 in either of the parent columns
uni.sel<-((uni.r["CD_S"]>0.1)&(uni.r["RIL4_S"]>0.1))  ##vertical line is 'or' symbol
##subsetting data based on the selection criteria from above
uni.subset<-uni.r[uni.sel,]

##Writing a for loop for log transformation because logs of multiple columns doesn't work

uni.log<-uni.subset
for (i in 19:22)
  uni.log[,i]<-log2(uni.subset[,i])
##writing csv as log2 values
write.csv(uni.log, file="E:/PhD/RNA Seq/RPKM/log2_uni.csv")
##selecting for higher than 1.5 fold expression
##subtract rather than divide because we have logs

###############################################
##differences between parents  have 1.5-fold difference
##this is logs so this is fold difference
uni.exP<-abs(uni.log[,"CD_S"]-uni.log[,"RIL4_S"])>1.5

sum(uni.exp,na.rm=T) #1735 with threshold 1.5

#exporting the data into a spreadsheet
write.csv(uni.log[uni.exp,], file="E:/PhD/RNA Seq/RPKM/log2_15fold_uni.csv")
deg<-read.csv("log2_15fold_uni_labelled.csv",as.is=T)

##GRAPHING
library("ggplot2")
library(reshape2)
```

```
 deg$category<-factor(deg$category, levels=c("all","2DS","2BS","2AS"), labels=c("rest of
genome","2DS", "2BS","2AS"))
 degs<-
 ggplot(data=deg,aes(x=RIL4_S, y=CD_S, colour=category, shape=category)) +
 geom_point(aes(size=category, alpha=category)) +
 xlab("Short NIL") +
 ylab("Tall NIL") +
 coord_cartesian(ylim = c(-4, 8), xlim=c(-4,8)) +
 scale_shape_manual(values=c(21,19,17,15),
                    breaks = c("rest of genome","2DS", "2BS","2AS"),
                    labels = c("rest of genome","2DS", "2BS","2AS")) +
 scale_colour_manual(values=c("#009E73","black","red","blue"),
                     breaks = c("rest of genome","2DS", "2BS","2AS"),
                     labels = c("rest of genome","2DS", "2BS","2AS")) +
 scale_size_manual(values=c(3,7,5,5)) +
 scale_alpha_manual(values=c(0.7,1,1,1)) +
 theme(panel.grid.minor.x=element_blank(), panel.grid.major.x=element_blank(),
       axis.title.y=element_text(vjust=1.5),
       plot.title = element_text(lineheight=.4, face="bold"),
       axis.title = element_text(size=28, face="bold", colour="black"),
       axis.text.y = element_text(size=24, colour="black"),
       axis.text.x = element_text(vjust=0.5, size=24,colour="black"),
       #strip.text.y=element_text(size =26, face="bold"),
       #strip.background=element_rect(colour = "black"),
       legend.text= element_text(size = 24),
       legend.title=element_blank(),
       legend.direction = "horizontal",
       legend.position=c(0.5,0.95)) +
 guides(fill=guide_legend(title=NULL))

##plotting
png("unigenes_exp.png", width=3000, height=3000, res=300)
plot(degs)
dev.off()

  ##v3.3 cDNAs, with same parameters, but overlap between the same expression in parent
NILs and bulks.
setwd("E:/PhD/RNA Seq/RPKM")
cdna<-read.csv("RPKM_cDNA_bowtie.csv",as.is=T)
#############
##selecting data, filter for expression higher than 0.1 in either of the parent columns
cdna.sel<-
((cdna["bowtie2_S24_RIL4_Spike_short"]>0.1)&(cdna["bowtie2_S13_CD_Spike_tall"]>0.1)&(cd
na["bowtie2_S246_bulk_Spike_tall"]>0.1)&(cdna["bowtie2_S135_bulk_Spike_short"]>0.1))
##vertical line is 'or' symbol

##subsetting data based on the selection criteria from abovebowtie2_S135_bulk_Spike_short
cdna.subset<-cdna[cdna.sel,]

##Writing a for loop for log transformation because logs of multiple columns doesn't work

cdna.log<-cdna.subset
for (i in 4:7)
  cdna.log[,i]<-log2(cdna.subset[,i])
##writing csv as log2 values
write.csv(cdna.log, file="E:/PhD/RNA Seq/RPKM/log2_cdna.csv")
##selecting for higher than 1.5 fold expression
##subtract rather than divide because we have logs

###############################################
##differences between parents and bulks have 1.5-fold difference
##this is logs so this is fold difference
```

```
cdna.exP<-(abs(cdna.log[,"bowtie2_S24_RIL4_Spike_short"]-
cdna.log[,"bowtie2_S13_CD_Spike_tall"])>1.5)&(abs(cdna.log[,"bowtie2_S135_bulk_Spike_sho
rt"]-cdna.log[,"bowtie2_S246_bulk_Spike_tall"])>1.5)
sum(cdna.exp,na.rm=T) #278 with threshold 1.5 common to both datasets

#exporting the data into a spreadsheet

write.csv(cdna.log[cdna.exp,], file="E:/PhD/RNA
Seq/RPKM/log2_15fold_cdna_parent_bulk.csv")

###expression skew is same way i.e. upreg in short across both or downregulated in short
across both
common<-read.csv("log2_15fold_cdna_parent_bulk.csv",as.is=T)

cdna.skew<-
(((common[,"bowtie2_S24_RIL4_Spike_short"]>common[,"bowtie2_S13_CD_Spike_tall"])&(co
mmon[,"bowtie2_S135_bulk_Spike_short"]>common[,"bowtie2_S246_bulk_Spike_tall"]))|((com
mon[,"bowtie2_S24_RIL4_Spike_short"]<common[,"bowtie2_S13_CD_Spike_tall"])&(common[,"
bowtie2_S135_bulk_Spike_short"]<common[,"bowtie2_S246_bulk_Spike_tall"])))
sum(cdna.skew)
##only 20 DEGs identified with skew in same direction across parent NIL and bulk
write.csv(common[cdna.skew,], file="E:/PhD/RNA
Seq/RPKM/log2_15fold_cdna_bias_parent_tall.csv")
```

Appendix to Chapter 3

	Church Farm									Reading			Lleida	
Yield (t DM/ha)	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N1)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)	2013	2014
par	8.97	8.74	8.80	8.53	12.24	11.05	11.25	7.39	9.53	5.70	7.25	9.16	6.78	6.27
Rht8	8.10	8.05	8.08	6.93	11.99	9.94	10.53	6.75	9.10	5.72	6.97	8.75	8.06	7.35
tall	8.99	8.35	8.59	9.34	11.90	10.80	10.33	7.25	9.00	5.47	7.39	9.22	7.35	6.35
P-value	*	*		*		***		**		*			NS (P=0.07)	
L.S.D.	0.64	0.50		0.72		0.67		0.82		0.39			1.46	
Rht8 (% of tall)	90	96	94	74	101	92	102	93	101	105	94	95	110	116
difference (%)	**-10**	**-4**	**-6**	**-26**	**1**	**-8**	**2**	**-7**	**1**	**5**	**-6**	**-5**	**10**	**16**

	Church Farm									Reading			Lleida	
Grains m^{-2}	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N1)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)	2013	2014
par	23758	19671	19865	21195	25187	19988	22768	13443	16420	11640	14964	18709	20241	16365
Rht8	22633	17185	17314	18231	23709	17284	20563	13032	16089	11770	13955	17775	24175	19328
tall	24330	18550	19484	22863	23951	18061	22019	13480	14498	10837	14908	18496	23711	17596
P-value	NS	**		*		*		**		*			NS	
L.S.D.	1465	1296		2132		2095		1500		883			4643	
Rht8 (% of tall)	93	93	89	80	99	96	93	97	111	109	94	96	102	110
difference (%)	**-7**	**-7**	**-11**	**-20**	**-1**	**-4**	**-7**	**-3**	**11**	**9**	**-6**	**-4**	**2**	**10**

	Church Farm									Reading			Lleida	
TGW (g)	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N1)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)	2013	2014
par	37.72	44.52	44.33	40.19	48.56	48.79	43.84	44.38	51.06	48.96	48.48	48.96	33.49	38.28
Rht8	35.81	46.90	46.74	38.17	50.59	49.80	44.11	45.26	49.78	48.56	50.00	49.20	33.44	37.99
tall	36.97	45.06	44.05	41.07	49.72	51.37	41.05	45.29	51.25	50.48	49.60	49.84	30.90	36.03
P-value	NS	*		NS		NS		NS		*			*	
L.S.D.	3.40	1.79		3.5		4.14		4.06		1.46			2.51	
Rht8 (% of tall)	97	104	106	93	102	97	107	100	97	96	101	99	108	105
difference (%)	**-3**	**4**	**6**	**-7**	**2**	**-3**	**7**	**0**	**-3**	**-4**	**1**	**-1**	**8**	**5**

	Church Farm									Reading		
Grain area (mm^2)	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N1)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)
par	17.27	20.30	20.26	19.52	21.06	22.29	20.99	21.05	22.8	21.59	21.50	21.42
Rht8	17.20	21.15	21.07	19.48	21.52	22.55	21.36	21.22	22.18	21.52	21.14	21.99
tall	17.13	20.60	20.49	20.07	21.54	22.89	20.79	21.07	22.86	21.83	21.53	21.64
P-value	NS	**		NS		NS		NS		NS		
L.S.D.	0.73	0.43		0.78		5.40		1.06		1.46		
Rht8 (% of tall)	100	103	103	97	100	99	103	101	97	99	98	102
difference (%)	**0**	**3**	**3**	**-3**	**0**	**-1**	**3**	**1**	**-3**	**-1**	**-2**	**2**

	Church Farm									Reading			Lleida	
Harvest index	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N1)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)	2013	2014
par	0.63	0.63	0.66	0.64	0.64	0.41	0.46	0.44	0.37	0.44	0.43	0.42	0.32	0.38
Rht8	0.63	0.65	0.66	0.62	0.65	0.40	0.41	0.42	0.38	0.41	0.42	0.41	0.36	0.37
tall	0.61	0.62	0.65	0.62	0.63	0.43	0.45	0.44	0.39	0.45	0.42	0.42	0.35	0.36
P-value	NS	*		NS		*		NS		NS			NS	
L.S.D.	0.04	0.02		0.013		0.05		0.04		0.03			0.04	
Rht8 (% of tall)	103	105	102	100	103	95	91	97	98	91	100	98	103	103
difference (%)	**3**	**5**	**2**	**0**	**3**	**-5**	**-9**	**-3**	**-2**	**-9**	**0**	**-2**	**3**	**3**

	Reading			Lleida	
Spikes/m^2	2014 (N1)	2014 (N2)	2014 (N3)	2013	2014
par	417	419	475	717	411
Rht8	300	391	383	738	516
tall	377	403	487	626	500
P-value	***			NS	
L.S.D.	66.0			187	
Rht8 (% of tall)	80	97	79	118	103
difference (%)	**-20**	**-3**	**-21**	**18**	**3**

A3.1: Yield components. Data shown as mean values. N1=40 kg N ha^{-1}, N2=100 kg N ha^{-1}, N3=200 kg N ha^{-1}, I = irrigated, UI = unirrigated (rainfed). UI is only indicated where there was a contrasting irrigation regime. The p-value refers to significant differences in height between genotypes within each experiment determined by the least significant difference (L.S.D.) test. * P<0.05, **P<0.01, ***P<0.001.

	Church Farm									Reading		
Grain length (mm)	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N1)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)
par	6.20	6.74	6.71	6.63	6.76	6.92	6.80	6.757	6.956	6.81	6.84	6.78
Rht8	6.17	6.89	6.85	6.66	6.82	6.97	6.87	6.843	6.959	6.84	6.71	6.89
tall	6.20	6.82	6.82	6.79	6.89	7.05	6.87	6.819	7.042	6.90	6.85	6.90
P-value	NS	***		NS		NS		NS		*		
L.S.D.	0.15	0.04		0.11		0.14		0.16		0.09		
Rht8 (% of tall)	99	101	100	98	99	99	100	100	99	99	98	100
difference (%)	**-1**	**1**	**0**	**-2**	**-1**	**-1**	**0**	**0**	**-1**	**-1**	**-2**	**0**

	Church Farm									Reading		
Grain width (mm)	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N1)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)
par	3.37	3.71	3.68	3.58	3.83	4.00	3.82	3.82	4.07	3.94	3.87	3.91
Rht8	3.30	3.79	3.79	3.55	3.87	4.03	3.84	3.81	3.95	3.87	3.90	3.94
tall	3.20	3.71	3.65	3.60	3.85	4.03	3.74	3.79	4.02	3.90	3.87	3.86
P-value	**	*		NS		NS		NS		NS		
L.S.D.	0.08	0.08		0.10		0.13		0.14		0.11		
Rht8 (% of tall)	103	102	104	99	101	100	103	101	98	99	101	102
difference (%)	**3**	**2**	**4**	**-1**	**1**	**0**	**3**	**1**	**-2**	**-1**	**1**	**2**

	Church Farm									Reading		
Spikelets spike^{-1}	2012	2013 (N3)	2013 (N2)	2013 (UI)	2013 (I)	2014 (N3 UI)	2014 (N3 I)	2014 (N1)	2014 (N2)	2014 (N1)	2014 (N2)	2014 (N3)
par	24.11	23.11	20.94	22.83	22.61	24.33	23.22	23.00	23.11	23.56	23.22	22.89
Rht8	25.00	23.00	20.89	23.22	22.78	23.44	25.44	22.11	22.56	24.89	23.22	23.56
tall	23.89	22.22	21	22.22	23.33	23.11	22.89	22.00	22.67	22.89	23.33	22.67
P-value	NS	NS		NS		NS		NS		NS		
L.S.D.	1.02	1.46		1.57		2.10		2.00		1.69		
Rht8 (% of tall)	105	104	99	105	98	101	111	101	100	109	100	104
difference (%)	**5**	**4**	**-1**	**5**	**-2**	**1**	**11**	**0**	**0**	**9**	**0**	**4**

A3.1 (continued)

Design: RCBD with three replicates
Analysis: Two-way ANOVA with randomised blocks
Lleida 2013 & 2014

ANOVA	Yield (t DM/ha)				TGW (g)				Height (cm)				Grains/m^2				Grains/spike				Harvest index				Spikes/m^2			
	d.f.	F-test	p-value		F-test	p-value			F-test	p-value			F-test	p-value			F-test	p-value			F-test	p-value			F-test	p-value		
year	1	4.32	0.401		62.25	0.007			1.95	0.235			0.75	0.059			0.75	0.436			7.86	0.145			16.67	0.015		
allele	2	3.91	0.072		6.53	0.024			20.78	<.001			0.53	0.103			0.53	0.607			1.55	0.223			0.63	0.481		
year.allele	2	0.17	0.856		0.07	0.933			16.31	0.002			2.7	0.741			2.7	0.127			2.67	0.098			0.95	0.348		
Means	**par**	**rht8**	**tall**		**par**	**rht8**	**tall**		**par**	**rht8**	**tall**		**par**	**rht8**	**tall**		**par**	**rht8**	**tall**		**par**	**rht8**	**tall**		**par**	**rht8**	**tall**	
2013	6.78	8.06	7.35		33.49	33.44	30.90		111.0	93.3	119.0		20241	24175	23711		28.6	33.2	39.0		0.32	0.36	0.35		717	738	626	
2014	6.27	7.35	6.35		38.28	37.99	36.03		119.0	114.7	115.3		16365	19328	17596		40.3	37.7	36.2		0.38	0.37	0.36		411	516	500	
mean S.E.D.		0.943				1.30				6.6				2504				6.3				0.02				86		
L.S.D. year.allele		2.202				2.88				16.7				5610				14.5				0.04				187		
allele		1.458				2.51				7.1				4643				10.2				0.04				187		

A3.2: ANOVA and experimental details of Lleida data. Shaded values are significant at the 95% confidence level.

Design: RCBD with three replicates
Analysis: One-way ANOVA with randomised blocks
Site: Church Farm 2012

Yield components

ANOVA	Yield (t DM/ha)			TGW (g)			Grains/m^2			Spikelets/ear			Grain area (mm^2)			Grain width (mm)			Grain length (mm)			Harvest index		
	d.f.	F-test	p-value	F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value	
allele	2	9.61	0.03	1.23	0.384		5.35	0.07		5.09	0.08		0.13	0.882		19	0.009		0.25	0.79		1.67	0.30	
Means	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**
	8.97	8.101	8.994	37.72	35.81	36.97	23758	22633	24330	24.11	25.00	23.89	17.27	17.20	17.13	3.37	3.30	3.20	6.20	6.17	6.20	0.63	0.63	0.61
mean S.E.D.	0.23			1.227			528			0.37			0.26			0.03			0.05			0.01		
L.S.D.	0.64			3.406			1465			1.02			0.73			0.08			0.15			0.04		

Height components

ANOVA	Height of plots (cm)			Height of tiller (cm)			Ear length (cm)			Peduncle (cm)			Internode-1 (cm)			Internode-2 (cm)		
	d.f.	F-test	p-value	F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value	
allele	2	5.24	0.076	8.52	0.036		2.16	0.23		9.19	0.032		6.66	0.053		2.61	0.188	
Means	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**
	102.0	99.0	108.0	93.4	87.8	99.0	11.2	10.3	11.5	29.3	28.3	32.7	19.3	17.8	21.2	13.0	13.1	13.7
mean S.E.D.	2.7			2.7			0.6			1.1			0.9			0.3		
L.S.D.	7.6			7.6			1.7			2.9			2.6			0.9		

A3.3: ANOVA and experimental details of Church Farm 2012 data. Shaded values are significant at the 95% confidence level.

A3.4: ANOVA and experimental details of Church Farm 2013 Nitrogen experiment. Shaded values are significant at the 95% confidence level. N2/N3=100/200 kg N ha^{-1}.

Design: RCBD with three replicates per N treatment (100 and 200kg N/ha)
Analysis: Two-way ANOVA with randomised blocks
Site: Church Farm 2013

Yield components

ANOVA	Yield (t DM/ha)			TGW (g)		Grains/m^2		Spikelets/ear		Grain area (mm^2)		Grain width (mm)		Grain length (mm)		Harvest index	
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
N treatment	1	0.12	0.745	4.11	0.113	0.35	0.585	11.67	0.027	0.86	0.407	7.56	0.061	1.21	0.333	16.47	0.015
allele	2	5.42	0.032	6.01	0.025	10.59	0.006	0.24	0.791	10.13	0.006	6.57	0.021	37.48	<.001	9.89	0.007
N treatment.allele	2	0.15	0.866	0.2	0.826	0.32	0.737	0.35	0.716	0.02	0.981	0.45	0.653	0.69	0.528	0.41	0.68

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N2	8.80	8.08	8.59	44.33	46.74	44.05	19865	17314	19484	20.94	20.89	21.00	20.26	21.07	20.49	3.68	3.79	3.65	6.71	6.85	6.82	0.66	0.66	0.65
N3	8.74	8.05	8.35	44.52	46.90	45.06	19671	17185	18550	23.11	23.00	22.22	20.30	21.15	20.60	3.71	3.79	3.71	6.74	6.89	6.82	0.63	0.65	0.62
mean S.E.D.		0.40			0.92			959			0.91			0.23			0.04			0.03			0.008	
L.S.D. allele		0.50			1.79			1296			1.46			0.43			0.08			0.04			0.016	

Height components

ANOVA	Height of plots (cm)			Height of tiller (cm)		Ear length (cm)		Peduncle (cm)		Internode-1 (cm)		Internode-2 (cm)	
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
N treatment	1	9.39	0.038	2.38	0.198	0.19	0.689	8.62	0.043	2.9	0.164	9.4	0.037
allele	2	58.96	<.001	71.45	<.001	25.03	<.001	85.8	<.001	17.81	0.001	10.6	0.006
N treatment.allele	2	0.52	0.616	2.19	0.174	0.15	0.866	2.55	0.139	0.38	0.698	1.23	0.341

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N2	103.7	94.3	107.3	102.2	92.1	104.7	10.7	8.8	10.7	40.9	36.8	44.0	22.0	20.7	23.0	14.7	12.6	14.1
N3	107.7	98.3	113.7	103.2	94.1	110.7	10.6	8.8	10.4	39.2	36.4	41.6	22.8	21.4	24.5	15.7	13.7	16.6
mean S.E.D.		2.2			2.4			0.4			0.8			0.8			0.8	
L.S.D. allele		3.2			2.9			0.7			1.1			1.1			1.2	

Developmental traits

ANOVA	Heading (^{0}C d)			Anthesis (^{0}C d)	
	d.f.	F-test	p-value	F-test	p-value
N treatment	1	2.57	0.184	4.22	0.109
allele	2	4.97	0.039	6.69	0.02
N treatment.allele	2	1	0.41	0.58	0.583

Means	par	rht8	tall	par	rht8	tall
N2	1498	1505	1515	1581	1592	1608
N3	1471	1493	1488	1564	1574	1579
mean S.E.D.		15			12	
L.S.D. allele		13			13	

Yield components

ANOVA	Yield (t DM/ha)			TGW (g)		Grains/m^2		Spikelets/ear		Grain area (mm^2)		Grain width (mm)		Grain length (mm)		Harvest index	
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
treatment	1	152.32	<.001	220.15	<.001	35.15	0.004	0.07	0.807	146.1	<.001	124.93	<.001	60.78	0.001	12.11	0.025
allele	2	7.7	0.014	0.3	0.749	4.26	0.054	0.09	0.912	1.15	0.365	0.13	0.884	4.26	0.055	2.89	0.113
treatment.allele	2	8.01	0.012	1.11	0.375	2.92	0.112	0.76	0.498	0.42	0.674	0.41	0.679	0.13	0.877	4.89	0.041

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
dry	8.53	6.93	9.34	40.19	38.17	41.07	21195	18231	22863	22.83	23.22	22.22	19.52	19.48	20.07	3.58	3.55	3.60	6.63	6.66	6.79	0.64	0.62	0.62
irrigated	12.24	11.99	11.90	48.56	50.59	49.72	25187	23709	23951	22.61	22.78	23.33	21.06	21.52	21.54	3.83	3.87	3.85	6.76	6.82	6.89	0.64	0.65	0.63
average S.E.D.	0.47			1.87			1221			0.97			0.42			0.06			0.06			0.008		
L.S.D. allele	0.72			3.5			2132			1.57			0.78			0.10			0.11			0.013		

Height components

ANOVA	Height of plots (cm)			Height of tiller (cm)			Ear length (cm)			Peduncle (cm)			Internode-1 (cm)			Internode-2 (cm)		
	d.f.	F-test	p-value	F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value	
treatment	1	5.14	0.086	4.19	0.11		0.8	0.421		0.42	0.551		0.78	0.428		0.55	0.498	
allele	2	52.32	<.001	64.65	<.001		20.51	<.001		36.56	<.001		0.34	0.723		0.04	0.961	
treatment.allele	2	0.27	0.769	0.21	0.814		0.12	0.887		0.15	0.862		0.77	0.495		1.49	0.281	

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
dry	106.8	97.3	112.7	102.7	93.7	108.9	10.7	9.1	11.2	39.5	34.4	42.1	21.5	24.7	22.7	15.3	18.8	15.6
irrigated	108.7	100.7	114.0	107.6	97.2	112.1	10.7	8.8	10.8	39.2	35.0	42.6	21.2	20.6	23.0	16.2	13.8	16.4
average S.E.D.	1.9			2.4			0.5			1.1			2.7			2.7		
L.S.D. allele	3.3			2.1			0.8			2.1			4.5			4.6		

Developmental traits

ANOVA	Heading (^{0}C d)			Anthesis (^{0}C d)	
	d.f.	F-test	p-value	F-test	p-value
treatment	1	0.21	0.674	0	0.965
allele	2	5.84	0.027	5.86	0.027
treatment.allele	2	0.97	0.42	0.69	0.527

Means	par	rht8	tall	par	rht8	tall
dry	1456	1478	1482	1551	1547	1569
irrigated	1458	1465	1471	1547	1556	1566
mean S.E.D.	17			12		
L.S.D. allele	14			14		

A3.5: ANOVA and experimental details of Church Farm 2013 Irrigation experiment. Shaded values are significant at the 95% confidence level.

A3.6: ANOVA and experimental details of Church Farm 2014 Nitrogen experiment. Shaded values are significant at the 95% confidence level.

Design: RCBD with three replicates per N treatment (40, 100 and 200 kg N/Ha)
Analysis: Two-way ANOVA with randomised blocks
Site: Church Farm 2014

Yield components

ANOVA	Yield (t DM/ha)			TGW (g)		Grains/m^2		Spikelets/ear		Grain area (mm^2)		Grain width (mm)		Grain length (mm)		Harvest index		Infertile spikelets	
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
treatment	2	79.68	<.001	13.53	0.006	49.28	<.001	2.07	0.208	15.39	0.004	14.93	0.005	11.99	0.008	13.53	0.028	8.04	0.02
allele	2	6.32	0.006	0.75	0.491	8.58	0.005	1.9	0.192	0.58	0.575	0.43	0.657	2.21	0.152	0.75	0.287	1.19	0.338
treatment.allele	4	1.06	0.396	0.38	0.821	4.6	0.017	0.11	0.976	0.71	0.599	0.83	0.529	0.38	0.817	0.38	0.861	0.97	0.46

Means	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**
N1	7.39	6.75	7.25	44.38	45.26	45.29	13443	13032	13480	23.00	22.11	22.00	21.05	21.22	21.07	3.82	3.81	3.79	6.757	6.843	6.819	0.438	0.42	0.44	3.11	2.56	3.44
N2	9.53	9.10	9.00	51.06	49.78	51.25	16420	16089	14498	23.11	22.56	22.67	22.8	22.18	22.86	4.07	3.95	4.02	6.956	6.959	7.042	0.374	0.38	0.39	1.22	1.56	1.39
N3	11.05	9.94	10.80	48.79	49.8	51.37	19988	17284	18061	24.33	23.44	23.11	22.29	22.55	22.89	4.00	4.03	4.03	6.922	6.969	7.053	0.415	0.4	0.43	2.22	1.11	1.78
mean S.E.D.		0.40			1.93			704			0.95			0.51			0.07			0.08			0.02			0.65	
L.S.D. allele		0.82			4.06			1500			2.00			1.06			0.14			0.16			0.04			1.37	

Height components

ANOVA	Height of tiller (cm)			Ear length (cm)		Peduncle (cm)		Internode-1 (cm)		Internode-2 (cm)	
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
treatment	2	15.55	0.004	18.42	0.003	2.92	0.13	4.4	0.067	23.46	0.001
allele	2	85.89	<.001	82.76	<.001	48.01	<.001	11.65	0.002	12.25	0.001
treatment.allele	4	0.48	0.752	1.11	0.397	1.33	0.313	0.92	0.485	1.61	0.235

Means	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**
N1	100.9	92.5	106.4	10.0	8.1	9.9	39.1	33.1	39.6	24.5	22.0	25.9	14.0	14.4	14.8
N2	102.8	96.5	110.5	12.0	9.2	11.2	41.1	34.4	42.0	22.3	21.2	25.1	14.8	15.3	16.0
N3	111.2	102.8	115.6	11.9	9.8	11.5	40.7	36.8	40.8	20.8	21.7	25.8	16.4	15.9	17.8
mean S.E.D.		2.3			0.4			1.4			1.3			0.5	
L.S.D. allele		5.0			0.9			2.9			2.8			1.0	

Developmental traits

ANOVA	Heading (^{0}C d)			Senescence (^{0}C d)		Lodging score (0-100)			Ground cover (%)		
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value		d.f.	F-test	p-value
treatment	2	0.19	0.832	32.94	<.001	55.94	<.001	**allele**	2	0.2	0.821
allele	2	11.34	<.001	3.79	0.037	23.77	<.001				
treatment.allele	4	0.04	0.997	3.16	0.032	6.52	0.001				

Means	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**	**par**	**rht8**	**tall**
N1	1766	1779	1782	1.5	2	1.8	0	0	0	26.9	26.3	27.0
N2	1769	1782	1782	0.3	0.4	0	34	0	58			
N3	1766	1779	1782	0	0	0	66	48	90			
mean S.E.D.		6			0.3			9			1.1	
L.S.D. allele		12			0.6			19			2.4	

ANOVA		Yield (t DM/ha)			TGW (g)			Grains/m^2			Spikelets/ear			Grain area (mm^2)			Grain width (mm)			Grain length (mm)			Harvest index			Infertile spikelets	
	d.f.	Wald	p-value		Wald	p-value		Wald	p-value		Wald	p-value		Wald	p-value		Wald	p-value		Wald	p-value		Wald	p-value		Wald	p-value
N	1	581.46	<0.001		0.39	0.536		258.68	<0.001		9.47	0.005		2.13	0.157		6.13	0.021		10.58	0.003		3.03	0.095		17.11	<0.001
water	1	4.81	0.033		9.25	0.006		18.9	<0.001		0.24	0.628		2.33	0.14		7.65	0.011		3.15	0.089		5.73	0.025		3.48	0.074
allele	2	21.65	<0.001		0.35	0.843		10.27	0.014		4.2	0.144		1.81	0.418		1.04	0.602		5.22	0.094		8.02	0.031		14.14	0.004
N.water	1	2.04	0.159		30.11	<0.001		13.78	0.001		0	0.965		0.01	0.926		36.08	<0.001		9.41	0.005		0	0.964		2.9	0.101
N.allele	2	2.33	0.32		0.32	0.852		2.81	0.264		2.09	0.368		2.34	0.327		0.47	0.793		0.71	0.705		0.06	0.972		0.04	0.98
water.allele	2	2.14	0.352		2.05	0.373		0.28	0.871		3.15	0.227		1.96	0.389		0.47	0.793		0.52	0.774		2.34	0.328		0.25	0.883
N.water.allele	2	3.09	0.223		2.21	0.348		0.69	0.712		1.91	0.398		2.65	0.284		1.52	0.48		0.74	0.693		0.43	0.807		2.25	0.341

Means		par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	dry	7.39	6.75	7.25	44.38	45.26	45.29	13443	13032	13480	23.00	22.11	22.00	21.05	21.22	21.07	3.82	3.81	3.79	6.76	6.84	6.82	0.44	0.42	0.44	3.11	2.56	3.44
	irrig	7.84	7.22	7.81	46.68	46.85	47.41	14103	12921	13721	23.11	22.56	22.00	21.53	14.94	21.69	3.90	3.88	3.89	6.82	6.85	6.85	0.47	0.43	0.47	2.89	1.56	2.22
N3	dry	11.05	9.94	10.80	48.79	49.80	51.37	19988	17284	18061	24.33	23.44	23.11	22.29	22.55	22.89	4.00	4.03	4.03	6.92	6.97	7.05	0.41	0.40	0.43	2.22	1.11	1.78
	irrig	11.25	10.53	10.33	43.84	44.11	41.05	22768	20563	22019	23.22	25.44	22.89	20.99	21.36	20.79	3.82	3.84	3.74	6.80	6.87	6.87	0.46	0.41	0.45	1.89	1.00	2.11
mean S.E.D.			0.33			2.01			1015			1.02			2.62			0.06			0.07			0.03			0.56	
L.S.D.			0.67			4.14			2095			2.10			5.40			0.13			0.14			0.05			1.15	

Height components

ANOVA		Height of tiller (cm)			Ear length (cm)			Peduncle (cm)			Internode-1 (cm)			Internode-2 (cm)		
	d.f.	Wald	p-value		Wald	p-value		Wald	p-value		Wald	p-value		Wald	p-value	
N	1	95.26	<0.001		15.66	<0.001		2.34	0.139		6.41	0.018		85.17	<0.001	
water	1	5.78	0.024		0.61	0.442		0.66	0.426		5.13	0.033		7.78	0.01	
allele	2	130.27	<0.001		18.12	0.001		39.22	<0.001		49.01	<0.001		12.89	0.006	
N.water	1	1.15	0.294		2.94	0.099		3.03	0.094		0.22	0.646		0	0.952	
N.allele	2	4.33	0.137		0.55	0.763		1.42	0.503		2.91	0.254		8.27	0.059	
water.allele	2	1.34	0.52		2.91	0.253		1.43	0.499		0.63	0.733		1.16	0.567	
N.water.allele	2	2.64	0.286		0.95	0.629		0.2	0.903		5.58	0.081		2.17	0.353	

Means		par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	dry	100.9	92.5	106.4	10.0	8.1	9.9	39.1	33.1	39.6	24.5	22.0	25.9	14.0	14.4	14.8
	irrig	103.3	96.3	109.9	10.6	9.3	10.4	38.8	35.2	39.7	23.9	23.4	27.5	14.5	15.3	15.5
N3	dry	111.2	102.8	115.6	11.9	9.8	11.5	40.7	36.8	40.8	20.8	21.7	25.8	16.4	15.9	17.8
	irrig	115.4	104.2	113.7	10.5	10.4	11.4	38.3	35.7	39.2	23.4	23.2	25.4	17.9	16.3	17.9
mean S.E.D.			2.2			0.7			1.6			1.1			0.6	
L.S.D.			4.6			1.5			3.3			2.3			1.2	

Developmental traits

ANOVA		Heading (°C d)			Senescence (°C d)			Lodging score (0-100)			Ground cover (%)		
	d.f.	Wald	p-value		Wald	p-value		Wald	p-value		d.f.	F-test	p-value
N	1	6.51	0.014		189.64	<0.001		184.92	<0.001	allele	2	0.18	0.835
water	1	41.16	<0.001		10.26	0.002		0.59	0.444				
allele	2	44.5	<0.001		4.67	0.108		11.91	0.005				
N.water	1	6.32	0.015		10.26	0.002		0.59	0.444				
N.allele	2	5.07	0.09		4.67	0.108		11.91	0.005				
water.allele	2	2.29	0.326		2.21	0.34		0.91	0.638				
N.water.allele	2	5.22	0.084		2.21	0.34		0.91	0.638				

Means		par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	dry	1766	1779	1782	1.5	2	1.8	0	0	0	29.27	28.55	28.92
	irrig	1777	1782	1800	1.1	1.5	0.7	0	0	0			
N3	dry	1766	1779	1782	0	0	0	67	48	91			
	irrig	1779	1812	1807	0	0	0	72	38	74			
mean S.E.D.			7			0.3			12			1.2	
L.S.D.			13			0.5			23			2.4	

Design: RCBD with three replicates per N (40 and 200 kg N/Ha) and water treatment
Analysis: Mixed model fitting N treatment, irrigation and allele as fixed effects: N*water*allele
Site: Church Farm 2014

A3.7: ANOVA and experimental details of Church Farm 2014 Nitrogen and Irrigation experiment. Shaded values are significant at the 95% confidence level.

A3.8: ANOVA and experimental details of Reading data. Shaded values are significant at the 95% confidence level.

Design: Split-plot with N (40, 100 and 200 kg N/Ha) as the whole plot treatment and allele as the sub-plot, with five replicates.
Analysis: Split-Plot ANOVA with N*allele as treatment structure
Site: Sonningham, Reading 2014

Yield components

ANOVA	Yield (t DM/ha)			TGW (g)		Grains/m^2		Spikelets/ear		Grain area (mm^2)		Grain width (mm)		Grain length (mm)		Harvest index		Infertile spikelets		Spikes/m^2	
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
treatment	2	302.3	<.001	0	0.997	263.4	<.001	1.1	0.415	1.2	0.391	0.98	0.451	2.52	0.195	3.11	0.153	24.25	0.006	6.81	0.019
allele	2	3.88	0.035	3.91	0.034	4	0.032	2.55	0.119	0.51	0.613	0.54	0.598	7.03	0.01	3.42	0.067	1.42	0.28	13.94	<.001
treatment.allele	4	3.35	0.026	1.5	0.235	3.89	0.014	1.06	0.416	1.73	0.209	0.67	0.623	5.04	0.013	0.86	0.514	2.9	0.068	2.17	0.103

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	5.70	5.72	5.47	48.96	48.56	50.48	11640	11770	10837	23.56	24.89	22.89	21.59	21.52	21.83	3.94	3.87	3.90	6.81	6.84	6.90	0.44	0.41	0.45	4.11	2.67	3.89	417.3	300.0	377.3
N2	7.25	6.97	7.39	48.48	50.00	49.60	14964	13955	14908	23.22	23.22	23.33	21.50	21.14	21.53	3.87	3.90	3.87	6.84	6.71	6.85	0.43	0.42	0.42	3.33	3.78	3.33	418.7	390.7	403.3
N3	9.16	8.75	9.22	48.96	49.20	49.84	18709	17775	18496	22.89	23.56	22.67	21.42	21.99	21.64	3.91	3.94	3.86	6.78	6.89	6.90	0.42	0.41	0.42	2.89	2.56	2.22	475.3	383.3	486.7
mean S.E.D.		0.19			0.72			428			0.79			0.31			0.05			0.04			0.01			0.41			32.0	
L.S.D.		0.39			1.46			883			1.69			0.67			0.11			0.09			0.03			0.88			66.0	

Height components

ANOVA	Height of plot			Height of tiller (cm)		Ear length (cm)		Peduncle (cm)		Internode-1 (cm)		Internode-2 (cm)	
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
treatment	2	52.38	<.001	86.41	<.001	19.51	0.009	2.46	0.201	0.72	0.54	0.65	0.568
allele	2	90.25	<.001	23.72	<.001	23.74	<.001	15.02	<.001	7.1	0.009	0.39	0.688
treatment.allele	4	1.6	0.207	0.62	0.655	2.77	0.077	1.68	0.219	0.15	0.957	0.81	0.543

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	97.5	93.0	105.1	104.8	98.0	112.5	9.9	9.1	9.9	38.1	32.9	41.7	25.9	24.6	27.7	14.5	17.7	15.6
N2	104.9	100.7	109.7	112.5	104.0	115.5	10.4	8.3	10.4	39.2	36.1	42.8	26.2	24.6	27.3	16.2	16.4	16.3
N3	108.2	103.9	112.2	114.9	109.4	118.4	10.9	10.2	11.3	37.7	38.6	41.4	25.5	24.9	26.9	17.2	16.4	17.3
mean S.E.D.		1.4			2.5			0.4			1.8			1.0			1.7	
L.S.D.		2.9			5.4			0.8			3.9			2.2			3.5	

Developmental traits

ANOVA	R:FR 30/10/13			R:FR 4/3/14		R:FR 7/7/14		R:FR 14/7/14		R:FR 24/7/14		PAR (%) 7/7/14		PAR (%) 14/7/14		PAR (%) 24/7/14		Ground cover (%)		
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	d.f.	F-test	p-value
treatment	2	2.35	0.157	0.01	0.988	43.64	<.001	173	<.001	1.26	0.335	98.57	<.001	142.82	<.001	10.31	0.006	2	0.89	0.421
allele	2	0.66	0.527	0.35	0.71	6.49	0.006	2.45	0.107	1.79	0.188	3.96	0.033	7.41	0.003	7.39	0.003			
treatment.allele	4	2.3	0.088	1.11	0.377	0.35	0.843	3.3	0.027	0.92	0.47	0.22	0.922	0.7	0.599	0.32	0.861			

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	1.04	1.06	1.06	0.49	0.49	0.50	0.57	0.59	0.54	0.76	0.73	0.71	0.84	0.86	0.82	82.2	80.2	82.1	81.8	77.8	79.7	84.0	81.3	84.0	48.3	47.1	46.1
N2	1.08	1.06	1.04	0.50	0.51	0.48	0.54	0.53	0.45	0.65	0.69	0.65	0.85	0.84	0.84	87.7	87.0	87.9	87.2	84.0	86.3	89.7	86.0	88.2			
N3	1.10	1.08	1.09	0.48	0.49	0.52	0.40	0.40	0.36	0.61	0.54	0.58	0.83	0.83	0.82	93.3	92.0	93.6	91.0	89.7	90.0	91.5	89.6	91.4			
mean S.E.D.		0.02			0.02			0.03			0.02			0.02			0.6			0.7			0.8			1.7		
L.S.D.		0.05			0.05			0.07			0.04			0.04			1.2			1.5			1.6			3.4		

Appendix to Chapter 4

Site: Church Farm 2014
Design: RCBD, 3 replicates per N (40 and 200 kg N/Ha) and water treatment
Analysis: Mixed model: N*water*allele

ANOVA	d.f.	Plant height (cm)		Spike length (cm)		Spike/PH		Spikelets/spike		Spike compaction (cm spikelet^{-1})	
		Wald	p-value	Wald	p-value	Wald	p-value	Wald	p-value	Wald	p-value
N	1	95.26	<0.001	15.66	<0.001	1.35	0.257	9.47	0.005	4.14	0.053
water	1	5.78	0.024	0.61	0.442	0.06	0.807	0.24	0.628	0.34	0.567
allele	2	130.27	<0.001	18.12	0.001	2.17	0.353	4.2	0.144	21.68	<0.001
N.water	1	1.15	0.294	2.94	0.099	1.81	0.191	0	0.965	2.46	0.13
N.allele	2	4.33	0.137	0.55	0.763	1.55	0.471	2.09	0.368	0.14	0.933
water.allele	2	1.34	0.52	2.91	0.253	3.24	0.219	3.15	0.227	0.43	0.807
N.water.allele	2	2.64	0.286	0.95	0.629	1.61	0.458	1.91	0.398	0.23	0.892

Means		par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	dry	100.9	92.5	106.4	10.0	8.1	9.9	0.100	0.088	0.093	23.00	22.11	22.00	0.437	0.366	0.450
	irrig	103.3	96.3	109.9	10.6	9.3	10.4	0.102	0.097	0.095	23.11	22.56	22.00	0.459	0.412	0.475
N3	dry	111.2	102.8	115.6	11.9	9.8	11.5	0.107	0.096	0.099	24.33	23.44	23.11	0.489	0.420	0.500
	irrig	115.4	104.2	113.7	10.5	10.4	11.4	0.092	0.100	0.101	23.22	25.44	22.89	0.458	0.410	0.499
mean S.E.D.			2.2			0.7			0.035			1.02			0.007	
L.S.D.			4.6			1.5			0.073			2.10			0.015	

Site: Church Farm 2014
Design: RCBD, 3 replicates per N treatment (40, 100 and 200 kg N/Ha)
Analysis: Two-way ANOVA, randomised blocks

ANOVA	d.f.	Plant height (cm)		Spike length (cm)		Spike/PH		Spikelets/spike		Spike compaction (cm spikelet^{-1})	
		F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
treatment	2	15.55	0.004	18.42	0.003	4.79	0.057	2.07	0.208	10.08	0.012
allele	2	85.89	<.001	82.76	<.001	23.26	<.001	1.9	0.192	33.29	<.001
treatment.allele	4	0.48	0.752	1.11	0.397	1.29	0.329	0.11	0.976	0.76	0.573

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	100.9	92.5	106.4	10.0	8.1	9.9	0.100	0.088	0.093	23.00	22.11	22.00	0.437	0.366	0.450
N2	102.8	96.5	110.5	12.0	9.2	11.2	0.117	0.095	0.102	23.11	22.56	22.67	0.517	0.406	0.496
N3	111.2	102.8	115.6	11.9	9.8	11.5	0.107	0.096	0.099	24.33	23.44	23.11	0.489	0.420	0.500
mean S.E.D.		2.3			0.4			0.005			0.95			0.022	
L.S.D.		5.0			0.9			0.010			2.00			0.046	

A4.1: ANOVA and experimental details of Church Farm 2012 - 2013 height and spike compaction data. Shaded values are significant at the 95% confidence level.

Site: Church Farm 2014
Design: RCBD, 3 replicates per N (40 and 200 kg N/Ha) and water treatment
Analysis: Mixed model: N*water*allele

ANOVA	d.f.	Plant height (cm)		Spike length (cm)		Spike/PH		Spikelets/spike		Spike compaction (cm spikelet^{-1})	
		Wald	p-value	Wald	p-value	Wald	p-value	Wald	p-value	Wald	p-value
N	1	95.26	<0.001	15.66	<0.001	1.35	0.257	9.47	0.005	4.14	0.053
water	1	5.78	0.024	0.61	0.442	0.06	0.807	0.24	0.628	0.34	0.567
allele	2	130.27	<0.001	18.12	0.001	2.17	0.353	4.2	0.144	21.68	<0.001
N.water	1	1.15	0.294	2.94	0.099	1.81	0.191	0	0.965	2.46	0.13
N.allele	2	4.33	0.137	0.55	0.763	1.55	0.471	2.09	0.368	0.14	0.933
water.allele	2	1.34	0.52	2.91	0.253	3.24	0.219	3.15	0.227	0.43	0.807
N.water.allele	2	2.64	0.286	0.95	0.629	1.61	0.458	1.91	0.398	0.23	0.892

Means		par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	dry	100.9	92.5	106.4	10.0	8.1	9.9	0.100	0.088	0.093	23.00	22.11	22.00	0.437	0.366	0.450
	irrig	103.3	96.3	109.9	10.6	9.3	10.4	0.102	0.097	0.095	23.11	22.56	22.00	0.459	0.412	0.475
N3	dry	111.2	102.8	115.6	11.9	9.8	11.5	0.107	0.096	0.099	24.33	23.44	23.11	0.489	0.420	0.500
	irrig	115.4	104.2	113.7	10.5	10.4	11.4	0.092	0.100	0.101	23.22	25.44	22.89	0.458	0.410	0.499
mean S.E.D.			2.2			0.7			0.035			1.02			0.007	
L.S.D.			4.6			1.5			0.073			2.10			0.015	

Site: Church Farm 2014
Design: RCBD, 3 replicates per N treatment (40, 100 and 200 kg N/Ha)
Analysis: Two-way ANOVA, randomised blocks

ANOVA	d.f.	Plant height (cm)		Spike length (cm)		Spike/PH		Spikelets/spike		Spike compaction (cm spikelet^{-1})	
		F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
treatment	2	15.55	0.004	18.42	0.003	4.79	0.057	2.07	0.208	10.08	0.012
allele	2	85.89	<.001	82.76	<.001	23.26	<.001	1.9	0.192	33.29	<.001
treatment.allele	4	0.48	0.752	1.11	0.397	1.29	0.329	0.11	0.976	0.76	0.573

Means	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1	100.9	92.5	106.4	10.0	8.1	9.9	0.100	0.088	0.093	23.00	22.11	22.00	0.437	0.366	0.450
N2	102.8	96.5	110.5	12.0	9.2	11.2	0.117	0.095	0.102	23.11	22.56	22.67	0.517	0.406	0.496
N3	111.2	102.8	115.6	11.9	9.8	11.5	0.107	0.096	0.099	24.33	23.44	23.11	0.489	0.420	0.500
mean S.E.D.		2.3			0.4			0.005			0.95			0.022	
L.S.D.		5.0			0.9			0.010			2.00			0.046	

A4.2: ANOVA and experimental details of Church Farm 2014 height and spike compaction data. Shaded values are significant at the 95% confidence level.

Site: Sonningham, Reading 2014
Design: Split-plot with N (40, 100 and 200 kg N/Ha) as whole plot treatment, allele as the sub-plot, 5 replicates.
Analysis: Split-Plot ANOVA, N*allele as treatment structure

ANOVA		Plant height (cm)			Spike length (cm)			Spike/PH			Spikelets/spike			Spike compaction (cm spikelet^{-1})		
	d.f.	F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value		F-test	p-value	
treatment	2	86.41	<.001		19.51	0.009		5.52	0.071		1.1	0.415		32.93	0.003	
allele	2	23.72	<.001		23.74	<.001		1.82	0.203		2.55	0.119		72.07	<.001	
treatment.allele	4	0.62	0.655		2.77	0.077		1.71	0.211		1.06	0.416		2.39	0.108	
Means		par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall	par	rht8	tall
N1		104.8	98.0	112.5	9.9	9.1	9.9	0.095	0.093	0.088	23.56	24.89	22.89	0.420	0.370	0.430
N2		112.5	104.0	115.5	10.4	8.3	10.4	0.093	0.080	0.090	23.22	23.22	23.33	0.450	0.360	0.450
N3		114.9	109.4	118.4	10.9	10.2	11.3	0.095	0.093	0.096	22.89	23.56	22.67	0.480	0.430	0.500
mean S.E.D.			2.5			0.4			0.004			0.79			0.012	
L.S.D.			5.4			0.8			0.009			1.69			0.027	

A4.3: ANOVA and experimental details of Reading height and spike compaction data. Shaded values are significant at the 95% confidence level.

Analysis: One-way ANOVA, N=20/allele

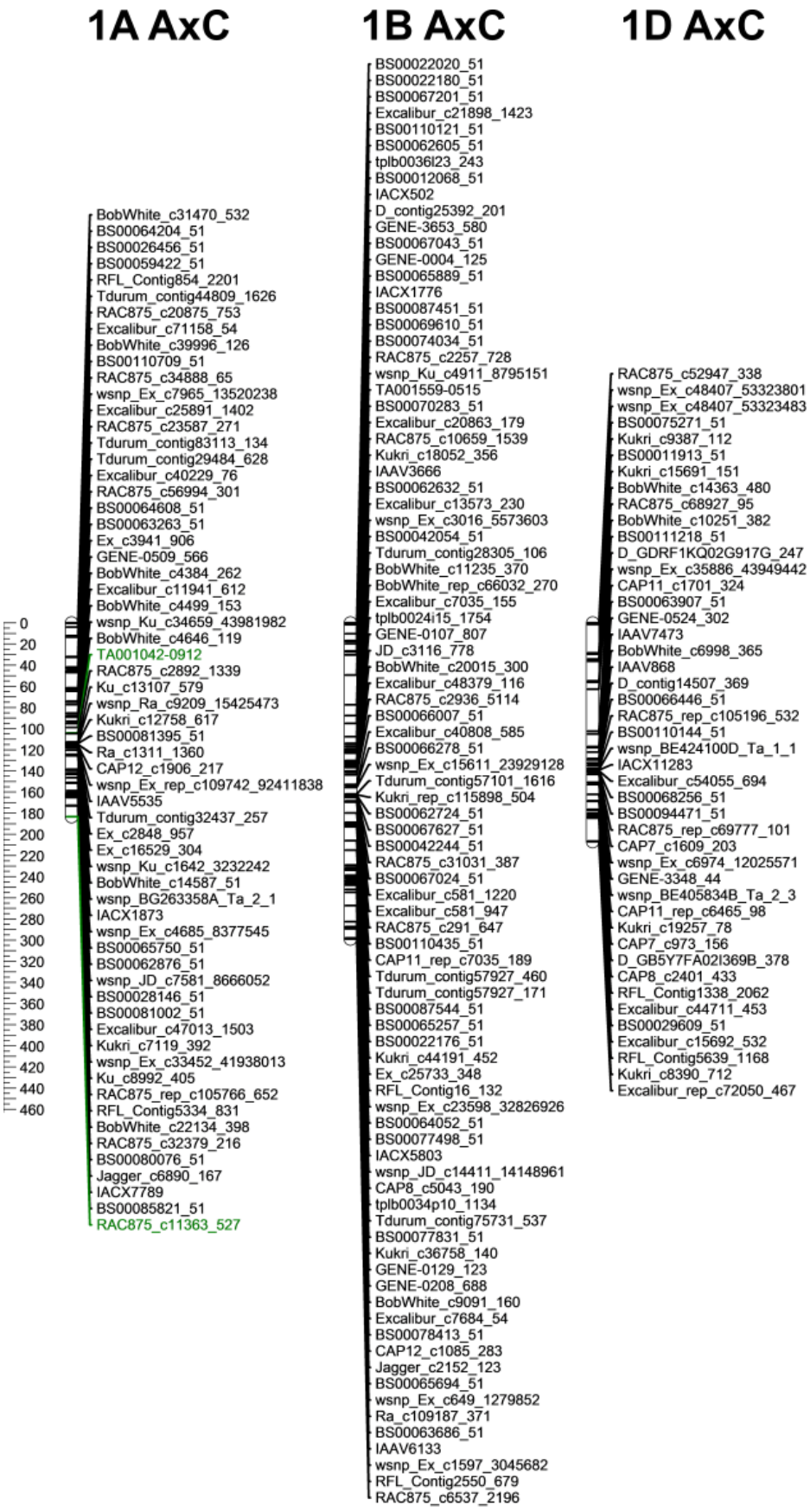
ANOVA		Spike compaction (cm spikelet^{-1})		Spike length (cm)		Spikelets spike^{-1}	
	d.f.	F-test	p-value	F-test	p-value	F-test	p-value
allele	1	1.08	0.305	0.45	0.505	3.09	0.087
Means		**short**	**tall**	**short**	**tall**	**short**	**tall**
		1.55	1.71	3.15	3.35	20.41	19.67
mean S.E.D.		0.15		0.30		0.42	
L.S.D.		0.30		0.60		0.86	

A4.4: ANOVA for spike compaction in the fine-mapping *Rht8* recombinants, grouped into short and tall types.

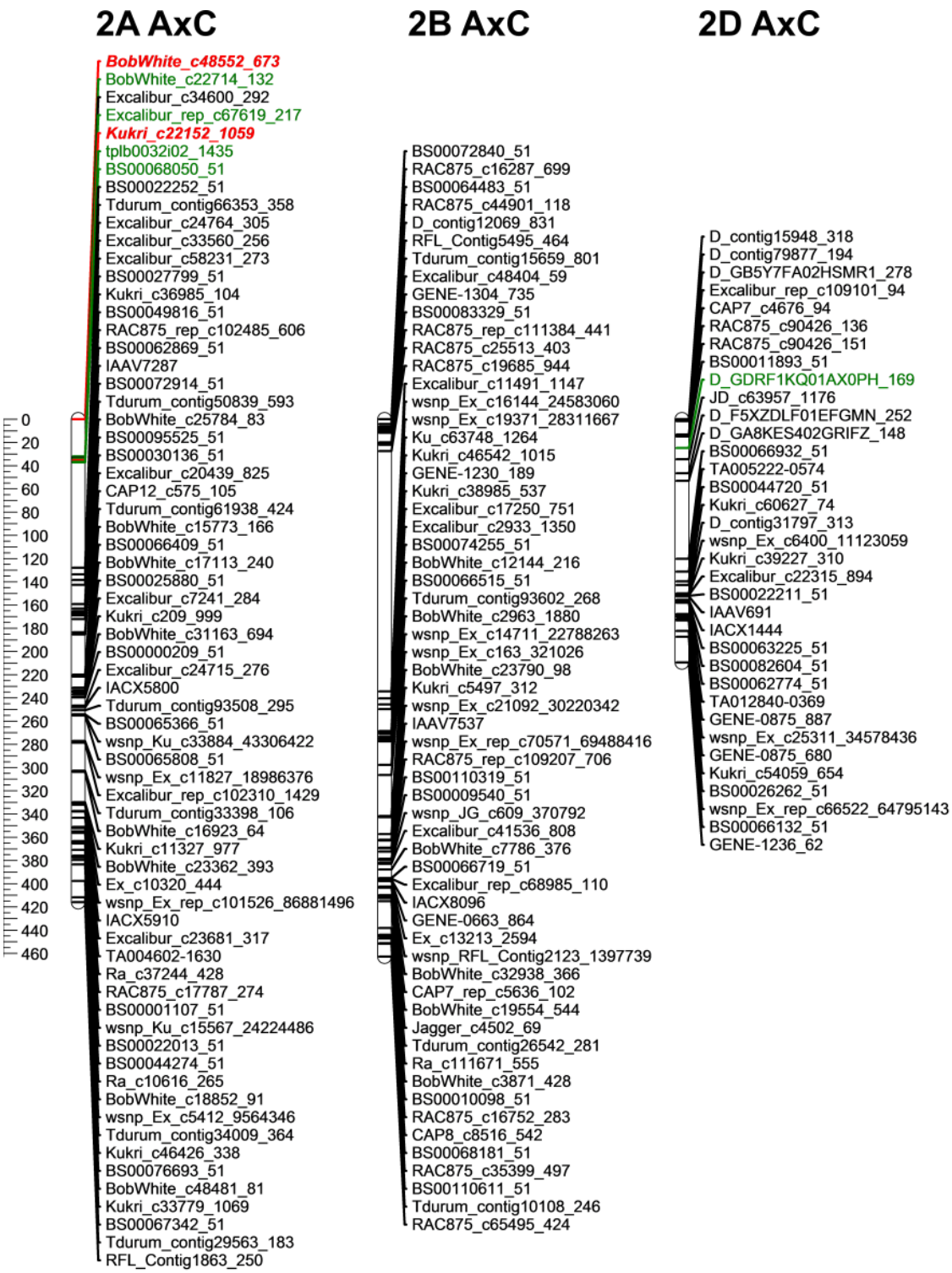
Appendix to Chapter 5

A5.1:

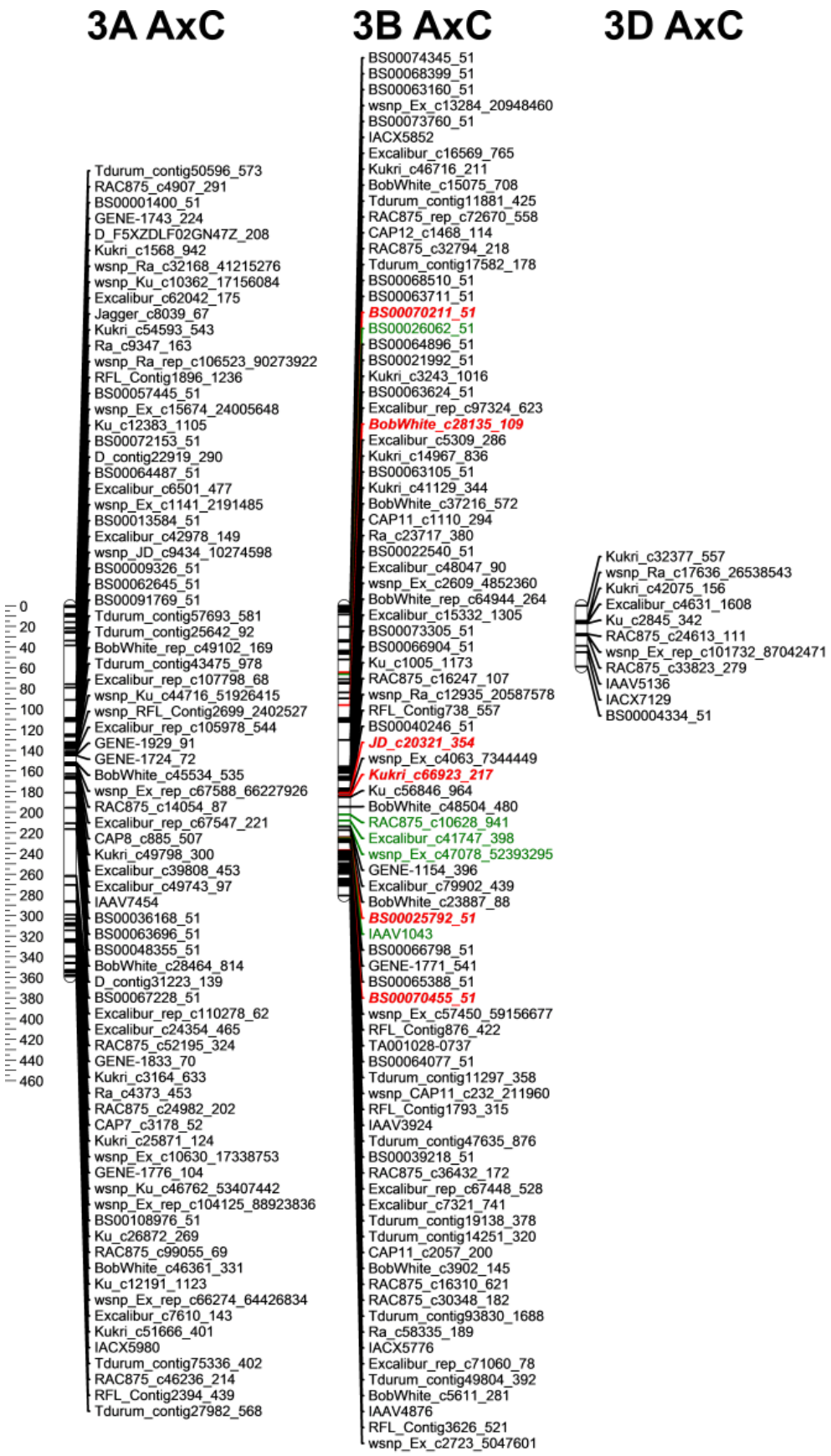
SNP probes across the entire iSelect 90K array were mapped in two ways. First, in the Avalon x Cadenza population and later, using the genetic positions from the Akhunov genetic map, named after the author (Krasileva et al., 2013). These probes had a genetic position in the array. SNPs between the parent NILs within the mapped probes are shown. Of the total 81,587 SNPs on the array, 9360 SNP probes had a genetic position in the AxC map. A total of 38,832 had a genetic position in the Akhunov map. Redundant SNP probes at the same genetic position and with the same allele call were removed. Probes coloured black have no polymorphism between the parent NILs. Probes coloured green are a polymorphism based on one of the parent NILs having a heterozygous call, ‘AB’ (described in the text in 5.3.4). Probes coloured red and italicised are more confident SNPs where the short parent NIL had a different (homozygous) call to the tall NIL.



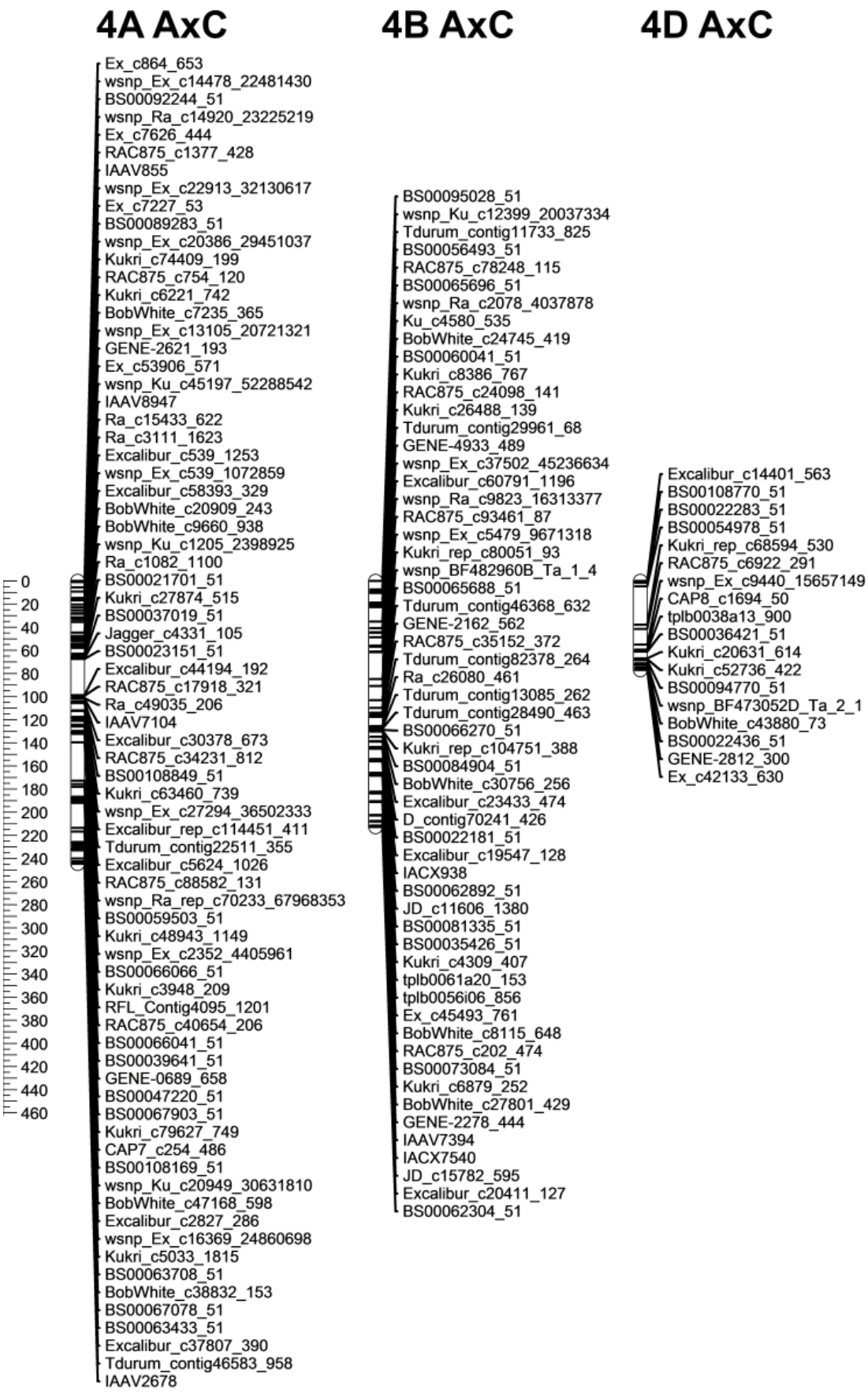
A5.1 (continued)



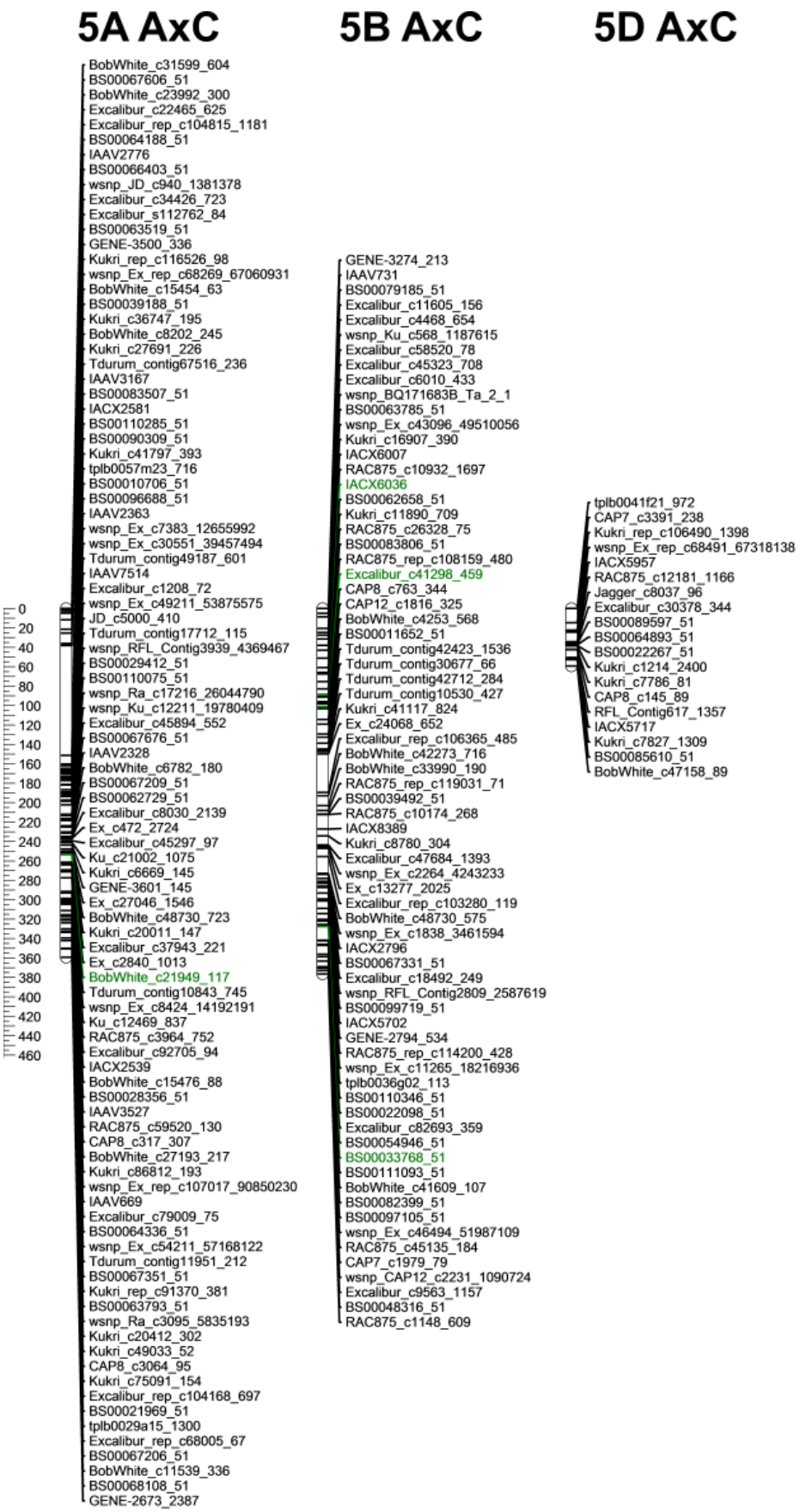
A5.1 (continued)



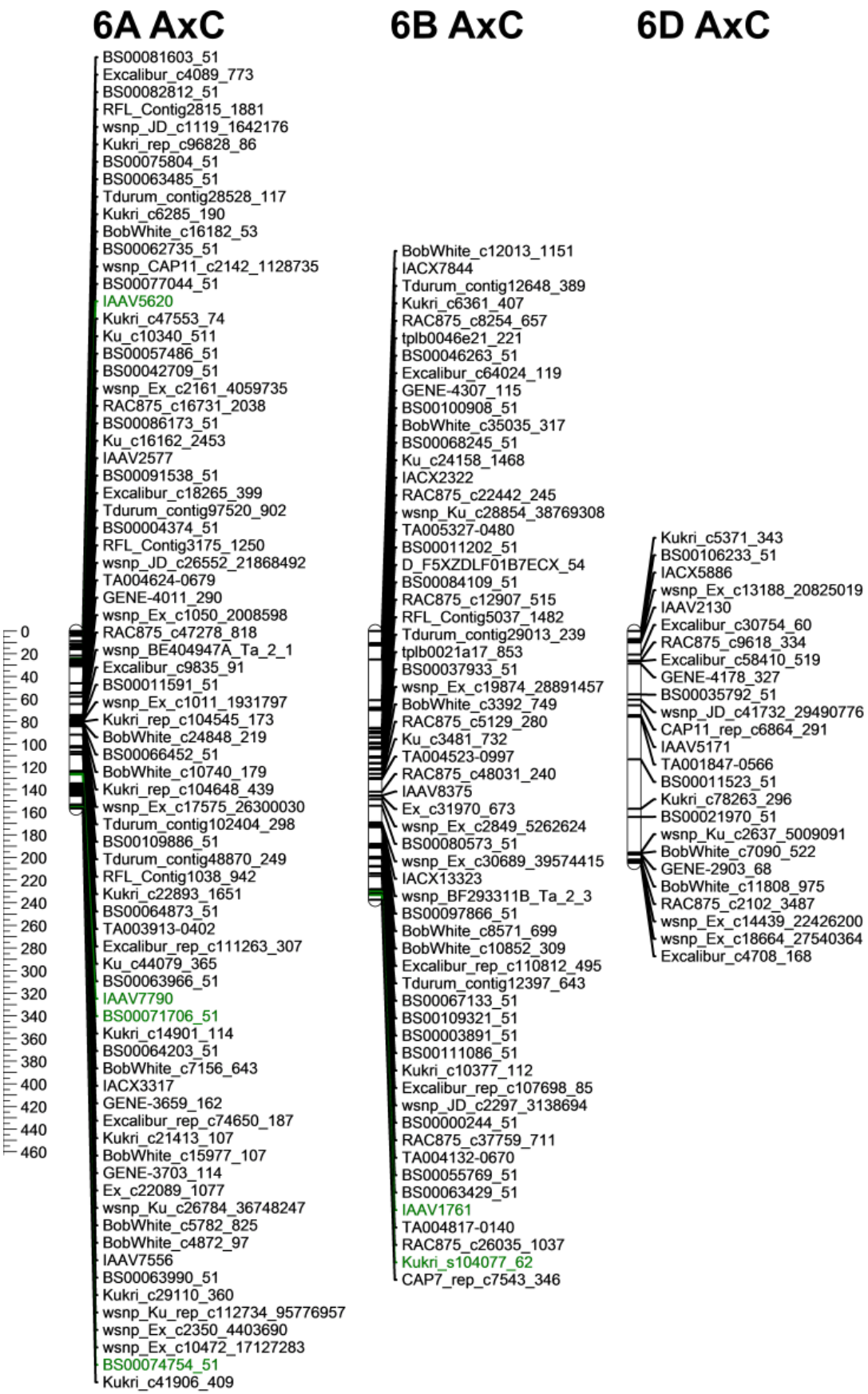
A5.1 (continued)



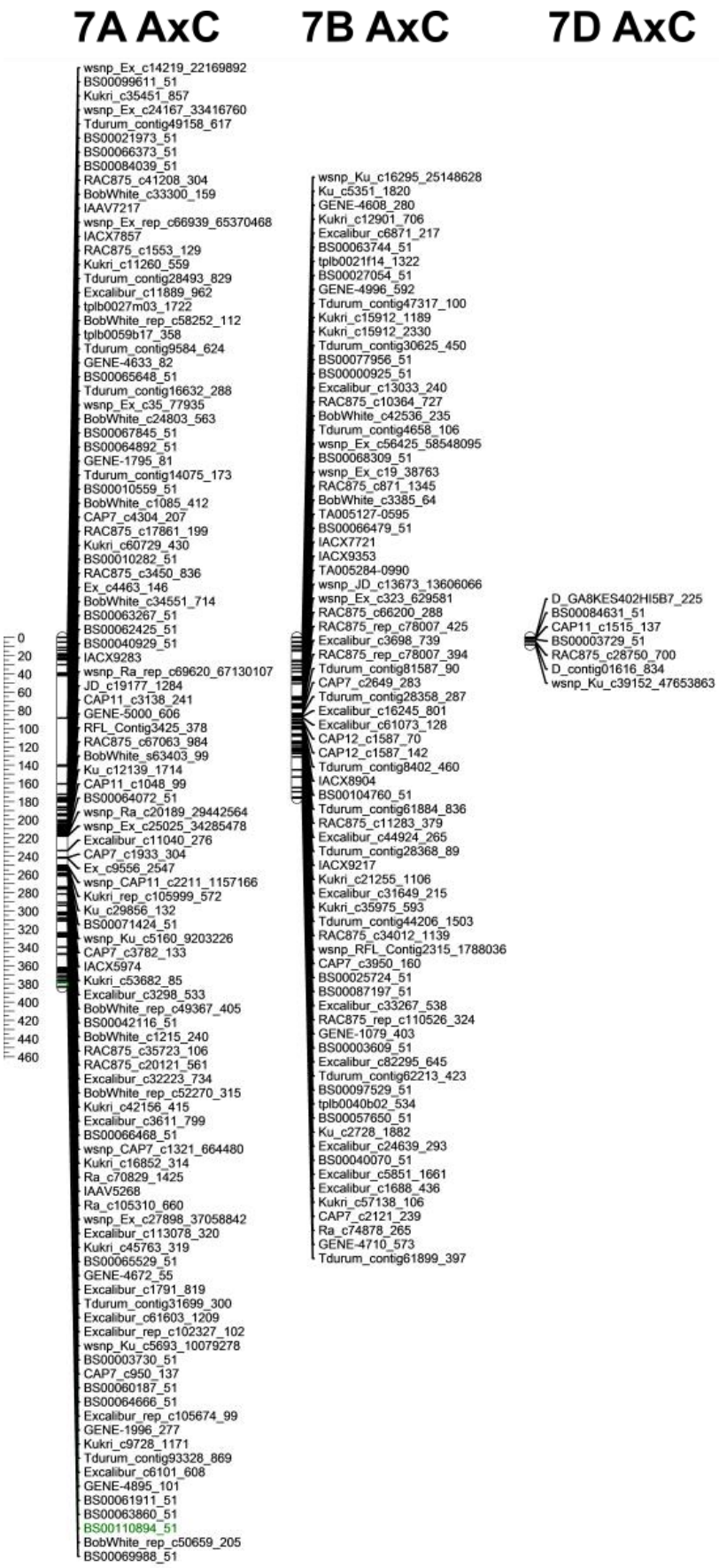
A5.1 (continued)



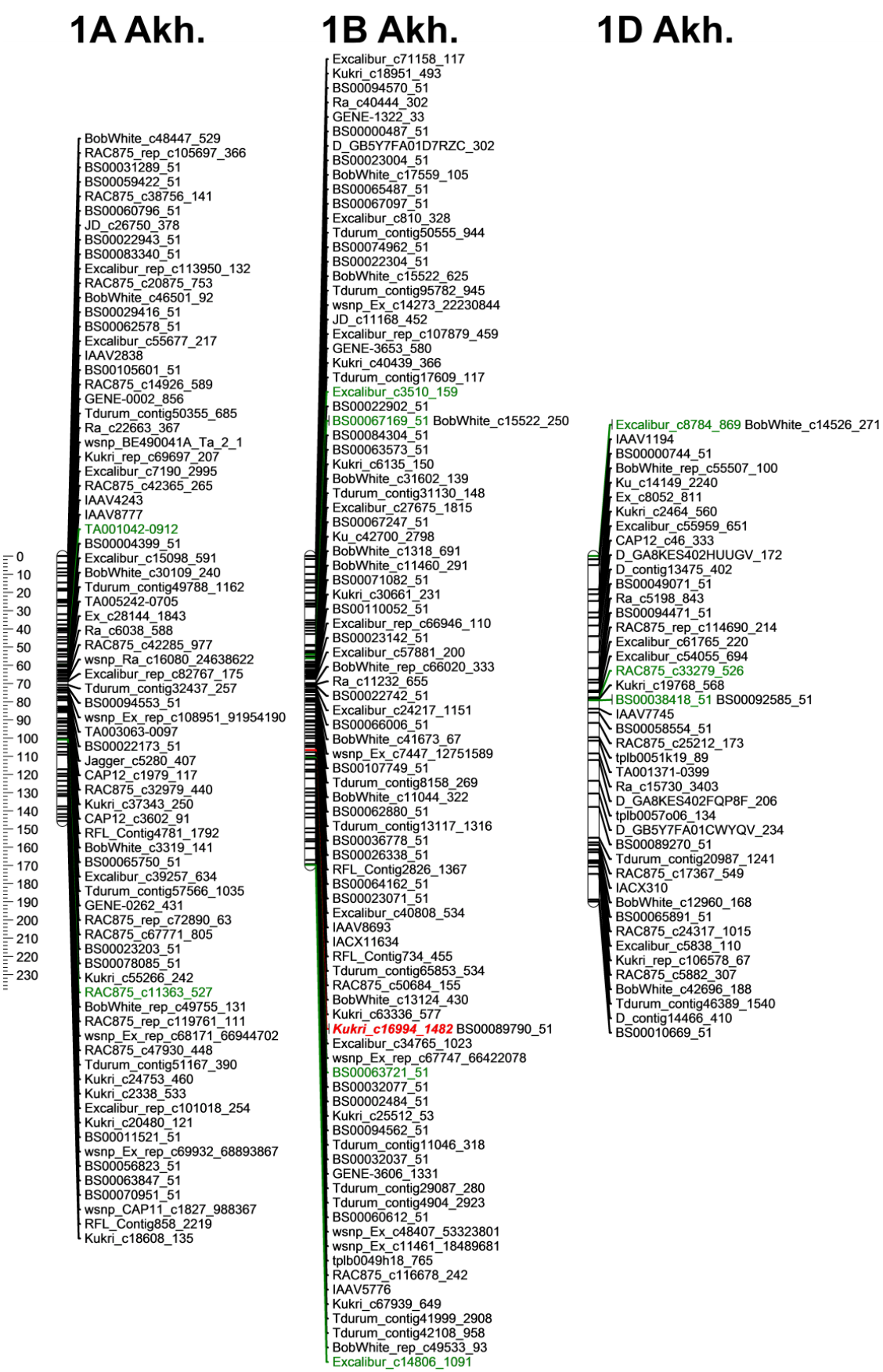
A5.1 (continued)



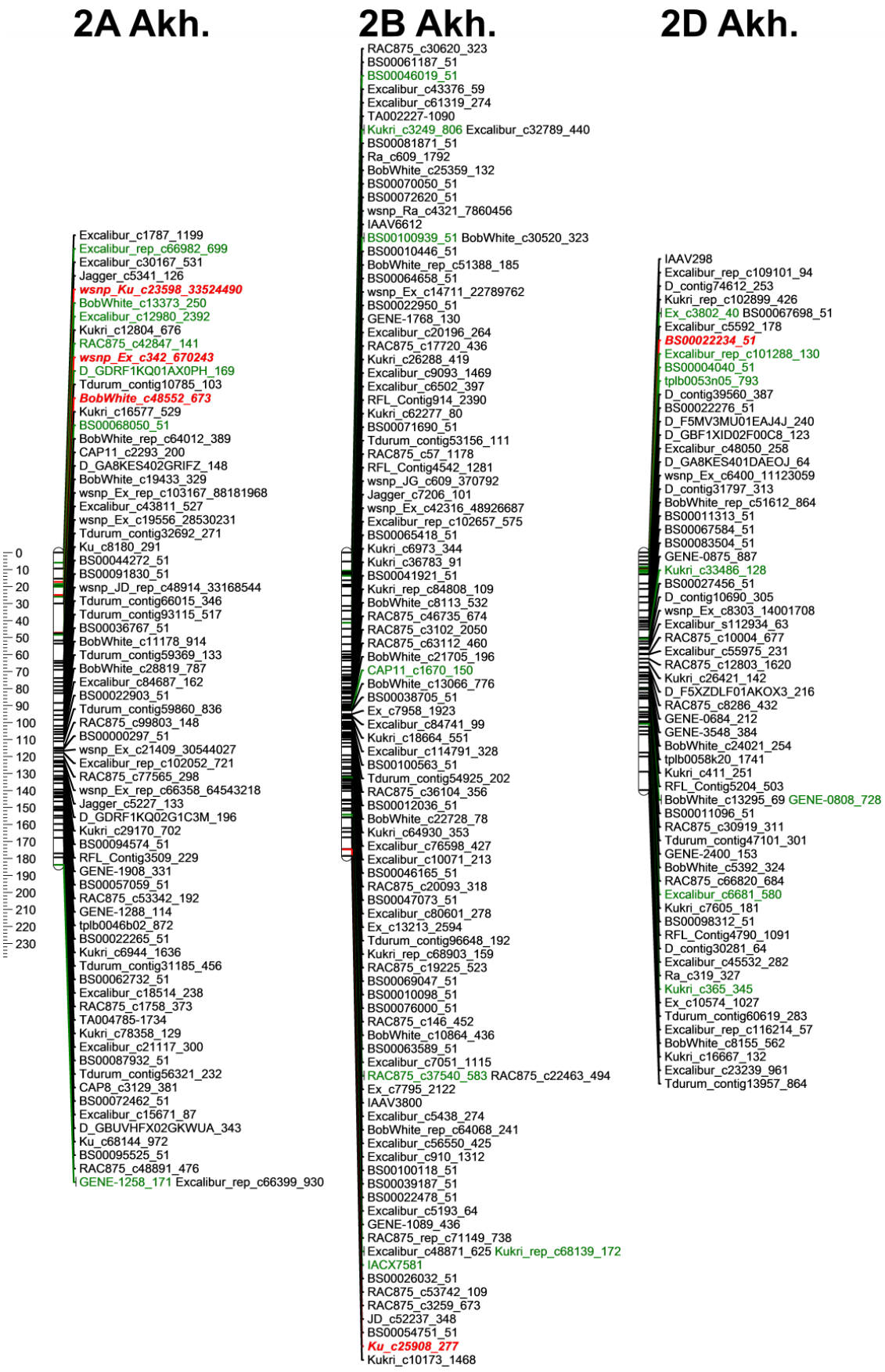
A5.1 (continued)



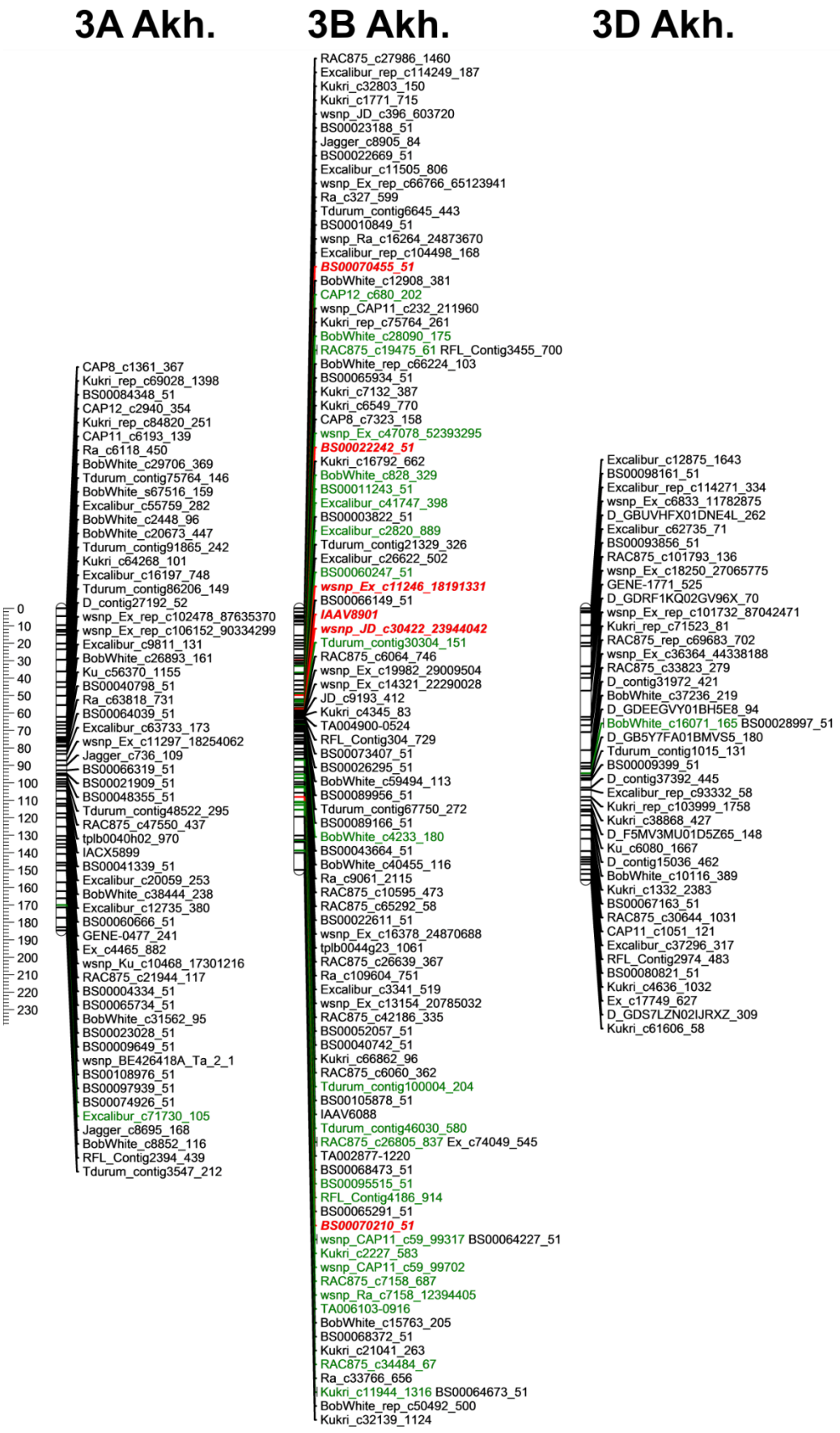
A5.1 (continued)



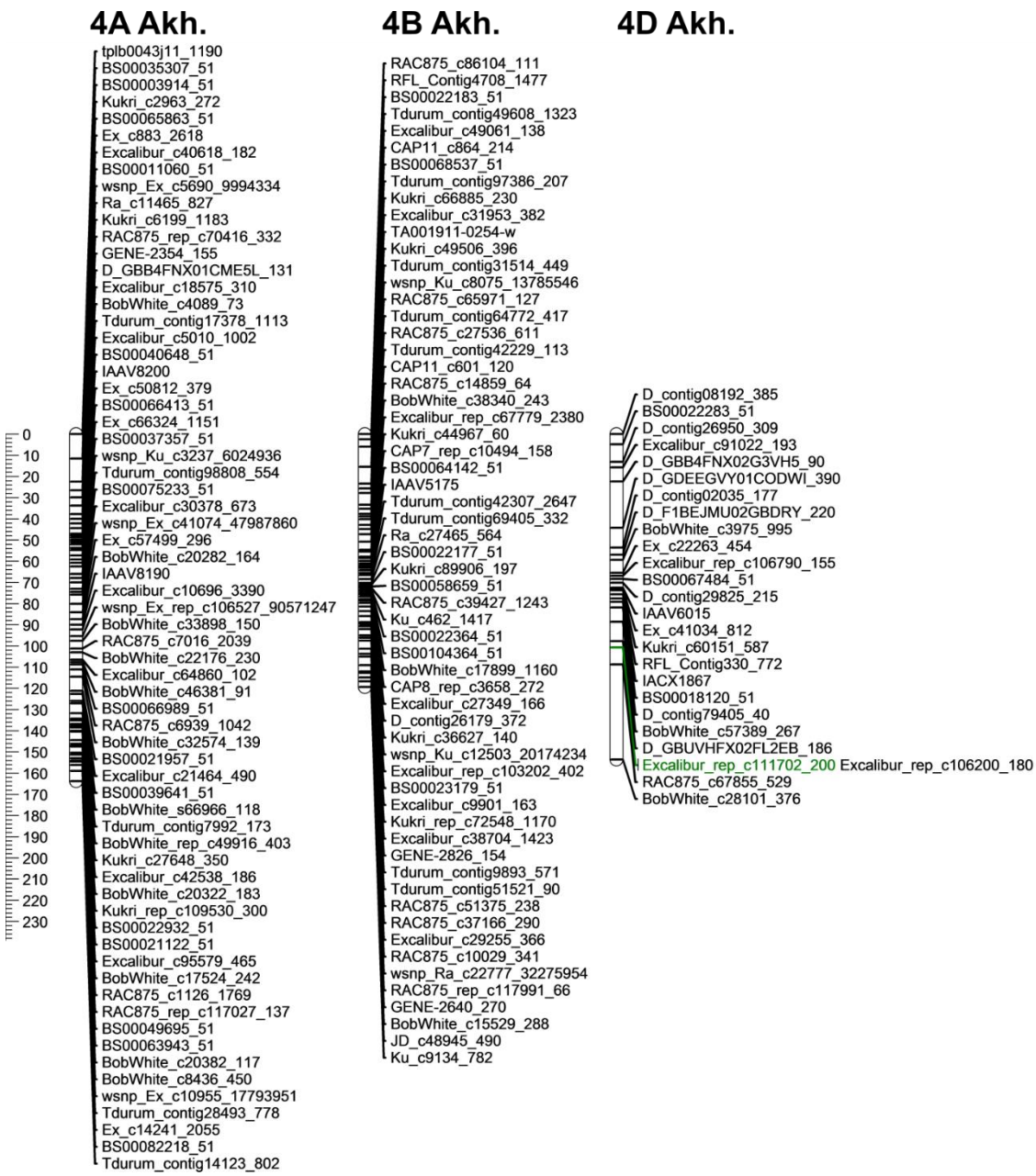
A5.1 (continued)



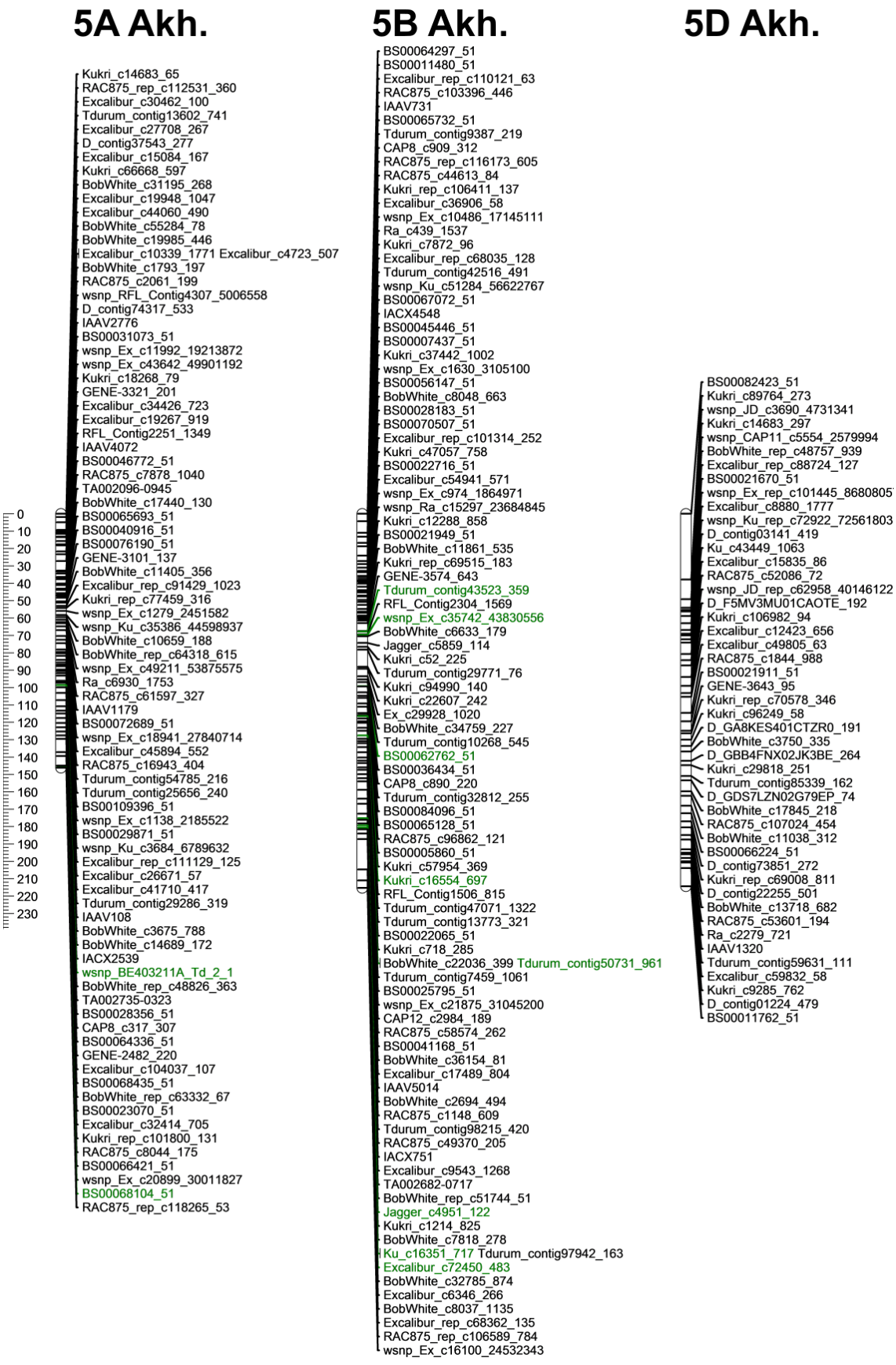
A5.1 (continued)



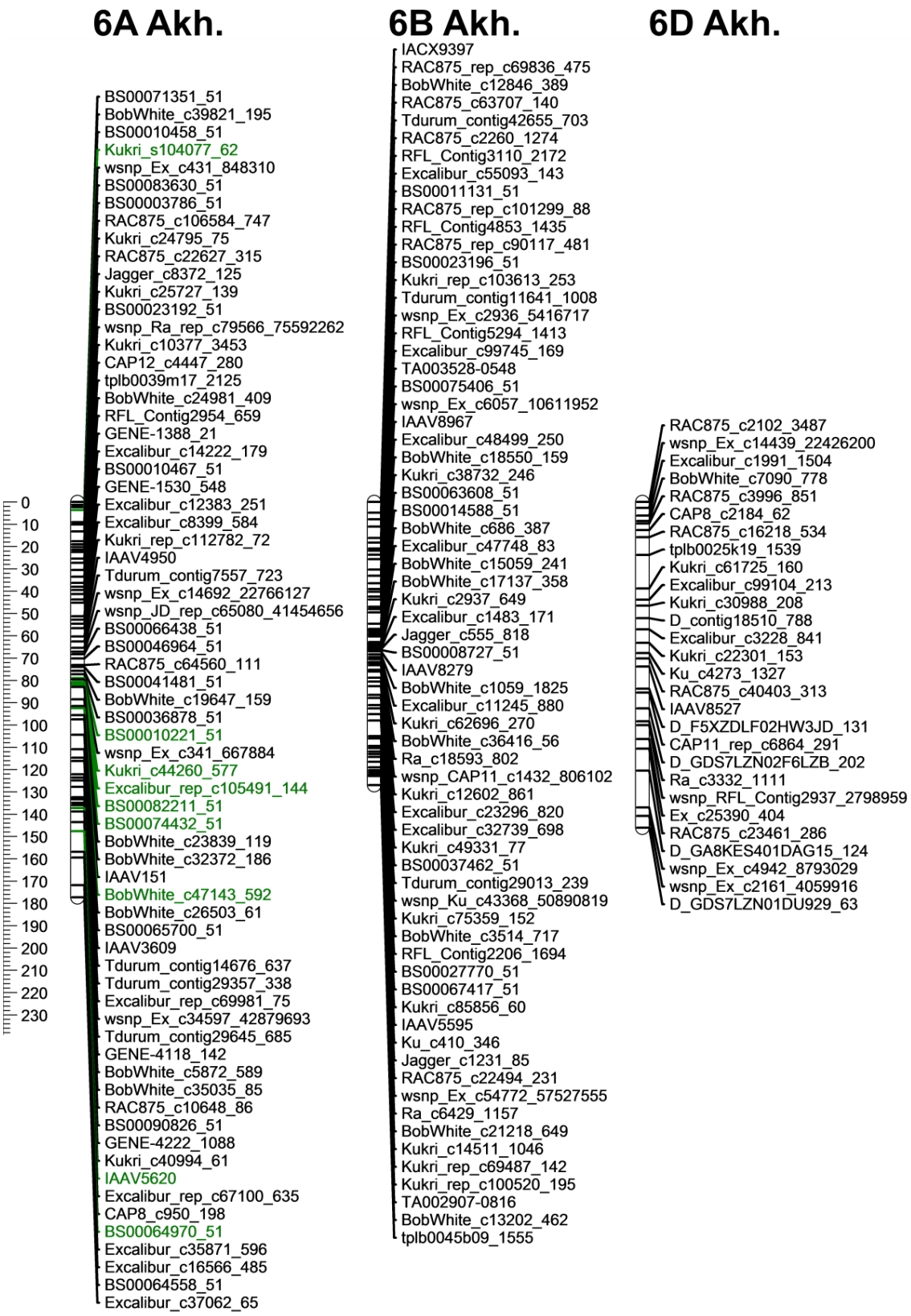
A5.1 (continued)



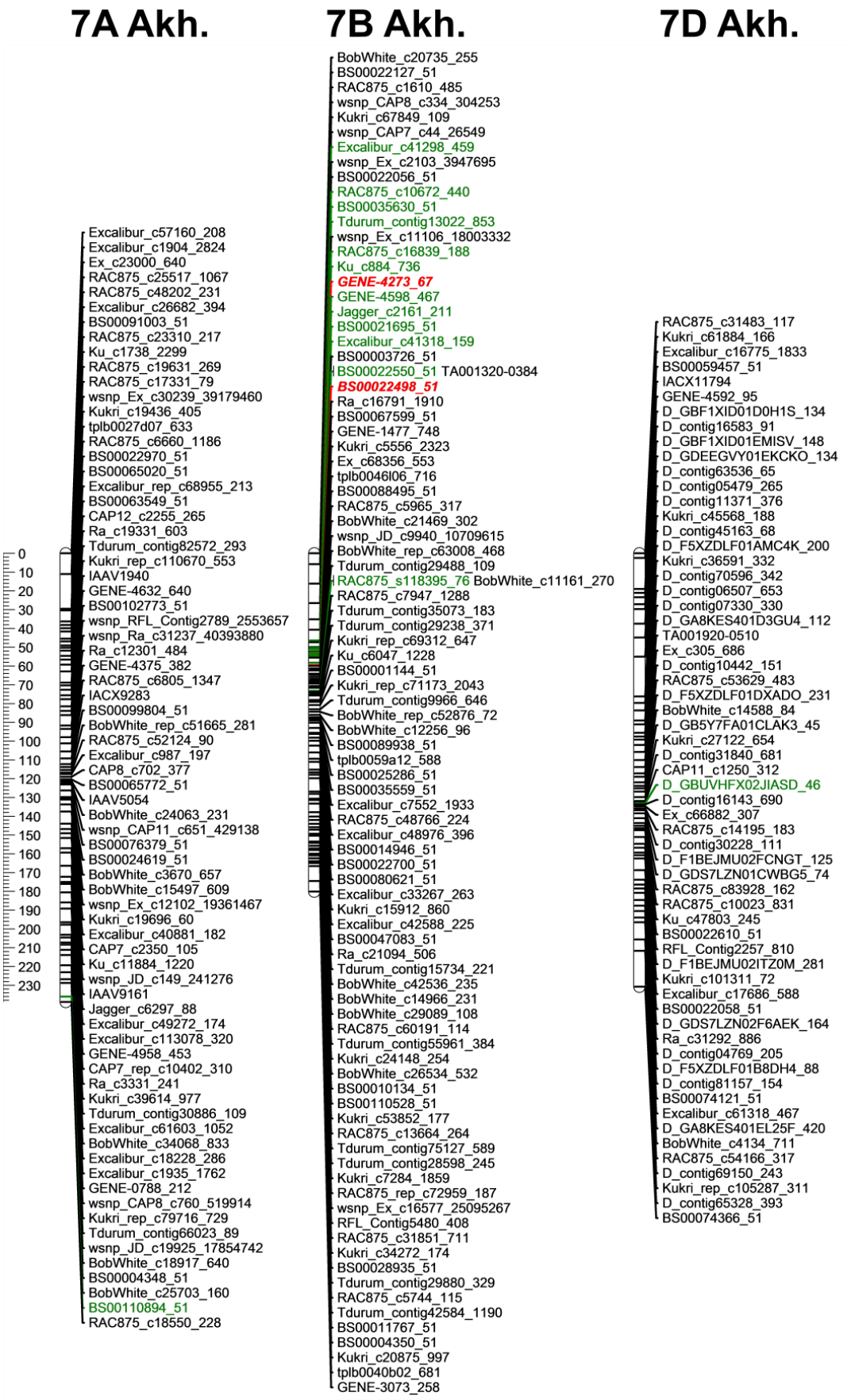
A5.1 (continued)



A5.1 (continued)



A5.1 (continued)



A5.1 (continued)

					KASP							
Barley			SSR		UniGene		v3.3		Array		BD	Rice
Gene	position	2DS hit	Tested	Marker	SNP	Marker	SNP	marker	iSelect	Axiom	Gene	Gene
MLOC_57507	15589004	2DS_3524440									Bradi5g03590	Os07g0534000
MLOC_57508	15592532	2DS_5354706									Bradi5g03600	Os04g0212200
MLOC_5957	15601547	2DS_5377037									Bradi5g03460	Os04g0209200
MLOC_62798	15618954	2DS_5321865									Bradi4g21260	Os11g0215100
MLOC_71963	15624566	2DS_5319467									Bradi2g42760	Os04g0208400
MLOC_11990	15656891	2DS_5352598									Bradi5g03810	Os04g0224900
MLOC_61793	17392210	2DS_5369305									Bradi5g04000	Os04g0227200
MLOC_58539	17416166	2DS_5390396									Bradi5g03960	Os06g0587200
MLOC_58540	17420938	2DS_5316778									Bradi5g03950	Os08g0110100
MLOC_52767	17428728	2DS_5390004									Bradi5g03860	Os05g0295300
MLOC_5849	17457414	2DS_5352525									Bradi5g03882	Os05g0553400
MLOC_63757	17478781	2DS_5327480									Bradi5g03710	Os04g0223500
MLOC_16798	17487159	2DS_5324300									Bradi3g20960	Os06g0203200
MLOC_67319	17502873	2DS_5389716									Bradi5g04030	Os04g0227500
MLOC_48019	17525526	2DS_5341587									Bradi5g03577	Os11g0691100
MLOC_74610	17556696	2DS_5363870									Bradi5g03697	Os04g0224600
MLOC_14804	17591628	2DS_5375380									Bradi5g03640	Os04g0223300
MLOC_21811	17599756	2DS_5339156									Bradi4g08800	Os12g0291400
MLOC_18415	17679159	2DS_5335120									Bradi5g04020	Os07g0406300
MLOC_10026	17688253	-									Bradi5g03977	Os02g0210700
MLOC_63015	17699255	2DS_5381312									Bradi5g04050	Os04g0228000
MLOC_63016	17703692	-									Bradi5g04057	Os04g0228100
MLOC_56367	17709336	2DS_5390981									Bradi1g21320	Os07g0622100
MLOC_65574	17713531	2DS_5384527									Bradi4g06970	Os10g0136100
MLOC_71561	17728883	2DS_5374739									Bradi5g04630	Os10g0150400
MLOC_37479	18924902	2DS_5379098									Bradi5g04340	Os10g0558900
MLOC_9931	18943132	2DS_5347513									Bradi5g04567	Os09g0252700
MLOC_56278	18975015	2DS_1805600									Bradi3g56020	Os04g0233400
MLOC_60079	18990753	2DS_5374739									Bradi3g37067	Os08g0436700
MLOC_69463	19016685	2DS_5390752									Bradi5g04550	Os04g0244400
MLOC_56660	19027873	-									Bradi5g04560	Os04g0244800
MLOC_62246	19048502	2DS_5320736									Bradi5g04590	-
MLOC_48245	19071820	-									Bradi5g04130	Os04g0229100
MLOC_81380	19149532	2DS_5374739									Bradi5g04630	Os10g0150400
MLOC_10439	19171497	-									Bradi4g36976	Os11g0684700
MLOC_4350	19440192	2DS_5366894									Bradi5g04673	Os04g0252200
MLOC_72777	19455074	2DS_5364728									Bradi5g04686	Os04g0252400
MLOC_59732	19468483	2DS_4338395									Bradi1g16097	Os03g0594700
MLOC_34868	19476019	-									Bradi1g16097	Os03g0594700
MLOC_58466	19491674	-									Bradi5g04660	Os02g0319800
MLOC_68321	19511497	2DS_5368388									Bradi3g36320	Os08g0427900
MLOC_12182	19521133	2DS_5366894									Bradi5g04673	Os04g0252200
MLOC_32207	19532277	2DS_5330382									Bradi5g04630	Os10g0150400
MLOC_81869	19752473	2DS_5388293									Bradi3g22850	Os02g0113200
MLOC_55119	19928531	-									Bradi5g04730	Os04g0258900
MLOC_55120	19945311	2DS_5388494									Bradi4g07480	Os04g0255600
MLOC_60943	19981189	2DS_5349408									Bradi5g04640	Os05g0304900

A5.2.1: Barley zipper used for mining variation on 2DS CSS contigs in the syntenic interval. The interval was defined as in Table 5.4. The annotation shows how the sequence space was searched. In the *SSR* columns, microsatellite variation identified on the CSS contig which could be tested is in green. No variation is shaded red. In the *marker* column a polymorphic marker between the parent NILs is shaded green; monomorphic markers are in red. The same colours apply to the *UniGene* and *v3.3 cDNA* columns. Orange cells indicate SNPs identified in the UniGene data where only one sample called a SNP (low concordance). Orange shading in the *iSelect* column indicates the corresponding SNP marker was on the array but monomorphic between the parent NILs. The shading in the *axiom* column indicates a marker on the array corresponding to the 2DS CSS contig in that row had a SNP between the *Rht8* NIL and Paragon.

				KASP							
Brachypodium		SSR		UniGene		v3.3		Array		Barley	Rice
Gene	2DS hit	Tested	Marker	SNP	Marker	SNP	marker	iSelect	Axiom	Gene	Gene
Bradi5g03460	2DS_5377037									MLOC_5957	Os04g0209200
Bradi5g03477	2DS_5368504									-	Os04g0209300
Bradi5g03520	-									MLOC_73392	Os11g0691400
Bradi5g03530	2DS_5292808									-	Os04g0220300
Bradi5g03540	-									MLOC_41224	
Bradi5g03550	2DS_5341587									MLOC_15303	Os04g0220300
Bradi5g03560	-									MLOC_8686	Os12g0156200
Bradi5g03577	2DS_5292808									-	Os04g0220300
Bradi5g03600	2DS_5354706									MLOC_57508	Os04g0212200
Bradi5g03627	-									MLOC_9245	-
Bradi5g03632	-									-	Os04g0220300
Bradi5g03640	2DS_5375380									MLOC_14804	Os04g0223300
Bradi5g03662	-									-	Os04g0376200
Bradi5g03720	2DS_5358023									MLOC_25063	Os04g0212450
Bradi5g03767	-									MLOC_37709	Os04g0276600
Bradi5g03780										MLOC_63786	Os02g0271000
Bradi5g03800	-									MLOC_50364	
Bradi5g03820	-									MLOC_55406	
Bradi5g03840	-									MLOC_67288	
Bradi5g03850	2DS_5363769									MLOC_45846	Os04g0274400
Bradi5g03882	2DS_5352525									MLOC_5849	-
Bradi5g03890	-									-	Os11g0289700
Bradi5g03897	2DS_5319959									MLOC_39510	Os10g0136400
Bradi5g03930	2DS_5319959									MLOC_39510	Os10g0136400
Bradi5g03960	2DS_5390396									MLOC_58539	Os04g0226340
	2DS_5385061										
Bradi5g03977	2DS_5343181									MLOC_61794	-
Bradi5g03990	-									MLOC_9079	
Bradi5g04000	2DS_5367475									MLOC_61793	Os04g0226800
	2DS_5369305										
Bradi5g04030	2DS_5389716									MLOC_67319	Os04g0227500
Bradi5g04057	2DS_5381312									MLOC_63016	Os04g0228100
Bradi5g04340	2DS_5390977									MLOC_37479	Os03g0856000
	2DS_5379098										
	2DS_5383642										
	2DS_5355519										
Bradi5g04540	2DS_5390977									MLOC_23980	Os04g0243700
Bradi5g04560	-									MLOC_56660	Os04g0244800
Bradi5g04577	2DS_5358467									MLOC_81817	Os04g0221600
Bradi5g04600	2DS_5341487										Os04g0250700
Bradi5g04630	2DS_5341846									MLOC_71561	Os10g0150300
	2DS_5374739									MLOC_81380	Os10g0150300
Bradi5g04640	2DS_5349408									MLOC_60943	Os05g0304900
Bradi5g04650	-									MLOC_10443	Os02g0288925
Bradi5g04660	2DS_5371750									MLOC_58466	Os02g0319800
Bradi5g04730	-									MLOC_55119	Os04g0258900
Bradi5g04750	-									MLOC_75680	Os01g0273900
Bradi5g05090	-									MLOC_5422	Os04g0284500

A5.2.2 Brachypodium zipper used for mining variation on 2DS CSS contigs in the syntenic interval. The interval was defined as in Table 5.4. The annotation shows how the sequence space was searched. In the *SSR* columns, microsatellite variation identified on the CSS contig which could be tested is in green. No variation is shaded red. In the *marker* column a polymorphic marker between the parent NILs is shaded green; monomorphic markers are in red. The same colours apply to the *UniGene* and *v3.3 cDNA* columns. Orange cells indicate SNPs identified in the UniGene data where only one sample called a SNP (low concordance). Orange shading in the *iSelect* column indicates the corresponding SNP marker was on the array but monomorphic between the parent NILs. The shading in the *axiom* column indicates a marker on the array corresponding to the 2DS CSS contig in that row had a SNP between the *Rht8* NIL and Paragon.

				KASP								
Rice		SSR		UniGene		v3.3		Array		Barley		BD
Gene	2DS hit	Tested	Marker	SNP	Marker	SNP	marker	iSelect	Axiom	Gene	Pos	Gene
Os04g0209200	2DS_5377037									MLOC_5957	15601547	Bradi5g03460
Os04g0212200	2DS_5354706									MLOC_57508	15592532	Bradi5g03600
Os04g0220300	2DS_5341587									MLOC_48019	17525526	Bradi5g03530
Os04g0223300	2DS_5375380									MLOC_14804	17591628	Bradi5g03640
Os04g0223500	-									MLOC_63757	17478781	Bradi3g20960
Os04g0224600	2DS_5363870									MLOC_74610	17556696	Bradi5g03697
Os04g0224900	2DS_5352598									MLOC_11990	15656891	Bradi5g03810
Os04g0226340	2DS_5390396									MLOC_58539	17416166	Bradi5g03960
Os04g0226800	2DS_5323988									MLOC_61793	17392210	Bradi5g04000
	2DS_5389048											
Os04g0227200	2DS_5367475									MLOC_61793	17392210	Bradi5g04000
	2DS_5369305											
Os04g0227500	2DS_5389716									MLOC_67319	17502873	Bradi5g04030
Os04g0228000	2DS_5381312									MLOC_63015	17699255	Bradi5g04050
Os04g0228100	2DS_5381312									MLOC_63016	17703692	Bradi5g04057
Os04g0229100	2DS_5318296									MLOC_48245	19071820	Bradi5g04130
Os04g0233400	2DS_5390752									MLOC_56278	18975015	
Os04g0244400	-									MLOC_69463	19016685	Bradi5g04550
Os04g0244800	2DS_5366894									MLOC_56660	19027873	Bradi5g04560
Os04g0252200	2DS_5364728									MLOC_4350	19440192	Bradi5g04673
Os04g0252400	2DS_5388494									MLOC_72777	19455074	Bradi5g04686
Os04g0255600	-									MLOC_55120	19945311	Bradi4g07480
Os04g0258900	2DS_5390725									MLOC_55119	19928531	Bradi5g04730
Os04g0266400	2DS_5390725									MLOC_4181	20626998	Bradi5g05225

A5.2.3: Rice zipper used for mining variation on 2DS CSS contigs in the syntenic interval. The interval was defined as in Table 5.4. The annotation shows how the sequence space was searched. In the *SSR* columns, microsatellite variation identified on the CSS contig which could be tested is in green. No variation is shaded red. In the *marker* column a polymorphic marker between the parent NILs is shaded green; monomorphic markers are in red. The same colours apply to the *UniGene* and *v3.3 cDNA* columns. Orange cells indicate SNPs identified in the UniGene data where only one sample called a SNP (low concordance). Orange shading in the *iSelect* column indicates the corresponding SNP marker was on the array but monomorphic between the parent NILs. The shading in the *axiom* column indicates a marker on the array corresponding to the 2DS CSS contig in that row had a SNP between the *Rht8* NIL and Paragon.

MIPS Genome Zipper 2015							iSelect array														
Marker	cM	Bd	Os	Contig - marker	Contig - Bd	Contig - Os	BS code	Pseudo	Ch	cM	HV	BD	OS	CD	RIL4	TB	T1	T2	T3	SB	S1	S2	S3
GDEEGVY01B0QLN	30.99	Bradi5g03460	Os04g0209200	2DS_5377037	2DS_5377037	2DS_5377037	BS00123480	929466	2Dx	12.39	-	-	-	NC	BB	NC	BB	BB	BB	BB	NC	NC	NC
contig21659	31.27	-	-	2DS_5318940	-	-	BS00120582	6154698	2Dx	12.87	-	-	-	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
contig10675	31.41	-	-	2DS_5375260	-	-	BS00120104	7783207	2Dx	12.31	2	22260307	-	AA	NC	AA	NC	NC	NC	NC	AA	AA	AA
contig19895	31.41	Bradi5g03477	Os04g0209300	2DS_5354001	2DS_5368504	2DS_3766997	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	Bradi5g03500	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	Bradi5g03530	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	Os04g0252000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F5XZDLF01C5J20	32.41	-	-	2DS_5340329	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F5XZDLF02IRQC9	33.83	-	-	2DS_5325019	-	-	BS00122201	-	-	-	-	-	-	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
GA8KES401CIV9Z	35.34	-	-	-	-	-	BS00122273	1277507	2Dx	15.62	2	19442600	Bradi5g04673	-	AA	AA	AA	AA	AA	AA	AA	AA	AA
contig14543	35.34	Bradi5g04673	Os04g0252200	2DS_5366894	2DS_5366894	2DS_5366894	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	Bradi5g04686	Os04g0252400	-	2DS_5364728	2DS_5364728	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	Bradi5g04710	Os04g0261400	-	2DS_5389857	2DS_5389857	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

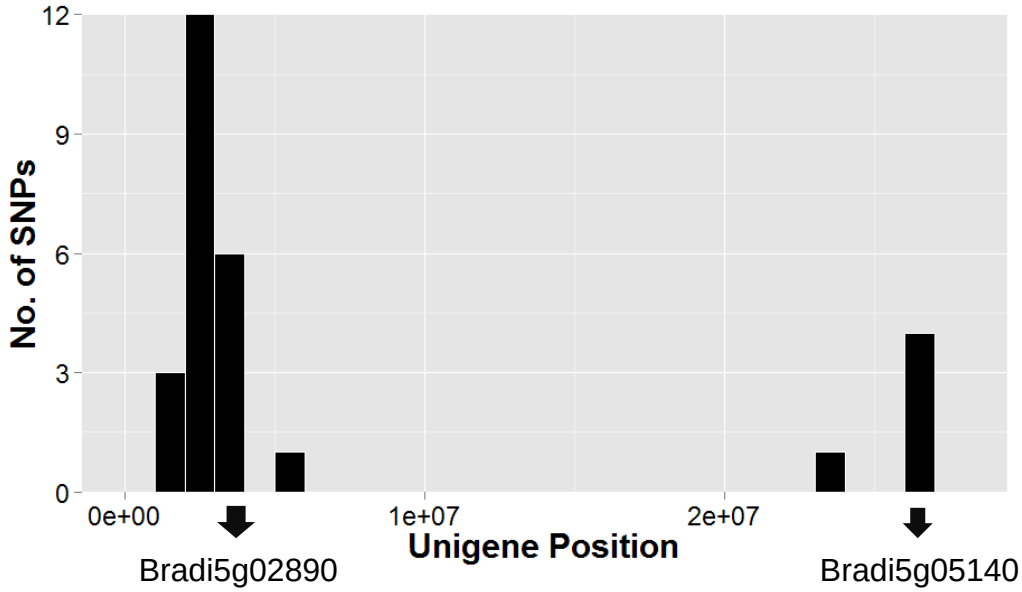
A5.3: Genome zipper based on IWGSC-2 and 90K iSelect array data (Wang et al., 2014a) compiled by MIPS and downloaded from URGI in March 2015 (URGI, 2015a). New 2DS CSS contigs anchored into the zipper that were searched for SSRs are shaded grey.

	SNP	Ch	Pos	Contig	HV		BD gene	OS gene	SB gene	CD				RIL4			
										P1	P3	P5	P7	P2	P4	P6	P8
Concordant	A_comp66086	2	1761095	2DS	-	-	Bradi5g01740	Os10g02210	Sb01g027135	T	T	T	T	C	C	C	C
	D_comp4534	2	2663320	2DS	2	12607798	Bradi5g02490	Os07g29220	Sb05g026900	T	T	T	T	A	A	A	A
	D_comp4534	2	2663320	2DS	2	12626022	Bradi5g02490	Os07g29220	Sb05g026900	G	G	G	G	C	C	C	C
	D_comp4534	2	2663320	2DS	2	12625891	Bradi5g02490	Os07g29220	Sb05g026900	C	C	C	C	T	T	T	T
	D_comp3280	2	2672958	2DS	2	11079891	Bradi5g02510	Os09g34130	Sb01g050630	A	A	A	A	G	G	G	G
	D_comp3280	2	2672958	2DS	2	11079891	Bradi5g02510	Os09g34130	Sb01g050630	C	C	C	C	T	T	T	T
	D_comp713	2	3150373	2DS	2	14137710	Bradi5g02890	Os12g17910	Sb09g014430	A	A	A	A	G	G	G	G
	D_comp554514	-	-	2DS	-	-	-	Os09g34130	-	C	C	C	C	T	T	T	T
Discordant	B_comp6379	2	5724062	2BS	2	19018455	Bradi5g04550	Os04g17050	Sb06g003270	T	T	A	-	T	T	T	A
	B_comp7893	2	5725742	2BS	2	19028489	Bradi5g04560	Os04g17100	Sb06g003280	G	C	G	G	G	C	G	G
	D_comp63460	2	5725896	2BS	2	19028525	Bradi5g04560	Os04g17100	Sb06g003280	T	C	C	T	T	T	T	C
	B_comp7893	2	5725742	2BS	2	19028489	Bradi5g04560	Os04g17100	Sb06g003280	G	C	G	G	G	C	G	G
	A_comp13702	4	35978921	2AS	2	19028516	Bradi5g04560	Os04g17100	Sb06g003280	C	C	C	C	T	C	C	C
	A_comp19945	2	5746743	2AS	2	18947226	Bradi5g04570	Os02g01170	Sb06g003290	A	T	T	T	A	T	T	A
	B_comp11282	2	5762103	2BS	-	-	Bradi5g04580	Os04g14510	Sb06g003300	T	T	T	T	C	T	T	T
	B_comp11282	2	5762103	2DS	-	-	Bradi5g04580	Os04g14510	Sb06g003300	A	G	A	A	A	G	G	A
	B_comp11282	2	5762103	2DS	-	-	Bradi5g04580	Os04g14510	Sb06g003300	G	A	G	G	G	A	A	G
	B_comp11282	2	5762103	2BS	-	-	Bradi5g04580	Os04g14510	Sb06g003300	C	C	A	A	A	A	A	C

A5.4: Prioritising high-confidence variants in SNPs between the parent NILs in the UniGene dataset according to concordance. The base call at each SNP position (Pos) in each of the CD (P1/3/5/7) and RIL4 samples (P2/4/6/8) is shown. The sample details are in Table 5.1. The UniGenes were annotated with information from syntenic species: HV: the barley chromosome and position; the Brachypodium (BD), rice (OS) and Sorghum (SB) genes.

												CD				RIL4		
SNP	Type	Validation	ID	Ch	Pos	Contig	HV	BD gene	OS gene	SB gene	P1	P3	P5	P7	P2	P4	P6	P8
D_comp554514_c0_seq1:82	non-hom	polymorphic	16	-	-	2DS_5364728	-	-	-	-	C	C	C	C	T	T	T	T
D_comp132046_c0_seq1:407	non-hom	polymorphic	50	-	-	2DS_4514573	-	-	-	-	G	G	G	G	A	A	A	A
D_comp16024_c0_seq50:546	non-hom	polymorphic	53	-	-	2DS_5330846	7	544549807	-	-	G	G	G	G	T	T	T	T
D_comp16925_c0_seq1:2569	non-hom	polymorphic	55	2	26162610	2DS_5375260	2	22256328	Bradi5g05140	Os03g58960	Sb06g004040	G	G	G	G	T	T	T	T
D_comp3280_c0_seq1:765	non-hom	polymorphic	57	2	2672958	2DS_5358861	2	11079891	Bradi5g02510	Os09g09320	Sb01g050630	A	A	A	A	G	G	G	G
D_comp3280_c0_seq1:875	non-hom	polymorphic	58	2	2672958	2DS_5358861	2	11079891	Bradi5g02510	Os09g09320	Sb01g050630	T	T	T	T	A	A	A	A
D_comp3280_c0_seq1:886	non-hom	polymorphic	59	2	2672958	2DS_5358861	2	11079891	Bradi5g02510	Os09g09320	Sb01g050630	C	C	C	C	T	T	T	T
D_comp375913_c0_seq1:199	non-hom	polymorphic	63	7	413600	2DS_5359909	-	-	-	Os10g21560	-	A	A	A	A	G	G	G	G
D_comp4534_c0_seq1:417	non-hom	polymorphic	66	2	2663320	2DS_5362023	2	12625891	Bradi5g02490	Os07g29220	Sb05g026900	C	C	C	C	T	T	T	T
D_comp4534_c0_seq1:570	non-hom	polymorphic	67	2	2663320	2DS_5362023	2	12626022	Bradi5g02490	Os07g29220	Sb05g026900	G	G	G	G	C	C	C	C
D_comp4534_c0_seq1:615	non-hom	polymorphic	68	2	2663320	2DS_5362023	2	12607798	Bradi5g02490	Os07g29220	Sb05g026900	T	T	T	T	A	A	A	A
D_comp4534_c0_seq1:661	non-hom	polymorphic	69	2	2663320	2DS_5362023	2	12607844	Bradi5g02490	Os07g29220	Sb05g026900	A	A	A	A	G	G	G	G
D_comp713_c0_seq1:154	non-hom	polymorphic	72	2	3150373	2DS_5343763	2	14138083	Bradi5g02890	Os12g17910	Sb09g014430	T	T	T	T	C	C	C	C
D_comp713_c0_seq1:245	non-hom	polymorphic	73	2	3150373	2DS_5343763	2	14137992	Bradi5g02890	Os12g17910	Sb09g014430	A	A	A	A	T	T	T	T
D_comp713_c0_seq1:527	non-hom	polymorphic	75	2	3150373	2DS_5343763	2	14137710	Bradi5g02890	Os12g17910	Sb09g014430	A	A	A	A	G	G	G	G
D_comp79633_c0_seq1:189	non-hom	polymorphic	76	-	-	2DS_5330846	-	-	-	-	-	T	T	T	T	C	C	C	C
A_comp20610_c0_seq1:1340	non-hom	monomorphic	46	2	23948502	2DS_5316996	-	-	Bradi1g59920	Os04g25550	Sb01g002390	T	T	T	T	C	C	C	C
A_comp66086_c0_seq1:116	non-hom	monomorphic	49	2	1761095	2DS_5335275	-	-	Bradi5g01740	Os10g02210	Sb01g027135	T	T	T	T	C	C	C	C
D_comp14446_c0_seq1:758	non-hom	monomorphic	51	5	3584417	2DS_5366459	-	-	Bradi4g04380	Os01g33160	Sb03g027060	C	C	C	C	G	G	G	G
D_comp16024_c0_seq55:350	non-hom	monomorphic	52	-	-	2DS_5330846	-	-	-	-	-	G	G	G	G	T	T	T	T
D_comp4534_c0_seq1:1233	non-hom	monomorphic	37	2	2663320	2DS_5362023	2	12631702	Bradi5g02490	Os07g29220	Sb05g026900	G	G	G	G	A	A	A	A
D_comp8227_c1_seq4:432	non-hom	monomorphic	77	-	-	2DS_5390061	-	-	-	-	-	C	C	C	C	G	G	G	G
D_comp8227_c1_seq4:434	non-hom	monomorphic	78	-	-	2DS_5390061	-	-	-	-	-	T	T	T	T	G	G	G	G
A_comp15095_c0_seq1:378	hom	-	-	2	1019583	2DS_5382931	-	-	Bradi5g01130	Os04g05050	Sb06g001410	T	T	T	T	G	G	G	G
A_comp15095_c0_seq1:737	hom	-	-	2	1019583	2DS_5382931	2	4921311	Bradi5g01130	Os04g05050	Sb06g001410	C	C	C	C	T	T	T	T
A_comp51540_c0_seq1:2137	hom	-	-	7	16310365	2DS_5389018	2	4911166	Bradi3g18920	Os08g14620	Sb07g007750	G	G	G	G	C	C	C	C
D_comp16925_c0_seq1:1606	hom	-	-	2	26162610	2DS_5375260	-	-	Bradi5g05140	Os03g58960	Sb06g004040	A	A	A	A	G	G	G	G
D_comp16925_c0_seq1:1673	hom	-	-	2	26162610	2DS_5375260	2	22260954	Bradi5g05140	Os03g58960	Sb06g004040	G	G	G	G	A	A	A	A
D_comp16925_c0_seq1:2332	hom	-	-	2	26162610	2DS_5375260	2	22256447	Bradi5g05140	Os03g58960	Sb06g004040	G	G	G	G	A	A	A	A
D_comp239028_c0_seq1:42	hom	-	-	2	5197739	2DS_5381312	2	17700625	Bradi5g04050	Os04g15790	Sb06g003200	C	C	C	C	T	T	T	T
D_comp3280_c0_seq1:898	hom	-	-	2	2672958	2DS_5358861	2	11079891	Bradi5g02510	Os09g09320	Sb01g050630	G	G	G	G	C	C	C	C
D_comp34324_c1_seq1:1130	hom	-	-	2	2673590	2DS_5358861	2	9985736	Bradi5g02520	-	Sb03g009340	G	G	G	G	A	A	A	A
D_comp34324_c1_seq1:869	hom	-	-	2	2673590	2DS_5358861	2	9986360	Bradi5g02520	-	Sb03g009340	A	A	A	A	G	G	G	G
D_comp375913_c0_seq1:939	hom	-	-	7	413600	2DS_5359909	-	-	-	Os10g21560	-	C	C	C	C	T	T	T	T
D_comp4534_c0_seq1:669	hom	-	-	2	2663320	2DS_5362023	2	12607852	Bradi5g02490	Os07g29220	Sb05g026900	T	T	T	T	C	C	C	C
D_comp713_c0_seq1:2027	hom	-	-	2	3150373	2DS_5356495	-	-	Bradi5g02890	Os12g17910	Sb09g014430	G	G	G	G	C	C	C	C
D_comp713_c0_seq1:281	hom	-	-	2	3150373	2DS_5343763	2	14137791	Bradi5g02890	Os12g17910	Sb09g014430	G	G	G	G	T	T	T	T
D_comp713_c0_seq1:404	hom	-	-	2	3150373	2DS_5343763	2	14137833	Bradi5g02890	Os12g17910	Sb09g014430	G	G	G	G	A	A	A	A

A5.5.1: The concordant SNPs mapping to 2DS from the UniGene data. The base call at the SNP position (Pos) in each of the CD (P1/3/5/7) and RIL4 samples (P2/4/6/8) is shown. The UniGenes were aligned to the IWGSC CSS contigs (Contig) and annotated with information from syntenic species: HV: the barley chromosome and position; the Brachypodium (BD), rice (OS) and Sorghum (SB) genes. The SNP-type (varietal or homoeologous) was output from the PolyMarker alignments to Chinese Spring and the SNP validated as polymorphic or monomorphic between the parent NILs, RIL4 and CD.



A5.5.1 (continued): UniGene position along chromosome 2DS of the 38 concordant SNPs from the previous page, identified in the UniGene data.

													CD				RIL4			
Hv	SNP	Type	Validation	ID	Ch	Pos	Contig	HV	BD gene	OS gene	SB gene		P1	P3	P5	P7	P2	P4	P6	P8
	D_comp375913_c0_seq1:540	non-hom	polymorphic	15	7	413600	2DS_5359909	2	15266465		-		T	T	C	T	C	C	C	C
	A_comp445854_c0_seq1:89	non-hom	monomorphic	23	2	3201457	2DS_5383636	2	14555516	Bradi5g02910	Os04g12480	Sb06g002250	G	G	A	A	G	G	-	A
	B_comp23635_c0_seq9:668	non-hom	monomorphic	24	2	3283635	2BS_5182158	2	14707569	Bradi5g02980	Os04g12560	Sb06g002210	C	C	T	C	T	T	T	T
	B_comp12234_c0_seq1:272	non-hom	monomorphic	25	2	4359872	2BS_5245578	2	15592537	Bradi5g03600	Os04g13470	Sb06g002040	C	C	C	T	C	C	C	C
	B_comp44417_c0_seq1:810	non-hom	monomorphic	26	2	4040646	2BS_4748675	2	15602745	Bradi5g03460	Os04g13210	Sb06g002080	A	A	A	A	A	G	-	A
	A_comp382707_c0_seq1:502	non-hom	monomorphic	27	2	4035376	2AS_5189613	2	15606207	Bradi5g03460	Os04g13210	Sb06g002080	G	G	G	A	G	G	A	G
	A_comp160609_c0_seq1:143	non-hom	monomorphic	28	2	5078283	2AS_5279086	2	17405146	-	Os04g15630	Sb06g003140	A	A	A	C	A	C	A	-
	A_comp160609_c0_seq1:115	non-hom	monomorphic	30	2	5078283	2AS_5279086	2	17405173	-	Os04g15630	Sb06g003140	C	T	C	T	T	T	C	-
	D_comp80183_c0_seq1:1557	non-hom	monomorphic	31	2	4883795	2DS_5390004	2	17428855	Bradi5g03860	Os05g22940	Sb06g003090	A	G	A	A	A	A	A	A
	D_comp8291_c0_seq1:3322	non-hom	monomorphic	32	2	4892454	2BS_5189287	2	17438247	Bradi5g03860	Os05g22940	Sb06g003090	A	A	A	C	A	A	C	A
	B_comp1151_c0_seq6:3306	non-hom	monomorphic	22	2	4892454	2BS_5189287	2	17438254	Bradi5g03860	Os05g22940	Sb06g003090	A	A	A	A	G	G	A	G
	B_comp1151_c0_seq6:2676	non-hom	monomorphic	34	2	4892454	2DS_5390004	2	17439079	Bradi5g03860	Os05g22940	Sb06g003090	T	T	T	C	T	T	T	T
	D_comp8291_c0_seq1:1731	non-hom	monomorphic	35	2	4892454	2AS_5222839	2	17440369	Bradi5g03860	Os05g22940	Sb06g003090	A	G	G	G	G	G	G	A
	D_comp8291_c0_seq1:1101	non-hom	monomorphic	37	2	4892454	2DS_5390004	2	17440999	Bradi5g03860	Os05g22940	Sb06g003090	A	G	A	A	A	A	A	A
	B_comp1151_c0_seq6:529	non-hom	monomorphic	38	2	4892454	2BS_5189287	2	17441567	Bradi5g03860	Os05g22940	Sb06g003090	G	G	G	G	G	G	C	C
	A_comp41913_c0_seq1:594	non-hom	monomorphic	39	2	4578684	2BS_5209315	2	17556695	Bradi5g03700	Os04g14760	Sb06g001950	T	C	C	T	T	C	C	C
	A_comp12_c1_seq2:212	non-hom	monomorphic	40	5	12995345	2BS_5246648	2	17599947	Bradi4g08800	Os12g17600	Sb05g003480	T	T	C	C	C	-	C	C
	D_comp42657_c0_seq1:1664	non-hom	monomorphic	41	2	24278186	2DS_5353317	2	17680017	Bradi5g04020	Os07g22350	Sb06g003160	A	G	A	A	A	G	A	G
	D_comp713_c0_seq1:404	hom	-	-	2	3150373	2DS_5343763	2	14137833	Bradi5g02890	Os12g17910	Sb09g014430	G	G	G	G	A	A	A	A
	A_comp445854_c0_seq1:93	hom	-	-	2	3201457	2DS_5383636	2	14555516	Bradi5g02910	Os04g12480	Sb06g002250	C	C	C	C	A	A	A	A
	A_comp305508_c0_seq1:222	hom	-	-	4	13653782	2DS_5372861	2	14646631	Bradi4g19450	Os11g26990	Sb05g012520	G	G	A	A	G	G	-	A
	D_comp20331_c0_seq2:1045	hom	-	-	5	28342712	2BS_5245089	2	14647031	Bradi1g07640	Os11g26990	Sb01g007140	C	T	C	C	T	C	T	T
	B_comp1041_c0_seq2:813	hom	-	-	2	4337476	2BS_5245578	2	15588991	Bradi5g03590	Os07g34950	Sb06g002030	C	C	C	T	C	C	C	C
	A_comp508552_c0_seq1:550	hom	-	-	2	4041209	2AS_5189613	2	15601647	Bradi5g03460	Os04g13210	Sb06g002080	G	C	C	C	G	G	G	G
	B_comp49772_c0_seq1:412	hom	-	-	2	4033280	2BS_4748675	2	15609211	Bradi5g03460	-	Sb06g002080	G	G	G	G	G	G	-	A
	B_comp5583_c0_seq1:311	hom	-	-	2	4777377	2AS_5285493	2	15660661	Bradi5g03810	Os04g14790	Sb06g003053	A	A	A	G	A	A	A	A
	A_comp123738_c0_seq3:354	hom	-	-	2	5167511	2DS_5369305	2	17392958	Bradi5g04000	Os04g15660	Sb06g003150	T	T	G	G	G	G	T	G
	D_comp122657_c0_seq1:653	hom	-	-	2	5077744	2BS_5242253	2	17405173	-	Os04g15660	Sb06g003140	C	C	C	G	C	G	C	C
	D_comp244592_c0_seq1:896	hom	-	-	2	5071645	2DS_5390396	2	17416791	Bradi5g03960	Os04g15660	Sb06g003120	A	A	G	G	A	A	-	G
	B_comp13811_c0_seq3:282	hom	-	-	2	5066871	2BS_5245210	2	17422790	Bradi5g03950	Os08g01859	Sb01g002970	T	T	T	T	C	T	T	C
	B_comp1151_c0_seq6:3082	hom	-	-	2	4892454	2BS_5189287	2	17438570	Bradi5g03860	Os05g22940	Sb06g003090	G	A	G	G	G	G	G	G
	B_comp1151_c0_seq6:2871	hom	-	-	2	4892454	2DS_5390004	2	17438884	Bradi5g03860	Os05g22940	Sb06g003090	C	T	T	T	T	T	T	T
	D_comp8291_c0_seq1:2001	hom	-	-	2	4892454	2AS_5222839	2	17440099	Bradi5g03860	Os05g22940	Sb06g003090	G	G	G	G	G	A	G	G
	B_comp8712_c0_seq5:123	hom	-	-	4	39958845	7DL_3360178	2	17443879	Bradi1g75720	Os04g14654	Sb03g025460	A	A	A	A	A	A	A	G
	A_comp61031_c0_seq10:1178	hom	-	-	7	13135096	2AS_5260516	2	17555209	Bradi3g21520	Os10g01380	Sb01g027460	C	C	C	C	C	C	C	T
	D_comp6_c0_seq1:238	hom	-	-	2	5280233	2DS_5339156	2	17599962	Bradi5g04080	Os12g19394	Sb05g003480	C	-	C	C	T	T	C	C
	B_comp30034_c0_seq1:146	hom	-	-	2	4859087	2BS_5209458	2	17674674	Bradi5g03840	-	Sb06g006100	C	C	T	-	C	C	C	T
	D_comp239028_c0_seq1:42	hom	-	-	2	5197739	2DS_5381312	2	17700625	Bradi5g04050	Os04g15790	Sb06g003200	C	C	C	C	T	T	T	T
	A_comp3833_c0_seq3:727	hom	-	-	2	10437335	2AS_5239327	2	17709330	-	Os07g42950	Sb02g039650	T	T	T	T	T	T	C	T

A5.5.2: The SNPs mapping to the barley chromosome 2:14,500,000-18,000,000 from the UniGene data. The base call at the SNP position (Pos) in each of the CD (P1/3/5/7) and RIL4 samples (P2/4/6/8) is shown. The UniGenes were aligned to the IWGSC CSS contigs (Contig) and annotated with information from syntenic species: HV: the barley chromosome and position; the Brachypodium (BD), rice (OS) and Sorghum (SB) genes. The SNP-type (varietal or homoeologous) was output from the PolyMarker alignments to Chinese Spring and the SNP validated as polymorphic or monomorphic between the parent NILs, RIL4 and CD.

		SNP	Type	Validation	ID	Ch	Pos	Contig		HV	BD gene	OS gene	SB gene	CD				RIL4			
														P1	P3	P5	P7	P2	P4	P6	P8
Relaxed		D_comp132046_c0_seq2:176	non-hom	polymorphic	21	2	1004884	2DS_5353487	-	-	-	-	-	G	A	G	G	A	A	A	A
Bd		D_comp3280_c0_seq1:727	non-hom	polymorphic	27	2	2672958	2DS_5358861	2	11079891	Bradi5g02510	Os09g09320	Sb01g050630	A	A	A	A	G	-	G	G
		A_comp25384_c0_seq1:1251	non-hom	monomorphic	1	2	5219313	2AS_5189660	-	-	Bradi5g04070	Os04g20164	Sb06g004290	A	G	G	A	-	A	A	A
		A_comp32519_c0_seq1:2281	non-hom	monomorphic	2	2	5762347	2AS_5199264	-	-	Bradi5g04580	Os04g14510	Sb06g003300	A	A	A	A	A	G	G	G
		B_comp31451_c0_seq3:355	non-hom	monomorphic	3	2	25682123	2BL_1035770	-	-	Bradi5g04790	Os04g25360	Sb09g000590	G	G	G	G	A	A	-	G
		B_comp3317_c0_seq2:4825	non-hom	monomorphic	4	6	117789	6DS_2071140	6	4404549	Bradi5g04570	Os02g01170	Sb06g003290	A	A	A	A	A	G	G	A
		B_comp6236_c0_seq1:1951	non-hom	monomorphic	5	6	120977	6BS_3045421	6	4398166	Bradi5g04570	Os02g01170	Sb06g003290	A	A	A	A	A	G	A	G
		B_comp6379_c0_seq1:997	non-hom	monomorphic	6	2	5724062	2BS_5242528	2	19020870	Bradi5g04550	Os04g17050	Sb06g003270	T	C	T	T	C	C	T	C
		B_comp6848_c0_seq2:1345	non-hom	monomorphic	7	2	25819499	2BL_8003360	2	307488738	Bradi5g04860	Os04g24170	-	-	A	A	C	C	C	C	C
		B_comp34971_c0_seq1:614	non-hom	monomorphic	8	2	4035383	2BS_4748675	2	15606328	Bradi5g03460	Os04g13210	Sb06g002080	G	G	A	A	G	G	G	G
		A_comp15355_c0_seq1:1664	non-hom	monomorphic	9	2	4777377	2AS_5285493	2	15660535	Bradi5g03810	Os04g14790	Sb06g003053	T	G	G	G	T	T	G	T
		A_comp25384_c0_seq1:480	hom	-	-	2	5219313	2AS_5209763	-	-	Bradi5g04070	Os04g20164	Sb06g004290	T	T	C	T	T	T	C	C
		B_comp11282_c0_seq4:2331	hom	-	-	2	5762103	2BS_5134070	-	-	Bradi5g04580	Os04g14510	Sb06g003300	T	T	T	T	T	T	A	A
		B_comp14433_c0_seq2:3252	hom	-	-	2	5900322	2AS_3262511	2	19521200	Bradi5g04670	Os04g18010	Sb06g003570	A	A	G	G	G	G	G	G
		B_comp3317_c0_seq1:3549	hom	-	-	2	5746743	2DS_5347513	2	18949503	Bradi5g04570	Os02g01170	Sb06g003290	C	C	T	T	C	C	C	C
		B_comp5402_c0_seq1:1892	hom	-	-	2	5219313	2AS_5209763	-	-	Bradi5g04070	Os04g20164	Sb06g004290	A	A	A	A	G	A	A	G
		B_comp7799_c0_seq9:2320	hom	-	-	3	16203222	5BS_2281486	-	-	Bradi5g03410	Os05g51070	Sb03g027400	G	G	A	A	A	A	A	A
		B_comp12234_c0_seq1:2856	hom	-	-	2	4359872	2AS_5269383	2	7986846	Bradi5g03600	Os04g13470	Sb06g002040	C	A	C	C	A	A	C	A
		A_comp24084_c0_seq2:413	hom	-	-	2	4744998	2BL_8087214	4	516423082	Bradi5g03770	Os04g20810	Sb06g006110	T	C	T	C	T	T	T	T
Informed		D_comp272596_c0_seq1:402	non-hom	polymorphic	8	5	1456476	2DS_5388088	-	-	Bradi4g02250	Os04g12560	Sb06g002210	G	-	G	G	A	A	G	A
		D_comp34324_c1_seq1:875	non-hom	monomorphic	5	2	2673590	2DS_5358861	2	9986354	Bradi5g02520	-	Sb03g009340	T	C	C	C	T	T	T	T
		B_comp38578_c0_seq2:793	non-hom	monomorphic	6	2	5891361	2DS_5371750	2	19492161	Bradi5g04660	Os02g21490	Sb08g003460	A	T	T	T	A	A	T	A
		D_comp211584_c0_seq2:332	non-hom	monomorphic	7	-	-	2DS_5375625	-	-	-	-	-	C	C	C	T	T	T	-	C
		A_comp17794_c0_seq1:575	non-hom	monomorphic	9	2	2514686	2DS_5388557	2	10041590	Bradi5g02390	Os04g01240	Sb06g000390	G	A	G	G	A	G	A	A
		D_comp3280_c0_seq1:933	hom	-	-	2	2672958	2DS_5358861	2	11079891	Bradi5g02510	Os09g09320	Sb01g050630	C	C	C	C	A	C	A	A
		D_comp6_c0_seq1:238	hom	-	-	2	5280233	2DS_5339156	2	17599962	Bradi5g04080	Os12g19394	Sb05g003480	C	-	C	C	T	T	C	C
		A_comp685_c0_seq1:1040	hom	-	-	2	3150373	2DS_5343763	2	14137475	Bradi5g02890	Os12g17910	Sb09g014430	C	C	T	T	C	C	T	T
		A_comp685_c0_seq1:662	hom	-	-	2	3150373	2DS_5343763	2	14137475	Bradi5g02890	Os12g17910	Sb09g014430	A	G	A	A	A	G	G	G
		A_comp685_c0_seq1:671	hom	-	-	2	3150373	2DS_5343763	2	14137475	Bradi5g02890	Os12g17910	Sb09g014430	A	G	G	G	A	A	A	G
		B_comp3317_c0_seq1:3549	hom	-	-	2	5746743	2DS_5347513	2	18949503	Bradi5g04570	Os02g01170	Sb06g003290	C	C	T	T	C	C	C	C

A5.5.3: The SNPs from the UniGene data identified from synteny to Brachypodium (Bd), relaxing stringency of concordance criteria (Relaxed) and retrospectively using the delimited 2D interval on the ordered v3.3 cDNAs (Informed). The base call at the SNP position (Pos) in each of the CD (P1/3/5/7) and RIL4 samples (P2/4/6/8) is shown. The UniGenes were aligned to the IWGSC CSS contigs (Contig) and annotated with information from syntenic species: HV: the barley chromosome and position; the Brachypodium (BD), rice (OS) and Sorghum (SB) genes. The SNP-type (varietal or homoeologous) was output from the PolyMarker alignments to Chinese Spring and the SNP validated as polymorphic or monomorphic between the parent NILs, RIL4 and CD.

	BS code	Type	Validation	ID	Pseudo		Ch	Arm	cM	Contig		HV		BD gene	OS gene	CD	RIL4	MARA	TB	T1	T2	T3	SB	S1	S2	S3
2DS	BS00049514	non-hom	polymorphic	4iS	-	-	-	-	-	2DS_5389775	2	14044468		-	-	AB	BB	BB	AB	BB	BB	BB	BB	AB	AB	AB
	BS00181365	non-hom	polymorphic	11iS	-	-	-	-	-	2DS_5343186	-	-		-	-	AB	AA	AA	AB	AA	AA	AA	AA	AB	AB	AB
	BS00134231	non-hom	polymorphic	5iS	-	-	-	-	-	2DS_5316382	-	-		-	-	AB	BB	BB	AB	BB	BB	BB	BB	AB	AB	AB
	BS00135852	non-hom	monomorphic	23iS	2DS	576661	2A	-	0	2DS_5335275	-	-		-	-	AA	AB	AB	AA	AA	AA	AA	AA	BB	AA	AA
	BS00132409	non-hom	monomorphic	24iS	-	-	-	-	-	2DS_5389158	-	-		-	-	BB	AB	AB	BB	BB	BB	BB	BB	BB	BB	BB
	BS00132410	non-hom	monomorphic	23iS	-	-	-	-	-	2DS_5389158	-	-		-	-	BB	AB	AB	BB	BB	BB	BB	BB	BB	BB	BB
	BS00181366	non-hom	monomorphic	24iS	-	-	-	-	-	2DS_3409803	-	-		-	-	BB	AB	AB	BB	AB	AB	AB	AB	BB	BB	BB
	BS00183818	non-hom	monomorphic	23iS	2DS	451435	-	-	-	2DS_5328979	-	-		-	-	BB	AB	AB	NC	BB	BB	BB	BB	AB	BB	BB
	BS00125131	non-hom	monomorphic	2iS	-	-	6A	-	82	2DS_5354297	-	-		-	-	AB	BB	BB	AB	BB	BB	BB	BB	AB	AB	AB
	BS00068050	non-hom	monomorphic	25iS	2DS	261222	6A	-	141	2DS_5352291	-	-		Bradi5g01070	-	AB	BB	BB	AB	AB	AB	AB	AB	BB	AB	AB
	BS00158900	hom	-	-	-	-	-	-	-	2DS_169559	2	12700112		-	-	BB	AB	AB	AB	AB	AB	AB	AB	BB	BB	BB
	BS00120319	hom	-	-	-	-	3B	-	100	2DS_5379411	2	14075742		-	-	AA	AB	AB	AA	AA	AA	AA	AA	AA	AA	AA
	BS00121305	hom	-	-	-	-	-	-	-	2DS_5389775	2	14096730		-	-	AA	AB	AB	AA	AB	AB	AB	AB	AA	AA	AA
	BS00123673	hom	-	-	-	-	2B	S	49	2DS_5333085	-	-		Bradi1g20185	-	AA	AB	AB	AA	NC	NC	NC	NC	AA	AA	AA
	BS00177016	hom	-	-	2DS	576661	2A	-	26	2DS_5335275	-	-		-	-	AA	AB	AB	AA	AA	AA	AA	AA	BB	AA	AA
	BS00004040	hom	-	-	2DS	843360	2B	-	68	2DS_5356495	-	-		Bradi5g02890	Os12g17910	AA	AB	AB	AA	AB	AB	AB	AB	AA	AA	AA
	BS00114244	hom	-	-	2DS	6747374	-	-	-	2DS_5390826	-	-		-	-	AA	AB	AB	AA	AA	AA	AA	AA	AA	AA	AA
	BS00142235	hom	-	-	-	-	-	-	-	2DS_2179719	-	-		Bradi5g02980	-	BB	AB	AB	BB	AB	AB	AB	AB	BB	BB	BB
	BS00126889	hom	-	-	2DS	845361	-	-	-	2DS_5343763	-	-		-	-	BB	AB	AB	BB	AB	AB	AB	AB	BB	BB	BB
	BS00181118	hom	-	-	2DS	845361	-	-	-	2DS_5343763	-	-		Bradi5g02890	Os12g17910	BB	AB	NC	BB	AB	AB	AB	AB	BB	BB	BB
Synteny	BS00164872	non-hom	polymorphic	52iS	2BS	1552551	1A	L	71	2BS_4748675	2	15603161		Bradi5g03460	Os04g13220	AB	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
	BS00136588	non-hom	monomorphic	1iS	2AS	1568103	1A	L	142	2AS_5249183	2	18951624		-	-	AB	BB	BB	BB	AB	AB	AB	AB	BB	BB	BB
	BS00162990	non-hom	monomorphic	27iS	2BS	1368763	-	-	-	2BS_5228991	2	9988309		Bradi5g02520	-	AB	BB	BB	BB	AB	AB	AB	AB	BB	BB	BB
	BS00148219	hom	-	-	-	-	-	-	-	2AS_5225942	2	12814931		-	-	BB	AB	AB	BB	AB	AB	AB	AB	BB	BB	BB
	BS00136528	hom	-	-	2	1651879	-	-	-	2AS_5198133	2	19494175		-	-	BB	AB	AB	AB	BB	BB	BB	BB	AB	AB	AB
	BS00179680	hom	-	-	2	10318562	6A	-	16	2AS_5279345	2	22257798		Bradi5g05140	Os03g58960	BB	AB	AB	BB	AB	AB	AB	AB	BB	BB	BB

A5.6.1: iSelect SNPs between parent NILs, selected on 2DS and synteny criteria as described in the text. Data from the short and tall recombinants was added later when available (shaded columns). Columns left to right: BS code: unique (Bristol code) identifier from the array; Type: the SNP type (homoeologous or varietal) as returned from PolyMarker alignments; Validation: the result of markers screened on RIL4 and CD and the ID of those markers tested; Pseudo: the wheat pseudomolecule chromosome and position from Martin Trick's UniGene reference; Akhunov assignment of chromosome (Ch), long (L) or short (S) arm (Arm) and genetic map position on the chromosome (cM); Contig: the IWGSC CSS best hit; HV: the barley chromosome and position; the Brachypodium (BD) and rice (OS) genes; TB: pooled DNA from T1/T2/T3 recombinants; SB = pooled DNA from S1/S2/S3 recombinants (details in Methods). Missing data in the SNPs is represented by 'NC' or otherwise '-'.

	BS code	Type	Validation	ID	Pseudo	Ch	Arm	cM	Contig	HV	BD gene	OS gene	CD	RIL4	MARA	TB	T1	T2	T3	SB	S1	S2	S3
Maps to 2DS	BS00022234	non-hom	polymorphic	3iS	-	2Dx	-	10	1BS_3467454	-	-	-	BB	AA	AA	BB	AA	AA	AA	AA	BB	BB	BB
	BS00185568	non-hom	polymorphic	6iS	-	2Dx	-	9	2DS_5354297	-	-	-	AB	BB	BB	AB	BB	BB	BB	BB	AB	AB	AB
	BS00160347	non-hom	monomorphic	7iS	-	2Dx	-	9	4DS_2212368	-	-	-	BB	BB	BB	AB	BB	BB	BB	BB	AB	AB	AB
	BS00170377	non-hom	monomorphic	8iS	-	2Dx	-	83	1AS_1061324	-	-	-	BB	BB	AB	NC	AB	AB	AB	AB	BB	AB	BB
	BS00062567	hom	-	-	-	2Dx	S	83	2DS_5359082	-	-	-	AA	AA	AA	AA	AB	AB	AB	AB	AA	AB	AA
	BS00146442	hom	-	-	-	2Dx	-	120	7DS_3896931	-	-	-	AA	AA	AB	AA	AB	AB	AB	AB	AA	AB	AA
	BS00130142	hom	-	-	-	2Dx	-	11	2AS_5264907	-	-	-	AA	AB	AB	AA	AB	AB	AB	AB	AA	AA	AA
	BS00179561	hom	-	-	2DS 13307081	2Dx	S	12	2BS_5219567	2	22260578	-	BB	AB	AB	BB	AB	AB	AB	AB	BB	BB	BB
	BS00086387	hom	-	-	2DS 845361	2Dx	S	11	2DS_5343763	2	7780308	-	NC	BB	BB	AA	BB	BB	BB	BB	AA	AA	AA
	BS00167427	hom	-	-	-	2Dx	-	1	2BS_5237275	2	1556924	-	NC	NC	AB	BB	AB	AB	AB	AB	BB	BB	BB
Consensus parent & bulk	BS00149099	non-hom	polymorphic	13iS	2DL 15796111	2Dx	L	101	2DL_9907588	-	-	-	AB	BB	BB	AB	BB	BB	BB	BB	BB	AB	AB
	BS00181544	non-hom	polymorphic	81iS	3BL 3822976	3B	-	62	3B_10607087	3	32713948	-	AA	BB	BB	AA	AA	AA	AA	AA	AA	AA	AA
	BS00184801	non-hom	polymorphic	79iS	3BL 2997663	3B	-	63	3B_10415228	3	67316972	Bradi2g05400	AB	BB	BB	AB	AB	AB	AB	AB	AB	AB	AB
	BS00170624	non-hom	polymorphic	83iS	6AL 15139301	-	-	-	6AL_5751146	-	-	Bradi3g56250 Os02g49600	AB	BB	BB	AB	AB	AB	AB	AB	AB	AB	AB
	BS00140536	non-hom	polymorphic	85iS	-	6A	-	82	6AL_5769197	6	464689564	Bradi3g52900	AB	BB	BB	AB	AB	AB	AB	AB	AB	AB	AB
	BS00022840	non-hom	polymorphic	86iS	6AL 11957482	6A	L	85	6AL_5803912	6	484432757	-	BB	AB	AB	BB	BB	BB	BB	BB	BB	BB	BB
	BS00138919	non-hom	polymorphic	84iS	6AL 13074686	6A	L	95	6AL_5755158	6	508376254	-	BB	AB	AA	AB	AB	AB	AB	AB	AB	AB	AB
	BS00145849	non-hom	polymorphic	88iS	6AL 11547448	6A	-	99	6AL_5833694	6	504956346	-	BB	AB	AB	BB	BB	BB	BB	BB	BB	BB	BB
	BS00079019	non-hom	polymorphic	15iS	7BS 6331692	7B	S	63	7BS_3032904	-	-	-	AA	BB	AA	AB	AA	AA	AA	AA	BB	AA	BB
	BS00138407	non-hom	polymorphic	14iS	7BS 6359550	7B	S	57	7BS_196497	7	84879395	-	AA	BB	AA	AB	AA	AA	AA	AA	BB	AA	BB
	BS00139230	non-hom	polymorphic	17iS	7BS 6329405	7B	-	63	7BS_3089157	-	-	-	AB	AA	BB	AB	BB	BB	BB	BB	AA	BB	AA
	BS00078214	non-hom	monomorphic	87iS	6AL 1992266	-	-	-	6AL_5830987	6	513902803	-	AB	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
	BS00184703	non-hom	monomorphic	82iS	3BS 7593240	3B	-	58	3B_10616876	3	22553008	Bradi2g12950 Os01g27040	AB	BB	BB	AB	AB	AB	AB	AB	AB	AB	AB
	BS00054881	non-hom	monomorphic	16iS	-	7B	S	56	7BS_3083309	-	-	-	BB	AA	BB	AB	BB	BB	BB	BB	AA	BB	AA
	BS00069075	non-hom	monomorphic	20iS	7DS 4699478	-	-	-	7DS_3926568	-	-	-	BB	AA	BB	NC	BB	BB	BB	BB	AA	BB	AA
	BS00108793	non-hom	monomorphic	21iS	7DS 10602445	7D	S	26	7DS_3965491	-	-	-	BB	BB	AA	BB	AA	AA	AB	AA	BB	BB	BB
	BS00110124	non-hom	monomorphic	22iS	7DS 563993	7D	S	27	7DS_402033	-	-	-	BB	BB	AA	BB	AA	AA	AA	AA	BB	BB	BB
	BS00110642	non-hom	monomorphic	23iS	7DS 563993	7D	S	27	7DS_402033	-	-	-	BB	BB	AA	BB	AA	AA	AA	AA	BB	BB	BB
	BS00120144	non-hom	monomorphic	18iS	-	7D	S	26	7DS_3858653	-	-	-	BB	BB	AA	BB	AA	AA	AA	AA	BB	BB	BB
	BS00124796	non-hom	monomorphic	19iS	-	7D	S	26	7DS_3888509	-	-	-	BB	BB	AA	BB	AA	AA	AA	AA	BB	BB	BB
	BS00141000	hom	-	-	7BS 6660865	7B	S	57	7BS_31214417	7	71286944	-	AB	BB	AB	NC	AA	AA	AA	AA	BB	AA	BB
	BS00114693	hom	-	-	7BS 6660865	7B	S	58	7BS_3121441	-	-	-	BB	AA	BB	AB	BB	BB	BB	BB	AA	BB	AA
	BS00180215	hom	-	-	7BS 6660865	7B	S	58	7BS_3121441	-	-	-	BB	AA	BB	AB	BB	BB	BB	BB	AA	BB	AA
	BS00140239	hom	-	-	-	7D	S	26	7DS_3927231	7	5906165	-	BB	BB	AA	BB	AA	AA	AA	AA	BB	BB	BB
	BS00150047	hom	-	-	7DS 14624836	7D	S	23	7DS_1593880	-	-	-	BB	BB	AB	BB	AA	AA	AA	AA	BB	BB	BB
	BS00184266	hom	-	-	7DS 2164194	7D	S	23	7DS_3965769	7	5047058	-	BB	BB	NC	BB	AA	AA	AA	AA	BB	BB	BB
	BS00144643	hom	-	-	-	7B	S	56	7BS_3083309	-	-	-	NC	AA	AB	NC	BB	BB	BB	BB	AA	BB	AA
	BS00178320	hom	-	-	2AL 27081575	2Dx	-	103	2AL_6433389	-	-	-	NC	AA	AA	NC	AA	AA	AA	AA	AB	NC	AB

A5.6.2: iSelect SNPs between the short and tall bulks, selected on 2DS and consensus between parent and bulk data on SNPs outside 2DS. Columns left to right: BS code: unique (Bristol code) identifier from the array; Type: the SNP type (homoeologous or varietal) as returned from PolyMarker alignments; Validation: the result of markers screened on RIL4 and CD and the ID of those markers tested; Pseudo: the wheat pseudomolecule chromosome and position from Martin Trick's UniGene reference; Akhunov assignment of chromosome (Ch), long (L) or short (S) arm (Arm) and genetic map position on the chromosome (cM); Contig: the IWGSC CSS best hit; HV: the barley chromosome and position; the Brachypodium (BD) and rice (OS) genes; TB: pooled DNA from T1/T2/T3 recombinants; SB: pooled DNA from S1/S2/S3 recombinants (details in Methods). Missing data in the SNPs is represented by 'NC' or otherwise '-'.

Barley synteny	BS code	Type	Validation	ID	Pseudo		Ch	Arm	cM	Contig		HV	BD gene	OS gene	CD	RIL4	MARA	TB	T1	T2	T3	SB	S1	S2	S3
	BS00135173	hom	-	-	-	-	2B	S	52	2DS_5337269	2	14580548	-	-	AB	AB	AB	AB	NC	AB	AB	AB	AB	AB	AB
	BS00176894	hom	-	-	-	-	2B	-	52	2BS_5210120	2	14580548	-	-	AB	AB	AB	AB	NC	AB	AB	AB	AB	AB	AB
	BS00164872	non-hom	polymorphic	52iS	2BS	1552551	1A	L	71	2BS_4748675	2	15603161	Bradi5g03460	Os04g13220	AB	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
	BS00136588	non-hom	monomorphic	1iS	2AS	1568103	1A	L	142	2AS_5249183	2	18951624	-	-	AB	BB	BB	BB	AB	AB	AB	AB	BB	BB	BB
	BS00147383	hom	untested	-	2BS	1851462	2B	-	68	2BS_5244622	2	17392689	-	-	AB	AB	AB	AB	NC	AB	AB	AB	AB	BB	BB
	BS00183870	non-hom	untested	-	2AS	1409368	-	-	-	2AS_5222839	2	17442377	-	-	AB	AB	AB	AB	NC	AB	AB	AB	AB	AB	AB
	BS00116967	hom	untested	-	-	-	2A	-	53	2BS_5203253	2	17710852	Bradi1g21320	Os07g42950	AB	AB	AB	AB	NC	AB	AB	AB	AB	BB	NC
	BS00164249	hom	-	-	-	1556934	2A	S	53	2AS_5218343	2	19021208	-	-	AA	AA	AA	AA	NC	AA	AA	AA	AA	AA	AA
	BS00181570	non-hom	untested	-	-	-	2B	S	66	2BS_5183319	2	19471391	-	-	AB	AB	AB	AB	NC	AB	AB	AB	AB	AB	BB
	BS00120104	non-hom	monomorphic	-	2DS	7783207	2Dx	-	12	2DS_5375260	2	22260307	-	-	AA	NC	NC	NC	AA	NC	AA	AA	AA	NC	NC

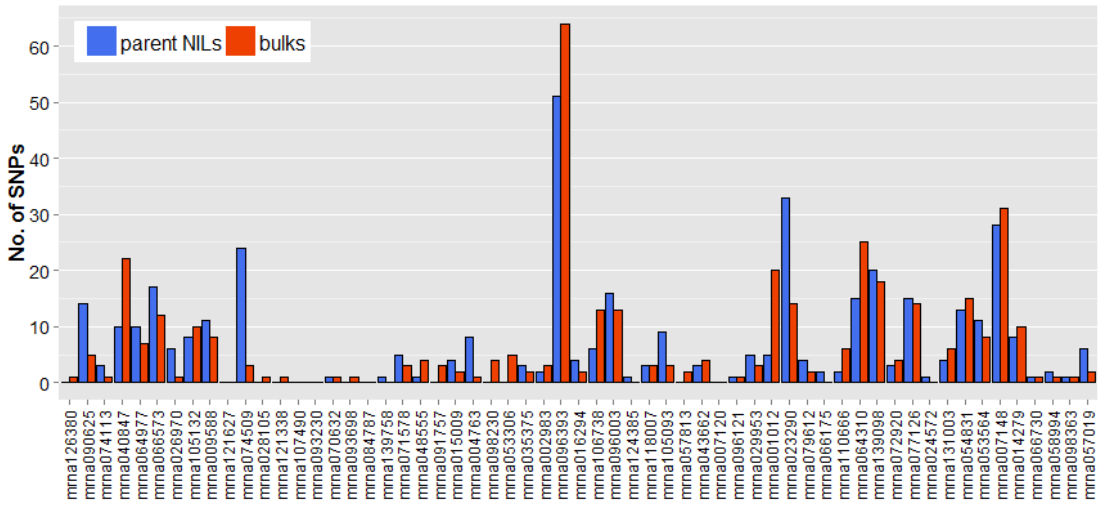
A5.6.3: iSelect SNPs in the barley syntenic interval. Columns left to right: BS code: unique (Bristol code) identifier from the array; Type: the SNP type (homoeologous or varietal) as returned from PolyMarker alignments; Validation: the result of markers screened on RIL4 and CD and the ID of those markers tested; Pseudo: the wheat pseudomolecule chromosome and position from Martin Trick's UniGene reference; Akhunov assignment of chromosome (Ch), long (L) or short (S) arm (Arm) and genetic map position on the chromosome (cM); Contig: the IWGSC CSS best hit; HV: the barley chromosome and position; the Brachypodium (BD) and rice (OS) genes; TB: pooled DNA from T1/T2/T3 recombinants; SB: pooled DNA from S1/S2/S3 recombinants (details in Methods). Missing data in the SNPs is represented by 'NC' or otherwise '-'.

	SNP	Ref	Var	Type	Validation	Marker	BFR	Inf	CO_1	CO_2	RA_1	RA_2	C_1	C_2	Contig	ID		HV	BD gene	2DS	Primer
2DS	mrna071136_396	t	C	non-hom	polymorphic	34	Infinity	RIL4	37	27	0.76	0	28	0	2AS_2899441	100	2	3995548	-	2DS_5340279	semispecific
	mrna035372_450	c	T	non-hom	polymorphic	41	9.33	RIL4	41	37	0.76	0.08	31	3	2AS_5265661	99.67	2	1460783	Bradi3g06027	2DS_5390412	semispecific
	mrna003675_756	g	C	non-hom	polymorphic	28	Infinity	RIL4	43	53	0.49	0	21	0	2AS_5223115	100	2	4370578	-	2DS_5334978	semispecific
	mrna083051_1080	c	T	non-hom	polymorphic	35	7.17	RIL4	103	123	0.23	0.03	24	4	2AS_5265661	100	2	1541445	-	2DS_5390412	semispecific
	mrna083050_40	g	A	non-hom	polymorphic	3	Infinity	RIL4	45	29	0.24	0	11	0	2DS_5390412	100	2	1541445	-	2DS_5390412	semispecific
	mrna034307_4498	g	A	non-hom	polymorphic	44	Infinity	RIL4	164	95	0.31	0	51	0	2AS_5247752	100	2	1370200	-	2DS_5311039	specific
	mrna070131_347	a	G	non-hom	polymorphic	47	Infinity	RIL4	122	91	0.57	0	70	0	2AS_5207752	99.38	7	596837756	-	2DS_5344849	semispecific
	mrna041371_249	c	T	non-hom	polymorphic	31	Infinity	RIL4	45	36	0.73	0	33	0	2AS_5188575	100	2	3867011	-	2DS_5352819	semispecific
	mrna051701_504	c	T	non-hom	polymorphic	49	Infinity	RIL4	262	266	0.43	0	113	0	2AS_5264433	98.63	2	3742954	Bradi5g01420	2DS_5387127	semispecific
	mrna120229_565	g	A	non-hom	polymorphic	38	Infinity	RIL4	68	68	0.66	0	45	0	2AS_5264433	100	2	3746497	-	2DS_5387127	semispecific
	mrna035372_1710	g	A	non-hom	polymorphic	42	Infinity	RIL4	136	95	0.24	0	32	0	2AS_5265661	99.67	2	1460783	Bradi3g06027	2DS_5390412	semispecific
	mrna042711_1045	g	A	non-hom	polymorphic	46	Infinity	RIL4	41	49	0.2	0	8	0	2BS_5158671	100	2	19492976	-	2DS_5371750	specific
Synteny	mrna047464_937	t	A	non-hom	monomorphic	47	Infinity	CD	21	22	0	0.26	0	12	2AS_5235474	99	2	22256333	Bradi5g05140	2DS_5375260	specific
	mrna057812_487	a	G	non-hom	monomorphic	50_2	Infinity	CD	33	41	0	0.25	0	15	2AS_5235188	99	2	19080187	Bradi4g40600	2DS_5390977	semispecific
	mrna035375_315	t	C	non-hom	monomorphic	46_2	7.26	RIL4	38	46	0.32	0.04	12	2	2AS_5277227	99.44	2	17496477	Bradi5g03850	2DS_5363769	semispecific
	mrna007632_834	c	G	non-hom	monomorphic	25_2	Infinity	CD	31	22	0	0.26	0	12	2BS_5161117	99.49	2	19511736	Bradi3g36320	2DS_5341683	semispecific
	mrna004339_1101	t	C	non-hom	monomorphic	11_2	7.54	CD	49	39	0.02	0.15	1	6	2BS_5242253	100	2	17404808	-	2DS_3730349	semispecific
	mrna113890_452	t	G	non-hom	monomorphic	37_2	Infinity	CD	70	74	0	0.27	0	39	2AS_5218356	100	2	19032963	-	2DS_5352693	semispecific

A5.7.1: Putative SNPs from v3.3 cDNA data selected on criteria of mapping to 2DS and synteny with barley and Brachypodium. Columns left to right: Ref: reference base at SNP position; Var: variant at SNP position; Validation: the result of markers screened on RIL4 and CD and the Marker name of those tested; BFR: calculated from RA_1/RA_2; Inf: the parent enriched at the SNP; CO: coverage, total number of reads mapping to the SNP; RA: ratio, relative contribution of reads that call the SNP from the total reads mapping to that position, C: count (coverage x ratio); Contig: the IWGSC CSS best hit; ID: % nucleotide identity of the gene to Contig; HV: barley chromosome and position; Bd: Brachypodium gene annotation; 2DS: 2DS contig for which PolyMarker designed primers to be specific to; Primer: outcome of specificity to 2DS contig.

High BFR outside 2DS	SNP	Ref	Var	Type	Validation	Marker	BFR	Inf	CO_1	CO_2	RA_1	RA_2	C_1	C_2	Contig	ID		HV	BD gene
	mrna084023_30	c	T	non-hom	polymorphic	11_2	26.2	RIL4	233	165	0.16	0.01	37	1	7BS_3121441	100	7	71287089	Bradi1g46980
	mrna053652_977	c	T	non-hom	monomorphic	25_2	22.04	RIL4	190	166	0.13	0.01	32	1	1BL_3805529	100	1	428574660	-
	mrna132275_73	g	T	non-hom	monomorphic	26_2	20.45	RIL4	256	56	0.73	0.04	187	2	1BS_3432767	100	-	-	Bradi2g33450
	mrna076686_337	g	A	non-hom	monomorphic	27_2	57.64	CD	1696	1788	0.01	0.3	9	877	1DL_2263471	100	-	-	Bradi4g24635
	mrna004363_1006	a	G	non-hom	monomorphic	1_2	79.25	RIL4	283	267	0.3	0	84	1	2AL_6416883	100	2	620908823	Bradi5g26730
	mrna028008_1105	a	T	non-hom	monomorphic	3_2	Infinity	CD	23	38	0	0.22	0	15	2AL_6413688	100	2	521463457	Bradi5g17070
	mrna040642_1336	a	G	non-hom	monomorphic	4_2	Infinity	CD	30	25	0	0.23	0	14	2AL_6358197	100	2	384884751	Bradi5g07930
	mrna116055_211	g	C	non-hom	monomorphic	5_2	Infinity	RIL4	21	27	0.27	0	16	0	2AL_6344313	100	2	413910011	Bradi5g09940
	mrna116326_470	t	G	non-hom	monomorphic	6_2	Infinity	CD	27	25	0	0.3	0	14	2AL_1213723	100	-	-	Bradi5g21570
	mrna021588_455	g	A	non-hom	monomorphic	8_2	19.11	RIL4	27	828	0.11	0.01	4	5	3B_10456072	100	3	528350533	Bradi2g59410
	mrna028532_741	g	C	non-hom	monomorphic	9_2	8.29	RIL4	176	316	0.15	0.02	36	6	3B_10633132	100	3	489637850	Bradi2g55180
	mrna044487_1070	a	T	non-hom	monomorphic	10_2	10.68	CD	70	54	0.01	0.15	1	12	3B_10410756	100	3	500192572	Bradi2g58050
	mrna052584_607	c	T	non-hom	monomorphic	12_2	9.28	CD	26	39	0.03	0.32	1	32	3B_10499179	100	3	337526827	-
	mrna092462_1537	g	T	non-hom	monomorphic	13_2	Infinity	RIL4	22	29	0.26	0	12	0	3B_10454372	100	3	154953373	Bradi2g10530
	mrna043262_1152	g	T	non-hom	monomorphic	14_2	Infinity	CD	26	35	0	0.23	0	16	5BL_10785963	100	5	513624725	Bradi1g06772
	mrna074551_265	c	T	non-hom	monomorphic	15_2	12.66	RIL4	209	168	0.3	0.02	63	4	5BL_10925653	100	5	450837532	-
	mrna081093_1293	t	G	non-hom	monomorphic	16_2	8.67	RIL4	60	39	0.67	0.08	40	3	5BL_10903966	100	5	519457307	-
	mrna085126_371	a	G	non-hom	monomorphic	17_2	Infinity	RIL4	57	37	0.26	0	35	0	5BL_10870740	100	5	476723446	-
	mrna118456_604	t	G	non-hom	monomorphic	18_2	11.28	RIL4	50	47	0.24	0.02	12	1	5BL_10903386	100	-	-	-
	mrna003152_848	t	G	non-hom	monomorphic	19_2	Infinity	RIL4	51	28	0.22	0	24	0	5DL_4564133	100	5	527895459	-
	mrna099437_158	g	C	non-hom	monomorphic	20_2	Infinity	CD	21	30	0	0.28	0	16	5DL_4560789	100	-	-	Bradi1g35960
	mrna003724_42	c	T	non-hom	monomorphic	21_2	72.39	RIL4	186	198	0.37	0.01	68	1	7AS_4246344	100	-	-	Bradi1g46137
	mrna063321_484	a	G	non-hom	monomorphic	22_2	16	CD	80	40	0.01	0.2	1	8	7AS_4249929	100	7	92407048	Bradi1g65740
	mrna020736_951	c	T	non-hom	monomorphic	23_2	6.46	RIL4	86	81	0.56	0.09	48	7	7BS_3095060	100	7	63878428	-
	mrna025486_860	a	T	non-hom	monomorphic	24_2	129.11	RIL4	304	223	0.58	0	176	1	7BS_3150559	100	7	34503295	-

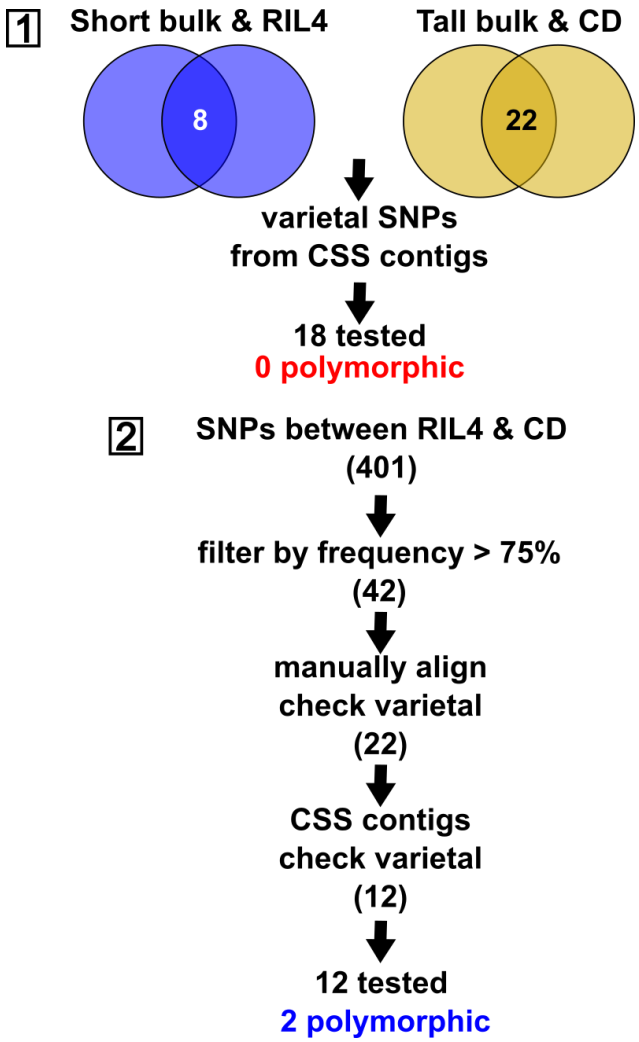
A5.7.2: Putative SNPs from v3.3 cDNA data selected on criteria of high BFR. Columns left to right: Ref: reference base at SNP position; Var: variant at SNP position; Validation: the result of markers screened on RIL4 and CD and the Marker name of those tested; BFR: calculated from RA_1/RA_2; Inf: the parent enriched at the SNP; CO: coverage, total number of reads mapping to the SNP; RA: ratio, relative contribution of reads that call the SNP from the total reads mapping to that position, C: count (coverage x ratio); Contig: the IWGSC CSS best hit; ID: % nucleotide identity of the gene to Contig; HV: barley chromosome and position; Bd: Brachypodium gene annotation.



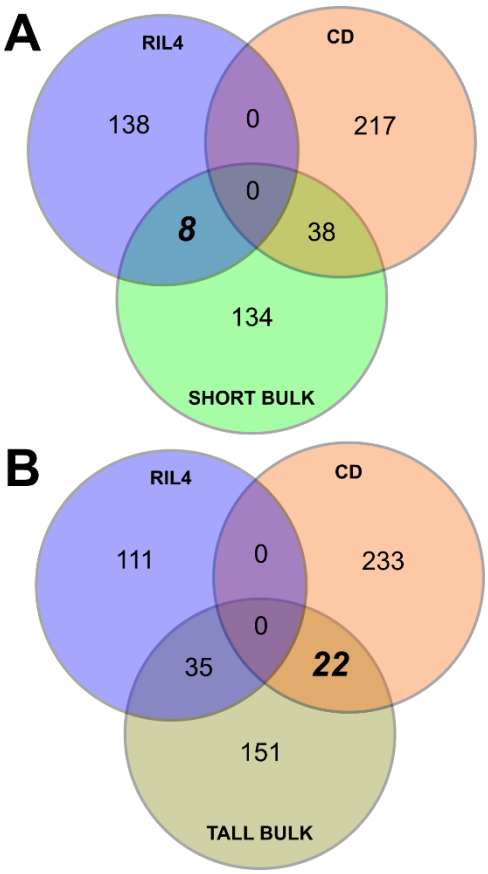
A5.8: Putative SNP distribution over the v3.3 cDNA 2D interval reference (a total of 59 cDNAs), with the genes ordered from left to right as they are anchored in the ordered v3.3 cDNA reference. The varietal SNPs between the parent NILs/ bulks and the reference are shown.

		Parent NILs				Bulks			
		Mapped reads			Coverage	Mapped reads			Coverage
Gene	Gene length	CD	RIL4	Combined		short	tall	Combined	
mrna126380	1056	44092	34935	79027	1273	14945	16693	31638	510
mrna090625	1015	18905	12844	31749	492	1778	2191	3969	61
mrna074113	732	3640	3238	6878	77	1463	1644	3107	35
mrna040847	2625	72138	51289	123427	4942	10341	11685	22026	882
mrna064977	621	6323	3491	9814	93	1151	1193	2344	22
mrna066573	282	1225	753	1978	9	532	557	1089	5
mrna026970	1404	67	405	472	10	35	72	107	2
mrna105132	1728	34918	8910	43828	1155	4362	4770	9132	241
mrna009588	1515	522	419	941	22	182	224	406	9
mrna121627	330	0	2	2	0.01	3	0	3	0.02
mrna074509	1674	183	110	293	7	67	68	135	3
mrna028105	1323	12956	10150	23106	466	6462	6315	12777	258
mrna121338	459	822	624	1446	10	401	384	785	5
mrna107490	363	229	18	247	1	8	0	8	0
mrna093230	295	1354	1095	2449	11	887	1042	1929	9
mrna070632	496	4277	2958	7235	55	2484	2634	5118	39
mrna093698	807	29555	18855	48410	596	5367	6079	11446	141
mrna084787	1230	119	37	156	3	15	8	23	0.43
mrna139758	2001	59124	61852	120976	3692	23993	25503	49496	1511
mrna071578	888	11254	16581	27835	377	2770	4181	6951	94
mrna048555	417	2769	2069	4838	31	2167	2216	4383	28
mrna091757	177	1323	844	2167	6	444	504	948	3
mrna015009	455	3177	1795	4972	35	58	109	167	1
mrna004763	1483	14171	9888	24059	544	4531	4776	9307	211
mrna098230	390	5668	4897	10565	63	2751	3116	5867	35
mrna053306	717	29591	23949	53540	586	11857	12582	24439	267
mrna035375	375	2214	1506	3720	21	1627	1602	3229	18
mrna002983	6936	379116	332147	711263	75244	106269	119995	226264	23936
mrna096393	3122	1789	1030	2819	134	953	916	1869	89
mrna016294	390	91	58	149	1	56	46	102	1
mrna106738	683	1892	1251	3143	33	1034	1106	2140	22
mrna096003	1038	3230	2046	5276	84	1536	1546	3082	49
mrna124385	355	6500	3461	9961	54	5647	4908	10555	57
mrna118007	271	795553	338824	1134377	4689	8089	11850	19939	82
mrna105093	573	5690130	2083915	7774045	67942	52859	76813	129672	1133
mrna057813	495	17502	13136	30638	231	9375	9785	19160	145
mrna043662	1704	67923	52517	120440	3130	25493	27019	52512	1365
mrna007120	315	47	46	93	0	9	13	22	0.11
mrna096121	1909	23127	15818	38945	1134	11159	11762	22921	667
mrna029953	435	4876	3706	8582	57	2247	2506	4753	32
mrna001012	633	38544	96913	135457	1308	964	3080	4044	39
mrna023290	1101	226	114	340	6	88	79	167	3
mrna079612	3980	72053	46633	118686	7205	35116	33653	68769	4175
mrna066175	618	3753	2970	6723	63	2444	2449	4893	46
mrna110666	321	5486	4444	9930	49	1947	1878	3825	19
mrna064310	2470	6572	6384	12956	488	3769	6522	10291	388
mrna139098	2557	8104	7684	15788	616	5436	8671	14107	550
mrna072920	465	834	674	1508	11	470	580	1050	7
mrna077126	2166	22912	11994	34906	1153	14854	13630	28484	941
mrna024572	753	6367	6037	12404	142	3125	3477	6602	76
mrna131003	504	1829	1687	3516	27	693	748	1441	11
mrna054831	2104	2185	1389	3574	115	947	1022	1969	63
mrna053564	369	1470	1110	2580	15	861	974	1835	10
mrna007148	2183	6151	4522	10673	355	3851	3826	7677	256
mrna014279	825	1799	1284	3083	39	893	920	1813	23
mrna066730	366	1052	793	1845	10	1406	1939	3345	19
mrna058994	732	9429	4746	14175	158	1682	1602	3284	37
mrna098363	165	15193	9486	24679	62	9657	10394	20051	50
mrna057019	168	40	34	74	0.19	33	30	63	0.16
	Total	7556391	3330367	10886758		413643	473887	887530	

A5.9: Mapped reads and coverage statistics for the 2D v3.3 cDNAs. Gene length refers to number of nucleotides in sequence. Coverage was calculated using: (combined mapped reads x gene length)/total length of reference (65,564).



A5.10*:* Schematic diagram of the strategies to prioritise SNPs for validation from putative SNPs identified by aligning to the 2D interval from the v3.3 cDNAs.



A5.11*:* Venn diagrams showing the overlap in SNPs from 2D v3.3 cDNAs between the filtered varietal SNPs in (A) the parent NILs and the short bulk (B) the parent NILs and the tall bulk. Numbers are SNPs including SNPs on duplicated genes. Italics indicate SNPs prioritised for validation.

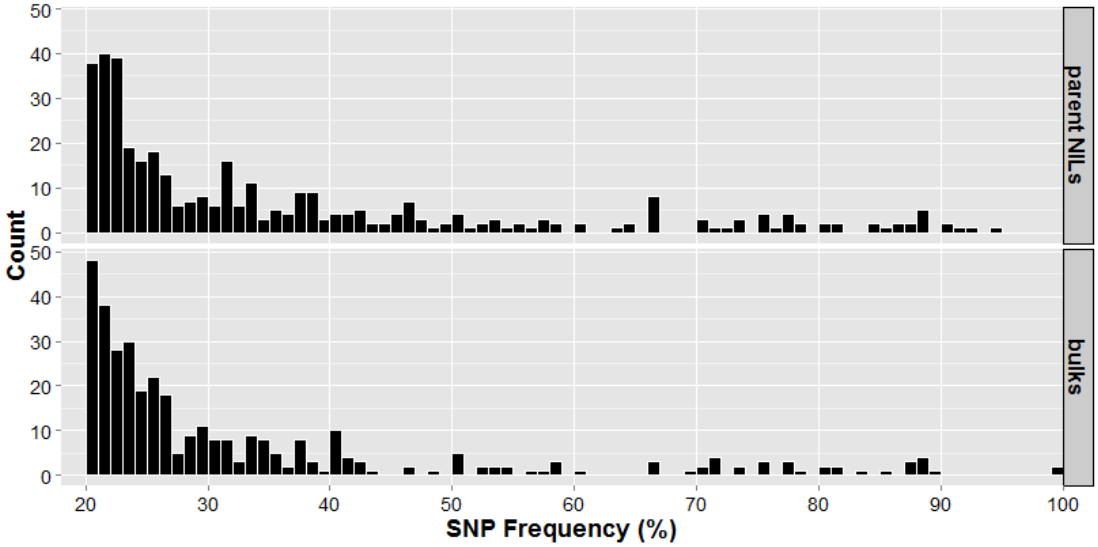
				Coverage		Reads1		Reads2		Frequency (%)	
Gene	Position	Ref	Var	RIL4	Short	RIL4	Short	RIL4	Short	RIL4	Short
mrna002983	1	A	G	143	17	107	10	33	6	23.08	35.29
mrna053564	111	T	C	216	131	163	104	53	27	24.54	20.61
mrna054831	1630	A	G	66	34	52	27	14	7	21.21	20.59
mrna064310	2179	G	A	262	214	204	159	58	55	22.14	25.70
mrna106738	85	C	T	236	248	185	196	51	52	21.61	20.97
mrna106738	610	G	A	53	55	40	44	13	11	24.53	20.00
mrna106738	627	G	T	40	43	26	28	9	9	22.50	20.93
mrna110666	144	C	T	1181	489	925	381	256	108	21.68	22.09

				Coverage		Reads1		Reads2		Frequency (%)	
Gene	Position	Ref	Var	CD	Tall	CD	Tall	CD	Tall	CD	Tall
mrna096003	492	G	A	450	214	358	164	92	50	20.44	23.36
mrna096393	2189	T	A	86	51	68	37	18	14	20.93	27.45
mrna007148	409	A	G	289	165	228	130	61	35	21.11	21.21
mrna105132	690	C	T	862	143	678	108	183	35	21.23	24.48
mrna077126	2051	T	C	964	821	757	646	206	174	21.39	21.19
mrna077126	917	T	C	952	475	748	375	204	100	21.43	21.05
mrna007148	169	G	A	280	136	219	107	61	29	21.79	21.32
mrna007148	168	T	C	278	138	217	108	61	30	21.94	21.74
mrna007148	352	C	T	340	178	264	142	76	36	22.35	20.22
mrna096393	2184	A	G	80	46	62	34	18	12	22.50	26.09
mrna007148	351	T	C	336	173	260	137	76	36	22.62	20.81
mrna096003	489	T	C	444	212	342	164	102	48	22.97	22.64
mrna096003	438	C	T	385	170	296	136	89	34	23.12	20.00
mrna007148	349	T	C	341	176	261	138	80	38	23.46	21.59
mrna066573	224	A	G	150	46	114	27	36	19	24.00	41.30
mrna096003	1002	C	T	56	33	42	22	14	11	25.00	33.33
mrna090625	232	C	G	754	75	561	57	193	18	25.60	24.00
mrna007148	2043	G	T	76	62	52	49	24	13	31.58	20.97
mrna098363	5	A	T	126	109	9	8	49	44	38.58	40.37
mrna053564	1	G	C	10	11	0	2	9	9	90.00	81.82
mrna057019	62	T	A	11	8	0	0	11	8	100.00	100.00
mrna057019	70	T	C	11	8	0	0	11	8	100.00	100.00

A5.12: The statistics of the overlapping putative varietal SNPs shown in A5.11 between (top) the short parent NIL and short bulk and (bottom) the tall parent NIL and tall bulk. Coverage: total depth of coverage; Reads1: number of reads supporting the reference; Reads2: number of reads supporting the SNP; Frequency: the SNP frequency from the read count (Reads2/total count).

						Coverage		Reads_Ref		Reads_Var		Freq (%)			
SNP	Ref	Var	Type	Validation	Marker	RIL4	Short	RIL4	Short	RIL4	Short	RIL4	Short	Contig	Primer
mrna002983_1	A	G	var	monomorphic	vcf_1	143	17	107	10	33	6	23.08	35.29	2DS_5390004	semispecific
mrna106738_85	C	T	var	monomorphic	vcf_2	236	248	185	196	51	52	21.61	20.97	2DS_5343181	semispecific
mrna106738_610	G	A	var	monomorphic	vcf_3	53	55	40	44	13	11	24.53	20.00	2DS_5343181	specific
mrna106738_627	G	T	var	monomorphic	vcf_4	40	43	26	28	9	9	22.50	20.93	2DS_5343181	specific
mrna053564_111	T	C	hom	-	-	216	131	163	104	53	27	24.54	20.61	2DS_5381947	specific
mrna054831_1630	A	G	hom	-	-	66	34	52	27	14	7	21.21	20.59	2DL_9887817	semispecific
mrna064310_2179	G	A	hom	-	-	262	214	204	159	58	55	22.14	25.70	2DS_5380297	specific
mrna110666_144	C	T	hom	-	-	1181	489	925	381	256	108	21.68	22.09	2DS_5351567	semispecific
						CD	Tall	CD	Tall	CD	Tall	CD	Tall		
mrna077126_917	T	C	var	monomorphic	vcf_5	952	475	748	375	204	100	21.43	21.05	2DS_4338395	specific
mrna007148_169	G	A	var	monomorphic	vcf_6	280	136	219	107	61	29	21.79	21.32	2DS_5381947	specific
mrna007148_352	C	T	var	monomorphic	vcf_7	340	178	264	142	76	36	22.35	20.22	2DS_5381947	specific
mrna096393_2184	A	G	var	monomorphic	vcf_8	80	46	62	34	18	12	22.50	26.09	2DS_5390396	specific
mrna007148_351	T	C	var	monomorphic	vcf_9	336	173	260	137	76	36	22.62	20.81	2DS_5381947	specific
mrna096003_489	T	C	var	monomorphic	vcf_10	444	212	342	164	102	48	22.97	22.64	2DS_5389048	specific
mrna007148_349	T	C	var	monomorphic	vcf_11	341	176	261	138	80	38	23.46	21.59	2DS_5381947	specific
mrna096003_1002	C	T	var	monomorphic	vcf_12	56	33	42	22	14	11	25.00	33.33	2DS_5323988	semispecific
mrna007148_2043	G	T	var	monomorphic	vcf_13	76	62	52	49	24	13	31.58	20.97	2DS_5381947	semispecific
mrna096003_492	G	A	var	monomorphic	vcf_14	450	214	358	164	92	50	20.44	23.36	2DS_5389048	specific
mrna096393_2189	T	A	var	monomorphic	vcf_15	86	51	68	37	18	14	20.93	27.45	2DS_5390396	specific
mrna007148_409	A	G	var	monomorphic	vcf_16	289	165	228	130	61	35	21.11	21.21	2DS_5381947	specific
mrna105132_690	C	T	var	monomorphic	vcf_17	862	143	678	108	183	35	21.23	24.48	2DS_5358861	semispecific
mrna077126_2051	T	C	var	monomorphic	vcf_18	964	821	757	646	206	174	21.39	21.19	2DS_4338395	semispecific
mrna007148_168	T	C	hom	-	-	278	138	217	108	61	30	21.94	21.74	2DS_5381947	specific
mrna066573_224	A	G	hom	-	-	150	46	114	27	36	19	24.00	41.30	2DS_5358861	semispecific
mrna090625_232	C	G	hom	-	-	754	75	561	57	193	18	25.60	24.00	2DS_5388557	specific
mrna096003_438	C	T	hom	-	-	385	170	296	136	89	34	23.12	20.00	2DS_5389048	specific

A5.13: SNPs identified for validation by steps shown in A5.10 (1). Columns left to right: Ref: reference base at SNP position; Var: variant at SNP position; Validation: the result of the marker screened on RIL4 and CD; Coverage: total number of reads mapping to the SNP; Reads: number of reads mapping with the reference or variant call at the SNP position; Freq: the SNP frequency from the read count (Reads of variant/total count); Contig: 2DS contig which the SNP maps to for which PolyMarker designed primers to be specific to; Primer: outcome of specificity to 2DS contig.



A5.14: The frequencies of the putative varietal SNPs, relative to the calls supporting the reference, from alignments to the 2D v3.3 cDNA interval. The total SNPs between parent NILs and the reference (top panel) and both short and tall bulks (bottom panel) are shown.

Source	SNP	Ref	Var	SNP type	Validation	SNP_ID	Manual aligning	Cov	R_Ref	R_Var	Freq (%)	Contig	Primer
CD	mrna009588_264	G	C	var	polymorphic	1_al	var	31	7	24	77.42	2DS_5338366	specific
RIL4	mrna026970_384	C	T	var	polymorphic	Freq_2	hom	10	0	10	100.00	2DS_5358861	semispecific
RIL4	mrna026970_1260	T	C	var	monomorphic	Freq_3	var	22	4	18	81.82	2DS_5358861	semispecific
CD	mrna074509_402	C	T	var	monomorphic	Freq_4	var	16	0	16	100.00	2DS_5342238	specific
CD	mrna090625_156	A	C	var	monomorphic	Freq_5	no B info	10	0	10	100.00	2DS_5388557	semispecific
CD	mrna090625_157	T	A	var	monomorphic	Freq_6	no B info	10	0	10	100.00	2DS_5346435	specific
CD	mrna096393_217	T	C	var	monomorphic	Freq_7	var	9	0	9	100.00	2DS_5390396	semispecific
CD	mrna096393_965	T	C	var	monomorphic	Freq_8	var	8	2	6	75.00	2DS_5390396	nonspecific
CD	mrna096393_1012	T	A	var	monomorphic	Freq_9	var	24	5	19	79.17	2DS_5390396	nonspecific
CD	mrna096393_1017	G	A	var	monomorphic	Freq_10	var	24	6	18	75.00	2DS_5390396	nonspecific
RIL4	mrna096393_1452	C	A	var	monomorphic	Freq_11	var	20	5	15	75.00	2DS_5390396	nonspecific
CD	mrna105132_303	A	C	var	monomorphic	Freq_12	no A info	910	128	774	85.05	2DS_5358861	semispecific

A5.15: SNPs identified for validation by steps shown in A5.10 (2). Columns left to right: Source: Informative parent of the variant call; Ref: reference base at SNP position; Var: variant at SNP position; Validation: the result of the marker (SNP_ID) screened on RIL4 and CD; Cov: total number of reads mapping to the SNP; R: number of reads mapping with the reference or variant call at the SNP position; Freq: the SNP frequency from the read count (Reads of variant/total count); Contig: 2DS contig which the SNP maps to for which PolyMarker designed primers to be specific to; Primer: outcome of specificity to 2DS contig.

	AX-2	2DS_278	2DS_242	2DS_280	2DS_138	2DS_15	2DS_89	2DS_293	2DS_192	2DS_6	2DS_94	2DS_210	AX-30	2DS_275	50_uni	6IS	5IS	1_al	52I	63_uni	72_uni	10IS	
	2DS_137		76_uni	2DS_235	2DS_66	2DS_26	2DS_95		2DS_187		AX-25	2DS_211			21_uni		4IS	27_uni		11IS			
	AX-4		2DS_243				2DS_105		2DS_201		2DS_145	2DS_149						Freq_2					
							55_uni		2DS_208		AX-28	2DS_217						66_uni					
							2DS_88				AX-29	2DS_215											
							2DS_540403				2DS_79	2DS_212											
							16_uni																
							BFR_46																
							2DS_222																
							2DS_4																
							2DS_5375260																
							2DS_3																
							2DS_1																
							2DS_223																
							8_uni																
							2DS_89																
							2DS_95																
							2DS_105																
							55_uni																
							2DS_88																
	No. markers	3	1	3	2	2	2	15	1	4	1	6	6	1	2	1	2	4	1	2	1	1	Total 62

A5.16: The 2D RIL marker classes based on graphical genotypes in A5.17.

																									Total
No. markers in class	3	1	3	2	2	2	5	1	4	10	1	6	6	1	1	2	1	2	4	1	2	1	1	62
2D RIL class	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	
RIL1	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b		b	
RIL2	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL3	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL4	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	a	b	
RIL5	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL6	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL7	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	
RIL8	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL9	b	b	b	b	a	a	b	-	a	b	b	b	b	b	a	b	b	b	b	b	b	b	b	
RIL10	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL11	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL12	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	
RIL13	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL14	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	
RIL15	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL16	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL17	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL18	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL19	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL20	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL21	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL22	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL23	a	a	a	b	b	b	b	b	b	b	b	b	b	a	b	a	b	b	b	b	b	b	b	
RIL24	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL25	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL26	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL27	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL28	b	b	b	b	b	b	b	a	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	
RIL29	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL30	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL31	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL33	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL34	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	b	
RIL35	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	b	a	b	b	a	
RIL36	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL37	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	b	b	
RIL38	b	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	b	a	a	a	a	a	a	
RIL39	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	
RIL40	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL41	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	b	a	b	b	a	
RIL42	a	a	a	a	b	b	b	b	b	b	b	a	b	a	b	a	a	a	a	b	a	b	b	
RIL43	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL44	a	a	a	b	b	b	b	b	b	b	b	b	b	a	a	a	b	b	b	b	b	b	b	
RIL45	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	
RIL46	b	b	b	b	a	a	a	a	a	a	a	a	a	b	a	b	b	a	b	a	b	a	a	
RIL47	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL48	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL49	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL50	a	a	a	a	b	b	b	b	b	b	b	a	b	a	b	a	a	a	a	b	a	a	a	
RIL51	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL52	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	
RIL53	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL54	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL55	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	
RIL56	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	
RIL57	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL59	a	b	b	b	a	b	b	b	b	b	b	b	b	a	b	a	b	b	b	b	b	b	b	
RIL60	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL61	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL62	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL63	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL64	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL65	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	
RIL66	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL67	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	
RIL68	b	b	b	b	a	a	a	a	a	a	a	b	b	b	a	b	b	b	b	a	b	b	b	
RIL70	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	
RIL71	a	a	a	a	b	b	a	b	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	
RIL73	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL74	b	b	b	b	a	a	a	a	a	a	a	b	a	b	a	b	b	b	b	a	b	a	b	
RIL75	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL77	a	a	a	a	a	a	a	b	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	
RIL79	a	a	a	a	a	a	a	b	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	
RIL80	a	a	a	a	b	a	a	b	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	
RIL81	a	a	a	a	b	b	b	b	b	b	b	b	b	a	b	a	b	b	b	b	b	b	b	
RIL82	b	b	b	a	a	a	a	a	a	a	a	a	a	b	a	b	a	a	a	a	a	a	a	
RIL83	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL84	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL85	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL86	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
RIL87	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
RIL88	b	b	b	b	a	a	a	a	a	a	a	b	a	b	a	b	b	b	b	a	b	b	b	

A5.17: The graphical genotype of each marker class to the 2D RIL population used for coarse mapping. The assignment of markers to classes is shown in A5.16. The total markers in each class are shown on the top row.

Marker	Source	2D RIL class	gwm class	RM class	POPSEQ bin	IWGSC contig	Barley		Brachypodium	Rice	v3.3 cDNAs	
2DS_137	IWGSC-1	1	7	-	-	2DS_5309868	-	-	-	-	mrna112293	-
AX-2	Axiom	1	7	-	-	2DS_5389432	MLOC_5618	9705271	Bradi5g02037	Os04g0107900	mrna105701	562484
AX-4	Axiom	1	7	-	-	2DS_5389432	MLOC_5618	9705271	Bradi5g02037	Os04g0107900	mrna105701	562484
2DS_278	IWGSC-2	2	10	-	33.06	2DS_5360680	-	-	-	-	-	-
2DS_242	v3.3 cDNAs	3	9	-	-	2DS_5388557	MLOC_66130	12806286	Bradi5g02400	Os04g0102500	mrna126380	704896
2DS_243	v3.3 cDNAs	3	9	-	-	2DS_5388557	MLOC_66130	12806286	Bradi5g02400	Os04g0102500	mrna126380	704896
76_uni	UniGenes	3	11	-	12.03	2DS_5330846	MLOC_15183	10865622	Bradi1g48432	Os07g0133100	mrna099460	-
2DS_235	v3.3 cDNAs	4	2	-	13.83	2DS_5365907	MLOC_62712	9685835	-	-	mrna074113	734103
2DS_280	IWGSC-2	4	12	-	18.14	2DS_5323734	AK365704	-	-	-	-	-
2DS_66	IWGSC-1	5	25	F	33.06	2DS_5357871	MLOC_26534	608666509	-	-	mrna102226	-
2DS_138	IWGSC-1	5	27	F	33.06	2DS_5318296	MLOC_48245	19071820	Bradi5g04130	Os04g0229100	-	-
2DS_15	IWGSC-1	6	25	F	33.06	2DS_5390752	MLOC_69463	19016685	Bradi5g04550	Os04g0244400	-	1189621
2DS_26	IWGSC-1	6	26	F	33.06	2DS_5390977	MLOC_37479	18924902	Bradi5g04340	Os03g0856000	mrna057813	1161159
2DS_88	IWGSC-1	7	17	A	17.34	2DS_5321770	MLOC_24124	15296824	Bradi5g03400	Os04g0206450	-	-
2DS_89	IWGSC-1	7	17	A	17.34	2DS_5321770	MLOC_24124	15296824	Bradi5g03400	Os04g0206450	-	-
2DS_95	IWGSC-1	7	17	A	17.34	2DS_5341322	MLOC_522	15276497	Bradi5g03340	-	-	-
55_uni	UniGenes	7	18	C	-	2DS_5375260	MLOC_52188	22256327	Bradi5g05140	Os03g0804300	mrna102641	7779113
2DS_105	IWGSC-1	7	23	A	17.34	2DS_5390456	-	-	Bradi5g02900	-	-	-
2DS_293	Chapman	8	30	-	33.06	2DS_5363870	MLOC_74610	-	Bradi5g03697	Os04g14760	-	-
2DS_187	Limagrain	9	28	E	17.34	2DS_5354335	-	-	Bradi5g15565	-	-	-
2DS_192	Limagrain	9	28	E	16.95	2DS_5378845	MLOC_76709	39057764	Bradi1g69730	Os10g0399100	-	-
2DS_201	Limagrain	9	28	E	17.34	2DS_5389660	MLOC_5590	Chr_1	-	-	-	-
2DS_208	Limagrain	9	28	E	17.34	2DS_5389857	MLOC_58453	18521524	-	Os04g0261400	mrna074899	-
2DS_3	IWGSC-1	7	13	A	-	2DS_5337443	MLOC_81137	15280265	Bradi5g03380	Os04g0204000	mrna103570	-
2DS_4	IWGSC-1	7	13	A	-	2DS_5337443	MLOC_81137	15280265	Bradi5g03380	Os04g0204000	mrna103570	-
2DS_1	IWGSC-1	7	14	A	17.34	2DS_5359909	-	-	-	Os10g21560	mrna008489	-

A5.18: The total markers developed in Chapter 5 which were likely to map to *Rht8* based on mapping with the 2D RIL population. These markers were used in Chapter 6 in the final step to fine-map *Rht8*. Markers are annotated by marker class at each mapping step and with comparative genomic data from syntenic species. Wheat annotation is shown in terms of the POPSEQ bin the 2DS CSS contig was mapped to in the IWGSC-2 and Chapman datasets. No information indicates the contig is not in the POPSEQ data. Each contig was also anchored on the v3.3 cDNAs where possible. Anchoring of *DG279* and *DG371* is shown fully in Chapter 5 but included here for ease of comparison between all markers.

A5.18 (continued)

Marker	Source	2D RIL class	gwm class	RM class	POPSEQ bin	IWGSC contig	Barley	Brachypodium	Rice	v3.3 cDNAs	
2DS_222	Limagrain	7	19	C	17.34	2DS_5342594	MLOC_70393	22198957	Bradi3g16570	Os04g0652600	-	-
2DS_223	Limagrain	7	19	C	17.34	2DS_5342594	MLOC_70393	22198957	Bradi3g16570	Os04g0652600	-	-
8_uni	UniGenes	7	21	A	-	2DS_5388088	-	-	Bradi4g02250	Os04g12560	-	881779
2DS_540403	IWGSC-2	7	22	A	17.34	2DS_540403	MLOC_81137	15280265	Bradi5g03380	Os04g0204000	-	-
2DS_5375260	IWGSC-2	7	18	D	-	2DS_5375260	MLOC_52188	22256327	Bradi5g05140	Os03g0804300	mrna102641	7779113
BFR_46	v3.3 cDNAs	7	25	F	17.34	2DS_5371750	MLOC_58466	19491674	Bradi5g04660	Os02g0319800	-	1264216
16_uni	UniGenes	7	28	E	17.34	2DS_5364728	MLOC_72777	19455074	Bradi5g04686	Os04g0252400	mrna103168	7646108
2DS_6	IWGSC-1	10	14	A	17.34	2DS_5321865	MLOC_62798	15618954	Bradi4g21260	Os11g0215100	mrna052919	-
2DS_145	IWGSC-1	11	3	-	12.93	2DS_5366858	MLOC_54461	12602333	-	Os08g0105600	mrna046956	-
2DS_79	IWGSC-1	11	3	-	18.14	2DS_5317970	MLOC_17561	14075639	Bradi5g02830	-	mrna095026	-
2DS_94	IWGSC-1	11	3	-	16.56	2DS_5338366	MLOC_56811	12508948	Bradi5g02860	-	mrna009588	801887
AX-25	Axiom	11	3	-	13.83	2DS_5358861	MLOC_38821	9984398	Bradi5g02520	Os03g0363600	mrna026970	766690
AX-28	Axiom	11	3	-	13.83	2DS_5358861	MLOC_38821	9984398	Bradi5g02520	Os03g0363600	mrna026970	766690
AX-29	Axiom	11	3	-	13.83	2DS_5358861	MLOC_38821	9984398	Bradi5g02520	Os03g0363600	mrna026970	766690
2DS_149	IWGSC-1	12	15	A	17.34	2DS_5319467	MLOC_43355	15293278	-	Os04g0208400	-	-
2DS_210	Limagrain	12	16	A	17.34	2DS_5337059	MLOC_16024	475472562	-	-	-	-
2DS_211	Limagrain	12	16	A	17.34	2DS_5337059	MLOC_16024	475472562	-	-	-	-
2DS_212	Limagrain	12	16	A	17.34	2DS_5337059	MLOC_16024	475472562	-	-	-	-
2DS_217	Limagrain	12	16	A	-	2DS_5319965	-	-	-	-	-	-
2DS_215	Limagrain	12	22	A	-	2DS_5379317	-	-	-	-	-	-
AX-30	Axiom	13	8	-	9.2	2DS_5388293	MLOC_81869	19752473	Bradi3g22850	Os02g0113200	-	-
2DS_275	IWGSC-2	14	29	G	33.06	2DS_5344159	-	-	-	-	-	-
50_uni	UniGenes	15	6	-	0	2DS_4514573	MLOC_10286	5751998	-	-	mrna095302	237336
21_uni	UniGenes	15	6	-	0	2DS_5353487	MLOC_53593	532316057	Bradi5g18940	Os04g0574100	-	-
6iS	iSelect	16	5	-	13.83	2DS_5354297	MLOC_38340	10177219	-	Os09g0479600	mrna123420	-
5iS	iSelect	17	2	-	17.34	2DS_5316382	MLOC_63202	14025380	Bradi5g02810	Os04g0188500	mrna060003	-
4iS	iSelect	17	2	-	17.34	2DS_5343763	MLOC_51927	-	Bradi5g02890	Os12g0277500		
1_al	2D v3.3 cDNAs	18	2	-	16.56	2DS_5338366	MLOC_56811	12508948	Bradi5g02860	-	mrna009588	801887
66_uni	UniGenes	18	2	-	13.83	2DS_5362023	MLOC_65493	12625788	Bradi5g02490	Os07g0474600	mrna040847	752563
27_uni	UniGenes	18	4	-	13.83	2DS_5358861	MLOC_38821	9984398	Bradi5g02520	Os03g0363600	mrna064977	766685
Freq_2	2D v3.3 cDNAs	18	4	-	13.83	2DS_5358861	MLOC_38821	9984398	Bradi5g02520	Os03g0363600	mrna026970	766690
52i	iSelect	19	20	B	-	2BS_4748675	MLOC_5957	15601547	Bradi5g03460	Os04g0209200	-	-
11iS	iSelect	20	2	-	16.56	2DS_5343186	-	-	-	-	mrna123373	-
63_uni	UniGenes	20	14	A	17.34	2DS_5359909	-	15266465	-	Os10g21560	mrna008489	-
72_uni	UniGenes	21	2	-	17.34	2DS_5343763	MLOC_51927	-	Bradi5g02890	Os12g0277500	-	-
10iS	iSelect	22	1	-	-	1BS_3467454	-	-	-	-	-	-

Gene	Base	Pos	SNP	BFR	Cov_1	Cov_2	Bulk	Ratio_1	Ratio_2	Count_1	Count_2	Contig	Identity	
mrna140874	t	1815	C	20.04	67	107	tall	0.01	0.3	1	32	5BL_10805975	100	
mrna019699	c	139	G	20.29	46	34	tall	0.02	0.44	1	15	1AL_3848105	100	
mrna132275	c	75	A	20.34	265	55	short	0.74	0.04	196	2	1BS_3432767	99.72	
mrna132275	g	73	T	20.45	256	56	short	0.73	0.04	187	2	1BS_3434134	99.72	
mrna119531	g	173	A	20.83	75	71	short	0.29	0.01	22	1	7BL_6742592	100	
mrna015649	c	429	T	21.17	254	42	tall	0.02	0.5	6	21	5BL_10789463	99.63	
mrna063321	a	504	G	21.72	157	159	tall	0.01	0.14	1	22	7AS_4249929	100	
mrna023899	t	636	G	21.74	68	1463	short	0.1	0	7	7	2DS_5382207	88.92	monomorphic
mrna041746	a	425	T	21.98	41	58	short	0.36	0.02	40	1	2BL_8075479	99.19	
mrna063321	t	489	G	22	147	147	tall	0.01	0.15	1	22	7AS_4249929	100	
mrna053652	c	977	T	22.04	190	166	short	0.13	0.01	32	1	1BL_3805529	100	
mrna022149	c	819	T	23.34	109	636	short	0.04	0	4	1	4AL_7159326	100	
mrna121159	g	237	A	23.5	681	705	short	0.1	0	68	3	7DS_3961038	100	
mrna022149	c	821	A	23.63	108	638	short	0.04	0	4	1	4AL_7159326	100	
mrna021588	g	455	C	23.89	27	828	short	0.11	0	4	4	3B_10456072	100	
mrna055972	c	192	T	24.29	56	85	short	0.29	0.01	16	1	7BS_3065836	100	
mrna065695	c	499	G	24.7	220	209	short	0.12	0	26	1	7DS_3906205	99.12	
mrna004362	g	556	A	25.29	328	237	short	0.11	0	35	1	2BL_8020296	100	
mrna015649	c	489	T	25.8	258	42	tall	0.04	1	10	42	5BL_10789463	99.63	
mrna074847	g	162	C	25.82	72	290	tall	0.01	0.36	1	104	4DL_14465849	100	
mrna016481	g	1383	A	26	1168	833	short	0.25	0.01	291	8	7BS_3139798	99.84	
mrna084023	c	30	T	26.2	233	165	short	0.16	0.01	37	1	7BS_3121441	98.35	BFR_11, group C
mrna056418	c	264	T	26.32	124	48	short	0.55	0.02	68	1	2BL_8033240	100	
mrna021868	t	2411	C	26.48	542	598	short	0.09	0	48	2	7BS_3139798	99.47	
mrna074551	a	1569	G	28.02	52	47	short	0.6	0.02	31	1	5BL_10925653	100	
mrna052020	c	1520	T	28.5	61	121	tall	0.02	0.47	1	57	4BL_7012682	100	
mrna052020	c	1522	G	28.95	61	118	tall	0.02	0.47	1	56	4BL_7012682	100	
mrna011230	a	269	G	29.73	83	67	tall	0.01	0.36	1	24	5AS_419734	100	
mrna127203	c	72	T	30.32	178	3598	short	0.29	0.01	51	34	1BS_3182035	99.42	
mrna013892	c	1518	A	30.68	93	74	tall	0.01	0.33	1	64	1AL_3889910	99.72	
mrna086098	c	1217	T	31.14	370	505	short	0.68	0.02	251	11	1BL_3867748	98.7	
mrna136711	a	1707	G	31.25	166	84	tall	0.01	0.19	1	16	7AL_4525468	99.33	
mrna132278	c	537	T	31.83	183	23	tall	0.01	0.17	1	4	1DS_1902038	99.72	
mrna036503	c	666	A	32.22	1150	25	tall	0	0.11	4	4	2AL_6401311	99.71	
mrna078757	c	183	T	35	361	259	short	0.14	0	80	1	3AS_3338466	70.97	
mrna049576	c	444	G	39.42	69	68	short	0.58	0.01	40	1	5BL_10896624	100	
mrna092193	t	562	C	39.79	237	205	short	0.19	0	46	1	2AL_6421201	100	
mrna017960	g	468	A	44.95	126	118	short	0.38	0.01	48	1	7BS_3164331	100	
mrna014319	a	348	G	45.58	300	318	short	0.14	0	43	1	7AS_4211701	99.48	
mrna098108	t	541	C	48.52	371	300	short	0.16	0	60	1	2AS_5237913	98.41	BFR_4, group A
mrna040324	g	621	A	56.9	552	487	short	0.23	0	129	2	1BS_3469575	100	
mrna076686	g	337	A	57.64	1696	1788	tall	0.01	0.3	9	877	1DL_2263471	99.62	
mrna112463	t	1116	C	67.89	454	557	short	0.37	0.01	166	3	2DL_9908256	99.78	
mrna003724	c	42	T	72.39	186	198	short	0.37	0.01	68	1	7AS_4246344	100	
mrna004363	a	1006	G	79.25	283	267	short	0.3	0	84	1	2AL_6416883	100	
mrna107518	t	420	C	79.31	64	94	short	0.84	0.01	54	1	1DL_2243205	100	
mrna025486	a	860	T	129.11	304	223	short	0.58	0	176	1	7BS_3150559	100	

A5.19: SNPs in the v3.3 cDNAs with highest non-infinity BFRs (with a cut-off at BFR>20). Only two out of 47 were localised to 2S (shaded in grey). One of the SNP assays to validate these was monomorphic, whilst the marker on the other SNP mapped outside the *Rht8* linkage group, to group A (Figure 5.11). One SNP assay was developed for the SNP on 7BS, which mapped to group C in Figure 5.11. Columns left to right: Base: reference base at SNP position; Pos: position on gene model; SNP: variant at SNP position; BFR: calculated from Ratio_1/Ratio_2; Coverage: total number of reads mapping to the SNP; Bulk: the bulk contributing the variant; Ratio: relative contribution of reads that call the SNP from the total reads mapping to that position; Count: coverage x ratio; Contig: IWGSC CSS best hit; Identity: % nucleotide identity of the gene to Contig.

Appendix to Chapter 6

			Glasshouse 2012-13				Church Farm 2013-14				Morley 2013 - 14			
F4 rec	BSA	Consensus Rht8 score	Height (cm)	Score	N	St. error	Height (cm)	Score	N	St. error	Height (cm)	Score	N	St. error
F4-1-1-7-3		a	82.3	a	8	3.0	97.5	a	5	1.8	109.2	a	5	1.3
F4-2-3-2-1		a*	71.0	b	8	2.7	96.7	a	5	2.2	-	-	-	-
F4-1-2-4-3	B2	a	84.1	a	8	1.8	96.6	a	5	2.0	108.0	a	4	1.5
F4-2-7-2-1		a	82.3	a	8	1.4	96.5	a	5	0.9	-	-	-	-
F4-1-1-12-1		a	80.6	a	9	1.4	96.1	a	5	0.8	111.0	a	4	1.4
F4-3-7-14-3		a	83.4	a	8	2.0	96.0	a	5	1.4	109.6	a	5	1.3
CD		a	82.3	a	16	1.2	95.8	a	35	0.7	110.5	a	49	0.4
F4-2-7-6-2		a	81.3	a	8	1.7	95.7	a	5	1.2	108.0	a	5	1.2
F4-1-6-16-1	B6	a	82.1	a	5	1.0	95.0	a	5	0.8	106.0	a	5	2.2
F4-2-7-9-2		a	81.3	a	8	1.0	95.0	a	5	1.3	110.1	a	5	0.7
F4-2-3-7-1		a*	79.7	a	9	1.5	94.8	a	5	1.6	98.2	b	5	1.1
F4-1-9-2-1		a	79.9	a	8	0.7	94.7	a	5	0.8	108.2	a	5	1.7
F4-1-9-3-1		a	78.6	a	8	1.1	94.7	a	5	1.9	108.6	a	5	2.0
F4-2-2-7-1		a	82.7	a	9	1.3	94.7	a	5	0.7	107.7	a	5	2.2
F4-2-2-2-3		a	79.1	a	8	0.9	94.4	a	5	2.6	106.3	a	5	1.5
F4-3-3-15-1		a	79.1	a	8	1.3	93.9	a	5	2.4	107.4	a	4	0.7
F4-1-7-7-1		a	78.8	a	8	1.0	93.7	a	5	1.3	109.3	a	5	1.7
F4-3-2-13-1	B2	a	82.3	a	8	1.4	93.7	a	5	2.2	109.2	a	5	1.0
F4-2-1-16-3	B6	a	82.6	a	7	1.7	93.6	a	5	2.3	112.5	a	5	1.1
F4-2-7-3-6	B6	a	82.9	a	8	1.1	93.6	a	5	1.9	109.6	a	4	1.4
F4-1-6-13-2	B4	a	85.2	a	7	1.2	93.5	a	5	1.8	109.8	a	5	1.7
F4-1-6-17-1	B4	a	83.9	a	8	1.2	93.4	a	5	1.8	104.5	a	4	2.2
F4-2-1-11-4		a	82.9	a	8	2.0	93.3	a	5	0.5	106.0	a	5	1.4
F4-3-7-9-1		a	80.0	a	8	1.5	93.3	a	5	1.0	108.6	a	5	1.8
F4-3-7-13-3		a	80.1	a	8	1.2	93.1	a	5	1.4	107.0	a	5	0.8
F4-1-7-4-1		a	83.6	a	8	1.3	92.8	a	5	0.5	112.0	a	5	0.8
F4-3-8-5-2		a	82.1	a	8	1.3	92.6	a	5	2.0	107.6	a	5	1.4
F4-1-6-19-2		a	79.6	a	8	1.0	92.2	a	5	1.9	109.1	a	5	0.8
F4-3-8-3-3		a	77.7	a	8	1.4	92.1	a	5	2.1	108.0	a	5	1.8
F4-3-2-7-2	B4	a	84.4	a	8	1.5	91.8	a	5	0.9	111.3	a	5	0.9
F4-2-8-6-1		a	80.8	a	8	0.9	91.3	a	5	3.2	106.8	a	5	1.6
F4-3-7-7-2		a	82.4	a	8	1.0	91.0	a	5	1.9	106.9	a	5	0.7
F4-3-7-6-1		a	79.8	a	8	1.7	90.8	a	5	1.3	106.0	a	4	1.3
F4-2-3-8-1		a	80.1	a	8	2.0	90.6	a	5	2.7	105.5	a	5	0.7
F4-1-2-9-1		a	79.5	a	8	1.4	90.5	a	5	1.7	106.1	a	5	0.6
F4-1-9-1-1	B2	a	82.0	a	8	1.6	90.4	a	5	0.8	109.1	a	5	1.4
F4-1-7-18-3		a	81.3	a	8	1.0	89.8	a	5	1.8	109.3	a	5	1.4
F4-1-6-11-1		a	80.8	a	8	1.5	89.7	a	5	1.2	105.0	a	5	1.6
F4-3-2-12-1		a	81.6	a	8	1.3	89.6	a	5	1.5	108.4	a	5	0.5
F4-3-8-2-2		a*	75.4	b	8	0.8	89.5	a	5	2.0	107.5	a	4	1.6
F4-3-1-3-6		a	83.7	a	8	2.0	89.3	a	5	1.4	-	-	-	-
F4-3-7-1-2		a*	78.2	a	7	1.3	81.6	b	5	2.1	104.8	a	4	0.8

A6.1

			Glasshouse 2012-13				Church Farm 2013-14				Morley 2013 - 14			
F4 rec	BSA	Consensus Rht8 score	Height (cm)	Score	N	St. error	Height (cm)	Score	N	St. error	Height (cm)	Score	N	St. error
F4-1-7-1-1	B3	b	75.0	b	7	1.8	88.4	b	5	0.8	-	-	-	-
F4-3-2-8-1	B5	b	72.5	b	9	1.3	87.9	b	5	1.6	102.1	b	4	1.5
F4-3-2-15-1		b*	76.4	b	8	2.0	87.8	b	5	1.0	105.6	a	5	3.3
F4-1-7-17-1		b	76.2	b	8	2.5	87.6	b	5	1.2	102.5	b	4	2.3
F4-2-2-6-1		b*	78.3	a	8	1.4	87.6	b	5	2.1	103.6	b	5	2.0
F4-3-8-6-3	B1	b	71.1	b	9	1.6	87.6	b	5	1.6	101.3	b	5	1.4
F4-2-7-4-1		b	75.9	b	8	0.9	87.4	b	5	1.5	101.8	b	5	2.0
F4-2-8-5-2		b*	77.3	a	8	1.5	87.1	b	5	1.9	101.5	b	5	2.2
F4-3-2-5-1		b	72.5	b	7	0.9	87.1	b	5	2.9	100.4	b	5	1.3
F4-3-1-2-6	B5	b	74.1	b	8	0.5	87.0	b	5	1.6	102.8	b	4	1.4
F4-1-6-12-1		b*	79.2	a	8	2.0	86.9	b	5	1.5	102.9	b	4	1.8
RIL4		b	75.7	b	14	1.2	86.9	b	40	0.5	100.8	b	53	0.3
F4-1-6-3-1		b	75.5	b	8	1.2	86.8	b	5	1.9	102.3	b	4	1.7
F4-3-7-5-3		b	74.5	b	8	2.4	86.5	b	5	1.3	-	-	-	-
F4-1-7-15-1		b*	78.1	a	8	1.9	86.3	b	5	1.4	102.7	b	5	0.9
F4-3-7-8-2		b*	78.4	a	8	1.6	86.2	b	5	1.7	102.0	b	4	1.9
F4-2-7-12-2	B1	b	73.9	b	8	1.6	86.1	b	5	1.1	103.2	b	5	1.8
F4-1-2-7-1		b	75.1	b	8	0.9	86.0	b	5	3.0	102.6	b	5	1.5
F4-2-2-10-1		b*	77.6	a	8	2.3	86.0	b	5	0.9	101.3	b	4	1.0
F4-3-1-1-6		b	75.9	b	8	1.1	86.0	b	5	2.3	102.4	b	5	1.6
F4-2-8-1-2		b*	74.3	b	9	1.5	85.8	b	5	1.7	106.2	a	5	1.6
F4-1-1-9-7	B5	b	74.8	b	8	1.2	85.6	b	4	2.0	101.9	b	4	1.0
F4-1-1-11-1		b	76.4	b	7	1.0	85.6	b	5	2.5	101.1	b	5	1.6
F4-1-6-9-1		b	75.7	b	8	2.0	85.6	b	5	1.9	100.9	b	5	1.0
F4-3-7-10-1		b	71.5	b	8	1.1	85.4	b	5	1.5	-	-	-	-
F4-3-8-1-1	B1	b	75.0	b	8	1.7	85.1	b	5	2.0	99.6	b	5	1.7
F4-2-1-4-1		b*	73.5	b	8	0.8	84.7	b	5	1.5	104.9	a	5	3.0
F4-2-1-12-1	B3	b	74.3	b	8	1.8	84.6	b	5	1.8	102.4	b	4	2.2
F4-2-8-4-1		b	75.3	b	8	1.1	84.3	b	5	1.1	100.9	b	5	1.2
F4-2-1-8-4		b	76.1	b	8	1.6	84.2	b	5	1.3	102.6	b	5	1.4
F4-3-2-16-1	B3	b	71.8	b	8	1.9	84.1	b	5	1.6	103.2	b	5	1.2
F4-3-2-2-1		b	73.8	b	8	2.6	82.8	b	5	1.2	100.4	b	5	1.7
F4-1-2-2-1		b	64.7	b	8	3.1	70.4	b	5	1.9	94.8	b	4	6.6

group	location	Height (cm)	N	St. error
short	gh	74.8	32	0.5
	cf	85.6	32	0.5
	mor	102.1	29	0.4
tall	gh	81.0	41	0.4
	cf	92.9	41	0.4
	mor	107.8	38	0.4
BSA	Short	69.3	204	0.4
	Tall	78.0	213	0.3

A6.1 continued: Mean heights of the fine-mapping recombinants and parent NILs grown at the three locations indicated. The score at each location is shown, along with the consensus score which was based on the score at two out of three locations where conflicts arose. The conflicts are coloured darker and marked with an asterisk.

data: subset of 32 recombinants from glasshouse 2012
ANOVA for linear model spike length ~ genotype + sterility

Response: spike length

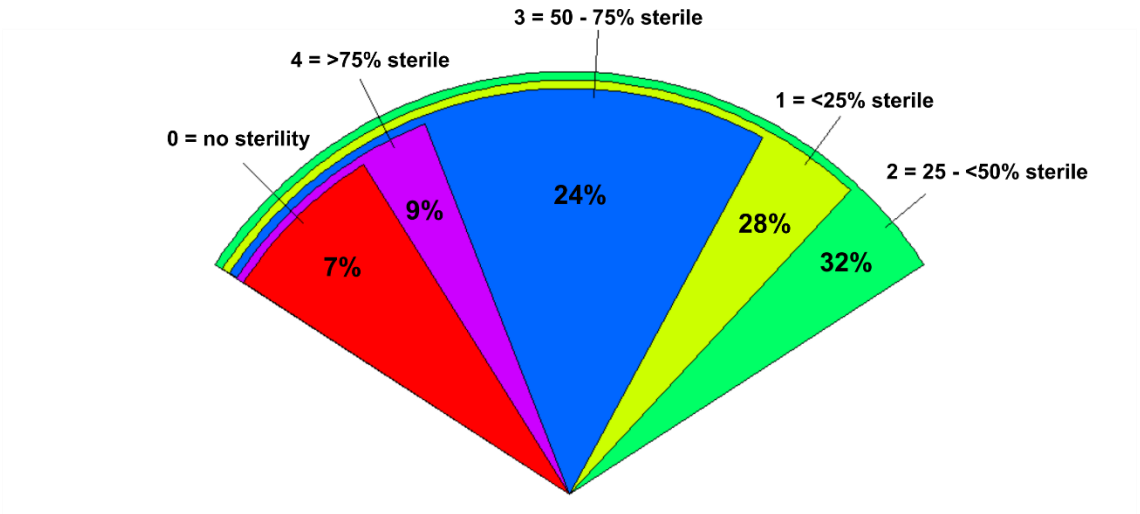
	Df	Sum Sq	Mean Sq	F-value	P-value	
genotype	31	27.936	27.9355	38.4907	1.99E-08	***
sterility	1	2.95	2.9501	4.0647	0.04698	*
Residuals	84	60.965	0.7258			

data: Total 73 recombinants from glasshouse 2013
ANOVA for linear model height ~ genotype*sterility

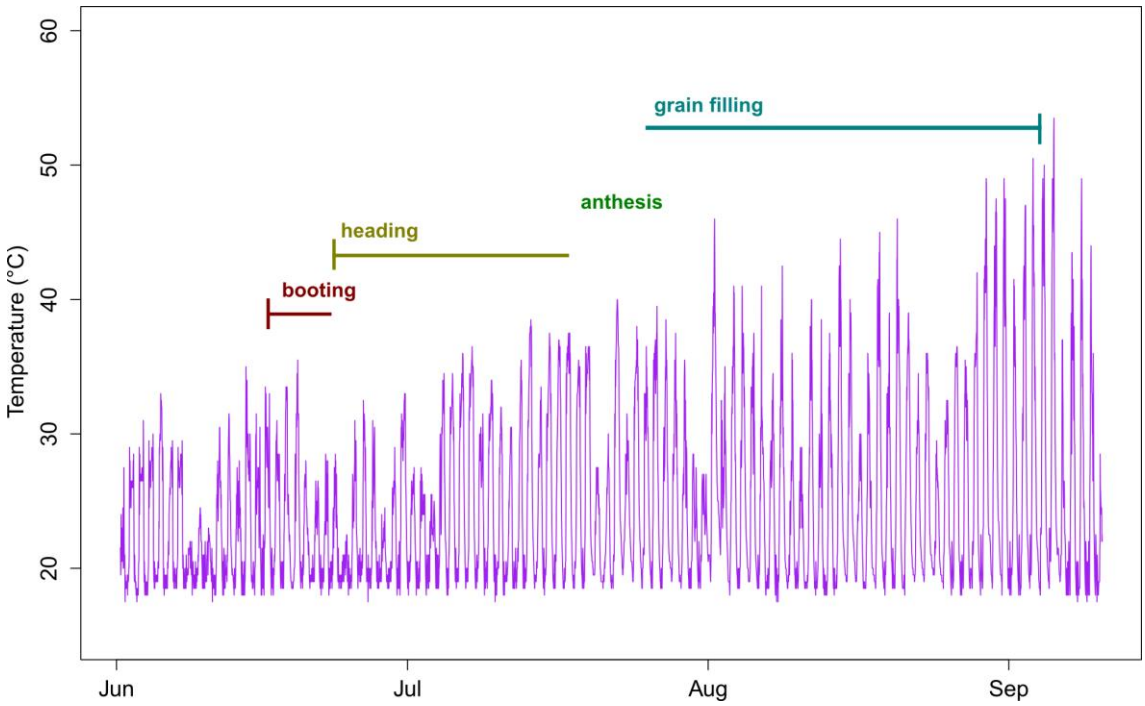
Response: height

	Df	Sum Sq	Mean Sq	F-value	P-value	
genotype	31	11648.6	375.76	20.7553	2.20E-16	***
sterility	1	370.5	370.54	20.4666	7.14E-06	***
genotype:sterility	31	886.6	28.6	1.5797	0.02464	*
Residuals	690	12492.1	18.1			

A6.2: ANOVA for sterility measured in glasshouse 2012 (top) and 2013 (bottom).



A6.3: Fan plot showing the sterility of the subset of 32 recombinants (N=880) grown in the glasshouse in 2013. A 0-5 scale was used (no plant was totally sterile (5)). The slices are re-arranged to overlap with each other and the radii have been modified so that each slice is visible, with the width of each slice relative to the sample size.



A6.4: Temperature recorded at canopy level in the glasshouse 2013 experiment. Temperature was measured at 30 minute intervals. Approximate growth stages of the plants, estimated from heading date data are superimposed.

gwm class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	10iS	2DS_235	2DS_79	Freq_2	6iS	21_uni	2DS_137	AX-30	2DS_243	2DS_278	76_uni	2DS_280	2DS_4	2DS_6	2DS_149
		11iS	2DS_94	27_uni		50_uni	AX-4		2DS_242				2DS_3	63_uni	
		4iS	2DS_145				AX-2							2DS_1	
		72_uni	AX-29												
		66_uni	AX-28												
		1_al	AX-25												
		5iS													
No. markers	1	7	6	2	1	2	3	1	2	1	1	1	2	3	1

gwm class	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	2DS_210	2DS_89	55_uni	2DS_223	52i	8_uni	2DS_215	2DS_105	BFR_46	2DS_66	2DS_26	2DS_138	2DS_187	2DS_275	2DS_293
	2DS_211	2DS_88	2DS_5375260	2DS_222			2DS_540403			2DS_15			2DS_201		
	2DS_212	2DS_95											2DS_192		
	2DS_217												2DS_208		
													16_uni		
No. markers	4	3	2	2	1	1	2	1	1	2	1	1	5	1	1

Total 62

A6.5: Gwm recs used in step 2 of fine-mapping. The assignment of markers to classes is shown above, along with the totals within each class. The graphical genotype of each marker class is shown overleaf. The score at *Rht8* was taken from Gasperini's (2010) assignation. Classes 1 – 12 were discarded, as described in the text.

No. markers in class		1	7	6	2	1	2	3	1	2	1	1	1		2	3	1	4	3	2	2	1	1	2	1		1	2	1	1	5	1			1	Total 62
gwm class	**Xgwm 261**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**DG279**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**Rht8**	**24**	**25**	**26**	**27**	**28**	**29**	**DG371**	**Xcfd 53**	**30**	
F3-1-1-1-2	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-1-1-5-10	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-1-2-5-3	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-1-6-18-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-1-6-2-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	b
F3-1-7-13-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-1-7-3-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-2-1-10-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	a	a	a	a
F3-2-1-1-6	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	a	a	a	-	b	b	a	a	a	a	a	a	a	a	a	a	b
F3-2-2-5-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-2-3-4-1	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-2-3-5-1	b	b	b	b	b	a	a	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	a	a	a	b
F3-2-7-10-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-2-7-1-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a
F3-2-7-11-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	a	a	a	b	b	b	a	a	a	a	a	a	a	a	a	b
F3-2-7-13-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b
F3-2-7-7-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-2-8-3-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	a	b	a	a	a
F3-3-1-4-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b
F3-3-2-14-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a
F3-3-2-4-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-3-2-9-1	b	b	b	b	b	a	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-3-3-3-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
F3-3-3-9-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	b	a	a	a	a	a	a	b	a	a	a	a	a	-	b	a	b	a	a	a
F3-3-7-15-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	a	a	a	b	b	b	a	a	a	a	a	a	a	a	a	a
F3-2-1-2-2	b	a	a	b	a	a	a	a	b	a	b	a	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b
F3-3-8-7-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	-	a	a	a
F3-1-1-13-11	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	a
F3-1-1-3-5	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	a	a	a	b	b	b	b	a	a	a	a	a	a	a	a	a
F3-3-7-4-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	a
F3-2-1-14-12	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	a	a	a	b	b	b	b	a	a	a	a	a	a	a	a	a
F3-2-1-6-1	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	b	a	a
F3-2-1-9-5	b	b	b	b	b	a	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	b	a	a	a
F3-2-3-6-1	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	a
F3-1-6-15-1	a	b	b	b	b	b	a	a	b	b	b	b	b	b	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	a
F3-2-1-7-1	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	a	b	b	b	a	b	b
F3-3-2-3-1	a	b	a	a	a	a	a	a	b	b	b	a	a	b	b	a	a	a	a	a	b	a	a	a	a	b	b	b	b	b	b	a	b	b	a
F3-3-2-6-1	a	a	a	a	a	a	a	a	b	b	b	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-1-2-3-1	a	a	a	a	a	a	a	a	b	b	b	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-2-7-5-1	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	-	b	a
F3-1-2-8-8	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-1-6-1-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	b	b	b	b	a	a	b	b	b	b	b	b	b	b	b	b	a
F3-1-6-14-1	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	b	b	b	b	b	b	b	b	b	a
F3-1-6-8-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	a	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	b	a
F3-1-7-10-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	b	b	b	a	a	a	b	b	b	b	b	b	b	b	b	a
F3-1-7-11-1	a	a	a	a	a	a	a	a	-	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-1-7-16-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-1-7-2-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-2-1-13-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
F3-2-1-15-2	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
F3-2-2-1-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-2-2-3-5	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	b	a
F3-2-2-4-3	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-2-3-10-1	a	a	a	a	a	a	a	a	-	a	a	a	a	b	a	a	a	a	a	b	b	b	a	a	a	b	b	b	b	b	b	b	b	b	a
F3-2-3-1-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	b	a
F3-2-3-11-1	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	a
F3-2-3-9-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	a
F3-2-7-8-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-3-3-10-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	b	a
F3-3-3-13-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	a	b	b	b	b	b	b	b	b	b	b	b	a
F3-3-3-14-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	a	a	a	a	b	b	b	a	a	a	b	b	b	b	b	b	b	b	b	a
F3-3-3-2-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-3-3-6-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-3-3-7-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-3-3-8-1	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	a
F3-3-8-4-1	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
F3-1-1-4-3	a	a	a	a	a	a	a	-	-	a	a	a	a	b	a	a	a	a	b	b	b	b	a	a	b	-	b	b	b	b	b	b	b	b	b
F3-3-7-11-1	a	a	a	a	a	a	a	b	a	b	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	-	b	b	b	b	b	b	b	b	a
F3-3-7-12-1	a	a	a	a	a	a	a	-	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b	b	-	b	b	b	b	b	b	b	b	a

A6.5 continued: Graphical genotypes of gwm rec marker classes.

FM class	A	B	C	D	G	F	E	
	2DS_1	52i	2DS_223	2DS_5375260	2DS_275	2DS_138	16_uni	
	2DS_95		2DS_222			2DS_66	2DS_192	
	2DS_149		55_uni			2DS_26	2DS_201	
	2DS_217					2DS_15	2DS_187	
	2DS_212					BFR_46	2DS_208	
	2DS_88							
	2DS_89							
	2DS_6							
	2DS_3							
	2DS_4							
	2DS_211							
	2DS_210							
	2DS_540403							
	2DS_105							
	2DS_215							
	8_uni							
	63_uni							Total
No. markers	17	1	3	1	1	5	5	33

A6.6: FM recombinants used in step 3 of fine-mapping. The assignment of markers to classes is shown above, along with the totals within each class. The graphical genotype of each marker class is shown overleaf. The score at *Rht8* was from the consensus score as described in 6.2.

No. markers in class		17	1	3	1		5		1	5	33	Total
FM class	**DG279**	A	B	C	D	**Rht8**	E	**DG371**	F	G		
F4-1-2-2-1	b	b	b	b	b	b	a	a	a	a		
F4-3-7-10-1	b	a	b	b	b	b	b	b	b	b		
F4-1-7-15-1	a	a	b	b	b	b*	b	b	b	b		
F4-1-6-3-1	a	a	a	b	b	b	b	b	b	b		
F4-3-8-1-1	a	a	a	b	b	b	b	b	b	b		
F4-1-7-17-1	a	a	a	a	b	b	b	b	b	b		
F4-3-2-2-1	a	a	a	a	a	b	b	b	b	b		
F4-3-2-16-1	a	a	a	a	a	b	b	b	b	b		
F4-2-1-8-4	a	a	a	a	a	b	b	b	b	b		
F4-2-8-4-1	a	a	a	a	a	b	b	b	b	b		
F4-2-1-12-1	a	a	a	a	a	b	b	b	b	b		
F4-1-1-11-1	a	a	a	a	a	b	b	b	b	b		
F4-1-6-9-1	a	a	a	a	a	b	b	b	b	b		
F4-1-1-9-7	a	a	a	a	a	b	b	b	b	b		
F4-1-2-7-1	a	a	a	a	a	b	b	b	b	b		
F4-3-1-1-6	a	a	a	a	a	b	b	b	b	b		
F4-2-7-12-2	a	a	a	a	a	b	b	b	b	b		
F4-3-7-5-3	a	a	a	a	a	b	b	b	b	b		
F4-3-1-2-6	a	a	a	a	a	b	b	b	b	b		
F4-3-2-5-1	a	a	a	a	a	b	b	b	b	b		
F4-2-7-4-1	a	a	a	a	a	b	b	b	b	b		
F4-3-8-6-3	a	a	a	a	a	b	b	b	b	b		
F4-3-2-8-1	a	a	a	a	a	b	b	b	b	b		
F4-1-7-1-1	a	a	a	a	a	b	b	b	b	b		
F4-3-7-8-2	a	a	a	a	a	b*	b	b	b	b		
F4-2-8-5-2	a	a	a	a	a	b*	b	b	b	b		
F4-3-2-15-1	a	a	a	a	a	b*	b	b	b	b		
F4-2-2-6-1	a	a	a	a	a	b*	b	b	b	b		
F4-2-1-4-1	a	a	a	a	a	b*	b	b	b	b		
F4-2-8-1-2	a	a	a	a	a	b*	b	b	b	b		
F4-2-2-10-1	a	a	a	a	a	b*	b	b	b	b		
F4-1-6-12-1	a	a	a	a	a	b*	b	b	a	b		
F4-3-8-2-2	a	a	a	a	a	a*	b	b	b	b		
F4-3-7-1-2	a	a	a	a	a	a*	b	b	b	b		
F4-1-6-11-1	a	a	a	a	a	a	b	b	b	b		
F4-1-7-18-3	b	b	b	a	a	a	a	a	a	a		
F4-3-7-13-3	b	b	b	a	a	a	a	a	a	a		
F4-1-2-9-1	b	b	b	b	a	a	a	a	a	a		
F4-1-7-4-1	b	b	b	b	a	a	a	a	a	a		
F4-3-2-13-1	b	b	b	b	a	a	a	a	a	a		
F4-3-7-9-1	b	b	b	b	b	a	a	a	b	b		
F4-2-3-7-1	b	b	b	b	b	a*	a	a	a	a		
F4-2-3-2-1	b	b	b	b	b	a*	a	a	a	a		
F4-3-1-3-6	b	b	b	b	b	a	a	a	a	a		
F4-3-2-12-1	b	b	b	b	b	a	a	a	a	a		
F4-1-9-1-1	b	b	b	b	b	a	a	a	a	a		
F4-2-3-8-1	b	b	b	b	b	a	a	a	a	a		
F4-3-7-6-1	b	b	b	b	b	a	a	a	a	a		
F4-3-7-7-2	b	b	b	b	b	a	a	a	a	a		
F4-2-8-6-1	b	b	b	b	b	a	a	a	a	a		
F4-3-2-7-2	b	b	b	b	b	a	a	a	a	a		
F4-3-8-3-3	b	b	b	b	b	a	a	a	a	a		
F4-1-6-19-2	b	b	b	b	b	a	a	a	a	a		
F4-3-8-5-2	b	b	b	b	b	a	a	a	a	a		
F4-2-1-11-4	b	b	b	b	b	a	a	a	a	a		
F4-1-6-17-1	b	b	b	b	b	a	a	a	a	a		
F4-1-6-13-2	b	b	b	b	b	a	a	a	a	a		
F4-2-1-16-3	b	b	b	b	b	a	a	a	a	a		
F4-2-7-3-6	b	b	b	b	b	a	a	a	a	a		
F4-1-7-7-1	b	b	b	b	b	a	a	a	a	a		
F4-3-3-15-1	b	b	b	b	b	a	a	a	a	a		
F4-2-2-2-3	b	b	b	b	b	a	a	a	a	a		
F4-1-9-2-1	b	b	b	b	b	a	a	a	a	a		
F4-1-9-3-1	b	b	b	b	b	a	a	a	a	a		
F4-2-2-7-1	b	b	b	b	b	a	a	a	a	a		
F4-1-6-16-1	b	b	b	b	b	a	a	a	a	a		
F4-2-7-9-2	b	b	b	b	b	a	a	a	a	a		
F4-2-7-6-2	b	b	b	b	b	a	a	a	a	a		
F4-3-7-14-3	b	b	b	b	b	a	a	a	a	a		
F4-1-1-12-1	b	b	b	b	b	a	a	a	a	a		
F4-2-7-2-1	-	b	b	b	b	a	a	a	a	a		
F4-1-2-4-3	b	b	b	b	b	a	a	a	a	a		
F4-1-1-7-3	b	b	b	b	b	a	a	a	a	a		

A6.6 continued: Graphical genotpyes of FM marker classes.

		SNP marker extended ctg			BAC ctg							
Chr	Linkage (cM)	ID	Start	End	ID	size	*B. distachyon* homologue	Comments	Rice homologue	Gene ontology	In SNP data?	DEG?
2D	30.174	AT2D1055_1	1	2739	ctg7056	1065	Bradi5g02990	2DS_86/7, 91/2	Os07g04160	FKBP-like peptidyl-prolyl cis-trans isomerase	Y	N
		AT2D1056_1	3175	5628			Bradi5g02980	2DS_5364496 2DS_5388088	Os04g12560	receptor-like protein kinase 1	N	N
	30.219	AT2D1057_1	1461	1994	ctg5606	397.5	Bradi4g21260	2DS_6	Os11g10870	Disease resistance-responsive dirigent-like protein	N	N
	30.401	AT2D1059_1	26	236	ctg494	1611	No Hit		No Hit			
		AT2D1059_2	270	591			No Hit		No Hit			
		AT2D1059_3	1484	2068			Bradi1g57240		Os01g29330	Unknown function (DUF679)	N	N
		AT2D1058_1	3476	6183			Bradi4g02240		Os07g38250	Concanavalin A-like lectin protein kinase	N	N
	30.447	AT2D1060_1	1	128			No Hit		No Hit			
		AT2D1060_2	223	555			Bradi5g03380	2DS_88	Os04g12980	UDP-glucosyltransferase 74F2	N	N
		AT2D1060_3	2068	2192			No Hit		No Hit			
		AT2D1060_4	2208	2256			No Hit		No Hit			
		AT2D1060_5	2335	2429			No Hit		No Hit			
		AT2D1060_6	2624	7604			Bradi5g03380	2DS_88	Os04g12960	UDP-glucosyltransferase 74F2	N	N
		AT2D1060_7	13032	13724			Bradi5g03380	2DS_88	Os04g12980	UDP-glucosyltransferase 74F2	N	N
		AT2D1060_8	15189	16696			Bradi5g03380	2DS_88	Os04g12960	UDP-glucosyltransferase 74F2	N	N
		AT2D1060_9	16833	16892			No Hit		No Hit			
		AT2D1060_10	18259	18489			Bradi4g39460		Os12g14080	Mitochondrial ribosomal protein L37	N	N
		AT2D1060_11	21382	21969			Bradi5g03380	2DS_88	Os04g12980	UDP-glucosyltransferase 74F2	N	N
		AT2D1060_12	22020	23030			Bradi5g03380	2DS_88	Os04g12960	UDP-glucosyltransferase 74F2	N	N
		AT2D1060_13	25084	25461			No Hit					
		AT2D1060_14	25500	25577			No Hit					
	30.539	AT2D1061_1	1393	2246			Bradi5g03390	2DS_88	Os04g12970	UDP-glucosyltransferase 74F2	N	N
		AT2D1061_2	2921	2993			No Hit		No Hit			
		AT2D1061_3	4020	4505			Bradi5g03380	2DS_4	Os04g12980	UDP-glucosyltransferase 74F2	N	N
	30.994	AT2D1062_1	4013	12146			Bradi5g03460	52i	Os04g13210	multidrug resistance-associated protein 10	Y, mono	N
		AT2D1062_2	12146	12247			No Hit		No Hit			
	31.272	AT2D1064_1	402	546			No Hit		No Hit			
		AT2D1064_2	3936	6541			Bradi3g16570		No Hit		N	N
	31.408	AT2D1065_1	1171	2949			Bradi5g05140	2DS_5375260	Os03g58960	DHHC-type zinc finger family protein	N	N
		AT2D1065_2	4063	4134			No Hit		No Hit			
		AT2D1065_3	4248	5876			Bradi5g09000	2DS_118-120	Os04g20880	Membrane bound O-acyl transferase (MBOAT) family protein	N	N
		AT2D1066_2	4882	6362			Bradi5g03477	2DS_103	Os04g13220	multidrug resistance-associated protein 4	N	N
		AT2D1066_3	8018	8158			No Hit		No Hit			
	32.412	AT2D1067_1	2823	2995			No Hit		No Hit			
		AT2D1067_2	4053	4973			Bradi2g20430		Os05g43252	Unknown function (DUF594)	N	N
		AT2D1067_4	7250	7501			No Hit		No Hit			
		AT2D1067_5	8454	8641			No Hit		No Hit			
	33.829	AT2D1068_1	2263	3280	ctg6164	360	Bradi4g27670		Os02g29810	Ribonuclease H-like superfamily protein	N	N

A6.7.1

		SNP marker extended ctg			BAC ctg							
Chr	Linkage (cM)	ID	Start	End	ID	size	*B. distachyon* homologue	Comments	Rice homologue	Gene ontology	In SNP data?	DEG?
2D	35.339	AT2D1070_1	236	8123	ctg1775	2280	Bradi5g04673	2DS_19/ 2DS_96	Os04g18010	cleavage and polyadenylation specificity factor 160	N	N
	35.385	AT2D1071_1	76	351			No Hit		No Hit			
		AT2D1071_2	4510	4680			No Hit		No Hit			
		AT2D1071_3	4739	5323			Bradi4g34520		Os03g63270	Nucleotide-diphospho-sugar transferase family protein	N	N
		AT2D1071_4	5787	6758			Bradi2g16396		Os05g49070	6-phosphogluconate dehydrogenase family protein	N	N
	35.657	AT2D1072_1	88	4559			Bradi5g04660	BFR_46	Os02g21490	Transducin/WD40 repeat-like superfamily protein	N	N
		AT2D1072_2	4915	5469			Bradi1g48292		Os12g14440	Mannose-binding lectin superfamily protein	N	N
	35.93	AT2D1073_1	2872	3131			Bradi3g01060		Os02g01940	Transcription factor jumonji (jmjC) domain-containing protein	Y, untested	N
	36.203	AT2D1074_1	5372	6522			Bradi5g03200		Os10g12620	protein serine/threonine kinase	N	N
		AT2D1074_2	8717	8974			Bradi4g22590		No Hit	plant U-box 13	N	N
	36.339	AT2D1075_1	2407	3136	ctg5773	1182	No Hit		No Hit			
		AT2D1075_5	4467	5144			Bradi5g20420		Os04g51090	P-loop containing nucleoside triphosphate hydrolase		Y , x1.87 short
	36.566	AT2D1076_1	215	566			Bradi4g14070		Os11g12410	Serine protease inhibitor (SERPIN) family protein	N	N
	37.112	AT2D1078_1	786	10484	ctg546	835.5	Bradi5g04567	2DS_14	Os02g01170	HEAT repeat, HECT-domain (ubiquitin-transferase)	N	N
		AT2D1078_2	10490	10828			Bradi5g04567	2DS_14	No Hit		N	N
		AT2D1077_1	2088	4075			Bradi5g04577	2DS_5358467	Os04g14510	RING/FYVE/PHD zinc finger superfamily protein	N	N
	37.294	AT2D1079_1	914	3648			Bradi5g04340	2DS_26	Os01g24980	2-oxoglutarate & Fe(II)-dependent oxygenase superfamily protein	N	N
	37.385	AT2D1080_2	8990	11232	ctg2121	2155.5	Bradi1g62430		Os03g22730	NOP56-like pre RNA processing ribonucleoprotein	N	N
		AT2D1080_3	12231	12473			Bradi1g36300		Os01g24980	2-oxoglutarate & Fe(II)-dependent oxygenase superfamily protein	N	N
		AT2D1081_1	913	1184			No Hit		No Hit			
		AT2D1081_2	1494	1765			No Hit		No Hit			
		AT2D1081_3	6154	7134			Bradi5g04340	2DS_26	Os01g24980	2-oxoglutarate & Fe(II)-dependent oxygenase superfamily protein	N	N
		AT2D1081_5	11376	11647			No Hit		No Hit			
	37.933	AT2D1082_1	1056	1428	ctg4363	420	Bradi2g09095		No Hit		N	N
	37.978	AT2D1083_1	2	229			Bradi2g58600		Os09g38170	Ribonuclease H-like superfamily protein	N	N
		AT2D1083_2	921	1997			Bradi5g04130	2DS_138	Os04g15920	cinnamyl alcohol dehydrogenase 9	N	N
	38.069	AT2D1084_1	119	319			No Hit		No Hit			
		AT2D1084_2	5420	6788			Bradi4g08800		Os12g19470	RuBisCO small chain 1A	Y, untested	N

A6.7.1 continued

List of genes surrounding the *Rht8* interval in the *Ae. tauschii* physical map with homology to syntenic species and gene ontology. The interval was delimited to the region between the bins highlighted in yellow. Comments: indicates attempt at marker development and variant identification. Shaded green indicates a successful (polymorphic marker). Shaded red are markers which were tested but were monomorphic between RIL4 and Cappelle-Desprez. If the gene was in the UniGene SNP data, this is indicated. Differential expression is also indicated (direction and magnitude as fold difference). The gene list was downloaded from http://probes.pw.usda.gov/WheatDMarker/downloads/.

Chr	Marker	Linkage (cM)	*B. distachyon* homologue	Comments	Rice homologue	In SNP data?	DEG?
2D	AT2D1061	30.539	-	-	-	-	-
	AT2D1062	30.994	Bradi5g03460	52i	Os04g0209200	Y, mono	N
	AT2D1064	31.27	-	-	-	-	-
	AT2D1066	31.408	Bradi5g03477	2DS_103	Os04g0209300	N	N
			Bradi5g03500	2DS_5339566 - no polymorphism	-	N	N
			Bradi5g03510	2DS_5343394 - contains microsatellite	-	N	N
			Bradi5g03530	2DS_5292808 - no polymorphism	-	N	N
			Bradi5g03550	2DS_51; Chapman 2DS_297	-	N	N
			Bradi5g03570	No 2DS contig	-	N	N
	AT2D1065	31.408	Bradi5g09000	2DS_118-120	-	N	N
			Bradi5g09010	2DS_5382800 - microsatellite, alignment 23 bases only	-	N	N
			Bradi5g09020	2DS_5367564 - no polymorphism	-	N	N
			Bradi5g09030	2DS_5367564 - no polymorphism	Os04g0406600	N	N
			Bradi5g09040	2DS_5327211 - no polymorphism	-	N	N
			Bradi5g09064	2DS_5333932 - no polymorphism	-	N	N
			-	-	Os04g0407800	-	-
			Bradi5g09090	2DS_5382207 - no polymorphism	Os04g0407500	Y, untested (2BL)	N
			Bradi5g04686	2DS_5364728 - no polymorphism	Os04g0252400	N	N
			-	-	Os04g0252300	-	-
			-	-	-	-	-
	AT2D1067	32.412	-	-	-	-	-
	AT2D1068	33.829	-	-	-	-	-
	AT2D1069	35.339	-	-	-	-	-
	AT2D1070	35.339	Bradi5g04673	2DS_19 & Chapman 2DS_296	Os04g0252200	N	N
			-	-	Os04g0252000	-	-
			-	-	Os04g0249700	-	-
			-	-	Os04g0249500	-	-
			Bradi5g04630	2DS_76 & 2DS_60	-	N	N
			Bradi5g04640	2DS_97-98	-	N	N
			-	-	-	-	-
			Bradi5g04650	2DS_5330217 - contains microsatellites	-	N	N
	AT2D1071	35.385	-	-	-	-	-
	AT2D1072	35.657	Bradi5g04660	BFR_46	-	N	N
	AT2D1073	35.930	-		-	-	-
	AT2D1074	36.203	-		-	-	-
	AT2D1075	36.339	Bradi5g20420		Os04g0599650	N	Y, x1.87 short

A6.7.2

Chr	Marker	Linkage (cM)	*B. distachyon* homologue	Comments	Rice homologue	In SNP data?	DEG?
	-	-	Bradi5g20427		-	N	N
	-	-	Bradi5g20440		-	N	N
	-	-	Bradi5g20450		-	N	N
	-	-	Bradi5g04610		-	N	N
	-	-	Bradi5g04600	2DS_67	-	N	N
	-	-	-		-	-	-
	-	-	Bradi5g04590	2DS_5320736 - no polymorphism	-	N	N
	-	-	-		Os04g0218100	-	-
	-	-	-		Os04g0218600	-	-
	-	-	-		Os04g0218900	-	-
	-	-	-		Os04g0219600	-	-
	-	-	-		Os04g0220300	-	-
	-	-	-		Os04g0220500	-	-
	-	-	-		Os04g0221000	-	-
	-	-	-		Os04g0221300	-	-
	-	-	-		-	-	-
	-	-	Bradi5g04577	2DS_5358467 - no polymorphism	Os04g0221600	N	N
	AT2D1076	**36.566**	-		-	-	-
	AT2D1077	**37.112**	-		-	-	-
2D	**AT2D1078**	**37.112**	Bradi5g04567	2DS_14	-	N	N
	-	-	Bradi5g04560	2DS_19 & Chapman 2DS_296	Os04g0244800	N	N
	-	-	-		Os04g0244900	-	-
	-	-	-		Os04g0245000	-	-
	-	-	-		Os04g0244500	-	-
	-	-	Bradi5g04550	2DS_15	Os04g0244400	N	N
	-	-	Bradi5g04540	2DS_26	Os04g0243700	N	N
	AT2D1079	**37.294**	-		-	-	-
	AT2D1081	**37.385**	-		-	-	-
	AT2D1080	**37.385**	Bradi5g04340	2DS_26	-	N	N
	AT2D1082	**37.933**	-		-	-	-
	AT2D1083	**37.978**	Bradi5g04130	2DS_138	Os04g0229100	N	N
	-	-	Bradi5g04120		Os04g0228400	N	N
	-	-	Bradi5g04090		-	N	N
	-	-	Bradi5g04080		-	N	N
	-	-	Bradi5g04070		-	N	N
	-	-	-		-	-	-
	-	-	-		-	-	-
	-	-	Bradi5g04057		Os04g0228100	N	N
	AT2D1084	**38.069**	-		-	-	-

A6.7.2 continued

Gene Zipper surrounding the *Rht8* interval in the *Ae. tauschii* physical map with homology to syntenic species. The interval was delimited to the region between the bins highlighted in yellow. Comments: indicates attempt at marker development and variant identification. Shaded green indicates a successful (polymorphic marker). Shaded red are markers which were tested but were monomorphic between RIL4 and Cappelle-Desprez. If the gene was in the UniGene SNP data, this is indicated. Differential expression is also indicated (direction and magnitude as fold difference). The zipper was downloaded from http://probes.pw.usda.gov/WheatDMarker/downloads/.

Chr	SNP_Name	cM	ctg #	BAC ctg length	*B. distachyon* chromosome (Mb)						Rice chromosome (Mb)											
					1	2	3	4	5		1	2	3	4	5	6	7	8	9	10	11	12
2D	AT2D1058	30.40	ctg494	**1611**				1.45									22.93					
	AT2D1059	30.40			55.96						16.43											
	AT2D1060	30.45							3.95					7.15								
	AT2D1061	30.54							3.95					7.15								
	AT2D1062	30.99							4.03					7.30								
	AT2D1063	31.04																				
	AT2D1064	31.27					14.75															
	AT2D1065	31.41							6.61				33.56									
	AT2D1066	31.41							4.06					7.31								
	AT2D1067	32.41				17.94									25.10							
	AT2D1068	33.83	ctg6164	**360**				32.83				17.73										
	AT2D1069	35.34	ctg1775	**2280**					5.89					9.88								
	AT2D1070	35.34							5.89					9.88								
	AT2D1071	35.39																				
	AT2D1072	35.66							5.89			12.74										
	AT2D1073	35.93					0.62					0.51										
	AT2D1074	36.20							3.70									7.04				
	AT2D1075	36.34	ctg5773	**1182**					23.30					30.06								
	AT2D1076	36.57						14.53													6.92	
	AT2D1077	37.11	ctg5546	**835.5**					5.76					8.11								
	AT2D1078	37.11							5.74			0.09										
	AT2D1079	37.29							5.54		14.08											
	AT2D1080	37.39	ctg2121	**2155.5**					5.54		14.08											
	AT2D1081	37.39							5.54		14.08											
	AT2D1082	37.93	ctg4363	**420**		7.40																
	AT2D1083	37.98							5.39					8.63								
	AT2D1084	38.07						7.94														11.32

A6.7.3 Comparative map of the linkage and physical maps of *Ae. tauschii* surrounding the *Rht8* interval. *B. distachyon* and rice genes with highest homology for each *Ae. tauschii* marker are shown. The delimited interval in *Ae. tauschii* is indicated by yellow. Cells containing *B. distachyon* and rice genes that are collinear with the *Ae. tauschii* genes are a shade of green. Cells that contained genes that were non-collinear received no colour. If no corresponding gene was detected, the cell was left empty. Light green was used for genes showing progression on the *B. distachyon* and rice pseudomolecules that were in the same orientation as on the linkage map of *Ae. tauschii*, either increasing or decreasing. Dark green was used for regions with changed progression due to an inversion or translocation. The comparative map was downloaded from http://probes.pw.usda.gov/WheatDMarker/downloads/.

A6.8.1 (continuing overleaf over multiple pages)

Differentially expressed genes between the parent NILs in the UniGene dataset. Genes are annotated if they overlapped with SNP data. Raw RPKM values are shown, as well as log-transformed data used to set the 1.5-fold threshold as a DEG. The data is sorted according to the magnitude of the fold difference, with the exception of the first row.

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp269261_c0_seq1	4	21331017	Bradi4g14000	Os11g39370	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor	0.11	0.35	-3.20	-1.53	1.67	short	N
D_comp85949_c0_seq1	7	13762683	Bradi3g21070	Os10g05720	lipid binding protein	0.63	145.21	-0.67	7.18	7.85	short	N
D_comp287929_c0_seq1	4	38272418	Bradi1g73800	Os03g07250	cytochrome P450 86A2	0.44	61.43	-1.18	5.94	7.12	short	N
B_comp3722_c1_seq14	2	12873029	Bradi1g24410	Os07g37850	-	0.16	20.19	-2.61	4.34	6.94	short	N
B_comp286_c0_seq1	6	23663505	Bradi3g56660	Os02g48900	aspartic proteinase nepenthesin-1 precursor	0.13	14.36	-2.89	3.84	6.74	short	N
D_comp276122_c0_seq1	4	19846598	Bradi4g15260	Os11g37280	protease inhibitor/seed storage/LTP family protein	0.14	12.80	-2.85	3.68	6.53	short	N
B_comp3228_c0_seq4	7	7979175	Bradi3g40290	Os08g40440	dihydroflavonol-4-reductase	1.37	87.02	0.45	6.44	5.99	short	N
B_comp50536_c0_seq1	5	11579581	Bradi4g25600	Os12g02110	conserved hypothetical protein	1.48	94.26	0.56	6.56	5.99	short	Y
B_comp68504_c0_seq1	5	21040141	Bradi4g34010	Os09g32020	ubiquitin fusion degradation protein	1.16	72.25	0.22	6.17	5.96	short	N
D_comp891_c0_seq1	4	12536436	Bradi4g20200	Os11g18690	beta-xylosidase putative	0.16	9.50	-2.65	3.25	5.90	short	N
D_comp27648_c0_seq1	3	11233321	Bradi2g12870	Os01g26000	acyltransferase/ catalytic	0.38	18.61	-1.39	4.22	5.61	short	N
D_comp572116_c0_seq1	6	4994809	Bradi3g06780	Os08g05620	cytochrome P450 89A2	0.38	16.60	-1.41	4.05	5.46	short	N
A_comp24242_c0_seq1	3	17156	Bradi2g00220	Os01g06580	fasciclin-like arabinogalactan protein 7 precursor	0.28	8.66	-1.85	3.12	4.97	short	N
D_comp4308_c0_seq1	1	26290168	Bradi3g14150	Os08g03350	LHT1	0.19	5.69	-2.37	2.51	4.88	short	N
B_comp85844_c0_seq1	4	41175127	Bradi1g77990	Os03g01800	xyloglucan endotransglucosylase/hydrolase protein 32 precursor	0.18	5.20	-2.49	2.38	4.87	short	N
D_comp35107_c0_seq1	1	9699531	Bradi3g27810	Os10g31330	glycine-rich cell w all structural protein 2 precursor	0.24	6.50	-2.09	2.70	4.79	short	N
B_comp736_c1_seq9	1	39365389	Bradi2g60500	Os05g41610	glucan endo-13-beta-glucosidase GVI precursor	0.20	5.15	-2.34	2.37	4.70	short	Y
B_comp88_c0_seq2	5	2997960	Bradi4g03730	Os12g38810	-	0.10	2.66	-3.28	1.41	4.69	short	N
B_comp6469_c0_seq5	3	10384146	Bradi2g12060	Os01g21630	carboxylic ester hydrolase	0.11	2.85	-3.13	1.51	4.64	short	N
A_comp92957_c0_seq1	3	17510688	Bradi2g44420	Os01g44130	aspartic proteinase oryzasin-1 precursor	0.20	4.88	-2.29	2.29	4.57	short	Y
D_comp228299_c0_seq1	7	32599870	Bradi1g43820	Os06g15760	aspartic proteinase Asp1 precursor	0.31	6.93	-1.70	2.79	4.50	short	N
D_comp552498_c0_seq1	6	11405087	Bradi3g44990	Os09g08072	beta-fructofuranosidase insoluble isoenzyme 7 precursor	0.14	2.94	-2.86	1.56	4.42	short	N
B_comp4421_c0_seq1	7	22152709	Bradi3g13030	Os04g58860	VAMP protein SEC22	0.53	11.17	-0.92	3.48	4.40	short	N
B_comp35694_c0_seq1	7	20957003	Bradi3g14430	Os08g04140	glucan endo-13-beta-glucosidase precursor	1.09	22.87	0.12	4.52	4.40	short	N
A_comp137166_c0_seq1	1	13899958	Bradi3g30870	Os10g37400	-	0.62	12.86	-0.69	3.69	4.38	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp61734_c0_seq1	7	34960256	Bradi1g45710	Os06g11490	blue copper protein precursor	0.78	15.94	-0.36	3.99	4.35	short	N
B_comp49996_c0_seq1	4	3054097	Bradi1g10270	Os03g51430	-	0.15	3.06	-2.71	1.61	4.32	short	N
B_comp27789_c0_seq1	5	3250661	Bradi1g32380	Os07g01214	potassium transporter 16	0.10	2.06	-3.26	1.04	4.30	short	N
A_comp9240_c0_seq1	2	40149034	Bradi5g17320	Os04g45960	serine proteinase	3.05	56.15	1.61	5.81	4.20	short	N
B_comp23174_c0_seq1	1	30285759	Bradi2g28380	Os05g28730	PIT1	0.16	2.94	-2.61	1.55	4.17	short	N
B_comp22716_c0_seq5	1	24379584	Bradi2g33380	Os07g26540	hexokinase-1	0.54	9.68	-0.88	3.28	4.16	short	Y
B_comp54013_c0_seq1	7	2769475	Bradi1g49010	Os06g06250	alpha-L-fucosidase 2 precursor	1.24	21.94	0.31	4.46	4.14	short	N
D_comp144488_c0_seq1	7	23802210	Bradi1g36990	Os06g39390	3-N-debenzoyl-2-deoxytaxol N-benzoyltransferase	0.12	2.01	-3.10	1.01	4.10	short	N
B_comp7760_c0_seq1	4	31772379	Bradi1g65780	Os03g17790	-	0.17	2.95	-2.53	1.56	4.09	short	Y
D_comp497706_c0_seq1	4	18490609	Bradi1g24840	Os10g17260	flavonoid 3-monooxygenase	0.10	1.65	-3.31	0.73	4.03	short	N
D_comp288047_c0_seq1	4	9341384	Bradi4g22250	Os11g10870	dirigent-like protein pDIR3	0.12	1.89	-3.09	0.92	4.01	short	N
D_comp72743_c0_seq1	7	21010579	Bradi3g14390	Os08g03682	flavonoid 3-monooxygenase	0.19	2.97	-2.38	1.57	3.96	short	N
A_comp707567_c0_seq1	2	7506671	Bradi1g17670	Os07g48229	vacuolar sorting receptor 1 precursor	0.18	2.78	-2.47	1.48	3.94	short	N
B_comp1755_c0_seq1	7	31388623	Bradi1g42970	-		1.41	21.28	0.50	4.41	3.92	short	N
B_comp371_c0_seq2	4	8706807	Bradi4g22660	Os11g10460	peroxidase 43 precursor	0.63	9.40	-0.67	3.23	3.90	short	N
D_comp176653_c0_seq1	2	6464928	Bradi1g22290	Os03g30830	glucan endo-13-beta-glucosidase 4 precursor	0.11	1.59	-3.23	0.67	3.90	short	N
A_comp14906_c0_seq1	3	2951804	Bradi2g04230	Os01g07420	hydrolase	0.12	1.77	-3.07	0.82	3.89	short	N
B_comp29229_c0_seq2	6	20231120	Bradi3g58560	Os02g52180	uclacyanin-2 precursor	0.13	1.86	-2.99	0.90	3.89	short	N
B_comp71716_c0_seq1	3	13506208	Bradi2g40620	Os01g36460	myb-related protein Hv33	0.67	9.76	-0.59	3.29	3.87	short	N
B_comp32069_c0_seq2	6	6941358	Bradi3g08820	Os02g13660	meiosis 5	0.31	4.49	-1.70	2.17	3.86	short	N
A_comp1592163_c0_seq1	3	28290748	Bradi2g57570	Os01g67090	SF16 protein	0.33	4.56	-1.61	2.19	3.79	short	N
A_comp429700_c0_seq1	4	40447265	Bradi1g76970	Os03g03164	homeobox protein HD1	0.10	1.43	-3.26	0.52	3.78	short	Y
D_comp19616_c0_seq1	1	30590069	Bradi2g28100	Os05g28830	-	0.29	3.82	-1.77	1.93	3.70	short	N
D_comp211234_c0_seq1	6	5592676	Bradi3g07450	Os02g10770	pre-mRNA-processing ATP-dependent RNA helicase PRP5	0.23	2.97	-2.12	1.57	3.69	short	N
A_comp522846_c0_seq1	5	11877780	Bradi4g44930	Os12g01490	DNA binding protein	0.23	2.90	-2.15	1.54	3.68	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp625666_c0_seq1	7	33231562	Bradi1g44170	Os06g14420	nudix hydrolase 4	1.42	0.12	0.50	-3.07	3.57	tall	N
B_comp15071_c0_seq6	2	1274379	Bradi5g01360	Os09g38239	flavonol sulfotransferase-like	0.22	2.50	-2.21	1.32	3.53	short	N
B_comp100451_c0_seq1	5	18119763	Bradi4g30540	Os09g25490	CESA9 - cellulose synthase	0.20	2.24	-2.36	1.16	3.52	short	N
A_comp71803_c0_seq1	2	44579682	Bradi5g22800	Os04g53920	leucoanthocyanidin reductase putative	0.28	3.16	-1.85	1.66	3.51	short	N
D_comp87309_c0_seq4	3	9643340	Bradi2g11380	Os01g19170	polygalacturonase precursor	0.12	1.34	-3.09	0.42	3.51	short	N
B_comp15763_c0_seq12	7	8818592	Bradi3g39300	Os08g38810	RAFTIN1a protein	3.18	35.20	1.67	5.14	3.47	short	N
D_comp2818_c0_seq1	1	24235837	Bradi2g33510	Os05g09280	ischemia related factor NYW-1	1.24	0.12	0.31	-3.11	3.42	tall	N
B_comp10437_c0_seq1	7	16176561	Bradi3g19020	Os08g15060	ATP binding protein	0.30	3.15	-1.76	1.65	3.41	short	N
B_comp1585_c0_seq1	5	31742116	Bradi1g02940	Os03g61670	calreticulin precursor	3.00	31.48	1.59	4.98	3.39	short	N
B_comp6072_c0_seq7	1	10452549	Bradi3g28350	Os10g32980	CESA7 - cellulose synthase	0.26	2.71	-1.94	1.44	3.38	short	N
A_comp61049_c0_seq1	5	12778912	Bradi3g06550	Os09g04710	anther-specific proline-rich protein APG precursor	0.78	8.17	-0.35	3.03	3.38	short	N
D_comp19734_c0_seq1	4	34784897	Bradi1g69220	Os03g12990	phytosulfokine precursor protein containing protein	0.15	1.53	-2.75	0.62	3.37	short	N
B_comp84187_c0_seq1	5	17974597	Bradi1g53470	Os09g25690	conserved hypothetical protein	0.32	3.18	-1.65	1.67	3.32	short	N
D_comp305238_c0_seq1	6	246519	Bradi3g00510	Os04g41490	DNA-directed RNA polymerase III largest subunit	2.23	0.22	1.16	-2.16	3.32	tall	N
D_comp97709_c0_seq1	5	27779787	Bradi1g08410	Os03g54050	circumsporozoite protein precursor	1.32	13.06	0.40	3.71	3.31	short	Y
D_comp58960_c0_seq2	3	7877732	Bradi2g09580	Os01g15770	-	0.23	2.29	-2.10	1.20	3.30	short	N
D_comp9793_c0_seq1	1	15163893	Bradi1g77140	Os08g02110	peroxidase 47 precursor	0.12	1.16	-3.09	0.21	3.30	short	N
A_comp11837_c0_seq1	2	20628906	Bradi1g56940	Os07g06800	3-oxo-5-alpha-steroid 4-dehydrogenase 2	0.17	1.65	-2.57	0.72	3.29	short	N
B_comp76815_c0_seq1	5	11523029	Bradi4g44390	Os11g02440	chalcone--flavonone isomerase	0.60	5.87	-0.73	2.55	3.29	short	N
B_comp2591_c0_seq2	2	1159118	Bradi5g01240	Os11g31090	transferase	0.29	2.82	-1.79	1.50	3.28	short	N
D_comp35664_c0_seq1	2	19270510	Bradi1g55770	Os07g07930	lipid transfer protein	0.10	0.99	-3.28	-0.01	3.27	short	N
A_comp20593_c4_seq4	2	31884324	Bradi5g09130	Os04g33450	-	1.23	11.82	0.29	3.56	3.27	short	N
D_comp186430_c0_seq1	2	15107801	Bradi1g26760	Os07g32620	anthocyanidin 53-O-glucosyltransferase	0.21	2.00	-2.23	1.00	3.23	short	N
B_comp27865_c0_seq1	6	20211751	Bradi3g58590	Os06g11610	heat shock 22 kDa protein mitochondrial precursor	0.14	1.27	-2.88	0.34	3.22	short	N
D_comp73459_c0_seq1	4	24456693	Bradi4g11540	Os11g42550	disease resistance response protein 206	0.58	5.39	-0.78	2.43	3.22	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp206156_c0_seq2	7	28492254	Bradi1g41000	Os06g35410	growth regulator	0.11	0.97	-3.24	-0.04	3.20	short	N
A_comp344259_c0_seq1	3	26120719	Bradi2g54680	Os01g62480	L-ascorbate oxidase precursor	0.16	1.47	-2.62	0.56	3.17	short	N
B_comp43777_c0_seq1	1	9154373	Bradi3g27430	Os10g30310	protein binding protein	0.16	1.47	-2.61	0.56	3.17	short	N
D_comp55423_c0_seq1	7	39815910	Bradi1g35570	Os02g09850	-	0.25	2.27	-1.97	1.18	3.15	short	N
A_comp81542_c0_seq1	2	32539187	Bradi5g09670	Os02g33550	VAMP protein SEC22	1.72	15.06	0.78	3.91	3.13	short	Y
B_comp49764_c0_seq1	4	27228235	Bradi2g03240	Os11g47809	metallothionein-like protein 1	0.11	0.92	-3.25	-0.12	3.13	short	N
A_comp1890793_c0_seq1	2	14384002	Bradi1g25990	Os07g34020	-	0.26	2.24	-1.95	1.17	3.11	short	N
D_comp37530_c0_seq1	5	11526288	Bradi4g44400	Os11g02424	-	1.57	13.38	0.65	3.74	3.09	short	N
A_comp353515_c0_seq1	3	22513163	Bradi2g49910	Os01g54620	CESA4 - cellulose synthase	0.45	3.76	-1.16	1.91	3.07	short	N
B_comp39660_c0_seq3	5	20917704	Bradi4g33890	Os09g31502	dihydroflavonol-4-reductase	0.47	3.99	-1.07	2.00	3.07	short	Y
B_comp76192_c0_seq1	1	39610420	Bradi2g20820	Os07g01370	peroxidase 1 precursor	0.17	1.42	-2.56	0.50	3.06	short	N
D_comp632320_c0_seq1	3	19795691	Bradi2g46770	Os01g49160	myb-like DNA-binding domain containing protein	0.13	1.09	-2.91	0.13	3.04	short	N
D_comp127_c0_seq1	5	3584347	Bradi4g04380	Os01g33204	retrotransposon protein unclassified	0.12	0.94	-3.11	-0.09	3.03	short	Y
B_comp31251_c0_seq1	3	5234025	Bradi2g07000	Os01g11810	-	0.36	2.96	-1.46	1.56	3.02	short	N
B_comp15235_c0_seq1	5	31799850	Bradi1g02820	Os09g32230	vignain precursor	0.31	2.56	-1.67	1.35	3.02	short	Y
B_comp482_c0_seq1	6	12571065	Bradi5g15980	Os02g14150	transposon protein Ac/Ds sub-class	0.18	1.47	-2.46	0.56	3.02	short	N
A_comp3243_c0_seq8	2	38852918	Bradi5g15850	Os06g30950	integral membrane protein like	3.81	30.72	1.93	4.94	3.01	short	N
D_comp567835_c0_seq1	5	9159998	Bradi4g41340	Os12g07580	disease resistance response protein 206	0.18	1.48	-2.44	0.57	3.01	short	N
D_comp269_c0_seq1	7	1497444	Bradi1g52210	Os02g01590	beta-fructofuranosidase 1 precursor	0.12	0.99	-3.00	-0.02	2.99	short	Y
B_comp10384_c0_seq1	7	3733596	Bradi1g50060	Os06g04090	NAM protein	0.16	1.27	-2.64	0.35	2.99	short	N
D_comp519283_c0_seq1	7	16575286	Bradi3g18680	Os11g17504	retrotransposon protein Ty1-copia subclass	0.11	0.90	-3.13	-0.15	2.98	short	N
A_comp540781_c0_seq1	5	18668835	Bradi4g30010	Os09g24520	pentatricopeptide repeat protein PPR868-14	2.12	0.27	1.09	-1.89	2.98	tall	N
B_comp3687_c0_seq1	2	7980041	Bradi1g18270	Os07g47400	SRC2	0.11	0.86	-3.18	-0.22	2.96	short	N
D_comp648133_c0_seq1	7	23470842	Bradi1g36790	Os10g36848	cytochrome P450 84A1	0.13	1.00	-2.95	0.00	2.95	short	N
D_comp418526_c0_seq1	2	35779899	Bradi2g40590	Os07g49360	peroxidase 21 precursor	0.25	1.89	-2.03	0.92	2.95	short	Y

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp32647_c0_seq5	5	24382988	Bradi1g22750	Os04g57490	cysteine protease 1 precursor	0.19	1.46	-2.39	0.55	2.94	short	N
D_comp27038_c0_seq1	7	40090217	Bradi1g35340	Os06g43044	anther-specific proline-rich protein APG	0.11	0.85	-3.16	-0.24	2.92	short	N
A_comp853438_c0_seq1	1	20566547	Bradi2g36910	Os05g04584	3-N-debenzoyl-2-deoxytaxol N-benzoyltransferase	0.22	1.67	-2.18	0.74	2.91	short	N
B_comp12157_c0_seq1	5	30450300	Bradi1g04930	Os03g58940	lipid binding protein putative	0.41	3.03	-1.29	1.60	2.89	short	Y
A_comp21677_c0_seq1	1	12065328	Bradi3g29310	Os10g34480	cytochrome P450 86A2	0.64	4.73	-0.65	2.24	2.89	short	N
B_comp1755_c0_seq2	3	649518	Bradi2g01050	Os01g02150	sulfated surface glycoprotein 185 precursor	0.16	1.21	-2.60	0.27	2.87	short	N
D_comp548087_c0_seq1	4	36242528	Bradi1g71040	Os03g10440	14-beta-xylanase	1.13	8.30	0.18	3.05	2.87	short	N
D_comp186644_c0_seq1	4	32582100	Bradi1g66720	Os03g16610	L-ascorbate oxidase precursor	0.13	0.95	-2.95	-0.08	2.87	short	N
A_comp172908_c0_seq2	2	5184615	Bradi5g04030	Os04g15690	DSBA-like thioredoxin domain containing protein	0.14	1.03	-2.83	0.04	2.87	short	Y
D_comp43281_c0_seq1	2	2616946	Bradi5g02460	Os04g01140	cytochrome P450 93A2	0.20	1.43	-2.32	0.51	2.83	short	Y
B_comp4796_c0_seq18	7	30587517	Bradi1g42470	Os06g30130	serine/threonine-protein kinase receptor precursor	0.12	0.89	-3.00	-0.17	2.83	short	N
B_comp21600_c0_seq3	1	2187276	Bradi2g37930	Os04g58830	ribosome biogenesis regulatory protein	1.01	0.14	0.02	-2.81	2.82	tall	N
B_comp28678_c0_seq4	1	16824218	Bradi3g33690	Os10g41490	serine/threonine-protein kinase SAPK3	0.14	0.99	-2.82	-0.01	2.81	short	N
D_comp7849_c0_seq1	4	3510376	Bradi1g09690	Os08g13920	xyloglucan endotransglycosylase/hydrolase protein 8 precursor	5.00	34.89	2.32	5.12	2.80	short	N
A_comp1221538_c0_seq1	2	45659660	Bradi5g23910	Os04g55250	-	0.71	4.90	-0.50	2.29	2.79	short	N
D_comp284076_c0_seq1	4	26124166	Bradi3g43670	Os02g29140	ankyrin-like protein	19.86	2.88	4.31	1.53	2.79	tall	N
B_comp27246_c1_seq2	3	11627536	Bradi2g13190	Os01g28030	peroxidase 24 precursor	0.89	6.12	-0.17	2.61	2.78	short	Y
B_comp20431_c0_seq1	5	29390782	Bradi1g06430	Os03g57220	hydroxyacid oxidase 1	0.12	0.82	-3.05	-0.29	2.76	short	Y
D_comp45307_c0_seq1	3	20970735	Bradi2g48050	Os01g51550	peroxidase family protein	0.23	1.56	-2.12	0.64	2.75	short	N
A_comp1030985_c0_seq1	2	47020593	Bradi5g16570	Os04g58570	elicitor-responsive protein 3	0.45	3.01	-1.16	1.59	2.75	short	Y
B_comp13012_c0_seq1	1	35728888	Bradi2g24330	Os05g35650	peptide transporter PTR2-B	0.10	0.68	-3.29	-0.56	2.72	short	N
D_comp52295_c0_seq1	7	24856893	Bradi1g37870	Os06g36560	inositol oxygenase	0.35	2.32	-1.51	1.21	2.72	short	N
D_comp863707_c0_seq1	7	10860683	Bradi4g04550	Os11g45400	glycerol-3-phosphate acyltransferase 1	0.18	1.15	-2.51	0.20	2.71	short	N
B_comp25969_c0_seq1	5	420912	Bradi4g00830	Os12g43750	-	0.16	1.02	-2.68	0.03	2.71	short	N
D_comp314734_c0_seq1	4	18489897	Bradi3g04750	Os10g16974	flavonoid 3-monooxygenase	0.25	1.63	-2.00	0.70	2.70	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
A_comp3243_c1_seq3	3	4438846	Bradi2g04510	Os05g07890	lipase/lipooxygenase PLAT/LH2	0.28	1.81	-1.85	0.86	2.70	short	Y
A_comp328999_c0_seq1	2	9661754	Bradi1g20460	Os07g43800	EF hand family protein	0.15	0.95	-2.77	-0.07	2.70	short	N
A_comp410675_c0_seq1	3	2622505	Bradi2g03650	Os05g01280	-	0.30	1.94	-1.74	0.96	2.70	short	N
A_comp299418_c0_seq1	2	8701758	Bradi1g19230	Os07g46350	serine carboxypeptidase 1 precursor	0.15	0.95	-2.77	-0.07	2.70	short	Y
B_comp16917_c0_seq1	6	13851258	Bradi3g47750	Os02g38290	cytochrome P450 94A2	0.22	1.44	-2.16	0.53	2.69	short	N
D_comp508294_c0_seq1	5	11593898	Bradi4g44510	Os12g02080	peroxidase 52 precursor	1.13	7.18	0.18	2.84	2.66	short	N
B_comp21388_c0_seq4	5	33174699	Bradi1g00710	Os03g64230	-	0.21	1.31	-2.26	0.39	2.65	short	N
A_comp20593_c3_seq4	2	19212934	Bradi1g55720	Os03g58110	systemin receptor SR160 precursor	0.12	0.75	-3.04	-0.41	2.63	short	N
D_comp282431_c0_seq1	2	8858877	Bradi1g19470	Os07g19000	-	0.10	0.62	-3.32	-0.69	2.63	short	Y
B_comp60866_c0_seq1	5	30123068	Bradi1g05310	Os03g58490	-	0.10	0.64	-3.27	-0.64	2.63	short	N
B_comp9200_c0_seq1	6	13335108	Bradi3g49250	Os05g35290	phenylalanine ammonia-lyase	1.58	9.63	0.66	3.27	2.61	short	N
D_comp379766_c0_seq1	3	19983764	Bradi2g47000	Os01g49510	esterase/lipase/thioesterase	0.10	0.62	-3.28	-0.68	2.60	short	N
B_comp25551_c0_seq1	6	15413474	Bradi3g49590	Os02g42310	lysosomal protective protein precursor	0.19	1.13	-2.42	0.18	2.60	short	N
D_comp196635_c0_seq1	4	37528205	Bradi1g72760	Os03g08100	catalytic/ hydrolase	0.54	3.22	-0.90	1.69	2.59	short	Y
B_comp21388_c0_seq1	5	14727403	Bradi4g28000	Os09g15330	sugar transport protein 14	0.27	1.60	-1.91	0.67	2.59	short	N
A_comp215512_c0_seq1	2	1760470	Bradi5g01740	Os10g02210	peptide transporter PTR2 putative	0.20	1.23	-2.29	0.29	2.58	short	N
D_comp23134_c0_seq1	7	8762076	Bradi3g39420	Os08g38920	caffeoyl-CoA O-methyltransferase 2	2.86	17.09	1.52	4.09	2.58	short	Y
D_comp31791_c0_seq1	1	31083116	Bradi2g27870	Os07g47700	UDP-glucuronic acid decarboxylase 1	0.16	0.93	-2.68	-0.10	2.58	short	N
D_comp392_c0_seq1	7	23240299	Bradi1g36590	Os06g40150	ethylene responsive element	0.69	4.13	-0.53	2.05	2.57	short	N
B_comp67227_c0_seq1	1	42644644	Bradi2g17070	Os05g48270	dopamine beta-monooxygenase	1.02	6.04	0.03	2.60	2.57	short	Y
D_comp260323_c0_seq1	7	6640904	Bradi3g42070	Os08g43290	LIM1 protein precursor	12.40	73.41	3.63	6.20	2.57	short	N
D_comp565894_c0_seq1	4	14331670	Bradi4g18910	Os11g30360	-	2.60	15.38	1.38	3.94	2.56	short	N
B_comp53515_c0_seq1	5	8303523	Bradi4g40100	Os12g12170	cytochrome b5 isoform 2	2.56	15.11	1.36	3.92	2.56	short	N
B_comp43428_c0_seq1	7	2734714	Bradi1g50970	Os06g05550	anther-specific proline-rich protein APG precursor	0.37	2.16	-1.42	1.11	2.54	short	N
D_comp97892_c0_seq1	3	11469772	Bradi2g13060	Os01g27360	glutathione S-transferase GSTF1	0.15	0.88	-2.73	-0.19	2.53	short	Y

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp305536_c0_seq1	4	38650168	Bradi1g74320	Os11g48060	monocopper oxidase-like protein SKS1 precursor	0.10	0.60	-3.27	-0.74	2.52	short	N
D_comp280274_c0_seq1	7	13797719	Bradi3g21020	Os02g14630	hydroquinone glucosyltransferase	0.14	0.78	-2.88	-0.36	2.52	short	N
B_comp53793_c0_seq1	6	20102111	Bradi3g58800	Os02g51770	TLD family protein	0.41	2.36	-1.27	1.24	2.51	short	N
B_comp59909_c0_seq1	5	13409075	Bradi4g27030	Os09g10740	protein translocase/ protein transporter putative	0.18	1.03	-2.46	0.04	2.50	short	N
B_comp56262_c0_seq1	7	15237436	Bradi4g08650	Os08g17500	dihydroflavonol-4-reductase	0.11	0.63	-3.16	-0.67	2.49	short	N
B_comp22775_c0_seq1	7	21578726	Bradi3g09520	Os08g02300	NAM like protein	0.25	1.38	-2.02	0.47	2.49	short	N
A_comp36143_c0_seq1	2	4585429	Bradi3g20960	Os04g14690	monooxygenase/ oxidoreductase	0.11	0.61	-3.21	-0.72	2.49	short	N
D_comp409261_c0_seq1	6	1180492	Bradi3g01960	Os02g02870	glycine-rich protein 2	0.34	1.91	-1.55	0.93	2.48	short	N
D_comp84864_c0_seq1	5	8719990	Bradi4g40680	Os12g08920	peroxidase 43 precursor	0.19	1.07	-2.38	0.10	2.48	short	N
D_comp36241_c0_seq4	4	31581938	Bradi1g65530	Os03g18140	-	0.22	1.24	-2.16	0.31	2.47	short	N
D_comp412827_c0_seq1	5	14786789	Bradi4g28070	Os10g09860	chalcone synthase 8 putative	0.27	1.50	-1.88	0.59	2.47	short	N
D_comp85890_c0_seq1	1	12540230	Bradi3g29710	Os10g34920	secretory protein	1.01	5.60	0.02	2.49	2.46	short	N
D_comp33251_c0_seq1	4	30987387	Bradi1g64710	Os03g19070	long cell-linked locus protein	0.84	0.15	-0.24	-2.71	2.46	tall	N
B_comp87048_c0_seq1	4	1562563	Bradi1g11990	Os03g48626	-	0.19	1.07	-2.36	0.10	2.46	short	N
A_comp261568_c0_seq1	3	16434092	Bradi2g43330	Os05g51050	acyl-protein thioesterase 2	0.35	1.92	-1.52	0.94	2.46	short	Y
D_comp5243_c0_seq1	3	20370676	Bradi2g47410	Os01g50160	multidrug resistance protein 8	0.12	0.66	-3.05	-0.60	2.46	short	N
B_comp60794_c0_seq1	1	16286879	Bradi3g33150	Os10g40730	beta-expansin 1a precursor	2.18	11.96	1.13	3.58	2.45	short	Y
A_comp76801_c0_seq1	1	35675912	Bradi2g24400	Os05g35500	mybHv5	0.13	0.69	-2.98	-0.54	2.44	short	N
A_comp208468_c0_seq1	2	23796961	Bradi1g22030	Os03g30250	phytochelatin synthetase-like conserved region family protein	0.32	1.72	-1.65	0.78	2.43	short	N
B_comp29655_c0_seq1	5	11596342	Bradi4g44530	Os11g02130	peroxidase 52 precursor	1.21	6.51	0.28	2.70	2.43	short	Y
B_comp81942_c0_seq1	4	555036	Bradi1g13090	Os03g45920	tubulin beta-8 chain	0.33	1.75	-1.61	0.81	2.42	short	N
B_comp85068_c0_seq1	5	4961587	Bradi4g05980	Os12g34524	peroxidase 24 precursor	0.37	1.99	-1.42	1.00	2.42	short	N
B_comp65814_c0_seq1	4	1367513	Bradi1g12190	Os03g48180	peptide transporter PTR2	4.06	21.55	2.02	4.43	2.41	short	Y
A_comp552861_c0_seq1	2	14316243	Bradi1g25930	Os01g10990	nodulin-like protein	0.38	2.02	-1.39	1.01	2.41	short	N
A_comp467430_c0_seq1	4	41075723	Bradi1g77840	Os03g02040	remorin	3.61	19.04	1.85	4.25	2.40	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp73457_c0_seq1	6	21940265	Bradi1g47580	Os02g55480	protein binding protein	0.31	1.65	-1.67	0.72	2.39	short	Y
D_comp726375_c0_seq1	3	26335001	Bradi2g54960	Os01g62970	-	0.12	0.61	-3.09	-0.70	2.39	short	N
D_comp930_c0_seq1	5	31391586	Bradi1g03500	Os03g60850	peptide transporter PTR2	0.13	0.67	-2.96	-0.58	2.38	short	N
B_comp57763_c0_seq1	7	37892013	Bradi3g21150	Os11g39420	jacalin-like lectin domain containing protein	0.16	0.84	-2.62	-0.25	2.37	short	N
B_comp423_c1_seq2	5	19271612	Bradi4g32220	Os09g28550	-	0.12	0.63	-3.03	-0.66	2.37	short	N
D_comp297189_c0_seq1	4	31450374	Bradi1g65340	Os03g18370	ATP binding protein	0.10	0.52	-3.30	-0.93	2.36	short	N
D_comp28929_c1_seq1	4	35458159	Bradi1g69910	Os03g11590	-	0.17	0.87	-2.55	-0.20	2.35	short	N
B_comp95118_c0_seq1	4	6201098	Bradi4g24460	Os11g07060	receptor protein kinase CLAVATA1 precursor	0.53	2.70	-0.91	1.44	2.34	short	N
A_comp742863_c0_seq1	1	17705819	Bradi3g34600	Os10g42900	peptide transporter PTR2	0.44	2.22	-1.19	1.15	2.34	short	N
B_comp9885_c0_seq1	1	41517321	Bradi3g14100	Os04g03579	ATP binding protein	0.32	1.59	-1.66	0.67	2.33	short	N
D_comp32215_c0_seq1	2	3121526	Bradi5g02360	Os06g47800	disease resistance protein RGA3	1.93	9.68	0.95	3.28	2.33	short	Y
A_comp28323_c1_seq1	2	36846706	Bradi1g22860	Os07g46980	sex determination protein tasselseed-2	0.26	1.29	-1.95	0.37	2.32	short	N
B_comp20797_c0_seq7	3	14981616	Bradi2g41940	Os01g40070	-	2.57	0.51	1.36	-0.96	2.32	tall	N
A_comp79766_c0_seq1	2	28186656	Bradi5g06620	Os04g28280	BHLH transcription factor	0.32	1.59	-1.64	0.67	2.31	short	N
B_comp43500_c0_seq1	4	29768767	Bradi1g63550	Os03g21230	protein kinase	0.34	1.67	-1.56	0.74	2.31	short	N
B_comp28882_c0_seq2	4	37310739	Bradi1g72510	Os03g08390	-	0.13	0.63	-2.97	-0.67	2.31	short	N
B_comp131_c0_seq2	7	37422489	Bradi1g47710	Os02g55560	DNA-binding protein phosphatase 2C	1.67	0.34	0.74	-1.56	2.31	tall	Y
D_comp4797_c0_seq1	6	19942422	Bradi3g59060	Os02g51570	FKBP-type peptidyl-prolyl cis-trans isomerase 4 chloroplast precursor	0.14	0.68	-2.87	-0.56	2.31	short	N
D_comp5318_c0_seq4	3	1035840	Bradi2g01480	Os01g02930	glycosyltransferase	0.11	0.56	-3.14	-0.84	2.30	short	N
D_comp915_c1_seq1	6	1954791	Bradi3g03060	Os02g04120	bifunctional coenzyme A synthase	0.85	4.20	-0.23	2.07	2.30	short	N
D_comp878527_c0_seq1	3	28844024	Bradi2g58310	Os01g68269	-	0.14	0.67	-2.86	-0.57	2.29	short	N
B_comp51319_c0_seq1	3	26161041	Bradi2g54740	Os01g62600	laccase putative	0.15	0.74	-2.73	-0.44	2.29	short	N
D_comp26897_c0_seq1	5	20519739	Bradi4g33490	Os09g30486	fasciclin-like arabinogalactan protein 7 precursor	0.50	2.44	-1.00	1.29	2.29	short	N
B_comp9810_c0_seq2	1	36940498	Bradi2g54690	Os05g38410	L-ascorbate oxidase precursor	0.44	2.16	-1.18	1.11	2.29	short	N
A_comp22036_c0_seq2	3	29887192	Bradi2g59660	Os01g70520	beta-glucosidase homolog precursor	0.30	1.48	-1.72	0.57	2.29	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp39701_c1_seq3	3	9223760	Bradi2g11010	Os05g13940	early nodulin 75 protein	0.79	3.83	-0.35	1.94	2.28	short	N
D_comp757_c0_seq1	3	1361089	Bradi2g02430	Os04g02280	F-box domain containing protein	0.18	0.86	-2.49	-0.21	2.28	short	N
D_comp444388_c0_seq1	3	12001504	Bradi2g13520	Os01g33420	alpha-galactosidase/ hydrolase hydrolyzing O-glycosyl compounds	0.18	0.87	-2.47	-0.19	2.27	short	N
B_comp8161_c0_seq1	1	470389	Bradi2g39930	Os05g01290	-	0.34	1.62	-1.57	0.70	2.27	short	Y
D_comp759259_c0_seq1	5	8206352	Bradi4g39980	Os04g41960	NADP-dependent oxidoreductase P1	0.65	3.14	-0.62	1.65	2.27	short	N
A_comp116782_c0_seq1	2	42055062	Bradi5g21590	Os04g52320	QRT3	2.76	13.28	1.46	3.73	2.27	short	N
B_comp25960_c0_seq5	6	15492015	Bradi3g49730	Os04g44530	-	0.39	1.87	-1.36	0.91	2.27	short	N
B_comp98223_c0_seq1	5	21311914	Bradi4g34300	Os09g32470	membrane protein	0.46	2.22	-1.11	1.15	2.26	short	N
B_comp16151_c0_seq1	6	16336175	Bradi3g50720	Os02g44102	remorin C-terminal region family protein	0.17	0.81	-2.56	-0.30	2.26	short	N
D_comp7140_c0_seq1	6	7587795	Bradi3g09470	Os02g15230	esterase precursor	1.39	6.65	0.47	2.73	2.26	short	N
D_comp12230_c0_seq1	3	7592711	Bradi2g09300	Os01g15320	rapid alkalinization factor 1 precursor	1.71	8.17	0.78	3.03	2.25	short	N
B_comp2117_c0_seq1	5	8212095	Bradi4g39990	Os12g12590	NADP-dependent oxidoreductase P1	0.20	0.94	-2.35	-0.09	2.25	short	N
B_comp97658_c0_seq1	5	6401246	Bradi4g07520	Os12g29400	ABA-responsive protein	0.12	0.58	-3.04	-0.79	2.25	short	Y
D_comp95949_c0_seq1	1	17338906	Bradi3g34140	Os10g42390	RING-H2 finger protein ATL1G	0.23	1.08	-2.13	0.12	2.25	short	Y
D_comp121909_c0_seq1	6	11316365	Bradi3g44870	Os02g32970	catalytic/ hydrolase	0.21	1.01	-2.22	0.02	2.24	short	N
D_comp101996_c0_seq1	7	42838092	Bradi1g32660	Os06g46270	NAC domain-containing protein 21/22	0.14	0.67	-2.82	-0.58	2.24	short	N
D_comp62589_c0_seq2	3	21553825	Bradi2g49490	Os01g53980	calmodulin binding protein	0.11	0.50	-3.24	-1.00	2.24	short	N
A_comp33929_c0_seq1	1	42251266	Bradi2g17540	Os05g47700	nonspecific lipid-transfer protein precursor	0.97	4.57	-0.04	2.19	2.23	short	N
D_comp23667_c0_seq2	3	18976213	Bradi2g45980	Os05g49040	-	0.42	1.98	-1.24	0.98	2.23	short	Y
A_comp941694_c0_seq1	1	36625224	Bradi2g23650	Os05g37880	axi 1 like protein	0.14	0.67	-2.80	-0.57	2.22	short	N
A_comp270863_c0_seq1	2	6987307	Bradi1g17050	Os12g24580	negatively light-regulated protein	2.11	9.83	1.08	3.30	2.22	short	N
B_comp13005_c0_seq3	4	38703457	Bradi1g74410	Os03g06570	calmodulin binding protein	0.46	2.11	-1.14	1.07	2.21	short	N
B_comp45144_c0_seq1	5	22401587	Bradi4g35710	Os09g35690	protein binding protein	0.12	0.54	-3.10	-0.89	2.21	short	N
D_comp414476_c0_seq1	6	920158	Bradi4g09460	Os11g39450	stem rust resistance protein	1.10	0.24	0.13	-2.07	2.20	tall	N
A_comp310675_c0_seq1	3	22635731	Bradi2g50090	Os01g54950	-	0.27	1.21	-1.92	0.27	2.19	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp41425_c0_seq1	4	28778654	Bradi1g62490	Os03g22680	RING finger and CHY zinc finger domain-containing protein 1	2.27	10.35	1.19	3.37	2.19	short	N
B_comp740_c0_seq1	6	19639247	Bradi3g59500	Os02g50990	RING-H2 finger protein ATL1Q	0.10	0.47	-3.26	-1.08	2.18	short	N
D_comp45143_c0_seq1	4	17526436	Bradi4g17230	Os11g32650	chalcone synthase	0.57	2.57	-0.80	1.36	2.17	short	N
D_comp507709_c0_seq1	5	1456476	Bradi4g02250	Os04g12560	receptor-like protein kinase	0.22	0.97	-2.22	-0.05	2.17	short	Y
D_comp89039_c0_seq1	5	16799813	Bradi1g44070	Os06g14670	odorant 1 protein	0.12	0.53	-3.09	-0.93	2.17	short	N
A_comp549410_c0_seq1	2	5280233	Bradi5g04080	Os12g19394	ribulose bisphosphate carboxylase small chain C chloroplast precursor	0.64	2.89	-0.63	1.53	2.16	short	Y
A_comp2011678_c0_seq1	1	22147668	Bradi2g35450	Os05g06140	lipase	0.69	3.09	-0.53	1.63	2.16	short	N
D_comp427260_c0_seq1	1	41864091	Bradi2g47590	Os05g46610	myb-related protein Hv33	0.16	0.71	-2.66	-0.50	2.16	short	N
D_comp120050_c0_seq1	4	33927334	Bradi1g68330	Os03g14030	thaumatin-like protein 1 precursor	0.45	2.03	-1.14	1.02	2.16	short	Y
B_comp83466_c0_seq1	5	11842840	Bradi4g44860	Os11g01570	-	2.04	9.11	1.03	3.19	2.16	short	N
B_comp6072_c0_seq6	1	16258915	Bradi3g33110	Os03g01260	beta-expansin 1a precursor	8.99	39.98	3.17	5.32	2.15	short	Y
A_comp1372120_c0_seq1	2	46789083	Bradi5g26170	Os04g58100	-	0.19	0.87	-2.36	-0.21	2.15	short	N
B_comp71039_c0_seq1	7	24267412	Bradi1g37450	Os06g37560	beta-galactosidase precursor	3.01	13.32	1.59	3.74	2.15	short	Y
A_comp508006_c0_seq1	5	24338582	Bradi4g37990	Os09g38777	mo-molybdopterin cofactor sulfurase	0.20	0.89	-2.31	-0.17	2.14	short	N
A_comp1183223_c0_seq1	2	35808572	Bradi5g12740	Os04g39150	major latex protein 22	23.79	104.93	4.57	6.71	2.14	short	N
A_comp250842_c0_seq1	2	7111111	Bradi1g17260	Os07g48750	alpha-N-arabinofuranosidase 1 precursor	0.16	0.72	-2.61	-0.48	2.14	short	N
B_comp51635_c0_seq1	4	31749237	Bradi1g65750	Os03g17850	beta3-glucuronyltransferase	1.34	5.89	0.42	2.56	2.14	short	N
A_comp44872_c0_seq1	4	37426699	Bradi1g72660	Os03g08250	-	0.25	1.09	-2.01	0.12	2.14	short	N
D_comp304007_c0_seq1	5	20617423	Bradi4g33550	Os08g39370	tonoplast dicarboxylate transporter	0.55	2.41	-0.86	1.27	2.13	short	Y
D_comp51840_c0_seq5	2	37272794	Bradi5g14300	Os04g41540	calmodulin	0.31	1.34	-1.71	0.42	2.13	short	N
B_comp30129_c0_seq2	6	16369655	Bradi3g50740	Os02g44108	beta-expansin 4 precursor	0.70	3.06	-0.52	1.61	2.13	short	N
D_comp90520_c0_seq1	1	14775098	Bradi2g47110	Os10g38720	glutathione S-transferase GSTU6	7.22	31.31	2.85	4.97	2.12	short	N
A_comp29980_c0_seq26	2	10996983	Bradi1g22040	Os07g41310	COBRA-like protein 4 precursor	0.66	2.83	-0.60	1.50	2.10	short	N
B_comp4172_c0_seq2	5	21848997	Bradi4g34930	Os09g33680	cyanogenic beta-glucosidase precursor	0.26	1.10	-1.96	0.14	2.10	short	N
A_comp706481_c0_seq1	2	45007727	Bradi5g24830	Os04g56500	-	0.20	0.84	-2.35	-0.26	2.09	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp386325_c0_seq1	3	25183692	Bradi2g53470	Os05g25640	trans-cinnamate 4-monooxygenase	0.65	2.78	-0.62	1.47	2.09	short	N
D_comp170811_c0_seq1	2	7281367	Bradi1g17450	Os07g48500	stress responsive protein	0.40	1.70	-1.33	0.76	2.09	short	N
D_comp162468_c0_seq1	4	33643334	Bradi1g68010	Os03g14590	calcium ion binding protein	0.19	0.81	-2.39	-0.30	2.09	short	Y
A_comp20_c0_seq30	2	47541900	Bradi4g09590	Os07g19320	stripe rust resistance protein Yr10 putative	0.18	0.75	-2.50	-0.42	2.09	short	N
A_comp1085031_c0_seq1	2	42929166	Bradi5g20400	Os04g51070	helix-loop-helix DNA-binding domain containing protein	1.02	4.31	0.03	2.11	2.08	short	N
B_comp17825_c0_seq1	6	3691330	Bradi3g05290	Os02g07480	transglycosylase SLT domain containing protein	0.43	1.81	-1.22	0.86	2.08	short	N
D_comp674392_c0_seq1	4	8608056	Bradi4g22730	Os11g10130	myb-like DNA-binding domain containing protein	0.13	0.56	-2.91	-0.84	2.08	short	N
A_comp3243_c0_seq11	2	37219855	Bradi5g14210	Os04g41340	4-nitrophenylphosphatase	1.34	5.66	0.42	2.50	2.08	short	N
A_comp745771_c0_seq1	3	28857239	Bradi2g58350	Os01g68300	-	1.52	6.42	0.61	2.68	2.08	short	N
D_comp491241_c0_seq1	7	4481788	Bradi3g44260	Os02g30910	MTN3	2.30	9.68	1.20	3.27	2.07	short	Y
A_comp315660_c0_seq2	3	18287969	Bradi2g45320	Os01g46350	-	1.17	4.92	0.23	2.30	2.07	short	N
B_comp64596_c0_seq1	5	19375815	Bradi4g32330	Os09g28730	gibberellin receptor GID1L2	0.62	2.63	-0.68	1.39	2.07	short	N
D_comp91511_c0_seq1	5	9846389	Bradi4g42070	Os12g06180	pathogenicity protein PATH531-like protein	0.13	0.56	-2.90	-0.83	2.07	short	N
B_comp63772_c0_seq1	7	6522372	Bradi3g42240	Os08g43700	OsSAUR36 - Auxin-responsive SAUR gene family member	1.25	5.23	0.32	2.39	2.07	short	N
D_comp229880_c0_seq1	7	2321601	Bradi2g62420	Os01g66830	carboxylic ester hydrolase	0.69	2.88	-0.54	1.52	2.07	short	N
D_comp454438_c0_seq1	4	27544816	Bradi4g08950	Os09g08190	flavonol 4-sulfotransferase	0.49	0.12	-1.03	-3.09	2.06	tall	Y
A_comp490021_c0_seq1	2	5823605	Bradi5g04640	Os05g23924	hydrolase hydrolyzing O-glycosyl compounds	0.41	1.70	-1.29	0.76	2.06	short	N
D_comp205768_c0_seq1	6	20052569	Bradi3g58910	Os02g51720	-	0.13	0.53	-2.97	-0.92	2.05	short	N
B_comp62722_c0_seq1	7	7430049	Bradi3g40990	Os09g32730	zinc finger-like protein	0.11	0.44	-3.24	-1.19	2.05	short	N
D_comp1082113_c0_seq1	5	11266851	Bradi3g06420	Os11g25330	nucleoside-triphosphatase	0.43	1.76	-1.22	0.82	2.04	short	N
D_comp69179_c0_seq1	3	20490974	Bradi2g47480	Os01g50370	mitogen-activated protein kinase kinase kinase 1	0.11	0.46	-3.15	-1.11	2.04	short	N
D_comp6584_c0_seq1	7	6646606	Bradi4g35680	Os08g43270	protein binding protein	1.07	4.37	0.09	2.13	2.03	short	N
D_comp17578_c0_seq1	7	1712408	Bradi3g16540	Os06g01960	retrotransposon protein unclassified	0.73	3.01	-0.44	1.59	2.03	short	N
B_comp4386_c1_seq1	6	7401873	Bradi3g09290	Os06g14310	potassium channel AKT1 putative	0.16	0.66	-2.63	-0.60	2.03	short	N
D_comp94454_c0_seq1	3	10176207	Bradi2g11860	Os01g21034	pectinesterase-2 precursor	1.87	7.65	0.90	2.94	2.03	short	Y

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp63329_c0_seq3	4	24142860	Bradi3g02290	Os11g42200	laccase LAC2-1	0.55	2.25	-0.86	1.17	2.03	short	N
D_comp359860_c0_seq1	2	15507770	Bradi1g27120	Os07g30970	nucleoside diphosphate kinase 1	1.81	7.39	0.86	2.88	2.03	short	N
B_comp56832_c0_seq1	7	22032091	Bradi3g13260	Os08g01600	polygalacturonase	0.31	1.27	-1.68	0.34	2.02	short	Y
B_comp32034_c0_seq1	5	28549102	Bradi1g07330	Os03g55980	-	0.10	0.41	-3.29	-1.27	2.02	short	N
D_comp12048_c0_seq1	5	2761440	Bradi4g03460	Os12g39220	zinc finger protein 7	0.34	1.39	-1.54	0.47	2.02	short	N
A_comp758825_c0_seq1	1	17483096	Bradi3g34280	Os10g42620	dihydroflavonol-4-reductase	3.14	12.69	1.65	3.67	2.01	short	N
D_comp106405_c0_seq4	5	14658840	Bradi4g27950	Os03g36080	-	0.14	0.57	-2.84	-0.82	2.01	short	N
A_comp8310_c0_seq1	1	13155943	Bradi3g30300	Os10g36170	nonspecific lipid-transfer protein precursor	2.36	9.53	1.24	3.25	2.01	short	N
D_comp116464_c0_seq1	7	19111653	Bradi3g16530	Os08g06100	quercetin 3-O-methyltransferase 1	10.89	43.95	3.44	5.46	2.01	short	N
D_comp330792_c0_seq1	6	14694377	Bradi3g48800	Os02g40260	protein binding protein	0.64	2.56	-0.65	1.36	2.01	short	Y
B_comp59071_c0_seq1	1	14952636	Bradi3g31970	Os10g38920	ATP binding protein	0.50	2.02	-1.00	1.01	2.01	short	N
D_comp3893_c0_seq1	3	29180623	Bradi2g58790	Os01g68890	wiscott-Aldrich syndrome C-terminal	0.53	2.13	-0.91	1.09	2.00	short	N
A_comp10324_c0_seq1	2	32832451	Bradi5g10030	Os04g34600	ABA/WDS induced protein	0.21	0.84	-2.25	-0.25	2.00	short	Y
D_comp382592_c0_seq1	2	24629990	Bradi1g60320	Os03g28330	sucrose synthase 2	24.89	99.73	4.64	6.64	2.00	short	N
A_comp2187_c0_seq1	2	41806613	Bradi5g19240	Os04g49194	naringenin2-oxoglutarate 3-dioxygenase	0.13	0.53	-2.92	-0.92	2.00	short	N
B_comp75905_c0_seq1	1	14874206	Bradi3g31870	Os10g38780	glutathione S-transferase GSTU6	0.12	0.48	-3.05	-1.05	2.00	short	N
D_comp141170_c0_seq1	7	2706319	Bradi1g50950	Os09g04210	hydrolase/ zinc ion binding protein	5.27	1.32	2.40	0.40	2.00	tall	Y
B_comp63675_c0_seq1	7	3731090	Bradi1g38730	Os08g40030	nam-like protein 16 putative	0.24	0.95	-2.06	-0.07	1.99	short	N
D_comp19205_c0_seq1	4	26598655	Bradi2g35770	Os04g53496	NBS-LRR disease resistance protein	0.46	1.85	-1.11	0.89	1.99	short	N
A_comp639293_c0_seq1	2	44590298	Bradi5g22830	Os04g53800	leucoanthocyanidin reductase	0.10	0.41	-3.27	-1.28	1.99	short	N
D_comp219169_c0_seq1	6	7186130	Bradi3g09080	Os02g14160	peroxidase 52 precursor	2.97	11.78	1.57	3.56	1.99	short	N
B_comp12069_c0_seq1	7	41476467	Bradi1g33860	Os06g48250	cell Division Protein AAA ATPase family	0.28	1.11	-1.83	0.15	1.98	short	N
D_comp10779_c0_seq1	1	9404181	Bradi3g27610	Os10g30560	cytokinin-O-glucosyltransferase 2	0.12	0.46	-3.10	-1.12	1.98	short	Y
D_comp217601_c0_seq1	6	21333076	Bradi3g60160	Os02g54254	alpha-aminoadipic semialdehyde synthase	0.34	1.33	-1.56	0.42	1.98	short	N
D_comp98557_c0_seq1	6	17281782	Bradi3g51780	Os02g45710	RING zinc finger protein-like	1.47	0.37	0.56	-1.42	1.98	tall	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp16872_c0_seq1	1	40397409	Bradi2g19860	Os12g25450	O-methyltransferase ZRP4	0.26	1.03	-1.93	0.05	1.98	short	N
B_comp67279_c0_seq1	4	5558469	Bradi4g25020	Os11g06020	BEL1-like homeodomain transcription factor	0.48	1.89	-1.05	0.92	1.97	short	Y
D_comp296330_c0_seq1	4	12574724	Bradi4g20180	Os11g19480	ATP binding protein	0.42	1.66	-1.24	0.73	1.97	short	N
A_comp101603_c0_seq1	2	15199243	Bradi1g26850	Os07g31720	ZAC	2.90	11.34	1.54	3.50	1.97	short	N
B_comp20451_c0_seq1	5	4511676	Bradi4g05450	Os01g01660	isoflavone reductase homolog IRL	14.42	56.28	3.85	5.81	1.96	short	N
B_comp40890_c0_seq1	4	37850540	Bradi1g73170	Os10g26470	sucrose transporter 1	1.03	4.00	0.04	2.00	1.96	short	Y
B_comp47325_c0_seq1	4	4850636	Bradi4g25560	Os11g05530	-	0.15	0.57	-2.78	-0.82	1.96	short	N
B_comp3707_c0_seq4	5	32852870	Bradi1g01310	Os03g63540	in cucumber hypocotyls	0.54	2.10	-0.89	1.07	1.96	short	N
B_comp80940_c0_seq1	7	19544967	Bradi3g16100	Os08g06640	-	0.16	0.62	-2.64	-0.68	1.96	short	N
A_comp224323_c0_seq1	3	25332892	Bradi2g53600	Os01g60800	-	0.22	0.85	-2.20	-0.24	1.96	short	N
D_comp449046_c0_seq1	7	24995038	Bradi1g37960	Os06g36270	receptor-like protein kinase 5 precursor	0.48	0.12	-1.07	-3.03	1.96	tall	N
B_comp9526_c0_seq4	6	17445617	Bradi3g51980	Os02g46100	RING-H2 finger protein ATL1R	0.26	1.02	-1.92	0.03	1.95	short	N
D_comp348637_c0_seq1	7	3747587	Bradi1g50070	Os04g01690	arginine decarboxylase	1.36	5.25	0.44	2.39	1.95	short	N
A_comp7974_c0_seq1	2	9518895	Bradi1g20280	Os07g44060	catalytic/ hydrolase	0.74	2.87	-0.43	1.52	1.95	short	N
D_comp914036_c0_seq1	2	40137340	Bradi5g17300	Os06g35650	reticuline oxidase precursor	0.12	0.46	-3.07	-1.12	1.94	short	N
B_comp5020_c0_seq1	4	36633459	Bradi1g71580	Os03g09850	dopamine beta-monooxygenase	0.44	1.70	-1.18	0.76	1.94	short	N
D_comp59091_c0_seq1	2	14579170	Bradi1g26220	Os07g33320	-	0.47	1.82	-1.08	0.86	1.94	short	N
B_comp14988_c0_seq3	3	5527809	Bradi2g06620	Os08g27840	phosphoenolpyruvate carboxylase 2	0.38	1.46	-1.39	0.54	1.93	short	N
B_comp16121_c0_seq2	5	27994738	Bradi1g08120	Os03g55070	UDP-glucose 6-dehydrogenase	16.42	62.62	4.04	5.97	1.93	short	N
A_comp46003_c0_seq1	2	39386302	Bradi5g16480	Os04g44750	arabinose-proton symporter	0.11	0.42	-3.17	-1.24	1.93	short	N
D_comp313443_c0_seq1	7	39762318	Bradi1g35600	Os06g42560	tryptophan synthase beta chain 2	0.59	2.24	-0.76	1.16	1.92	short	N
B_comp22537_c0_seq1	2	33026673	Bradi5g10210	Os04g35140	subtilisin-like protease precursor	0.11	0.42	-3.16	-1.24	1.92	short	N
B_comp7797_c0_seq8	7	6121818	Bradi3g43000	Os08g44620	-	0.37	1.41	-1.42	0.49	1.91	short	N
A_comp318255_c0_seq1	3	6697123	Bradi2g08300	Os08g44270	vignain precursor	0.15	0.54	-2.78	-0.88	1.91	short	Y
B_comp30921_c0_seq1	5	22154774	Bradi4g35360	Os09g34230	indole-3-acetate beta-glucosyltransferase	0.16	0.60	-2.65	-0.74	1.91	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp417618_c0_seq1	3	27440492	Bradi2g56520	Os11g07890	transposon protein CACTA En/Spm sub-class	0.24	0.89	-2.08	-0.18	1.91	short	N
D_comp6252_c0_seq1	2	16480283	Bradi1g28060	Os07g29600	RING-H2 finger protein ATL2B	0.43	1.62	-1.21	0.69	1.90	short	Y
B_comp14070_c0_seq2	4	7730497	Bradi4g23280	Os11g09150	-	0.26	0.97	-1.94	-0.04	1.90	short	N
D_comp835674_c0_seq1	6	3390667	Bradi3g04930	Os12g42090	inner envelope membrane protein chloroplast precursor	3.44	12.85	1.78	3.68	1.90	short	Y
B_comp10632_c0_seq1	5	18604835	Bradi4g30110	Os06g01610	vacuolar processing enzyme precursor putative	0.14	0.51	-2.86	-0.96	1.90	short	N
A_comp104971_c0_seq1	2	33433385	Bradi5g10590	Os04g35580	-	0.59	2.21	-0.75	1.14	1.90	short	N
D_comp401857_c0_seq1	1	39304298	Bradi1g29940	Os10g37620	retrotransposon protein unclassified	0.19	0.72	-2.37	-0.47	1.89	short	N
B_comp13109_c0_seq1	6	4475680	Bradi3g06200	Os02g09080	-	0.23	0.85	-2.12	-0.23	1.89	short	N
B_comp52945_c0_seq2	4	30383559	Bradi1g64120	Os03g20120	galactinol synthase 3	0.65	2.41	-0.62	1.27	1.89	short	N
B_comp58710_c0_seq1	5	24128740	Bradi4g37670	Os09g38510	ATPUP5	0.11	0.42	-3.14	-1.26	1.88	short	N
D_comp33971_c0_seq1	7	36712782	Bradi1g47220	Os07g23150	anthocyanin 5-aromatic acyltransferase	0.22	0.80	-2.20	-0.32	1.88	short	N
B_comp15763_c0_seq30	1	3032176	Bradi1g14350	Os10g11860	transparent testa 12 protein	0.50	1.84	-1.00	0.88	1.88	short	N
D_comp79096_c0_seq1	5	11374740	Bradi4g44170	Os12g02760	-	0.18	0.67	-2.47	-0.59	1.88	short	N
B_comp19519_c0_seq1	5	32465323	Bradi1g01850	Os03g62850	acyl-CoA synthetase-like protein	0.13	0.47	-2.96	-1.08	1.88	short	N
D_comp23_c5_seq1	6	23893607	Bradi3g60470	Os02g58260	Zn-dependent hydrolases including glyoxylases	0.10	0.37	-3.30	-1.43	1.87	short	N
A_comp199297_c0_seq1	2	42904382	Bradi5g20420	Os04g51090	-	2.90	10.59	1.54	3.40	1.87	short	Y
D_comp221030_c0_seq1	2	34937153	Bradi5g11900	Os04g37760	-	0.18	0.64	-2.51	-0.64	1.87	short	N
B_comp67534_c0_seq1	1	36744264	Bradi2g23570	Os05g37950	guanylyl cyclase	0.11	0.41	-3.16	-1.29	1.87	short	N
A_comp14999_c0_seq47	3	23130189	Bradi2g50790	Os07g02200	blue copper protein precursor	0.34	1.25	-1.54	0.33	1.87	short	N
D_comp464196_c0_seq1	3	17739827	Bradi2g44820	Os01g45110	cytokinin-O-glucosyltransferase 1	0.23	0.84	-2.12	-0.25	1.87	short	Y
B_comp21732_c0_seq1	3	11204609	Bradi2g12830	Os01g26120	plant integral membrane protein TIGR01569 containing protein	0.88	3.22	-0.18	1.69	1.87	short	Y
A_comp24186_c0_seq1	3	4467335	Bradi2g04440	Os01g07720	dolichyl-P-Man Man-PP-dolichyl mannosyltransferase	3.59	0.99	1.84	-0.02	1.86	tall	Y
A_comp3243_c1_seq7	2	42603463	Bradi1g21920	Os07g41460	flavonol 4-sulfotransferase	2.02	0.56	1.02	-0.84	1.86	tall	N
D_comp111674_c0_seq1	7	44514004	Bradi1g30910	Os06g44040	DOMON domain containing protein	0.16	0.56	-2.69	-0.83	1.86	short	N
D_comp1078126_c0_seq1	1	39484138	Bradi2g21010	Os08g08980	germin-like protein subfamily 1 member 7 precursor	0.15	0.55	-2.73	-0.87	1.86	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp33537_c0_seq2	4	35130406	Bradi1g69540	Os03g12260	cytochrome P450 86A2	0.39	1.42	-1.35	0.50	1.86	short	N
A_comp76698_c0_seq3	5	5454509	Bradi4g06490	Os12g32640	haemolysin-III related family protein	0.28	0.99	-1.86	-0.01	1.85	short	N
B_comp4082_c0_seq3	5	11816391	Bradi4g44810	Os11g01730	L-ascorbate oxidase precursor putative	0.11	0.41	-3.14	-1.29	1.85	short	N
A_comp69771_c0_seq1	5	16372527	Bradi4g29530	Os09g21710	AN1-type zinc finger protein 2B	0.15	0.55	-2.70	-0.85	1.85	short	Y
D_comp598953_c0_seq1	7	35554941	Bradi1g46190	Os03g20290	aspartic proteinase nepenthesin-1 precursor	0.92	3.33	-0.12	1.73	1.85	short	N
A_comp416_c0_seq1	2	9494006	Bradi1g20250	Os07g44090	myb-related protein Hv33	0.63	2.28	-0.66	1.19	1.85	short	Y
A_comp72814_c0_seq1	1	19968679	Bradi2g14250	Os05g51670	UDP-glucose 4-epimerase GEPI48	1.46	5.23	0.55	2.39	1.84	short	N
D_comp390388_c0_seq1	1	25367503	Bradi2g32590	Os01g10890	CBL-interacting serine/threonine-protein kinase 15	0.27	0.95	-1.91	-0.08	1.84	short	Y
B_comp10636_c0_seq1	5	11223041	Bradi4g43960	Os11g03420	zinc finger homeodomain protein 1	0.61	2.19	-0.70	1.13	1.84	short	N
A_comp3243_c1_seq4	3	29894877	Bradi2g59690	Os01g70550	-	0.12	0.43	-3.05	-1.22	1.84	short	Y
D_comp15327_c0_seq1	1	997830	Bradi2g38790	Os08g42670	resistance protein	0.52	1.84	-0.95	0.88	1.84	short	N
A_comp926345_c0_seq1	3	18236795	Bradi2g45280	Os01g46210	esterase precursor putative	0.20	0.72	-2.31	-0.48	1.84	short	Y
B_comp16121_c0_seq5	4	32018733	Bradi1g66100	Os03g29920	-	0.81	2.87	-0.31	1.52	1.83	short	N
D_comp640957_c0_seq1	5	7675245	Bradi4g39350	Os10g01110	serine carboxypeptidase 1 precursor	0.11	0.41	-3.12	-1.29	1.83	short	Y
A_comp76801_c0_seq2	1	14568748	Bradi3g31560	Os10g37870	hypothetical protein	0.60	2.13	-0.74	1.09	1.83	short	N
B_comp33566_c0_seq6	6	15913626	Bradi3g50200	Os02g43280	aldehyde dehydrogenase 3B1	1.37	4.87	0.46	2.28	1.83	short	N
B_comp468_c0_seq5	7	7984783	Bradi3g40270	Os08g40420	ternary complex factor MIP1	0.93	3.30	-0.10	1.72	1.83	short	N
A_comp631960_c0_seq1	5	17671995	Bradi4g30990	Os09g26370	-	7.19	25.48	2.85	4.67	1.83	short	N
A_comp540052_c0_seq1	2	40662509	Bradi5g17990	Os04g46940	copper-transporting ATPase 3	0.37	1.32	-1.42	0.40	1.82	short	N
A_comp570175_c0_seq1	5	18585080	Bradi4g30130	Os09g24620	-	0.29	1.03	-1.78	0.04	1.82	short	N
D_comp165610_c0_seq1	2	34382565	Bradi5g11280	Os05g20050	ras-related protein RGP2	0.30	1.06	-1.73	0.09	1.82	short	N
D_comp1153_c0_seq1	7	9911667	Bradi3g38060	Os08g37040	gibberellin receptor GID1L2	0.12	0.42	-3.08	-1.26	1.82	short	Y
B_comp27953_c0_seq1	5	5982745	Bradi4g07100	Os12g31000	maternal protein pumilio	0.34	1.19	-1.57	0.25	1.82	short	Y
B_comp24014_c0_seq3	7	7546171	-	Os08g41280	membrane protein	0.75	0.21	-0.41	-2.23	1.82	tall	N
A_comp1364396_c0_seq1	2	39647582	Bradi5g16820	Os04g45170	ATP binding protein	0.48	1.68	-1.07	0.75	1.81	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
A_comp667017_c0_seq1	5	31153064	Bradi1g03840	Os03g60509	-	1.19	4.17	0.25	2.06	1.81	short	N
A_comp46479_c0_seq1	2	38499178	Bradi5g15590	Os04g43560	NAC domain-containing protein 21/22	0.24	0.84	-2.07	-0.26	1.81	short	N
D_comp8085_c0_seq2	4	27779754	Bradi1g61320	Os03g25790	glycosyl hydrolases family 17 protein	0.31	1.09	-1.68	0.13	1.81	short	N
B_comp30649_c0_seq1	5	19700224	Bradi4g32690	Os08g04370	uclacyanin-2 precursor putative	0.33	1.16	-1.59	0.22	1.81	short	N
D_comp226017_c0_seq1	5	25406189	Bradi1g14020	Os06g24404	anther-specific proline-rich protein APG precursor	2.47	8.64	1.30	3.11	1.81	short	Y
B_comp36787_c0_seq1	4	32413128	Bradi1g66590	Os03g16860	heat shock cognate 70 kDa protein 2	7.33	25.61	2.87	4.68	1.80	short	Y
B_comp26577_c0_seq1	2	46321512	Bradi5g25440	Os04g57200	metal ion binding protein	0.11	0.37	-3.24	-1.44	1.80	short	N
B_comp27756_c0_seq2	6	17211495	Bradi3g51660	Os02g45520	beta-lactamase class A	3.21	0.92	1.68	-0.12	1.80	tall	N
B_comp42428_c0_seq1	7	41516156	Bradi1g33840	Os06g48200	xyloglucan endotransglucosylase/hydrolase protein 23 precursor	7.20	25.03	2.85	4.65	1.80	short	N
D_comp6730_c0_seq5	5	33025107	Bradi1g01000	Os11g31900	acyl carrier protein 2 chloroplast precursor	2.38	8.29	1.25	3.05	1.80	short	N
D_comp67431_c0_seq1	7	13003430	Bradi3g21680	Os02g47110	ADP-ribosylation factor	0.18	0.63	-2.45	-0.66	1.80	short	Y
B_comp22812_c0_seq8	6	22594878	Bradi3g55020	Os02g56800	ATPP2-B2 putative	0.59	2.06	-0.75	1.05	1.80	short	N
B_comp100962_c0_seq1	4	30766206	Bradi1g64480	Os03g19452	-	0.85	2.96	-0.23	1.56	1.79	short	Y
D_comp277554_c0_seq1	1	410432	Bradi2g39790	Os05g01470	methionine S-methyltransferase	0.45	0.13	-1.14	-2.93	1.79	tall	N
D_comp21994_c0_seq1	6	20668411	Bradi3g58020	Os02g52650	chlorophyll a-b binding protein 4 chloroplast precursor	0.17	0.58	-2.57	-0.78	1.79	short	N
D_comp445643_c0_seq1	4	24088611	Bradi3g02300	Os01g61160	L-ascorbate oxidase precursor	0.35	1.21	-1.51	0.28	1.79	short	N
D_comp139229_c0_seq1	6	8563900	Bradi3g10370	Os02g18070	NBS-LRR type disease resistance protein Hom-B	2.49	0.72	1.32	-0.47	1.79	tall	Y
B_comp46745_c0_seq2	1	25426943	Bradi2g32520	Os05g11950	esterase precursor	0.58	1.99	-0.80	0.99	1.79	short	N
A_comp984152_c0_seq1	2	32840943	Bradi5g10050	Os04g34610	-	0.43	0.13	-1.20	-2.99	1.79	tall	N
B_comp1741_c0_seq2	7	5975495	Bradi3g43150	Os02g26720	inositol-tetrakisphosphate 1-kinase 1	0.14	0.49	-2.82	-1.04	1.79	short	N
D_comp62133_c0_seq1	5	9323826	Bradi4g41550	Os11g06900	N-acylethanolamine amidohydrolase	2.09	7.19	1.06	2.85	1.78	short	Y
B_comp3925_c0_seq1	7	38372850	Bradi1g48700	Os06g06440	multidrug resistance-associated protein 14	0.66	2.29	-0.59	1.19	1.78	short	N
D_comp622478_c0_seq1	1	18216045	Bradi3g35120	Os08g28970	-	0.27	0.94	-1.88	-0.10	1.78	short	N
A_comp311499_c0_seq2	3	21880373	Bradi2g49100	Os01g53420	anthocyanidin 53-O-glucosyltransferase	0.37	0.11	-1.44	-3.22	1.78	tall	N
B_comp23340_c0_seq8	4	40950254	Bradi1g77610	Os03g02240	AT-GTL1	1.53	5.27	0.62	2.40	1.78	short	Y

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp49139_c0_seq1	1	20830772	Bradi2g36730	Os05g04820	MYB2	0.19	0.64	-2.42	-0.64	1.78	short	N
B_comp3980_c0_seq1	5	2256356	Bradi4g03000	Os12g40070	B3 DNA binding domain containing protein	0.43	0.13	-1.21	-2.99	1.78	tall	N
B_comp29233_c0_seq1	5	19565357	Bradi4g32570	Os09g29120	FIP1	0.39	1.34	-1.35	0.42	1.78	short	N
B_comp8119_c0_seq5	4	37143876	Bradi1g72350	Os03g08600	transferase transferring glycosyl groups	0.38	1.30	-1.40	0.38	1.78	short	Y
B_comp17413_c0_seq1	4	3581435	Bradi1g09610	Os03g52239	homeodomain protein JUBEL1	0.42	1.44	-1.25	0.52	1.77	short	N
D_comp10420_c1_seq1	6	3055368	Bradi3g04460	Os02g06380	plant-specific domain TIGR01627 family protein	0.10	0.35	-3.28	-1.51	1.77	short	Y
D_comp133241_c0_seq1	5	8956680	Bradi4g23340	Os12g08160	conserved hypothetical protein	0.44	1.50	-1.18	0.59	1.77	short	N
A_comp8261_c0_seq1	1	35996456	Bradi2g24090	Os01g64660	fructose-16-bisphosphatase cytosolic	0.11	0.38	-3.18	-1.41	1.77	short	N
A_comp194334_c0_seq1	2	11463836	Bradi1g22660	Os07g40620	saccharopine dehydrogenase	0.14	0.49	-2.79	-1.02	1.77	short	N
A_comp29421_c0_seq7	2	40252044	Bradi5g17420	Os04g46110	fibroin heavy chain precursor putative	4.55	15.44	2.19	3.95	1.76	short	Y
A_comp1877225_c0_seq1	1	14875540	Bradi3g31880	Os10g38740	glutathione S-transferase GSTU6	0.25	0.83	-2.02	-0.26	1.76	short	N
B_comp41209_c0_seq1	4	205300	Bradi4g38460	Os09g39410	male sterility protein 2	0.33	1.10	-1.62	0.14	1.76	short	N
B_comp52893_c0_seq1	5	28782089	Bradi1g07060	Os03g56430	-	0.12	0.41	-3.03	-1.28	1.76	short	Y
A_comp1172672_c0_seq1	2	10979721	Bradi1g21990	Os07g41360	alpha-14-glucan-protein synthase 1	2.97	10.04	1.57	3.33	1.76	short	Y
A_comp501953_c0_seq1	2	21271443	Bradi1g57510	Os07g03180	-	0.47	1.58	-1.09	0.66	1.76	short	N
D_comp514690_c0_seq1	5	25677777	Bradi2g51490	Os03g43010	-	0.48	1.62	-1.06	0.70	1.76	short	N
D_comp90422_c0_seq8	1	17660753	Bradi3g34520	Os01g32830	-	0.22	0.73	-2.22	-0.46	1.76	short	N
A_comp79766_c0_seq2	6	646988	Bradi3g01100	Os02g01980	anther-specific proline-rich protein APG precursor	2.34	7.90	1.23	2.98	1.75	short	N
B_comp5888_c0_seq1	5	2097480	Bradi4g02850	Os12g40419	WAK-like kinase	0.12	0.41	-3.05	-1.30	1.75	short	N
B_comp45870_c0_seq1	7	4383637	Bradi3g44290	Os02g31030	glycerophosphodiester phosphodiesterase	29.14	8.65	4.86	3.11	1.75	tall	Y
B_comp1770_c0_seq3	3	1972990	Bradi2g02800	Os01g04920	glycosyl transferase group 1 family protein	2.60	0.77	1.38	-0.37	1.75	tall	N
D_comp326848_c0_seq1	4	509947	Bradi5g21880	Os04g52670	OsSAUR21 - Auxin-responsive SAUR gene family member	1.76	5.91	0.81	2.56	1.75	short	N
D_comp756003_c0_seq1	6	8739781	Bradi3g10560	Os02g37010	disulfide oxidoreductase/monooxygenase	0.40	1.36	-1.30	0.44	1.75	short	N
A_comp268616_c0_seq1	3	22013408	Bradi2g48940	Os01g53090	pathogen-related protein	0.13	0.42	-2.99	-1.25	1.75	short	N
B_comp22084_c0_seq1	6	15028040	Bradi3g49160	Os02g41480	OsWAK12 - OsWAK receptor-like cytoplasmic kinase (OsWAK-RLCK)	0.16	0.54	-2.63	-0.89	1.74	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
A_comp1027983_c0_seq1	3	27609751	Bradi2g56750	Os01g65650	receptor-like protein kinase 5 precursor	0.35	1.15	-1.53	0.21	1.74	short	Y
B_comp4566_c0_seq3	5	30768254	Bradi1g04340	Os03g59360	remorin	0.62	2.07	-0.69	1.05	1.74	short	N
B_comp26290_c0_seq7	3	28712195	Bradi2g58090	Os01g67950	BCL-2 binding anthanogene-1	0.13	0.42	-3.00	-1.26	1.74	short	N
D_comp211217_c0_seq1	6	4736608	Bradi1g35390	Os02g09630	nitrate-induced NOI protein	0.68	2.26	-0.56	1.18	1.74	short	N
B_comp22716_c0_seq1	1	8881916	Bradi3g27180	Os02g58610	serine/threonine-protein kinase NAK	0.15	0.49	-2.76	-1.02	1.73	short	N
A_comp54243_c0_seq1	2	23306672	Bradi1g59180	Os07g02060	OsWRKY29 - Superfamily of rice TFs having WRKY and zinc finger domains	0.78	2.60	-0.35	1.38	1.73	short	N
D_comp30014_c0_seq1	7	36055379	Bradi1g46540	Os06g09900	-	0.15	0.50	-2.74	-1.01	1.73	short	N
A_comp216644_c0_seq4	3	27325843	Bradi2g56390	Os01g65200	peptide transporter PTR2	0.11	0.35	-3.24	-1.50	1.73	short	N
B_comp17581_c0_seq1	6	18016927	Bradi3g52650	Os02g47450	-	0.12	0.40	-3.06	-1.34	1.73	short	N
B_comp34319_c0_seq1	5	14994445	Bradi4g28280	Os09g16510	OsWRKY74 - Superfamily of rice TFs having WRKY and zinc finger domains	1.92	0.58	0.94	-0.78	1.73	tall	N
D_comp65978_c0_seq1	5	7599228	Bradi4g39260	Os12g16410	isoflavone reductase	0.86	2.84	-0.22	1.51	1.73	short	Y
B_comp4760_c0_seq9	3	24362895	Bradi2g52370	Os01g58420	ethylene-responsive transcription factor 4	0.57	0.17	-0.80	-2.53	1.73	tall	N
D_comp305871_c0_seq1	1	16563287	Bradi5g22380	Os09g03860	retrotransposon protein unclassified	0.27	0.88	-1.91	-0.18	1.73	short	N
D_comp284932_c0_seq1	4	40045079	Bradi1g75600	Os03g04890	protein binding protein	0.11	0.37	-3.17	-1.45	1.72	short	N
A_comp509935_c0_seq1	1	18441061	Bradi3g35390	Os08g30770	ATATH6	0.21	0.69	-2.26	-0.54	1.72	short	N
D_comp84830_c0_seq1	2	42312426	Bradi5g21170	Os06g11270	anthocyanidin 3-O-glucosyltransferase	0.18	0.59	-2.49	-0.77	1.72	short	N
D_comp139322_c0_seq1	3	13769911	Bradi2g40870	Os05g51590	N-rich protein	1.78	5.85	0.83	2.55	1.72	short	Y
D_comp636527_c0_seq1	3	18026119	Bradi2g45060	Os01g45700	-	1.67	5.49	0.74	2.46	1.72	short	Y
D_comp114034_c0_seq1	7	37178957	Bradi3g18790	Os06g08310	ATPase 8 plasma membrane-type	0.14	0.46	-2.83	-1.11	1.72	short	Y
D_comp65233_c0_seq1	4	7983311	Bradi4g23170	Os11g09180	-	0.18	0.58	-2.50	-0.78	1.72	short	N
A_comp24674_c0_seq1	1	16047966	Bradi3g32880	Os10g40360	proline oxidase mitochondrial precursor	2.74	9.00	1.46	3.17	1.71	short	N
D_comp328221_c0_seq1	6	20611917	Bradi3g58110	Os02g52480	cyclin-dependent kinase inhibitor 2	0.43	1.41	-1.22	0.50	1.71	short	Y
D_comp165815_c0_seq1	2	29795303	Bradi5g07760	Os08g30150	desacetoxyvindoline 4-hydroxylase	0.11	0.37	-3.16	-1.45	1.71	short	N
B_comp7750_c0_seq1	7	35143872	Bradi1g45880	Os06g11280	12-oxophytodienoate reductase 2	0.11	0.36	-3.17	-1.46	1.71	short	N
B_comp41872_c0_seq1	6	23758584	Bradi3g56790	Os02g54030	endo-polygalacturonase precursor	1.68	5.50	0.75	2.46	1.71	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp91277_c0_seq1	7	30752022	Bradi1g42540	Os06g30480	-	0.19	0.61	-2.42	-0.71	1.71	short	N
D_comp349931_c0_seq1	7	8216829	Bradi3g40010	Os09g31410	beta-glucosidase chloroplast precursor	0.96	3.12	-0.07	1.64	1.71	short	N
B_comp11944_c0_seq4	5	19471665	Bradi4g32450	Os08g37150	plant-specific domain TIGR01570 family protein	0.17	0.56	-2.54	-0.84	1.71	short	N
A_comp302900_c0_seq1	3	24726891	Bradi2g52870	Os01g59360	calcium-dependent protein kinase isoform 2	0.40	1.31	-1.32	0.39	1.70	short	N
D_comp34716_c0_seq1	5	8505396	Bradi4g40400	Os12g10320	plant-specific domain TIGR01627 family protein	1.68	5.48	0.75	2.45	1.70	short	N
D_comp1057964_c0_seq1	6	19998591	Bradi3g58980	Os02g51670	AP2 domain-containing protein	0.15	0.49	-2.73	-1.02	1.70	short	N
A_comp443500_c0_seq1	1	24898106	Bradi2g32950	Os05g11250	-	0.15	0.50	-2.71	-1.01	1.70	short	N
D_comp354002_c0_seq1	1	24482488	Bradi2g33320	Os05g10330	HAD superfamily phosphatase containing protein	0.30	0.97	-1.75	-0.05	1.70	short	N
D_comp51840_c0_seq3	7	26625249	Bradi1g39110	Os06g34830	cationic amino acid transporter 4	0.15	0.50	-2.70	-1.00	1.70	short	N
D_comp5232_c0_seq1	3	4442154	Bradi2g04490	Os01g07770	peroxidase 25 precursor putative	0.28	0.91	-1.83	-0.13	1.70	short	Y
D_comp272596_c0_seq1	7	8771345	Bradi3g39400	Os08g38910	caffeoyl-CoA O-methyltransferase 2	0.50	1.63	-0.99	0.71	1.70	short	N
B_comp65047_c0_seq1	7	41524609	Bradi1g33830	Os06g48180	xyloglucan endotransglucosylase/hydrolase protein 23 precursor	0.79	2.54	-0.35	1.35	1.69	short	N
B_comp21834_c0_seq1	6	7198475	Bradi3g09090	Os02g14170	peroxidase precursor	0.11	0.35	-3.22	-1.52	1.69	short	N
D_comp849145_c0_seq1	1	8317589	Bradi3g26790	Os10g28030	acylamino-acid-releasing enzyme	0.30	0.98	-1.72	-0.03	1.69	short	N
B_comp9085_c0_seq7	2	36663277	Bradi5g13610	Os04g40560	nucleotide binding protein	0.15	0.47	-2.78	-1.09	1.69	short	N
D_comp10867_c0_seq1	7	3762277	Bradi1g50090	Os06g04200	granule-bound starch synthase 1 chloroplast precursor	0.81	2.62	-0.30	1.39	1.69	short	N
D_comp118744_c0_seq1	5	9063194	Bradi4g41190	Os12g07810	aldehyde dehydrogenase dimeric NADP-preferring	1.50	4.83	0.58	2.27	1.69	short	N
B_comp15107_c0_seq1	7	8481097	Bradi3g39750	Os08g39300	serine--glyoxylate aminotransferase	0.22	0.71	-2.18	-0.49	1.69	short	N
D_comp1039907_c0_seq1	1	43509408	Bradi4g37210	Os01g41710	chlorophyll a-b binding protein 2 chloroplast precursor	0.30	0.96	-1.74	-0.06	1.69	short	N
B_comp48560_c0_seq1	4	34894099	Bradi1g69380	Os03g12414	cyclin N-terminal domain containing protein	1.00	0.31	-0.01	-1.69	1.68	tall	N
A_comp3243_c1_seq10	2	6801007	Bradi1g16760	Os03g18770	w ound-induced protein 1	1.96	6.27	0.97	2.65	1.68	short	N
B_comp1520_c0_seq1	4	27934847	Bradi1g61470	Os03g25440	-	0.28	0.89	-1.84	-0.16	1.68	short	N
A_comp310201_c0_seq1	3	26827973	Bradi2g55690	Os01g64170	glucan endo-13-beta-glucosidase 7 precursor	0.58	1.85	-0.79	0.88	1.68	short	N
A_comp330929_c0_seq1	2	38784338	Bradi3g49260	Os02g41650	phenylalanine ammonia-lyase	0.45	1.44	-1.15	0.53	1.68	short	N
D_comp34300_c0_seq1	5	30517259	Bradi1g04830	Os03g59080	acyl-activating enzyme 18	0.12	0.37	-3.11	-1.44	1.67	short	Y

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp20406_c0_seq1	5	15884388	Bradi3g35610	Os09g20480	carbohydrate transporter/ sugar porter/ transporter	0.39	0.12	-1.37	-3.04	1.67	tall	N
B_comp21726_c0_seq5	4	33527494	Bradi1g67870	Os03g14880	-	0.75	2.39	-0.42	1.26	1.67	short	N
B_comp27263_c0_seq14	4	41375097	Bradi1g78240	Os10g40510	cortical cell-delineating protein precursor	34.33	109.29	5.10	6.77	1.67	short	N
D_comp725208_c0_seq1	3	30737536	Bradi4g30530	Os11g07960	anthranilate N-benzoyltransferase protein 1	0.16	0.49	-2.69	-1.02	1.67	short	N
B_comp8538_c0_seq1	3	17992550	Bradi2g45010	Os01g45620	OsMPK21-2 - putative MAPK based on amino acid sequence homology	0.11	0.36	-3.16	-1.49	1.67	short	N
B_comp2501_c0_seq1	5	29567748	Bradi1g06200	Os03g57690	aldehyde oxidase 1	0.13	0.42	-2.91	-1.24	1.67	short	N
D_comp117361_c0_seq1	1	41988583	Bradi2g17860	Os05g46720	phosphatidylinositol transfer-like protein III	1.10	3.51	0.14	1.81	1.67	short	Y
B_comp91354_c0_seq1	6	12197929	Bradi3g45700	Os02g34810	OsAPx8 - Thylakoid-bound Ascorbate Peroxidase encoding gene	0.86	2.73	-0.22	1.45	1.66	short	Y
A_comp661022_c0_seq1	2	9611233	Bradi1g20370	Os07g43940	glucan endo-13-beta-glucosidase 4 precursor	0.59	1.88	-0.75	0.91	1.66	short	N
D_comp566816_c0_seq1	5	13859231	Bradi4g27360	Os06g27560	HGA4	0.14	0.45	-2.81	-1.15	1.66	short	N
D_comp255515_c0_seq1	3	9475321	Bradi2g11220	Os05g04550	CBL-interacting serine/threonine-protein kinase 1	0.48	1.53	-1.05	0.62	1.66	short	N
D_comp26808_c0_seq1	5	18188111	Bradi3g45580	Os11g37300	F-box domain containing protein	0.23	0.74	-2.09	-0.43	1.66	short	N
A_comp27887_c0_seq1	2	46594650	Bradi5g25880	Os04g57730	uracil-DNA glycosylase	10.06	3.18	3.33	1.67	1.66	tall	N
B_comp65453_c0_seq1	6	21778650	Bradi3g54010	Os02g55134	cytochrome c oxidase copper chaperone	0.10	0.32	-3.31	-1.65	1.66	short	N
A_comp147709_c0_seq1	2	47817383	Bradi5g27530	Os04g59520	phosphoribosylanthranilate transferase	0.40	1.27	-1.31	0.35	1.66	short	Y
A_comp1660323_c0_seq1	2	46098222	Bradi5g23280	Os04g54390	nuclease PA3	0.85	2.68	-0.24	1.42	1.66	short	N
B_comp8259_c0_seq4	5	19019937	Bradi4g31920	Os09g27990	fiber annexin	0.17	0.53	-2.58	-0.92	1.66	short	N
D_comp653301_c0_seq1	3	15040443	Bradi2g41970	Os01g40190	-	0.12	0.39	-3.03	-1.38	1.66	short	N
B_comp14075_c0_seq2	3	12946371	Bradi2g40400	Os01g34700	protein usf	0.73	2.29	-0.46	1.20	1.66	short	Y
B_comp1117_c0_seq1	6	4757966	Bradi3g06580	Os02g09720	multidrug resistance protein 13	0.22	0.70	-2.17	-0.51	1.65	short	N
D_comp656987_c0_seq1	2	25149440	Bradi1g60800	Os03g27210	LOL3	0.11	0.36	-3.13	-1.48	1.65	short	Y
D_comp139988_c0_seq1	7	37846593	Bradi1g48250	Os09g20980	RING-H2 finger protein ATL5I	0.12	0.37	-3.08	-1.42	1.65	short	N
B_comp423_c1_seq5	7	25623863	Bradi1g38360	Os01g12190	-	0.30	0.93	-1.75	-0.10	1.65	short	N
B_comp26971_c0_seq1	4	28008021	Bradi2g11620	Os03g25150	flavonoid 35-hydroxylase 2	0.26	0.83	-1.93	-0.28	1.65	short	N
D_comp591749_c0_seq1	2	46381671	Bradi5g25570	Os04g57340	ethylene-responsive transcription factor 3	1.85	0.59	0.89	-0.76	1.65	tall	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
A_comp480408_c0_seq1	4	37386331	Bradi1g72600	Os03g08320	pnFL-2	0.18	0.56	-2.48	-0.83	1.65	short	N
A_comp537028_c0_seq1	2	41401196	Bradi5g18750	Os04g48130	membrane protein	0.14	0.43	-2.88	-1.23	1.65	short	N
D_comp625975_c0_seq1	2	40868427	Bradi5g18280	Os04g47360	prolyl endopeptidase	0.41	1.30	-1.27	0.38	1.65	short	Y
A_comp221946_c0_seq1	3	28520771	Bradi2g57860	Os01g67530	4-coumarate--CoA ligase 2	0.13	0.41	-2.92	-1.27	1.65	short	Y
B_comp29229_c0_seq4	6	24110169	Bradi3g60790	Os06g51270	CALS1	0.17	0.54	-2.54	-0.90	1.65	short	N
B_comp27578_c0_seq1	3	29873310	Bradi2g59640	Os08g03700	glycerol-3-phosphate acyltransferase 1 putative	0.23	0.73	-2.09	-0.45	1.65	short	Y
B_comp7827_c0_seq2	6	18395087	Bradi3g53140	Os06g30400	-	0.90	2.83	-0.15	1.50	1.64	short	Y
B_comp2856_c0_seq1	5	12938427	Bradi4g08760	Os09g04790	plastid-lipid-associated protein 2 chloroplast precursor	1.10	3.44	0.14	1.78	1.64	short	N
B_comp94493_c0_seq1	4	39656950	Bradi1g76180	Os03g04110	receptor-like GPI-anchored protein 2	0.19	0.58	-2.43	-0.79	1.64	short	Y
D_comp258287_c0_seq1	7	8211849	Bradi3g00650	Os09g31430	non-cyanogenic beta-glucosidase precursor	22.05	68.71	4.46	6.10	1.64	short	Y
D_comp101371_c0_seq1	7	969368	Bradi1g52750	Os07g17330	B12D protein	1.07	3.32	0.09	1.73	1.64	short	N
B_comp28678_c0_seq1	1	9141819	Bradi3g27410	Os05g39220	anther-specific proline-rich protein APG	0.25	0.77	-2.01	-0.37	1.64	short	N
D_comp5707_c0_seq1	6	2936967	Bradi3g04270	Os01g32330	farnesylated protein 2	0.12	0.37	-3.08	-1.44	1.64	short	N
D_comp58187_c0_seq1	6	19691094	Bradi3g59390	Os02g51110	aquaporin NIP4.2	0.21	0.65	-2.25	-0.61	1.64	short	N
D_comp3581_c0_seq1	6	47768	Bradi3g00250	Os02g01060	VAMP protein SEC22	0.12	0.38	-3.02	-1.38	1.64	short	N
D_comp38760_c0_seq1	5	4964768	Bradi4g05990	Os05g03760	zinc finger transcription factor-like protein	2.48	0.80	1.31	-0.32	1.63	tall	Y
D_comp352859_c0_seq1	7	43535713	Bradi1g31990	Os06g45500	copper-transporting ATPase RAN1	0.82	2.54	-0.29	1.34	1.63	short	Y
B_comp28396_c0_seq2	4	13632358	Bradi1g30610	Os11g26880	RALFL33	1.60	4.97	0.68	2.31	1.63	short	N
B_comp7484_c1_seq5	5	23249049	Bradi4g36740	Os02g55020	-	0.44	0.14	-1.17	-2.80	1.63	tall	N
D_comp241064_c0_seq1	7	37181029	Bradi1g47560	Os06g08300	oxidoreductase	0.27	0.83	-1.90	-0.27	1.63	short	Y
B_comp55358_c0_seq1	5	19342003	Bradi4g32270	Os09g28600	-	0.11	0.33	-3.23	-1.60	1.63	short	N
D_comp81954_c0_seq1	2	29994144	Bradi3g44010	Os02g30320	drought-induced protein 1	0.19	0.58	-2.41	-0.78	1.63	short	N
A_comp789196_c0_seq1	1	23017351	Bradi2g34790	Os01g22640	alpha-L-fucosidase 2 precursor	1.88	5.83	0.91	2.54	1.63	short	Y
A_comp929600_c0_seq1	2	47836181	Bradi5g27540	Os04g59540	phosphatidylinositol-4-phosphate 5-Kinase family protein	0.23	0.71	-2.12	-0.50	1.63	short	Y
D_comp460260_c0_seq1	2	33158982	Bradi5g10360	Os04g35280	neutral/alkaline invertase	0.12	0.37	-3.08	-1.45	1.63	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp28216_c0_seq4	5	11629057	Bradi4g44550	Os11g37870	stripe rust resistance protein Yr10	0.16	0.49	-2.65	-1.02	1.63	short	N
D_comp79740_c0_seq1	6	15688141	Bradi5g16700	Os02g42860	ATP-dependent RNA helicase dhh1	0.75	0.24	-0.42	-2.04	1.63	tall	N
D_comp497843_c0_seq1	5	5047966	Bradi4g06080	Os03g44880	rhicadhesin receptor precursor	0.24	0.75	-2.04	-0.41	1.63	short	N
D_comp353824_c0_seq1	4	34512187	Bradi1g68980	Os03g12570	DNA cytosine methyltransferase MET2a	0.48	0.16	-1.05	-2.68	1.63	tall	N
A_comp16514_c0_seq1	1	32705829	Bradi2g26520	Os05g32590	methylase	4.69	1.52	2.23	0.61	1.63	tall	N
D_comp7129_c0_seq1	3	22915247	Bradi2g50410	Os01g55410	-	0.71	2.20	-0.48	1.14	1.62	short	N
D_comp23979_c0_seq5	3	10245159	Bradi2g11930	Os01g21240	MLA6 protein	0.43	1.31	-1.23	0.39	1.62	short	Y
D_comp53967_c0_seq1	4	22237060	Bradi4g13290	Os11g40570	plant viral-response family protein	0.38	1.16	-1.40	0.22	1.62	short	N
D_comp285659_c0_seq1	1	23685831	Bradi2g34120	Os05g08420	-	0.88	0.29	-0.18	-1.80	1.62	tall	N
A_comp268755_c0_seq1	3	6298838	Bradi2g07940	Os01g13210	DREPP4 protein	5.78	17.79	2.53	4.15	1.62	short	Y
D_comp94389_c0_seq1	5	7102971	Bradi4g08130	Os12g22284	ATP-binding cassette sub-family G member 2	0.28	0.87	-1.82	-0.20	1.62	short	N
B_comp1375_c1_seq5	7	37483120	Bradi1g47820	Os06g07941	iron/ascorbate-dependent oxidoreductase putative	0.57	1.75	-0.81	0.80	1.62	short	N
D_comp750138_c0_seq1	2	7139244	Bradi1g17310	Os01g16120	zinc finger C3HC4 type family protein	0.25	0.77	-2.00	-0.39	1.62	short	N
B_comp15202_c0_seq1	7	13037707	Bradi3g21600	Os04g48230	ankyrin protein kinase-like	1.64	5.04	0.71	2.33	1.62	short	N
B_comp18604_c0_seq1	5	18995799	Bradi4g31870	Os06g10670	aspartic proteinase nepenthesin-1 precursor	4.09	12.55	2.03	3.65	1.62	short	N
D_comp230166_c0_seq1	5	4291592	Bradi1g18080	Os07g47620	universal stress protein	0.14	0.42	-2.86	-1.25	1.62	short	N
D_comp60979_c0_seq1	5	7646138	Bradi4g39310	Os12g16200	glutathione synthetase chloroplast precursor	0.17	0.53	-2.54	-0.93	1.61	short	N
B_comp85537_c0_seq1	5	7975034	Bradi4g39620	Os12g13300	ATP binding protein	0.49	1.51	-1.02	0.60	1.61	short	Y
A_comp128339_c0_seq1	2	40699423	Bradi5g18050	Os04g47010	-	0.38	0.12	-1.39	-3.01	1.61	tall	N
B_comp16121_c0_seq1	4	38520073	Bradi1g74170	Os03g06760	exocyst complex subunit Sec15-like family protein	0.13	0.39	-2.96	-1.35	1.61	short	N
D_comp48991_c0_seq1	4	33573398	Bradi1g67930	Os03g14730	gibberellin receptor GID1L2	0.33	1.02	-1.58	0.03	1.61	short	N
B_comp20108_c0_seq4	5	7317752	Bradi4g38860	Os08g44850	SRC2	0.14	0.43	-2.83	-1.22	1.61	short	N
D_comp43032_c0_seq1	5	7085895	Bradi1g09300	Os12g22030	serine hydroxymethyltransferase mitochondrial precursor	3.92	11.95	1.97	3.58	1.61	short	N
A_comp9799_c0_seq1	7	30588485	Bradi3g53070	Os02g48080	serine/threonine-protein kinase receptor precursor	0.11	0.33	-3.22	-1.61	1.61	short	N
D_comp731388_c0_seq1	7	36722582	Bradi1g47240	Os06g08600	versicolorin reductase	0.12	0.37	-3.04	-1.43	1.61	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp22812_c0_seq2	4	2967628	Bradi1g10410	Os03g51180	-	0.20	0.60	-2.33	-0.73	1.61	short	N
D_comp38175_c0_seq1	4	32555749	Bradi1g66700	Os11g42500	disease resistance response protein 206	1.03	3.15	0.05	1.65	1.61	short	Y
B_comp15931_c0_seq2	5	23234933	Bradi4g36720	Os09g37012	aspartic-type endopeptidase/ pepsin A	0.54	1.63	-0.90	0.71	1.61	short	N
A_comp119189_c0_seq1	2	13655163	Bradi1g25110	Os07g36630	CSLF8 - cellulose synthase-like family F; beta13;14 glucan synthase	0.11	0.32	-3.25	-1.64	1.61	short	N
D_comp3591_c0_seq2	7	3621010	Bradi2g10300	Os01g16750	disulfide oxidoreductase/ monooxygenase	0.28	0.86	-1.83	-0.22	1.60	short	N
A_comp39052_c0_seq18	1	35391218	Bradi2g24680	Os05g35340	-	0.20	0.61	-2.32	-0.72	1.60	short	Y
B_comp6420_c0_seq1	5	16113365	Bradi4g29290	Os09g20880	ATP binding protein	0.11	0.33	-3.19	-1.59	1.60	short	N
D_comp5963_c0_seq1	6	661664	Bradi3g01110	Os02g02000	cytochrome P450 74A4	0.24	0.72	-2.07	-0.47	1.60	short	N
A_comp44568_c0_seq1	4	31740185	Bradi1g65740	Os06g11190	acyl-peptide hydrolase-like	0.13	0.38	-2.99	-1.39	1.60	short	N
B_comp9925_c0_seq1	5	22154509	Bradi4g35350	Os09g34250	indole-3-acetate beta-glucosyltransferase	0.20	0.60	-2.34	-0.74	1.60	short	N
A_comp1720606_c0_seq1	2	10273161	Bradi1g21100	Os07g42740	calmodulin binding protein	0.17	0.51	-2.56	-0.96	1.60	short	N
B_comp59951_c0_seq1	7	40374176	Bradi1g35010	Os06g49860	ankyrin-like protein	0.21	0.62	-2.28	-0.68	1.60	short	N
A_comp722_c1_seq1	2	769185	Bradi5g00830	Os04g10350	1-aminocyclopropane-1-carboxylate oxidase	0.23	0.70	-2.11	-0.51	1.60	short	N
A_comp234423_c0_seq1	1	14736394	Bradi5g19400	Os10g38189	glutathione S-transferase GSTU6	0.47	1.41	-1.10	0.50	1.60	short	Y
D_comp13204_c0_seq1	4	17891382	Bradi4g17020	Os11g33120	respiratory burst oxidase protein D	0.18	0.55	-2.47	-0.87	1.60	short	N
A_comp536848_c0_seq1	5	20313948	Bradi4g33180	Os09g30200	-	0.33	0.11	-1.59	-3.19	1.59	tall	N
A_comp151986_c0_seq1	2	42649735	Bradi5g20760	Os04g51400	zinc finger C3HC4 type family protein	0.17	0.52	-2.54	-0.95	1.59	short	Y
D_comp724157_c0_seq1	6	12136376	Bradi3g45610	Os01g59840	cyanogenic beta-glucosidase precursor	1.92	0.64	0.94	-0.65	1.59	tall	N
B_comp2312_c0_seq1	1	724758	Bradi2g39220	Os05g01940	zinc finger RING-type	1.54	0.51	0.63	-0.96	1.59	tall	N
A_comp245571_c0_seq1	3	19149144	Bradi2g46120	Os01g47780	fasciclin-like arabinogalactan protein 7 precursor	1.16	3.48	0.21	1.80	1.59	short	N
B_comp3794_c0_seq38	1	972695	Bradi2g38850	Os03g58764	F-box domain containing protein	0.12	0.36	-3.05	-1.46	1.59	short	N
A_comp932825_c0_seq1	2	13960173	Bradi1g25440	Os07g35940	beta-amylase putative	7.13	21.47	2.83	4.42	1.59	short	N
A_comp6024_c0_seq1	2	40591514	Bradi5g17900	Os04g46780	structural molecule	0.19	0.57	-2.39	-0.80	1.59	short	N
B_comp35213_c0_seq1	4	31172259	Bradi1g64920	Os03g18850	pathogenesis-related protein 1	0.27	0.81	-1.89	-0.30	1.59	short	Y
D_comp188633_c0_seq1	1	43313591	Bradi2g16550	Os05g48930	OsGrx_S2 - glutaredoxin subgroup III	0.11	0.32	-3.22	-1.63	1.59	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp31719_c0_seq1	4	37261015	Bradi1g72450	Os03g08500	ethylene-responsive element binding protein 2	6.09	2.03	2.61	1.02	1.59	tall	Y
A_comp329423_c0_seq1	1	32737293	Bradi2g26510	Os05g32600	CDC2+/CDC28-related protein kinase R2	35.16	11.70	5.14	3.55	1.59	tall	N
A_comp621932_c0_seq1	5	17368112	Bradi4g31260	Os09g26810	chlorophyll a-b binding protein chloroplast precursor	0.24	0.72	-2.06	-0.47	1.59	short	N
A_comp275727_c0_seq1	3	29712605	Bradi2g59410	Os01g70180	secondary cell w all-related glycosyltransferase family 47	1.60	4.81	0.68	2.27	1.59	short	Y
A_comp222347_c0_seq1	2	32507021	Bradi5g09610	Os04g33920	lipid binding protein	2.54	7.63	1.35	2.93	1.58	short	N
A_comp1762213_c0_seq1	2	23798251	Bradi1g59880	Os07g41320	phytochelatin synthetase-like conserved region family protein	0.23	0.67	-2.15	-0.57	1.58	short	N
D_comp3866_c0_seq3	6	9787089	Bradi3g11340	Os02g22380	glycosyltransferase	0.56	1.66	-0.85	0.73	1.58	short	Y
A_comp11432_c0_seq1	2	37583167	Bradi5g14580	Os04g41970	glycoside transferase six-hairpin subgroup	0.91	2.73	-0.13	1.45	1.58	short	N
D_comp455014_c0_seq1	7	18701008	Bradi3g16940	Os08g05820	monocopper oxidase-like protein SKS1 precursor	2.50	7.48	1.32	2.90	1.58	short	N
D_comp37973_c0_seq1	2	38783872	Bradi5g15830	Os04g43800	phenylalanine ammonia-lyase	0.73	2.18	-0.46	1.12	1.58	short	N
D_comp6987_c0_seq1	2	8406349	Bradi1g18890	Os08g04100	hypothetical protein	0.11	0.34	-3.13	-1.55	1.58	short	N
D_comp23979_c0_seq7	7	19564189	Bradi3g16060	Os06g19130	conserved hypothetical protein	0.32	0.11	-1.63	-3.21	1.58	tall	Y
D_comp89169_c0_seq1	2	13299040	Bradi1g24770	Os01g22710	carbohydrate transporter/ sugar porter	1.02	3.06	0.03	1.61	1.58	short	N
D_comp415476_c0_seq1	3	22400193	Bradi2g48420	Os01g52230	phosphatase phospho1	5.05	1.69	2.34	0.76	1.58	tall	N
A_comp37581_c0_seq1	4	32333791	Bradi1g66470	Os03g16920	heat shock cognate 70 kDa protein	0.79	2.35	-0.35	1.23	1.58	short	N
D_comp320850_c0_seq1	6	2918417	Bradi3g04250	Os06g07932	flavonol synthase/flavanone 3-hydroxylase	0.20	0.61	-2.30	-0.72	1.58	short	N
B_comp22221_c0_seq2	5	435387	Bradi4g00860	Os12g43720	RXW8	0.60	0.20	-0.73	-2.31	1.58	tall	N
B_comp22372_c0_seq3	6	19960076	Bradi3g59020	Os02g51620	beta-D-xylosidase	3.63	10.85	1.86	3.44	1.58	short	N
B_comp21675_c0_seq35	3	5405079	Bradi2g06770	Os01g11340	CYP710A1	0.19	0.58	-2.36	-0.78	1.58	short	N
B_comp16761_c0_seq1	6	3816280	Bradi3g05410	Os02g07690	VQ motif family protein	0.11	0.34	-3.13	-1.55	1.58	short	N
D_comp34422_c0_seq19	7	37443492	Bradi1g47750	Os06g08110	nodulin-like protein	0.88	0.30	-0.18	-1.76	1.58	tall	N
D_comp15126_c0_seq1	1	40396965	Bradi2g19850	Os11g33300	O-methyltransferase ZRP4 putative	0.61	1.81	-0.72	0.86	1.58	short	N
B_comp9937_c0_seq1	5	17568637	Bradi4g31080	Os09g26590	OsSAUR37 - Auxin-responsive SAUR gene family member	0.90	2.68	-0.15	1.42	1.58	short	N
B_comp24014_c0_seq1	2	20257682	Bradi1g56640	Os07g07060	prolyl-tRNA synthetase	0.40	0.13	-1.32	-2.89	1.57	tall	N
D_comp198343_c0_seq1	3	17359916	Bradi2g44260	Os05g25780	rhodanese like protein	0.33	0.11	-1.61	-3.18	1.57	tall	N

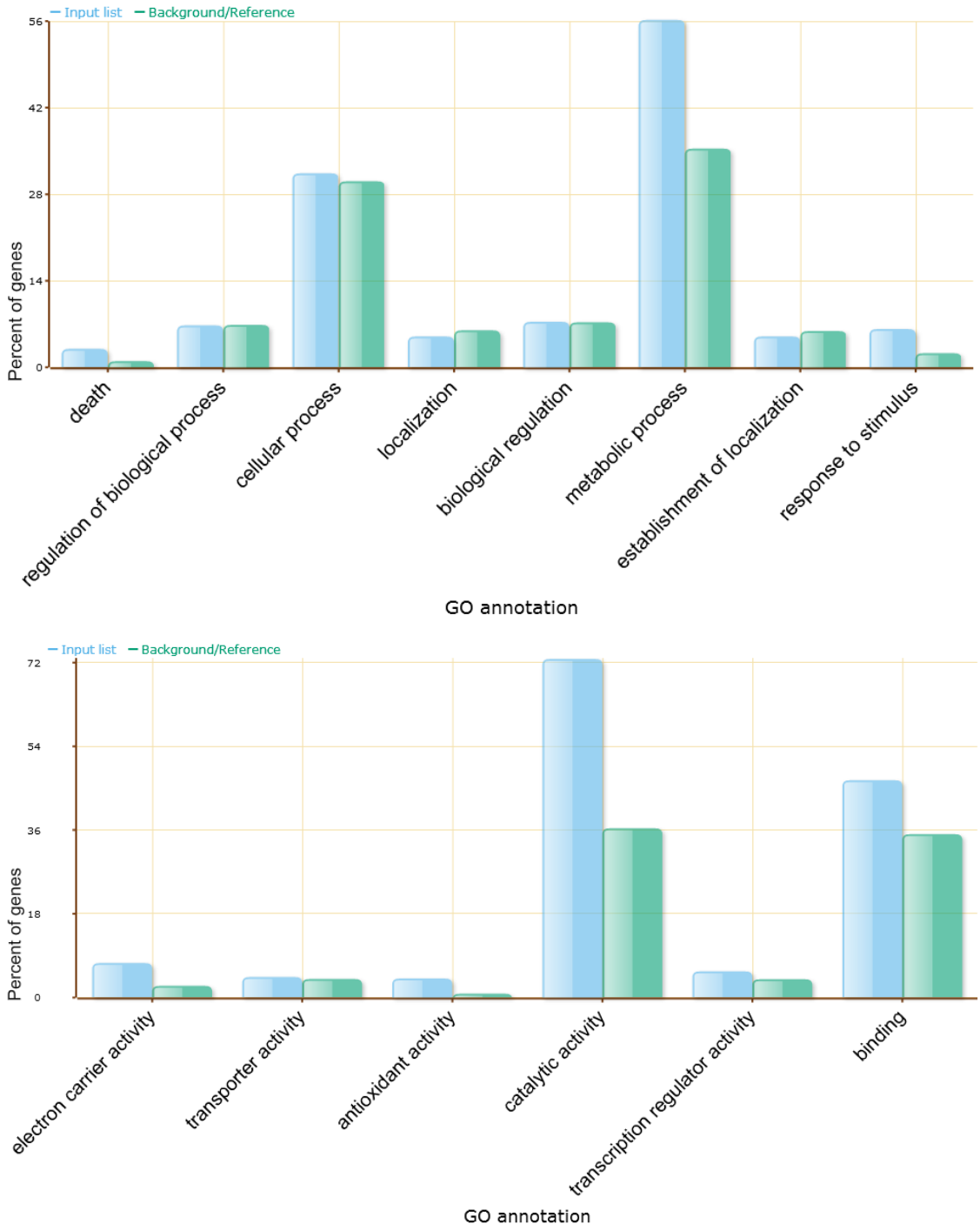
Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp385941_c0_seq1	3	20144596	Bradi2g47210	Os06g51050	basic endochitinase 1 precursor	2.55	7.60	1.35	2.93	1.57	short	N
D_comp13319_c0_seq1	5	28721823	Bradi1g23770	Os03g51600	tubulin alpha-3 chain	0.37	1.11	-1.43	0.15	1.57	short	N
B_comp169_c0_seq1	3	1408499	Bradi2g02320	Os01g04300	endo-14-beta-xylanase	0.70	2.09	-0.51	1.06	1.57	short	Y
A_comp54735_c0_seq1	5	33279204	Bradi1g00540	Os03g64340	copper ion binding protein	0.46	1.37	-1.12	0.45	1.57	short	N
A_comp289011_c0_seq3	3	27508886	Bradi2g56600	Os05g35360	beta-galactosidase/ sugar binding protein	1.13	3.36	0.18	1.75	1.57	short	N
B_comp59906_c0_seq2	4	27249256	Bradi4g09250	Os11g35580	disease resistance protein RPM1 putative	0.15	0.46	-2.69	-1.12	1.57	short	N
A_comp196510_c0_seq1	2	22318538	Bradi1g58530	Os07g05880	kelch motif family protein	0.69	2.03	-0.55	1.02	1.57	short	Y
B_comp736_c1_seq4	1	25409263	Bradi2g32540	Os05g11910	esterase precursor	0.57	1.68	-0.82	0.75	1.57	short	Y
B_comp89002_c0_seq1	7	42161989	Bradi1g33150	Os06g47130	calcium lipid binding protein-like	0.42	0.14	-1.26	-2.83	1.57	tall	Y
B_comp32021_c0_seq2	6	14443983	Bradi3g48530	Os02g39850	anthranilate N-benzoyltransferase protein 1	0.59	1.75	-0.76	0.81	1.57	short	N
B_comp7998_c0_seq1	6	4355589	Bradi3g06070	Os02g08440	OsWRKY71 - Superfamily of rice TFs having WRKY and zinc finger domains	2.89	0.98	1.53	-0.03	1.57	tall	N
A_comp16217_c0_seq1	3	20281521	Bradi2g47330	Os01g50080	multidrug resistance protein 4	0.13	0.39	-2.93	-1.37	1.56	short	N
B_comp41547_c0_seq1	1	43987602	Bradi2g15650	Os05g50100	-	0.15	0.45	-2.71	-1.15	1.56	short	N
B_comp34088_c0_seq1	5	23823142	Bradi1g54640	Os06g13110	retrotransposon protein unclassified	0.44	1.30	-1.18	0.38	1.56	short	N
D_comp15298_c0_seq1	7	21952156	Bradi3g13390	Os08g01830	receptor protein kinase CRINKLY4 precursor	0.12	0.37	-3.01	-1.44	1.56	short	N
A_comp1046997_c0_seq1	3	5558174	Bradi2g06590	Os01g10950	peptide-N4-asparagine amidase A	0.88	2.59	-0.19	1.37	1.56	short	N
A_comp53426_c0_seq2	5	10823655	Bradi4g43410	Os01g24430	-	0.51	1.51	-0.97	0.60	1.56	short	N
D_comp370751_c0_seq1	3	18511879	Bradi2g45530	Os01g46870	ap2 domain protein	0.88	0.30	-0.18	-1.75	1.56	tall	N
D_comp85705_c0_seq1	5	6264582	Bradi4g07380	Os09g17740	chlorophyll a-b binding protein 1 chloroplast precursor	0.39	1.15	-1.36	0.21	1.56	short	N
D_comp832187_c0_seq1	2	16509004	Bradi1g28090	Os07g29410	electron transporter/ thiol-disulfide exchange intermediate	1.17	3.46	0.23	1.79	1.56	short	N
A_comp35775_c0_seq3	3	23862304	Bradi2g51690	Os01g57040	-	1.31	3.87	0.39	1.95	1.56	short	N
D_comp271282_c0_seq1	6	2703132	Bradi3g03990	Os02g05660	ATP-dependent RNA helicase DDX41	0.14	0.40	-2.88	-1.32	1.56	short	N
A_comp7797_c0_seq1	2	35834313	Bradi3g47050	Os02g37070	-	0.50	1.48	-1.00	0.56	1.56	short	N
A_comp44274_c0_seq5	4	35803685	Bradi1g70350	Os03g11290	-	0.19	0.55	-2.43	-0.87	1.56	short	Y
B_comp20801_c0_seq8	7	33632332	Bradi1g44470	Os06g13760	transferase transferring glycosyl groups	2.29	0.78	1.19	-0.36	1.56	tall	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
D_comp165585_c0_seq1	4	28593926	Bradi1g62270	Os12g42910	K-exchanger-like protein	0.22	0.65	-2.18	-0.63	1.56	short	N
D_comp98340_c0_seq1	2	11737990	Bradi1g22980	Os07g39920	actin binding protein	0.61	1.79	-0.72	0.84	1.56	short	Y
B_comp9806_c0_seq1	6	8929449	Bradi3g10730	Os02g22020	DNA binding protein	4.09	1.39	2.03	0.48	1.55	tall	N
D_comp848455_c0_seq1	6	23039630	Bradi3g55670	Os02g57480	anthocyanin 5-aromatic acyltransferase	0.47	1.39	-1.08	0.47	1.55	short	Y
D_comp7654_c0_seq1	6	24170551	Bradi3g60920	Os02g58790	-	0.40	1.18	-1.31	0.24	1.55	short	Y
D_comp302784_c0_seq1	4	40729936	Bradi1g53090	Os03g62410	phospholipase D gamma 3	0.13	0.39	-2.91	-1.35	1.55	short	N
B_comp19224_c0_seq11	1	42520129	Bradi2g17230	Os05g48060	phosphatidylserine synthase 2	0.62	1.83	-0.69	0.87	1.55	short	N
D_comp710_c0_seq1	5	21041831	Bradi4g34000	Os05g31010	hypothetical protein	0.52	1.54	-0.93	0.62	1.55	short	Y
D_comp548173_c0_seq1	1	3668665	Bradi3g23360	Os03g09070	protein binding protein	0.12	0.36	-3.01	-1.46	1.55	short	N
B_comp10709_c0_seq1	7	45465251	Bradi1g30010	Os06g50910	serine/threonine-protein kinase ATR	0.42	0.14	-1.24	-2.79	1.55	tall	N
B_comp18062_c0_seq1	4	40741817	Bradi1g77290	Os03g02710	hydroxymethylglutaryl-CoA synthase	5.18	15.18	2.37	3.92	1.55	short	Y
A_comp807118_c0_seq1	6	7609231	Bradi3g56990	Os02g57940	F-box domain containing protein	0.17	0.51	-2.53	-0.98	1.55	short	N
B_comp74740_c0_seq1	3	26771149	Bradi2g55570	Os01g64020	transcription factor HBP-1b	0.43	0.15	-1.23	-2.78	1.55	tall	N
B_comp12944_c0_seq2	1	3227081	Bradi3g23150	Os10g13940	antiporter/ drug transporter	0.28	0.82	-1.84	-0.29	1.55	short	N
D_comp456000_c0_seq1	6	21433496	Bradi3g53610	Os02g48870	aspartic proteinase nepenthesin-2 precursor	0.13	0.38	-2.95	-1.40	1.55	short	N
D_comp302704_c0_seq1	5	23758624	Bradi3g42620	Os01g52240	chlorophyll a-b binding protein 2 chloroplast precursor	1.16	3.39	0.21	1.76	1.55	short	Y
D_comp139920_c0_seq1	5	28834029	Bradi1g06980	Os03g56540	mitochondrial import inner membrane translocase subunit TIM14	0.16	0.48	-2.61	-1.06	1.55	short	Y
B_comp64820_c0_seq1	1	9868841	Bradi3g27900	Os10g31850	RING finger and CHY zinc finger domain-containing protein 1	0.47	1.37	-1.09	0.46	1.55	short	N
A_comp162096_c0_seq1	2	44435313	Bradi5g22550	Os06g49360	NBS-LRR disease resistance protein	0.10	0.30	-3.27	-1.72	1.55	short	N
A_comp651751_c0_seq1	2	46809770	Bradi5g26190	Os04g58130	katanin p80 WD40-containing subunit B1 homolog 1	0.12	0.35	-3.05	-1.50	1.54	short	N
B_comp4852_c0_seq1	6	15706670	Bradi3g50000	Os02g42880	remorin	0.56	1.63	-0.84	0.71	1.54	short	N
A_comp64791_c0_seq1	5	15523604	Bradi4g28770	Os09g19820	aminopeptidase-like protein	0.15	0.44	-2.71	-1.17	1.54	short	N
B_comp20108_c0_seq3	2	25068817	Bradi1g60680	Os03g27370	phospholipase D alpha 1	0.13	0.37	-2.97	-1.43	1.54	short	N
D_comp322420_c0_seq1	4	18763206	Bradi4g16280	Os11g35400	acyl CoA synthetase	3.76	10.95	1.91	3.45	1.54	short	Y
B_comp38245_c0_seq1	6	15917722	Bradi3g50210	Os02g43290	protein kinase APK1A chloroplast precursor	0.25	0.72	-2.03	-0.48	1.54	short	N

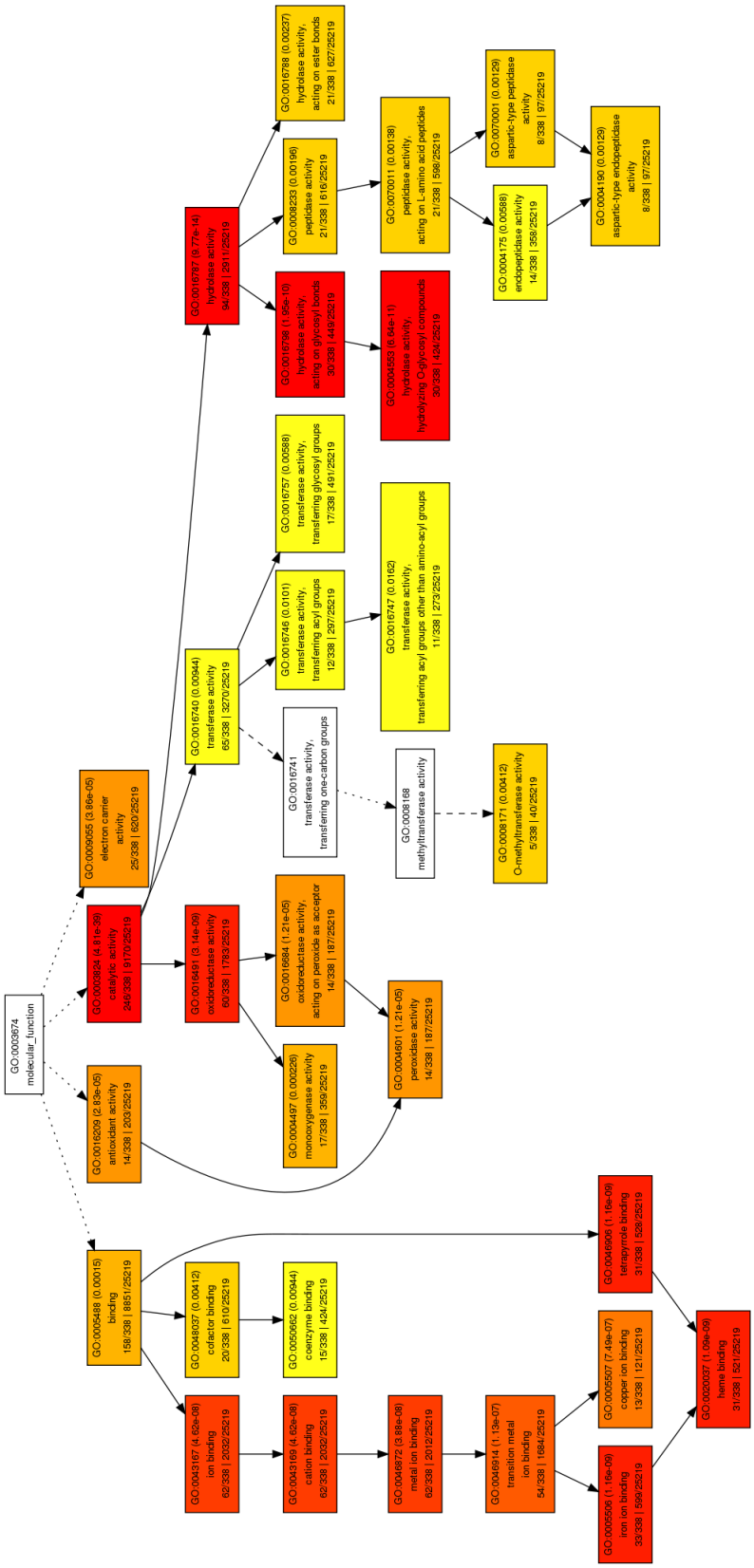
Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp31628_c0_seq2	5	27683514	Bradi1g08550	Os03g53790	periplasmic beta-glucosidase precursor	0.22	0.64	-2.19	-0.65	1.54	short	Y
B_comp29017_c0_seq1	6	3185127	Bradi3g04640	Os02g06580	formin homology 2 domain-containing protein 5	0.11	0.33	-3.13	-1.59	1.54	short	N
A_comp303859_c1_seq1	3	14344005	Bradi2g41510	Os01g39000	-	7.07	20.56	2.82	4.36	1.54	short	N
B_comp60276_c0_seq1	1	31890893	Bradi2g27180	Os05g31020	eukaryotic peptide chain release factor subunit 1-1	0.11	0.31	-3.22	-1.68	1.54	short	N
B_comp77118_c0_seq1	5	212451	Bradi1g12080	Os03g48480	acyl-CoA thioesterase/ catalytic/ hydrolase acting on ester bonds	0.37	1.08	-1.43	0.11	1.54	short	Y
B_comp26928_c0_seq2	4	39254691	Bradi1g76770	Os12g23280	GMFP5	0.23	0.68	-2.09	-0.56	1.54	short	N
D_comp663158_c0_seq1	4	189005	Bradi4g38490	Os09g39440	pyrimidine-specific ribonucleoside hydrolase rihB	0.51	1.49	-0.97	0.57	1.54	short	N
D_comp89024_c0_seq1	4	29414335	Bradi1g63120	Os07g48150	plant-specific domain TIGR01568 family protein	0.33	0.95	-1.61	-0.07	1.54	short	N
A_comp12998_c0_seq1	2	44079646	Bradi5g22060	Os04g52880	-	9.75	28.27	3.29	4.82	1.54	short	Y
B_comp25805_c0_seq1	3	27310686	Bradi2g56360	Os01g65110	POT family protein	0.16	0.47	-2.64	-1.10	1.53	short	N
B_comp14969_c0_seq1	7	4548348	Bradi3g44220	Os11g06190	-	0.41	0.14	-1.30	-2.84	1.53	tall	N
D_comp285548_c0_seq1	1	32759411	Bradi2g26490	Os01g27890	cytochrome P450 71D10	0.11	0.31	-3.22	-1.69	1.53	short	N
D_comp535818_c0_seq1	1	18369518	Bradi3g35310	Os08g31030	protein HOTHEAD precursor	1.21	3.51	0.28	1.81	1.53	short	N
D_comp408389_c0_seq1	5	11366968	Bradi2g58760	Os11g02820	-	0.13	0.37	-2.98	-1.45	1.53	short	N
A_comp345322_c0_seq1	1	27878364	Bradi2g30390	Os05g27790	start codon	0.21	0.62	-2.22	-0.69	1.53	short	N
D_comp724367_c0_seq1	7	47004350	Bradi1g07770	Os07g02460	-	0.12	0.35	-3.06	-1.53	1.53	short	N
B_comp64277_c0_seq1	3	26195269	Bradi2g54800	Os01g62670	avr9/Cf-9 rapidly elicited protein 137	0.15	0.42	-2.78	-1.25	1.53	short	N
A_comp20066_c0_seq1	2	10629734	Bradi1g21510	Os07g42324	-	11.94	4.13	3.58	2.05	1.53	tall	N
A_comp208261_c0_seq1	2	7983031	Bradi1g18280	Os03g14620	-	0.48	1.38	-1.07	0.47	1.53	short	Y
A_comp296991_c0_seq1	3	27542190	Bradi2g56650	Os01g65510	F-box protein interaction domain containing protein	0.12	0.36	-3.00	-1.47	1.53	short	N
B_comp7846_c0_seq2	5	16053897	Bradi4g29260	Os09g20830	heat shock factor-binding protein 1	0.41	0.14	-1.29	-2.82	1.53	tall	N
D_comp500288_c0_seq1	6	12388751	Bradi3g45940	Os02g35190	chloride channel protein CLC-c	0.21	0.61	-2.24	-0.71	1.53	short	N
D_comp106993_c0_seq1	2	22046707	Bradi1g58200	Os06g12100	mTERF family protein	1.46	0.51	0.55	-0.98	1.53	tall	N
B_comp21129_c0_seq7	7	34399257	Bradi1g45200	Os06g12310	aquaporin NIP4.1	0.17	0.48	-2.58	-1.06	1.53	short	N
A_comp24039_c0_seq1	3	16809426	Bradi2g43730	Os01g43140	triacylglycerol lipase	1.30	3.75	0.38	1.91	1.53	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
A_comp309104_c0_seq1	3	31215969	Bradi2g61590	Os01g73250	bundle sheath cell specific protein 1	0.40	1.15	-1.33	0.20	1.53	short	Y
B_comp47777_c0_seq1	7	37469749	Bradi1g47790	Os06g08060	leucoanthocyanidin dioxygenase	0.17	0.49	-2.57	-1.04	1.52	short	Y
D_comp20997_c0_seq2	7	400752	Bradi1g53310	Os07g14590	IAA-amino acid hydrolase ILR1 precursor	0.52	1.50	-0.94	0.58	1.52	short	N
A_comp3243_c0_seq7	3	17613377	Bradi2g44550	Os01g44960	catalytic/ hydrolase	0.13	0.37	-2.94	-1.42	1.52	short	N
A_comp285784_c0_seq1	3	17481079	Bradi2g55290	Os01g44069	glycerol-3-phosphate acyltransferase 1	0.11	0.31	-3.23	-1.70	1.52	short	N
B_comp2550_c2_seq15	2	43541454	Bradi5g19720	Os04g49748	ATPUP11	0.13	0.37	-2.97	-1.45	1.52	short	Y
B_comp28179_c0_seq3	5	22566856	Bradi4g35940	Os09g36080	cytochrome P450 71A2	0.13	0.38	-2.90	-1.38	1.52	short	N
B_comp44975_c0_seq2	5	14804849	Bradi4g28100	Os09g15670	protein phosphatase 2C	0.25	0.71	-2.02	-0.50	1.52	short	N
A_comp10039_c2_seq7	2	41556572	Bradi5g18880	Os04g48390	beta-lactamase class A	1.74	4.98	0.80	2.32	1.52	short	N
B_comp33583_c0_seq2	2	10058792	Bradi1g20870	Os07g42440	hydroxyacid oxidase 1	0.35	1.00	-1.52	0.00	1.52	short	N
B_comp37925_c0_seq3	4	6256099	Bradi4g24410	Os11g07260	leucine-rich repeat receptor protein kinase EXS precursor	0.57	1.64	-0.80	0.72	1.52	short	N
D_comp40099_c0_seq1	4	39548276	Bradi1g76330	Os03g03910	catalase-1	0.20	0.58	-2.31	-0.79	1.52	short	Y
D_comp12706_c0_seq1	7	35038857	Bradi1g45790	Os06g11310	copper ion binding protein	0.13	0.36	-2.98	-1.46	1.52	short	N
B_comp1851_c0_seq3	3	28834491	Bradi2g58290	Os01g68140	-	0.62	1.77	-0.70	0.82	1.52	short	N
B_comp42092_c0_seq1	6	928259	Bradi3g01470	Os02g02400	catalase isozyme A	6.38	18.24	2.67	4.19	1.52	short	N
A_comp3243_c0_seq1	1	30620984	Bradi2g28040	Os05g28980	fiber protein Fb2	0.47	0.17	-1.08	-2.59	1.51	tall	N
D_comp9741_c0_seq1	3	23673345	Bradi2g51480	Os05g43310	photosystem II reaction center W protein chloroplast precursor	0.35	1.00	-1.51	0.00	1.51	short	N
D_comp416079_c0_seq1	2	40649738	Bradi1g03760	Os03g60580	actin-depolymerizing factor 4	3.08	8.80	1.62	3.14	1.51	short	N
A_comp77030_c1_seq10	5	4491489	Bradi4g05440	Os12g43490	alpha-amylase/trypsin inhibitor	0.16	0.45	-2.68	-1.17	1.51	short	N
D_comp66207_c0_seq1	4	13268470	Bradi4g19660	Os11g22404	catalytic/ protein phosphatase type 2C	0.50	0.18	-1.00	-2.51	1.51	tall	N
A_comp3243_c0_seq3	3	3835909	Bradi2g05200	Os01g08670	-	0.16	0.46	-2.65	-1.14	1.51	short	N
B_comp31266_c0_seq2	3	7113546	Bradi2g08790	Os01g14670	nectarin-1 precursor	1.62	4.63	0.70	2.21	1.51	short	N
B_comp8751_c0_seq1	4	3391360	Bradi1g09820	Os03g52080	SNG2	0.78	2.23	-0.35	1.16	1.51	short	Y
B_comp64797_c0_seq1	5	5174598	Bradi4g06170	Os02g55550	ubiquitin-protein ligase	0.35	0.12	-1.53	-3.04	1.51	tall	N
D_comp723038_c0_seq1	3	28066222	Bradi2g25150	Os08g02160	NAC domain-containing protein 68	0.18	0.51	-2.47	-0.96	1.51	short	N

Unigene	Chr	Pos	Brachy	Rice	Annotation	CD raw	RIL4 raw	Log2 (CD)	Log2 (RIL4)	Fold diff	Upreg	In SNP data?
B_comp55250_c0_seq1	5	6477854	Bradi4g07570	Os12g29220	MTN3	0.37	1.05	-1.43	0.07	1.51	short	N
D_comp103726_c0_seq2	7	7514289	Bradi3g40860	Os08g41320	BHLH transcription factor	0.16	0.44	-2.68	-1.17	1.51	short	N
B_comp12069_c0_seq5	7	38084766	Bradi1g48510	Os06g06760	protein kinase	0.46	0.16	-1.13	-2.64	1.51	tall	N
B_comp423_c1_seq3	3	22954141	Bradi2g50480	Os11g04010	caspase	0.45	1.27	-1.16	0.34	1.51	short	N
B_comp23373_c0_seq1	7	45155493	Bradi1g30390	Os06g43520	cytochrome P450 71D7	0.21	0.60	-2.24	-0.73	1.51	short	N
B_comp3169_c1_seq6	5	3366862	Bradi4g04140	Os12g38100	-	0.79	2.25	-0.34	1.17	1.50	short	N
A_comp289011_c0_seq23	3	28253025	Bradi2g57520	Os01g66970	zinc finger protein	0.61	1.72	-0.72	0.78	1.50	short	N
D_comp629373_c0_seq1	3	6176104	Bradi2g07830	Os01g13130	aquaporin TIP4.1	0.11	0.30	-3.22	-1.72	1.50	short	N
D_comp418048_c0_seq1	4	22621034	Bradi4g12950	Os11g41120	pistil-specific extensin-like protein precursor	0.60	1.70	-0.74	0.77	1.50	short	Y
B_comp47912_c0_seq1	3	24066717	Bradi2g51940	Os01g57510	receptor protein kinase	0.11	0.32	-3.16	-1.66	1.50	short	N
B_comp23761_c0_seq2	2	42273173	Bradi5g21250	Os04g52090	ethylene-responsive transcription factor 4	1.40	0.49	0.48	-1.02	1.50	tall	Y



A6.8.2: Annotation of the 1735 differentially expressed genes in the UniGene dataset in A6.8.1 by biological process (top) and molecular function (bottom). The DEG count is in blue bars. A relative measure is made against the reference library of 25,219 annotated Brachypodium genes (green bars) (Du et al., 2010).



A6.8.3: Enrichment of most prominent GO molecular function categories in the DEG dataset from A6.8.1. Most highly represented categories are highlighted in red, with the lowest in white. Networks of related function are linked by arrows. A relative measure is made against the reference library of 25,219 annotated Brachypodium genes (Du et al., 2010).

Unigene	CSS contig	Chr	Pos	Brachy	Rice	CD raw	RIL4 raw	Tall_bulk raw	Short_bulk raw	Log2 (CD)	Log2 (RIL4)	Log2 (Tall_bulk)	Log2 (Short_bulk)	Fold Diff Parent	Fold Diff Bulk	Upreg	In SNP data?
mrna044873	1AL_3928868	1A	3408472	-	-	1.03	311.56	1.47	9.98	0.04	8.28	0.55	3.32	8.24	2.77	short	N
mrna044872	1BL_3915343	1B	1910009	-	-	1.31	270.89	2.18	11.37	0.39	8.08	1.12	3.51	7.69	2.38	short	N
mrna123201	6DS_2119761	6D	932349	-	-	0.24	27.28	0.40	2.42	-2.08	4.77	-1.34	1.28	6.85	2.61	short	N
mrna123200	6DS_2086692	6Du	2484187	-	-	0.27	29.03	0.34	2.64	-1.91	4.86	-1.56	1.40	6.77	2.96	short	N
mrna004986	6AS_4422147	6Au	7523111	-	-	0.49	17.56	0.45	1.94	-1.03	4.13	-1.15	0.95	5.16	2.10	short	N
mrna070193	5BL_10899388	5Bu	21483172	-	-	0.17	4.75	0.10	0.79	-2.53	2.25	-3.27	-0.34	4.78	2.93	short	N
mrna133163	2BL_8041230	2B	25270285	-	-	0.20	3.68	0.21	0.95	-2.32	1.88	-2.25	-0.08	4.20	2.17	short	N
mrna117914	2BL_7909877	1Bu	3959895	-	-	0.27	2.55	0.13	0.40	-1.89	1.35	-2.98	-1.34	3.24	1.64	short	N
mrna017473	5AL_2752773	5Au	6766900	-	-	0.97	0.11	1.72	0.38	-0.04	-3.13	0.79	-1.41	3.09	2.19	tall	N
mrna069155	7DS_3914567	5B	7983196	-	-	1.40	0.30	0.41	0.13	0.49	-1.76	-1.30	-2.91	2.25	1.61	tall	N
mrna059751	3AS_3440107	3A	476382	-	-	0.25	1.15	0.24	0.91	-1.98	0.20	-2.04	-0.14	2.18	1.90	short	N
mrna055680	6BS_872410	6B	1576235	Bradi3g05210	-	0.70	0.16	1.00	0.20	-0.51	-2.60	0.01	-2.29	2.09	2.29	tall	N
mrna032096	1AL_3870748	1Au	6052684	Bradi2g02720	Os07g575000	1.00	0.25	10.83	1.86	0.00	-2.02	3.44	0.90	2.02	2.54	tall	N
mrna017060	2AS_5285710	2A	5714415	Bradi1g23990	Os07g0572400	0.49	1.77	1.14	4.15	-1.02	0.82	0.19	2.05	1.84	1.86	short	N
mrna111603	2AS_5306691	2Au	7845050	-	-	0.15	0.51	0.15	0.65	-2.77	-0.98	-2.71	-0.63	1.80	2.08	short	N
mrna059953	4DL_14391067	1Au	10955187	-	Os03g0669100	0.43	0.13	0.53	0.15	-1.21	-2.99	-0.90	-2.71	1.78	1.81	tall	N
mrna089773	1AS_3255897	1Au	790349	-	-	0.28	0.97	0.10	0.53	-1.82	-0.04	-3.25	-0.92	1.77	2.33	short	N
mrna084749	2BS_5246502	2B	1301163	-	-	0.32	0.10	2.61	0.89	-1.64	-3.31	1.39	-0.17	1.68	1.55	tall	N
mrna040738	5BL_10855125	5Bu	16266605	Bradi2g37430	-	0.44	0.15	4.36	1.06	-1.18	-2.75	2.12	0.08	1.56	2.04	tall	N
mrna072310	7DL_3391820	7Bu	7895777	-	-	2.31	0.79	2.79	0.96	1.21	-0.34	1.48	-0.05	1.54	1.53	tall	N

A6.8.4: Differentially expressed genes between the parent NILs and BSA in the v3.3 cDNAs. Genes are annotated if they overlapped with SNP data. Raw RPKM values are shown, as well as log-transformed data used to set the 1.5-fold threshold as a DEG.

A6.9 (continued across three pages)

Wheat gene models from CSS contigs anchored into the 17.3 cM bin by POPSEQ data. Gene models were extracted from *Ensembl*Plants and annotated. Genes which were duplicated (based on BLASTP results) are annotated in the 'Dup' column. Genes which were tested with markers are indicated. Shaded red are markers which were developed and were monomorphic between the parents to the fine-mapping population. Genes on contigs which had no variation are shaded red. The BLASTP Annotation is from the NCBI database, with the top hit used for the annotation in terms of peptide identity and accession number, using the peptides to the genes as queries. The species from which the annotation derives is indicated. Tauschii: *Ae. tauschii*; BD: Brachypodium; Urartu: *T. Urartu*; Aestivum: *T. aestivum*. The Interpro, GO annotations and synteny information were obtained from *Ensembl*Plants. If the Brachypodium gene had a putative SNP between parent NILs in the UniGene data, this is indicated. The gene is also marked if it was in the differentially expressed genes (DEGs) in A6.8.1. Finally, if the Brachypodium gene could be anchored in the *Ae. tauschii* gene list/zipper in A6.7, the bin is indicated in the final column. Shaded red means a cM bin that was outside the *Rht8* interval. Shaded green is a cM bin within the defined interval.

A6.9.1: Genes within the 17.3 cM bin and anchored within the *Rht8* interval in *Ae. tauschii* or not found in the *Ae. tauschii* data.

A6.9.2: Genes within the 17.3 cM bin and anchored outside the *Rht8* interval in *Ae. tauschii*.

A6.9.1

Wheat gene (Traes_)	Dup	Previously tested		BlastP NCBI Annotation	Species	% Query	% Id	GenBank Accession	InterPro	GO term name	HV gene	HV pos (Mb)		BD gene	Os gene	SNP ?	DEG?	Tauschii
2DS_5D0D6CE9A 2DS_77EE9D5CB 2DS_31BABEFBA 2DS_87ABEEA3D 2DS_CC0A98FBC 2DS_6BFAA38F7 2DS_B1CD8F65A	7	yes	2DS_46, 51, 60, 76	Proline-rich protein 4-like	BD	93	69	XP_003571659	Pollen Ole e 1 allergen/extensin	-	MLOC_32207	2	18.55	Bradi5g04630	Os10g0150800	N	N	35.34
2DS_0E850F4E2		yes	2DS_19, 2DS_196	Ccleavage/polyadenylation specificity factor subunit 1	Tauschii	97	95	EMT25361	Cleavage/polyadenylation specificity factor	cytosol	MLOC_12182	2	18.49	Bradi5g04673	Os04g0252200	N	N	35.339
2DS_2B31E631C		yes	2DS_118-20	Acyl-CoA--sterol O-acyltransferase 1-like	BD	95	78	XP_003581182	-	-	MLOC_77446	7	104.77	Bradi5g09000	Os04g0277400	N	N	31.41
2DS_7478A1CD1		2DS_5364728		Probable protein transport Sec1a	BD	99	86	XP_010239737	Sec1-like protein	vesicle docking involved in exocytosis	MLOC_72777	2	17.52	Bradi5g04686	Os04g0252400	N	N	31.41
2DS_1E247C42B		yes	2DS_103	MRP-like ABC transporter	OS	99	79	CAD59594	ABC transporter-like	ATP binding	-	-	-	Bradi5g03477	Os04g0209300	N	N	31.41
2DS_C50053547				Momilactone A synthase-like	BD	99	59	XP_003581367	Short-chain dehydrogenase SDR	oxidoreductase activity	-	-	-	Bradi1g22860	Os07g0665000	N	short x2.32	-
2DS_8B6DB2397				LRR extensin-like protein 2	BD	37	70	XP_003559819	C2 domain	protein binding	MLOC_69091	2	20.26	Bradi1g18280	Os07g0670300	N	short x1.53	-
2DS_05DCEB6801		yes	2DS_68, 2DS_104	Reticuline oxidase-like protein	Tauschii	99	96	EMT18490	FAD linked oxidase	flavin adenine dinucleotide binding	MLOC_63480	2	14.56	Bradi5g02900	Os06g0549600	N	N	-
2DS_985CFD29C				Cytochrome P450 71D11	Tauschii	94	99	EMT05295	Cytochrome P450	iron ion binding	-	-		Bradi4g09000	Os06g0642300	N	N	-
2DS_7EDB434AD				Cytochrome P450 71D8	Tauschii	98	92	EMT05296	Cytochrome P450	iron ion binding	-	-		Bradi3g15020	Os06g0641800	N	N	-
2DS_5C01AE66B		yes	2DS_72, 2DS_114	Isoflavone 3'-hydroxylase-like	BD	82	80	XP_010238949	Cytochrome P450	iron ion binding	MLOC_55120	2	19.95	Bradi4g07480	Os04g0255600	N	N	-
2DS_8917CF955				UDP-glycosyltransferase 74E1	Urartu	99	96	EMS60126	UDP-glucuronosyl/ UDP-glucosyltransferase	transferase, transferring hexosyl grps	MLOC_37045	2	602.21	Bradi5g02780	Os04g0206700	N	N	-
2DS_203132448				Wall-associated receptor kinase 3	Tauschii	100	100	EMT31734	Protein kinase-like domain	ATP binding	-	-	-	-	-	-	-	
2DS_CB771B9DF				Receptor-like Ser/Thr-protein kinase SD1-8	BD	96	75	XP_010227301	Protein kinase domain	ATP binding	MLOC_60096	2	14.60	-	Os07g0301500	-	-	-
2DS_8E5665A3E		2DS_5342238		Vacuolar amino acid transporter 1	Tauschii	99	99	EMT21150	Amino acid transporter, transmembrane	integral component of membrane	MLOC_61723	2	15.26	Bradi5g02920	Os04g0201800	N	N	-
2DS_0F3296166				Bidirectional sugar transporter SWEET6a	BD	68	71	XP_003569319	-	-	-	-	-	-	Os09g0259200	-	-	-

A6.9.1 (continued)

Wheat gene (Traes_)	Dup	Previously tested		BlastP NCBI Annotation	Species	% Query	% Id	GenBank Accession	InterPro	GO term name	HV gene	HV pos (Mb)		BD gene	Os gene	SNP ?	DEG?	Tauschii
2DS_86B38D3CF 2DS_7E56EEE5F	2			Cystathionine gamma-synthase, chloroplastic	Tauschii	99	100	EMT09613	Cys/Met metabolism, pyridoxal pho-dependent	catalytic activity	-	-	-	Bradi1g69730	Os10g0399200	N	N	-
2DS_562B4B7BF				Cyclopropane-fatty-acyl-phospholipid synthase	Urartu	99	77	EMS55075	Mycolic acid cyclopropane synthase	lipid biosynthetic process	-	-	-	Bradi5g09110	-	N	N	-
2DS_BD1D507B5				Putative methyltransferase DDB_G0268948	BD	56	89	XP_003567596	-	-	MLOC_3492	2	11.79	Bradi2g11740	Os01g0307686	N	N	-
2DS_59A6D76F6				Cytosolic sulfotransferase 5-like	BD	99	69	XP_003575593	Sulfotransferase domain	sulfotransferase activity	MLOC_7130	2	22.43	Bradi3g09500	-	N	N	-
2DS_4D186BBBF				Caffeic acid 3-O-methyltransferase	Tauschii	99	90	EMT13959	O-methyltransferase	O-methyltransferase	-	-	-	Bradi1g14870	Os08g0157500	N	N	-
2DS_6C351FD1B				Cation-chloride cotransporter 1-like	BD	99	91	XP_003565703	Amino acid permease	integral component of membrane	MLOC_5064	2	19.95	Bradi2g11652	Os01g0304100	N	N	-
2DS_677AECF17				Pleiotropic drug resistance protein 5-like	BD	99	89	XP_003563012	ABC transporter-like	membrane	-	-	-	-	Os02g0318450	-	-	-
2DS_ABAA92C8B				THO complex subunit 4	Urartu	99	98	EMS62008	-	-	-	-	-	-	-	-	-	-
2DS_9AE3D3114				Endo-1,4-beta-xylanase Z	Tauschii	55	100	EMT25110	-	-	-	-	-	-	Os05g0305000	-	-	-
2DS_E88D10B93				Putative nuclease HARBI1	Urartu	100	92	EMS68915	-	-	-	-	-	-	-	-	-	-
2DS_8C3F6A558 2DS_E2F2E9EAF	2	yes	2DS_46, 2DS_60	Protein FEZ-like	BD	99	78	XP_003574551	NAC domain	DNA binding	MLOC_60079	2	17.55	Bradi3g37067	Os08g0436700	N	N	-
2DS_FA94FA8BC				Xaa-Pro dipeptidase	Tauschii	100	84	EMT28716	Peptidase M24, structural domain	aminopeptidase activity	MLOC_37956	6	176.76	Bradi3g08420	Os02g0224400	N	N	-
2DS_01082F3EE		2DS_5341322		Senescence-specific cysteine protease SAG39-like	BD	76	72	XP_003581047	Peptidase C1A	cysteine-type peptidase activity	-	-	-	Bradi5g03340	-	N	N	-
2DS_A181DDC76				Disease resistance protein RPM1	Tauschii	-	-	-	-	-	-	-	-	Bradi4g09577	-	N	N	-
2DS_F632F903F				Disease resistance protein RPM1	Tauschii	100	100	EMT03313	-	-	-	-	-	-	-	N	N	-
2DS_D264DB414				Dirigent protein 5-like	BD	87	68	XP_003562603	Plant disease resistance response protein	-	MLOC_18373	2	78.22	Bradi1g20185	Os07g0638500	N	N	-
2DS_FB2531322				Dirigent protein 22-like	BD	89	67	XP_003578856	Plant disease resistance response protein	-	MLOC_12325	3	544.03	Bradi4g41300	Os12g0174700	N	N	-
2DS_FDF701AA7				no annotation	-	-	-	-	Uncharacterised protein	-	-	-	-	Bradi2g13280	Os01g0389200	N	N	-
2DS_F3B9AEBCA				no annotation	-	-	-	-	-	methylation-dependent chromatin silencing	MLOC_59733	2	18.22	Bradi1g16097	Os03g0594700	N	N	-
2DS_5B2DCCCBD				no annotation	-	-	-	-	-	-	MLOC_43355	2	15.34	-	-	-	-	-
2DS_E298FC27B				no annotation	-	-	-	-	-	-	MLOC_58453	2	18.52	Bradi5g04710	Os04g0261400	N	N	-
2DS_85F3A47F8				no annotation	-	-	-	-	Nucleotide-diphospho-sugar transferase	-	MLOC_61444	2	609.30	Bradi4g34530	-	N	N	-
2DS_7392B6ED9				no annotation	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2DS_DD6E4E269				no annotation	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2DS_17C0B823A1				no annotation	-	-	-	-	-	-	-	-	-	-	-	-	-	-

A6.9.2

Wheat gene		Previously tested		BlastP NCBI Annotation	Species	% Query	% Id	GenBank Accession	InterPro	GO term name	HV gene	HV pos (Mb)		BD gene	Os gene	SNP ?	DEG?	Tauschii
2DS_6384C3FE1		2DS_5364496 2DS_5388088		G-type lectin S-receptor-like Ser/Thr-protein kinase RLK1	BD	98	98	XP_003581087	Protein kinase domain	ATP binding	MLOC_75639	2	1.66	Bradi5g02980	Os04g0202500	N	N	30.17
2DS_CA46F5E71 2DS_EC4BBF5CF	2			Ubiquitin thioesterase otubain-like isoform X3	BD	99	72	XP_010234388	Peptidase C65, otubain	-	MLOC_70393	2	16.07	Bradi3g16570	Os04g0652600	N	N	31.23
2DS_7ED349BCA1		yes	2DS_6	Dirigent protein 21-like	BD	99	64	XP_003577661	Plant disease resistance response protein	-	MLOC_62798	2	15.49	Bradi4g21260	Os11g0215100	N	N	30.22
2DS_5CE0A969D 2DS_48FB7EC2D	2	yes	2DS_88	UDP-glycosyltransferase 74E2	Urartu	99	86	EMS63799	UDP-glucuronosyl/ UDP-glucosyltransferase	transferase, transferring hexosyl grps	MLOC_24124	2	15.42	Bradi5g03400	Os04g0203600	N	N	30.45
2DS_44A7F70FB		yes	52i	ABC transporter C family member 14-like	BD	99	94	XP_003581195	ABC transporter-like	ATP binding	MLOC_5957	2	15.45	Bradi5g03460	Os04g0209200	N	N	30.99
2DS_A8B23EC52				ABC transporter G family member 24-like	BD	99	91	XP_003581016	ABC transporter-like	ATP binding	MLOC_52698	2	12.65	Bradi5g02870	Os04g0194500	Y	Y	27.23
2DS_62F241B7E		2DS_5371750		WD repeat-containing protein 25 isoform X1	BD	97	86	XP_003579310	WD40 repeat	Photoperiodism	MLOC_58466	2	17.56	Bradi5g04660	Os02g0319800	Y	N	35.66
2DS_FB16AB9A8		yes	2DS_86/7, 91/2	Peptidyl-prolyl cis-trans isomerase FKBP19	BD	98	99	XP_003581090	Peptidyl-prolyl cis-trans isomerase, FKBP-type	protein folding	MLOC_6969	2	15.51	Bradi5g02990	Os07g0133700	Y	N	30.17
2DS_3D2C53D93				RuBisCO large subunit-binding protein subunit alpha	Aestivum	99	99	P08823	Chaperonin Cpn60	cytoplasm	MLOC_51927	2	14.66	Bradi5g02890	Os12g0277500	N	N	27.41

A6.10 (continued across three pages)

Wheat gene models from CSS contigs anchored into the 18.1-33.1 cM bins by POPSEQ data. Gene models were extracted from *Ensembl*Plants and annotated. Genes which were duplicated (based on BLASTP results) are annotated in the 'Dup' column. Genes which were tested with markers are indicated. Shaded red are markers which were developed and were monomorphic between the parents to the fine-mapping population. Genes on contigs which had no variation are shaded red. The BLASTP Annotation is from the NCBI database, with the top hit used for the annotation in terms of peptide identity and accession number, using the peptides to the genes as queries. The species from which the annotation derives is indicated. Tauschii: *Ae. tauschii*; BD: Brachypodium; Urartu: *T. Urartu*; Aestivum: *T. aestivum*. The Interpro, GO annotations and synteny information were obtained from *Ensembl*Plants. The gene is also marked if it was in the differentially expressed genes (DEGs) in A6.8.1. Finally, if the Brachypodium gene could be anchored in the *Ae. tauschii* gene list/zipper in A6.7, the bin is indicated in the final column. Shaded red means a cM bin that was outside the *Rht8* interval.

Wheat gene	Previously tested		BlastP NCBI Annotation	Species	% Query	% Id	GenBank	InterPro	GO term name	HV gene	HV pos (Mb)		BD gene	OS gene	DEG?	Tau
2DS_5EDDE822C			Cys-rich receptor-like protein kinase 10	Tauschii	61	100	EMT21144	Bulb-type lectin domain	-	-	-	-	-	-	N	-
2DS_45F8CE226			F-box only protein 7	Urartu	72	93	EMS67273	F-box domain	protein binding	-	-	-	Bradi2g38850	Os03g0802100	short x1.59	3D
2DS_476DD6E63			no annotation	-	0	0	-	-	-	-	-	-	Bradi5g10350	-	N	107.4
2DS_6182EF594	yes	2DS_54	no annotation	-	0	0	-	-	-	-	-	-	Bradi5g02830	-	N	27.2
2DS_212E10376	yes	2DS_280	Brassinosteroid-regulated protein BRU1	Tauschii	95	56	EMT00455	-	-	-	-	-	-	-	N	-
2DS_0BFF3B23D			Cys-rich receptor-like protein kinase 6	Tauschii	100	100	EMT24053	Protein kinase domain	ATP binding	-	-	-	-	Os07g0301500	N	-
2DS_45F92FDF3	yes	2DS_76, 121-123, 124, 125	Ubiquitin thioesterase otubain-like isoform X1	BD	71	63	XP_003581187	Peptidase C65, otubain	-	MLOC_8298	2	22.47	Bradi5g09076	Os04g0652600	N	-
2DS_A57A221DE	2DS_5341122		Cytochrome P450 71D7	Tauschii	78	96	EMT07521	Cytochrome P450	iron ion binding	MLOC_21415	7	323.57	-	-	N	-
2DS_554CD5259	2DS_5292808		Wall-associated receptor kinase 3	Tauschii	52	99	EMT08949	EGF-type aspartate hydroxylation site	ATP binding	MLOC_48019	2	16.12	Bradi5g03577	Os04g0220300	N	40.9
2DS_625B3FD77	2DS_5319489		IAA-amino acid hydrolase ILR1-like protein 8	Tauschii	100	98	EMT05755	Peptidase M20	hydrolase activity	MLOC_37301	2	22.36	-	Os07g0249800	N	-
2DS_071650798	2DS_5342673		NAD-dependent deacetylase sirtuin-6	Tauschii	98	85	EMT21754	Sirtuin family	NAD+ binding	MLOC_295	2	22.53	Bradi5g02940	Os04g0271000	N	44.9
2DS_DE678DC0D			FBD-associated F-box protein At3g50710	BD	98	62	XP_010229605	-	-	MLOC_60987	2	22.24	-	-	N	-
2DS_2E5286F5D			Germin-like protein 8-14	Urartu	100	99	EMS51159	Germin	extracellular region	MLOC_55453	2	22.16	Bradi3g37680	Os08g0460000	N	-
2DS_3F5D36630	yes	2DS_8-10	Transcription factor RAX2-like	BD	100	67	XP_003581397	SANT/Myb domain	DNA binding	MLOC_5849	2	17.22	Bradi5g03882	-	N	112.4
2DS_B6AFE40D4	yes	2DS_7, 32-35	Glycerol-3-pho dehydrogenase SDP6, mitochondrial	Tauschii	100	100	EMT06126	FAD-dependent glycerol-3-pho dehydrogenase	glycerol-3-pho dehydrogenase activity	MLOC_11990	2	16.13	Bradi5g03810	Os04g0225001	N	39.3
2DS_05F454C37	yes	2DS_151-2	Thylakoid lumenal 15.0 kDa protein 2, chloroplastic	Tauschii	94	99	EMT07626	Vps54-like	retrograde transport, endosome to Golgi	MLOC_57508	2	15.72	Bradi5g03600	Os04g0212200	N	41.3
2DS_BEDDCAA0F	yes	2DS_26	S-norcoclaurine synthase 1	Tauschii	90	98	EMT07772	Oxoglutarate/iron-dependent dioxygenase	oxidoreductase activity	-	-	-	Bradi5g04340	Os03g0856000	N	37.3
2DS_DB37EBC5E			Acyl-carrier-desaturase, chloroplastic	Tauschii	100	99	EMT08285	Fatty acid desaturase, type 2	fatty acid metabolic process	MLOC_62967	2	22.21	Bradi3g17670	Os08g0199400	N	44.1
2DS_A7B538273			no annotation	-	-	-	-	-	-	MLOC_6116	2	568.49	-	-	N	-
2DS_7CF3C36BA			no annotation	-	-	-	-	F-box domain	protein binding	-	-	-	Bradi3g61000	Os02g0288925	N	3D
2DS_DE390194C	2DS_5358467		no annotation	-	-	-	-	-	-	MLOC_81817	2	18.23	Bradi5g04577	Os04g0221600	N	37.1

A6.10 (continued)

Wheat gene	Previously tested		BlastP NCBI Annotation	Species	% Query	% Id	GenBank	InterPro	GO term name	HV gene	HV pos (Mb)		BD gene	OS gene	DEG?	Tau
2DS_A3DACD282			Strictosidine synthase	Tauschii	97	84	EMT28724	Six-bladed beta-propeller, TolB-like	biosynthetic process	-	-	-	Bradi4g40305	Os08g0442200	N	5D
2DS_708D33A0C			ATP-dependent helicase	Tauschii	100	85	EMT33029	-	-	-	-	-	-	-	N	-
2DS_6C34B489A	2DS_5363769		INO80 complex subunit C	BD	100	90	XP_003581371	YL1 nuclear, C-terminal	regulation of transcription	MLOC_45846	2	16.21	Bradi5g03850	Os04g0274400	N	39.3
2DS_0D126BAD7	yes	2DS_40, 293-5	no annotation	-	-	-	-	His phosphatase superfamily, clade-1	-	MLOC_74610	2	17.29	Bradi5g03697	Os04g0224600	N	40.9
2DS_C4237A91B	2DS_5364388		Transketolase, chloroplastic	Tauschii	100	90	EMT08283	Transketolase, N-terminal	catalytic activity	MLOC_21709	2	22.64	Bradi5g07190	Os04g0266900	N	44.1
2DS_2F8D8BB67			Glutamate dehydrogenase	Aestivum	83	94	ADW95819	Glu/Leu/Val dehydrogenase, C-terminal	oxidoreductase activity	MLOC_69020	5	518.04	Bradi1g05680	Os03g0794500	N	5D
2DS_F37207649			Zn finger BED domain, RICESLEEPER 2-like	BD	67	82	XP_010237819	HAT dimerisation domain, C-terminal	nucleic acid binding	-	-	-	Bradi3g15846	-	N	7D
2DS_915D15B39	yes	2DS_36-7	LRR receptor-like Ser/Thr-protein kinase	Tauschii	99	94	EMT21841	Protein kinase domain	ATP binding	MLOC_61793	2	17.20	Bradi5g04000	Os04g0227200	N	38.6
2DS_E593E738D	yes	2DS_53	L-ascorbate peroxidase 3	Tauschii	100	100	EMT31421	Haem peroxidase	heme binding	MLOC_14804	2	16.17	Bradi5g03640	Os04g0223300	N	41.1
2DS_64BD8DBC0	2DS_5390981		LRR receptor-like Ser/Thr-protein kinase	Tauschii	100	99	EMT21840	LRR	protein binding			-	-	-	N	-
2DS_8993BF910	yes	2DS_150	GDT1-like protein 2, chloroplastic	BD	98	85	XP_010236049	Uncharacterised protein	membrane	MLOC_68294	2	15.58	Bradi5g03610	Os11g0544500	N	40.9
2DS_F04F1D341			Cys-rich receptor-like protein kinase 29	Tauschii	34	98	EMT16696	Protein kinase domain	ATP binding			-	-	Os11g0212900	N	-
2DS_E070E0B04	yes	2DS_58,59	no annotation	-	-	-	-		histone phosphorylation	MLOC_63016	2	17.25	Bradi5g04057	Os04g0228100	N	38.2
2DS_DF1680D79	yes	D_comp239028_c0 (hom)	Aspartic proteinase-like protein 1	Tauschii	100	99	EMT21844	Aspartic peptidase	aspartic-type endopeptidase	MLOC_63015	2	17.24	Bradi5g04050	Os04g0228000	N	38.2
2DS_F43C1EB35			Lipoamide acyltransferase	Tauschii	100	92	EMT06110	2-oxoacid dehydrogenase acyltransferase, catalytic	transferase activity, transferring acyl groups	MLOC_55450	2	22.15	Bradi2g11900	Os01g0314100	N	43.9
2DS_E29509E47	yes	2DS_288-9	Disease resistance protein RPM1	Tauschii	100	99	EMT11677	NB-ARC	ADP binding	MLOC_65574	2	18.31		Os10g0136100	N	-
2DS_3017E946B	2DS_5385535		Primary amine oxidase-like	BD	99	83	XP_003581494	Copper amine oxidase	quinone binding	MLOC_17390	2	22.29	Bradi5g04070	Os04g0269600	N	44.5
2DS_9DA0399BC			Obtusifoliol 14-alpha demethylase	Tauschii	80	91	EMT25993	Cytochrome P450	iron ion binding	MLOC_59386	1	10.93	Bradi1g24340	Os05g0211100	N	97.8
2DS_698DAD811	yes	2DS_47-50, 308-9	no annotation	-	-	-	-	Thioredoxin-like fold		MLOC_67319	2	17.42	Bradi5g04030	Os04g0227500	short x2.87	38.2
2DS_0D00D7A76			no annotation	-	-	-	-	PMR5 N-terminal domain				-		Os05g0587700	N	-
2DS_10E998B46			Protein trichome birefringence-like 3	BD	98	81	XP_003566882	PC-Esterase		MLOC_20162	2	22.59	Bradi2g43447	Os01g0614300	N	3D
2DS_53A082B2C	yes	D_comp6_c0 (hom)	RuBisCO small chain PW9, chloroplastic	Tauschii	100	100	EMT21846	RuBisCO domain	chloroplast	MLOC_21811	2	18.58	Bradi4g08500	Os12g0291200	N	5D

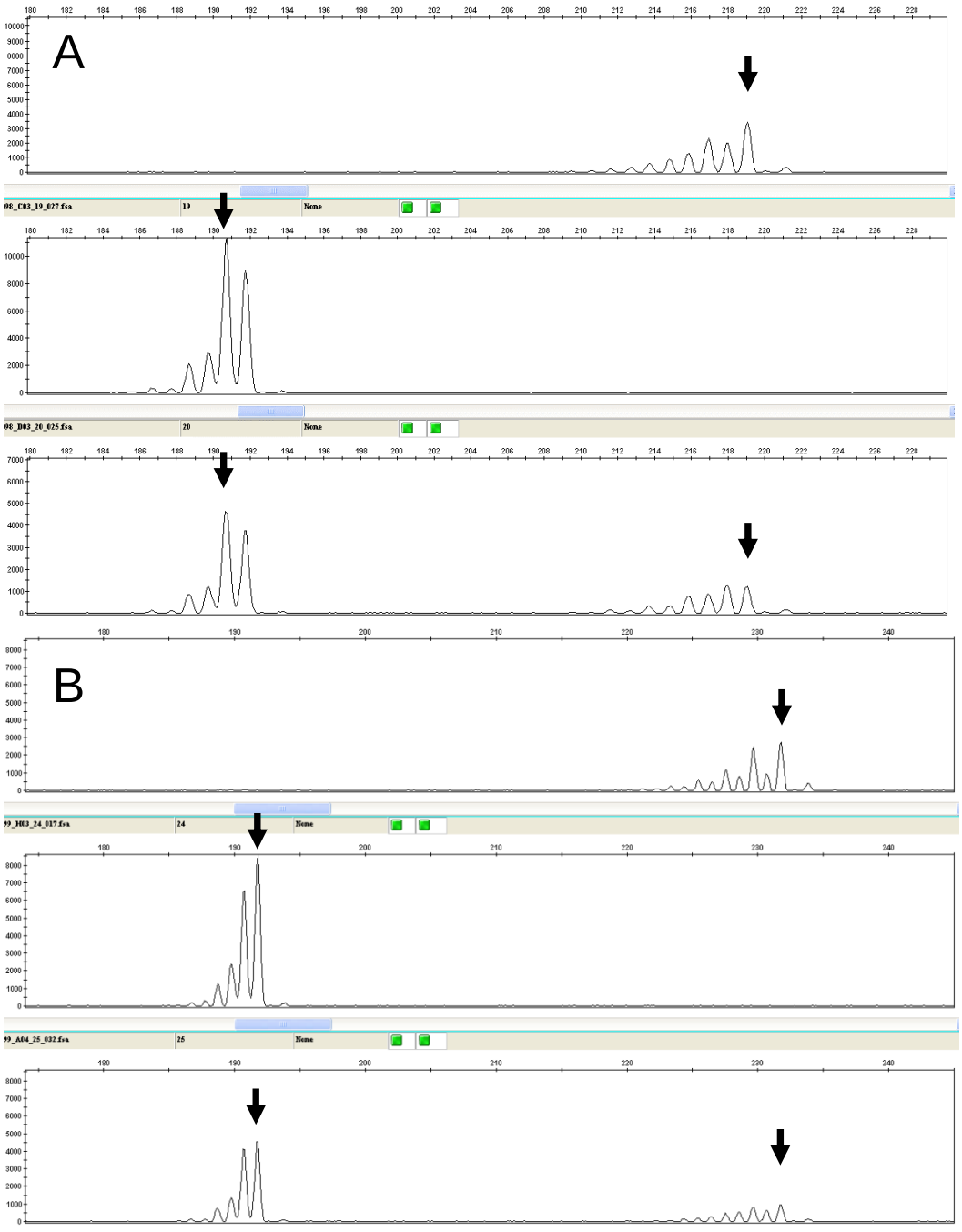
A6.10 (continued)

Wheat gene	Previously tested		BlastP NCBI Annotation	Species	% Query	% Id	GenBank	InterPro	GO term name	HV gene	HV pos (Mb)		BD gene	OS gene	DEG?	Tau
2DS_B28D63DC4			SWI/SNF-related regulator	Tauschii	100	99	EMT07824	SNF2-related	DNA binding	MLOC_71318	2	15.76	Bradi5g00770		N	12.9
2DS_032B64A46	2DS_5390396	D_comp244592_c0 (hom)	LRR receptor-like Ser/Thr-protein kinase	Tauschii	100	93	EMT30323	Protein kinase domain	ATP binding	MLOC_58539	2	16.03	Bradi5g03960		N	38.6
2DS_AB825FECC	2DS_5390725		AHM2 Cd/Zn transporting ATPase 2	Aestivum	67	92	BAA89308	HMG-I/HMG-Y, DNA-binding	DNA binding	MLOC_4181	2	22.42	Bradi5g05225	Os04g0266400	N	1D
2DS_46AF4C8FE	yes	2DS_15	Prostaglandin E synthase 2	Tauschii	88	91	EMT02260	Glutathione S-transferase	protein binding	MLOC_69463	2	18.34	Bradi5g04550	Os04g0244400	N	37.1
2DS_EE20139F8	yes	no	RING finger protein 165	Tauschii	100	99	EMT32136	Zn finger, RING-type	Zn ion binding	MLOC_23980	2	18.47	Bradi5g04540	Os04g0243700	N	37.1
2DS_4A0DEDDA9			mRNA-capping enzyme-like	BD	99	93	XP_003578804	Dual specificity phosphatase, catalytic domain	nucleus			-	Bradi4g40600	Os12g0193200	N	5D
2DS_A441EECD9	yes	no	40S ribosomal protein S6	Tauschii	98	99	EMT11004	Ribosomal protein S6e	structural constituent of ribosome	MLOC_56367	2	17.57	Bradi1g21320	Os07g0622100	N	90.9
2DS_EFFD49EF0	2DS_2728767		Cytochrome P450 71D7	Tauschii	89	99	EMT28726	Cytochrome P450	iron ion binding	-	-	-	-	-	N	-
2DS_8213F86B4			Nitrate transporter	Tauschii	82	97	EMT06111	Major facilitator superfamily domain	-	-	-	-	-	Os02g0112100	N	-
2DS_F9104B72E			FACT complex subunit SPT16	Tauschii	99	99	EMT18870	Peptidase M24, structural domain	-	MLOC_78789	2	16.11	Bradi1g59940	Os08g0404350	N	7D
2DS_875B9CA1A	2DS_2780239		Cinnamyl alcohol dehydrogenase 6	BD	61	85	XP_003581549	Alcohol dehydrogenase superfamily, Zn-type	Zn ion binding	MLOC_48245	2	17.58	Bradi5g04130	Os04g0229100	N	38.0
2DS_82C34F5AF			NADP-dependent oxidoreductase P1	Tauschii	100	91	EMT16695	Alcohol dehydrogenase superfamily, Zn-type	Zn ion binding	-	-	-	Bradi4g39980	Os12g0225900	short x2.27	39.5
2DS_795D042B8	yes	2DS_11, 43-4	Flavin-containing monooxygenase 1	Tauschii	100	99	EMT31048	Dimethylaniline monooxygenase	N,N-dimethylaniline monooxygenase activity	MLOC_63757	2	17.23	Bradi5g03710	Os04g0223901	N	40.9
2DS_072F43960			F-box/LRR-repeat protein 23	BD	99	62	XP_010234168	F-box domain	protein binding	-	-	-	Bradi3g10990	Os02g0317300	N	40.9
2DS_86F47AEE8			no annotation	-	-	-	-	Unknown function	-	-	-	-	-	Os12g0242200	N	-
2DS_4D393CDD0			Poly(A) polymerase	Tauschii	76	70	EMT04732	Nucleotidyl transferase domain	nucleus	-	-	-	-	-	N	-
2DS_67E848060			Chloroplast envelope membrane protein	Tauschii	50	97	YP_008474311	Chloroplast envelope membrane protein, CemA	integral component of membrane	-	-	-	-	-	N	-
2DS_86C131DEE	2DS_5339566		GTP-binding protein	Aestivum	100	99	ABF48401	Elongation factor, GTP-binding domain	GTPase activity	MLOC_70274	2	22.38	Bradi5g05200	Os04g0270100	N	44.6
2DS_7E4AADF01			O-glucosyltransferase 1-like	BD	97	89	XP_003569968	Lipopolysaccharide-modifying protein	-	-	-	-	Bradi3g50030	Os02g0642700	N	120.1
2DS_1208A147B			Rho GTPase-activating protein 22	Tauschii	100	91	EMT11006	CRIB domain	intracellular	MLOC_5009	2	14.77	Bradi1g52540	Os07g0408500	N	-
2DS_F2445CE9E	yes	D_comp42657_c0 (monomorphic)	G6PD 2, chloroplastic	Tauschii	100	99	EMT21843	G6PD	NADP binding	MLOC_18415	2	16.14	Bradi5g04020	Os07g0406300	N	38.4

Marker	Gene	Pos	Barley			Brachypodium		Rice	DEG?	Tauschii bin (cM)	IWGSC-2 bin (cM)
52i	mrna020368	929999..929511	MLOC_5957	2	15.61	Bradi5g03460	xenobiotic-transporting ATPase activity	Os04g13210	multidrug resistance-associated protein 4	N	30.99	17.34
	mrna070632	934335..937458	MLOC_37835	2	283.33	Bradi5g03510	single-stranded DNA specific 3'-5' exodeoxyribonuclease activity	Os04g23830	oligoribonuclease	N	31.41	-
	mrna093698	942301..940006	MLOC_54824	2	283.30	Bradi1g55690	tyrosine-tRNA ligase activity	Os04g23820	multisynthetase complex auxiliary component p43	N	-	-
	mrna084787	952651..950993	MLOC_48019	2	17.53	Bradi5g03580	-	-	-	N	-	-
	mrna139758	967144..975027	MLOC_57508	2	15.60	Bradi5g03600	-	Os04g13470	expressed protein	N	41.01	18.34 - 33.06
	mrna071578	980992..982656	MLOC_14804	2	17.59	Bradi5g03640	cytochrome-c peroxidase activity	Os04g14680	OsAPx3 - Peroxisomal Ascorbate Peroxidase	N	41.06	18.34 - 33.06
	mrna048555	987628..992477	MLOC_72300	-	-	-	-	-	-	-	-	-
	mrna091757	996752..998090	MLOC_78870	2	433.62	Bradi5g11360	-	Os04g36062	expressed protein	N	107.99	-
	mrna015009	1008566..1009383	MLOC_16798	2	17.49	Bradi3g20960	YUCCA-like flavin monooxygenase	Os04g14690	monooxygenase/ oxidoreductase	short x2.49	41.01	-
	mrna004763	1014979..1022844	MLOC_57069	2	293.51	Bradi5g03740	FAD binding	Os04g20990	tRNA-dihydrouridine synthase A	N	39.35	-
	mrna098230	1034572..1032265	MLOC_11990	2	15.66	Bradi5g03810	-	Os04g14790	glycerol-3-phosphate dehydrogenase, mitochondrial precursor	N	39.35	18.34 - 33.06
	mrna053306	-	-	2	17.55	Bradi5g03830	-	Os05g22970	-	N	40.38	-
	mrna035375	1052274..1048977	MLOC_45846	2	17.50	Bradi5g03850	-	Os04g20590	nucleus protein	N	39.35	18.34 - 33.06
	mrna002983	1067514..1055341	MLOC_52767	2	17.44	Bradi5g03860	ATP binding	Os05g22940	acetyl-coenzyme A carboxylase	N	39.3	-
	mrna096393	1078530..1075332	MLOC_58539	2	17.42	Bradi5g03960	leucine-rich repeat protein kinase, subfamily LRR-XII	Os04g15560	receptor kinase-like protein, putative	N	38.57	18.34 - 33.06
	mrna016294	1097699..1098161	MLOC_10026	2	17.64	-	-	-	-	-	-	-
	mrna106738	1109640..1110322	MLOC_61794	2	17.69	Bradi5g03980	-	-	-	N	-	-
	mrna096003	1117860..1116717	MLOC_61793	2	17.40	Bradi5g04000	leucine-rich repeat protein kinase, subfamily LRR-XII	Os04g15660	receptor-like protein kinase 5 precursor	N	38.57	18.34 - 33.06
	mrna124385	1128085..1128538	MLOC_63016	2	17.70	Bradi5g04060	-	Os04g15800	expressed protein	N	-	-
	mrna118007	1133627..1133074	MLOC_21811	2	17.60	Bradi4g08500	RuBisCO activity	Os12g17600	RuBisCO small chain C	N	5D	18.34 - 33.06
	mrna105093	1141782..1137361	MLOC_64679	5	308.97	Bradi4g08800	RuBisCO activity	Os12g17600	RuBisCO small chain C	N	38.16	-
2DS_26	mrna057813	1178953..1175902	MLOC_10084	2	19.08	Bradi4g40600	protein tyrosine/serine/threonine phosphatase activity	Os12g09120	mRNA capping enzyme	N	5D	18.34 - 33.06

A6.11: Delimited interval in v3.3 cDNAs. The markers used to delimit the interval are highlighted in grey. The gene is marked if it was in the differentially expressed genes (DEGs) in A6.8.1. If the Brachypodium gene could be anchored in the *Ae. tauschii* gene list/zipper in A6.7, the bin is indicated. Shaded red are cM bins outside the *Rht8* interval. Shaded green is a cM bin within the defined interval in *Ae. tauschii*. The final column indicates if the Brachypodium gene was in the 17.34 cM or 18.3 – 33.1 cM bin in the IWGSC data, shown in A6.9 and A6.10. Shaded green is the 17.3 cM bin within the defined *Rht8* interval in IWGSC-2. Shaded red is outside the IWGSC-2 *Rht8* interval.

Appendix to Chapter 7



A7.1: Allelic variation at *Xgwm261*. The arrows show the calculated size of each DNA fragment. (A) Maringa x Mercia 4-2: top = donor (Maringa, 219-bp), middle = parent (Mercia, 191-bp), bottom = heterozygous. (B) Klein 157 x Mercia 4-4: top = donor (Klein 157, 231-bp), middle = parent (Mercia, 191-bp), bottom = heterozygous.

References

Abramoff, M. D., Magalhaes, P. J. & Ram, S. J. 2004. Image Processing with ImageJ. *Biophotonics International,* 11**,** 36-42.

Acreche, M. M. & Slafer, G. A. 2006. Grain weight response to increases in number of grains in wheat in a Mediterranean area. *Field Crops Research,* 98**,** 52-59.

Acreche, M. M. & Slafer, G. A. 2011. Lodging yield penalties as affected by breeding in Mediterranean wheats. *Field Crops Research,* 122**,** 40-48.

Adamski, N. M., Bush, M. S., Simmonds, J., Turner, A. S., Mugford, S. G., Jones, A., Findlay, K., Pedentchouk, N., Von Wettstein-Knowles, P. & Uauy, C. 2013. The Inhibitor of wax 1 locus (Iw1) prevents formation of - and OH--diketones in wheat cuticular waxes and maps to a sub-cM interval on chromosome arm 2BS. *Plant Journal,* 74**,** 989-1002.

Addisu, M., Snape, J. W., Simmonds, J. R. & Gooding, M. J. 2009a. Effects of reduced height (Rht) and photoperiod insensitivity (Ppd) alleles on yield of wheat in contrasting production systems. *Euphytica,* 172**,** 169-181.

Addisu, M., Snape, J. W., Simmonds, J. R. & Gooding, M. J. 2009b. Reduced height (Rht) and photoperiod insensitivity (Ppd) allele associations with establishment and early growth of wheat in contrasting production systems. *Euphytica,* 166**,** 249-267.

Ahmad, M. & Sorrells, M. E. 2002. Distribution of microsatellite alleles linked to Rht8 dwarfing gene in wheat. *Euphytica,* 123**,** 235-240.

Akpinar, B. A., Lucas, S. J., Vrana, J., Dolezel, J. & Budak, H. 2015. Sequencing chromosome 5D of Aegilops tauschii and comparison with its allopolyploid descendant bread wheat (Triticum aestivum). *Plant Biotechnology Journal,* 13**,** 740-752.

Alghabari, F., Lukac, M., Jones, H. E. & Gooding, M. J. 2014. Effect of Rht Alleles on the Tolerance of Wheat Grain Set to High Temperature and Drought Stress During Booting and Anthesis. *Journal of Agronomy and Crop Science,* 200**,** 36-45.

Allen, A. M., Barker, G. L. A., Berry, S. T., Coghill, J. A., Gwilliam, R., Kirby, S., Robinson, P., Brenchley, R. C., D'amore, R., Mckenzie, N., Waite, D., Hall, A., Bevan, M., Hall, N. & Edwards, K. J. 2011. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum aestivum L.). *Plant Biotechnology Journal,* 9**,** 1086-1099.

Allen, A. M., Barker, G. L. A., Wilkinson, P., Burridge, A., Winfield, M., Coghill, J., Uauy, C., Griffiths, S., Jack, P., Berry, S., Werner, P., Melichar, J. P. E., Mcdougall, J., Gwilliam, R., Robinson, P. & Edwards, K. J. 2013. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (Triticum aestivum L.). *Plant Biotechnology Journal,* 11**,** 279-295.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W. & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389-3402.

Ariyadasa, R., Mascher, M., Nussbaumer, T., Schulte, D., Frenkel, Z., Poursarebani, N., Zhou, R., Steuernagel, B., Gundlach, H., Taudien, S., Felder, M., Platzer, M., Himmelbach, A., Schmutzer, T., Hedley, P. E., Muehlbauer, G. J., Scholz, U., Korol, A., Mayer, K. F. X., Waugh, R., Langridge, P., Graner, A. & Stein, N. 2014. A Sequence-Ready Physical Map of Barley Anchored Genetically by Two Million Single-Nucleotide Polymorphisms. *Plant Physiology,* 164**,** 412-423.

Asplund, L., Leino, M. W. & Hagenblad, J. 2012. Allelic variation at the Rht8 locus in a 19th century wheat collection. *The Scientific World Journal,* 2012**,** #385610.

Asseng, S., Foster, I. & Turner, N. C. 2011. The impact of temperature variability on wheat yields. *Global Change Biology,* 17**,** 997-1012.

Austin, R. S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N. J., Bonetta, D., Zhang, J., Fung, P., Gong, Y., Wang, P. W., Mccourt, P. & Guttman, D. S. 2011. Next-generation mapping of Arabidopsis genes. *Plant Journal,* 67**,** 715-725.

Babiker, E. M., Gordon, T. C., Chao, S., Newcomb, M., Rouse, M. N., Jin, Y., Wanyera, R., Acevedo, M., Brown-Guedira, G., Williamson, S. & Bonman, J. M. 2015. Mapping resistance to the Ug99 race group of the stem rust pathogen in a spring wheat landrace. *Theoretical and Applied Genetics,* 128**,** 605-612.

Badaeva, E. D., Dedkova, O. S., Gay, G., Pukhalskyi, V. A., Zelenin, A. V., Bernard, S. & Bernard, M. 2007. Chromosomal rearrangements in wheat: their types and distribution. *Genome,* 50**,** 907-926.

Bai, G. H., Das, M. K., Carver, B. F., Xu, X. Y. & Krenzer, E. G. 2004. Covariation for microsatellite marker alleles associated with Rht8 and coleoptile length in winter wheat. *Crop Science,* 44**,** 1187-1194.

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S. & Werck-Reichhart, D. 2011. Cytochromes P450. *The Arabidopsis Book / American Society of Plant Biologists,* 9**,** e0144.

Bakshi, S. & Bhagwat, S. G. 2012. Allelic Variations at Xgwm261 Locus, Sequence Determination and Agronomic Evaluation in Indian Bread Wheat Genotypes. *Cereal Research Communications,* 40**,** 34-43.

Barraclough, P. B., Howarth, J. R., Jones, J., Lopez-Bellido, R., Parmar, S., Shepherd, C. E. & Hawkesford, M. J. 2010. Nitrogen efficiency of wheat: Genotypic and environmental variation and prospects for improvement. *European Journal of Agronomy,* 33**,** 1-11.

Barrett, B. A. & Kidwell, K. K. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Science,* 38**,** 1261-1271.

Beales, J., Turner, A., Griffiths, S., Snape, J. W. & Laurie, D. A. 2007. A pseudo-response regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik,* 115**,** 721-33.

Bennett, D., Izanloo, A., Edwards, J., Kuchel, H., Chalmers, K., Tester, M., Reynolds, M., Schnurbusch, T. & Langridge, P. 2012. Identification of novel quantitative trait loci for days to ear emergence and flag leaf glaucousness in a bread wheat (Triticum aestivum L.) population adapted to southern Australian conditions. *Theoretical and Applied Genetics,* 124**,** 697-711.

Berkman, P. J., Lai, K., Lorenc, M. T. & Edwards, D. 2012a. Next-generation sequencing applications for wheat crop improvement. *American Journal of Botany,* 99**,** 365-371.

Berkman, P. J., Skarshewski, A., Manoli, S., Lorenc, M. T., Stiller, J., Smits, L., Lai, K., Campbell, E., Kubalakova, M., Simkova, H., Batley, J., Dolezel, J., Hernandez, P. & Edwards, D. 2012b. Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. *Theoretical and Applied Genetics,* 124**,** 423-432.

Berry, P. M. & Spink, J. 2012. Predicting yield losses caused by lodging in wheat. *Field Crops Research,* 137**,** 19-26.

Berry, P. M., Spink, J. H., Gay, A. P. & Craigon, J. 2003. A comparison of root and stem lodging risks among winter wheat cultivars. *Journal of Agricultural Science,* 141**,** 191-202.

Berry, P. M., Sterling, M., Spink, J. H., Baker, C. J., Sylvester-Bradley, R., Mooney, S. J., Tams, A. R. & Ennos, A. R. 2004. Understanding and reducing lodging in cereals. *Advances in Agronomy, Vol 84,* 84**,** 217-271.

Berry, P. M., Sylvester-Bradley, R. & Berry, S. 2007. Ideotype design for lodging-resistant wheat. *Euphytica,* 154**,** 165-179.

Biesiekierski, J. R. & Iven, J. 2015. Non-coeliac gluten sensitivity: piecing the puzzle together. *United European Gastroenterology Journal,* 3**,** 160-165.

Bivand, R. & Lewin-Koh, N. 2015. maptools: Tools for Reading and Handling Spatial Objects. . R package version 0.8-36 ed.

Bolser, D. M., Kerhornou, A., Walts, B. & Kersey, P. 2015. Triticeae Resources in Ensembl Plants. *Plant and Cell Physiology,* 56.

Borner, A., Worland, A. J., Plaschke, J., Schumann, E. & Law, C. N. 1993. Pleiotropic effects of genes for reduced height (rht) and day-length insensitivity (ppd) on yield and its components for wheat grown in middle europe. *Plant Breeding,* 111**,** 204-216.

Borojevic, K. & Borojevic, K. 2005. The transfer and history of "reduced height genes" (Rht) in wheat from Japan to Europe. *The Journal of heredity,* 96**,** 455-9.

Borrill, P. 2014. *The NAM-B1 transcription factor and the control of grain composition in wheat.*

Borrill, P., Adamski, N. & Uauy, C. 2015. Genomics as the key to unlocking the polyploid potential of wheat. *New Phytologist***,** 4, 1008-22.

Bottley, A., Xia, G. M. & Koebner, R. M. D. 2006. Homoeologous gene silencing in hexaploid wheat. *Plant Journal,* 47**,** 897-906.

Botwright, T. L., Rebetzke, G. J., Condon, A. G. & Richards, R. A. 2005. Influence of the gibberellin-sensitive Rht8 dwarfing gene on leaf epidermal cell dimensions and early vigour in wheat (Triticum aestivum L.). *Annals of botany,* 95**,** 631-9.

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

Brouns, F. J. P. H., Van Buul, V. J. & Shewry, P. R. 2013. Does wheat make us fat and sick? *Journal of Cereal Science,* 58**,** 209-215.

Brown, P. R., Singleton, G. R., Tann, C. R. & Mock, I. 2003. Increasing sowing depth to reduce mouse damage to winter crops. *Crop Protection,* 22**,** 653-660.

Bucksch, A., Burridge, J., York, L. M., Das, A., Nord, E., Weitz, J. S. & Lynch, J. P. 2014. Image-Based High-Throughput Field Phenotyping of Crop Roots. *Plant Physiology,* 166**,** 470-486.

Burney, J. A., Davis, S. J. & Lobell, D. B. 2010. Greenhouse gas mitigation by agricultural intensification. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 12052-12057.

Busemeyer, L., Mentrup, D., Moeller, K., Wunder, E., Alheit, K., Hahn, V., Maurer, H. P., Reif, J. C., Wuerschum, T., Mueller, J., Rahe, F. & Ruckelshausen, A. 2013. BreedVision - A Multi-Sensor Platform for Non-Destructive Field-Based Phenotyping in Plant Breeding. *Sensors,* 13**,** 2830-2847.

Butler, J. D., Byrne, P. F., Mohammadi, V., Chapman, P. L. & Haley, S. D. 2005. Agronomic performance of Rht alleles in a spring wheat population across a range of moisture levels. *Crop Science,* 45**,** 939-947.

Cabello-Hurtado, F., Zimmerlin, A., Rahier, A., Taton, M., Derose, R., Nedelkina, S., Batard, Y., Durst, F., Pallett, K. E. & Werckreichhart, D. 1997. Cloning and functional expression in yeast of a cDNA coding for an obtusifoliol 14 alpha-demethylase (CYP51) in wheat. *Biochemical and Biophysical Research Communications,* 230**,** 381-385.

Casal, J. J., Sanchez, R. A. & Deregibus, V. A. 1987. Tillering responses of lolium-multiflorum plants to changes of red far-red ratio typical of sparse canopies. *Journal of Experimental Botany,* 38**,** 1432-1439.

Casson, S. A. & Hetherington, A. M. 2012. GSK3-Like Kinases Integrate Brassinosteroid Signaling and Stomatal Development. *Science Signaling,* 5.

Cattivelli, L., Rizza, F., Badeck, F.-W., Mazzucotelli, E., Mastrangelo, A. M., Francia, E., Mare, C., Tondelli, A. & Stanca, A. M. 2008. Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Research,* 105**,** 1-14.

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G. L., Akhunova, A., See, D., Bai, G., Pumphrey, M., Tomar, L., Wong, D., Kong, S., Reynolds, M., Da Silva, M. L., Bockelman, H., Talbert, L., Anderson,

J. A., Dreisigacker, S., Baenziger, S., Carter, A., Korzun, V., Morrell, P. L., Dubcovsky, J., Morell, M. K., Sorrells, M. E., Hayden, M. J. & Akhunov, E. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 8057-8062.

Cecilia Ugarte, C., Ariel Trupkin, S., Ghiglione, H., Slafer, G. & Jose Casal, J. 2010. Low red/far-red ratios delay spike and stem growth in wheat. *Journal of Experimental Botany,* 61**,** 3151-3162.

CerealsDB. 2015a. Available: http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/blast_WGS.php.

Cerealsdb. 2015b. *820K Axiom® Array* [Online]. Available: http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/axiom_820K_search.php.

Chao, S., Sharp, P. J., Worland, A. J., Warham, E. J., Koebner, R. M. D. & Gale, M. D. 1989. RFLP-based genetic maps of wheat homologous group-7 chromosomes. *Theoretical and Applied Genetics,* 78**,** 495-504.

Chapman, J. A., Mascher, M., Buluc, A., Barry, K., Georganas, E., Session, A., Strnadova, V., Jenkins, J., Sehgal, S., Oliker, L., Schmutz, J., Yelick, K. A., Scholz, U., Waugh, R., Poland, J. A., Muehlbauer, G. J., Stein, N. & Rokhsar, D. S. 2015. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology,* 16**,** 17.

Chapman, S. C., Mathews, K. L., Trethowan, R. M. & Singh, R. P. 2007. Relationships between height and yield in near-isogenic spring wheats that contrast for major reduced height genes. *Euphytica,* 157**,** 391-397.

Charmet, G. 2011. Wheat domestication: Lessons for the future. *Comptes Rendus Biologies,* 334**,** 212-220.

Chauveau, P., Fouque, D., Combe, C. & Aparicio, M. 2013. Evolution of the diet from the paleolithic to today: Progress or regress? *Nephrologie & Therapeutique,* 9**,** 202-208.

Chebotar, S. V., Korzun, V. N. & Sivolap, Y. M. 2001. Allele distribution at locus WMS261 marking the dwarfing gene Rht8 in common wheat cultivars of southern Ukraine. *Russian Journal of Genetics,* 37**,** 894-898.

Chen, L., Hao, L., Condon, A. G. & Hu, Y.-G. 2014. Exogenous GA3 Application Can Compensate the Morphogenetic Effects of the GA-Responsive Dwarfing Gene Rht12 in Bread Wheat. *Plos One,* 9.

Chen, L., Phillips, A. L., Condon, A. G., Parry, M. a. J. & Hu, Y.-G. 2013. GA-Responsive Dwarfing Gene Rht12 Affects the Developmental and Agronomic Traits in Common Bread Wheat. *Plos One,* 8.

Chen, S. L., Gao, R. H., Wang, H. Y., Wen, M. X., Xiao, J., Bian, N. F., Zhang, R. Q., Hu, W. J., Cheng, S. H., Bie, T. D. & Wang, X. E. 2015. Characterization of a novel reduced height gene (Rht23) regulating panicle morphology and plant architecture in bread wheat. *Euphytica,* 203**,** 583-594.

Chevalier, D. & Walker, J. C. 2005. Functional genomics of protein kinases in plants. *Briefings in Functional Genomics & Proteomics,* 3**,** 362-371.

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., Pingault, L., Sourdille, P., Couloux, A., Paux, E., Leroy, P., Mangenot, S., Guilhot, N., Le Gouis, J., Balfourier, F., Alaux, M., Jamilloux, V., Poulain, J., Durand, C., Bellec, A., Gaspin, C., Safar, J., Dolezel, J., Rogers, J., Vandepoele, K., Aury, J.-M., Mayer, K., Berges, H., Quesneville, H., Wincker, P. & Feuillet, C. 2014. Structural and functional partitioning of bread wheat chromosome 3B. *Science,* 345.

Cloutier, S., Mccallum, B. D., Loutre, C., Banks, T. W., Wicker, T., Feuillet, C., Keller, B. & Jordan, M. C. 2007. Leaf rust resistance gene Lr1, isolated from bread wheat (Triticum aestivum L.) is a member of the large psr567 gene family. *Plant Molecular Biology,* 65**,** 93-106.

Colmsee, C., Beier, S., Himmelbach, A., Schmutzer, T., Stein, N., Scholz, U. & Mascher, M. 2015. BARLEX - the Barley Draft Genome Explorer. *Molecular Plant,* 8**,** 964-966.

Crook, M. J. & Ennos, A. R. 1995. The effect of nitrogen and growth-regulators on stem and root characteristics associated with lodging in 2 cultivars of winter-wheat. *Journal of Experimental Botany,* 46**,** 931-938.

Cui, F., Ding, A., Li, J., Zhao, C., Wang, L., Wang, X., Qi, X., Li, X., Li, G., Gao, J. & Wang, H. 2012. QTL detection of seven spike-related traits and their genetic correlations in wheat using two related RIL populations. *Euphytica,* 186**,** 177-192.

Daoura, B. G., Chen, L., Du, Y. Y. & Hu, Y. G. 2014. Genetic effects of dwarfing gene Rht-5 on agronomic traits in common wheat (Triticum aestivum L.) and QTL analysis on its linked traits. *Field Crops Research,* 156**,** 22-29.

David, C., Abecassis, J., Carcea, M., Celette, F., Friedel, J. K., Hellou, G., Hiltbrunner, J., Messmer, M., Narducci, V., Peigné, J., Samson, M. F., Schweinzer, A., Thomsen, I. K. & Thommen, A. 2012. Organic Bread Wheat Production and Market in Europe. *In:* LICHTFOUSE, E. (ed.) *Sustainable Agriculture Reviews.* Springer Netherlands.

De Punder, K. & Pruimboom, L. 2013. The Dietary Intake of Wheat and other Cereal Grains and Their Role in Inflammation. *Nutrients,* 5**,** 771-787.

Dean, J. V. & Delaney, S. P. 2008. Metabolism of salicylic acid in wild-type, ugt74f1 and ugt74f2 glucosyltransferase mutants of Arabidopsis thaliana. *Physiologia Plantarum,* 132**,** 417-425.

DEFRA 2015. Agriculture in the United Kingdom 2014.

Deng, W., Nickle, D. C., Learn, G. H., Maust, B. & Mullins, J. I. 2007. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics,* 23**,** 2334-2336.

Devos, K. M., Dubcovsky, J., Dvorak, J., Chinoy, C. N. & Gale, M. D. 1995. Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theoretical and Applied Genetics,* 91**,** 282-288.

Devos, K. M. & Gale, M. D. 1992. The use of random amplified polymorphic DNA markers in wheat. *Theoretical and Applied Genetics,* 84**,** 567-572.

Doebley, J. F., Gaut, B. S. & Smith, B. D. 2006. The molecular genetics of crop domestication. *Cell,* 127**,** 1309-1321.

Dolezel, J., Vrana, J., Capal, P., Kubalakova, M., Buresova, V. & Simkova, H. 2014. Advances in plant chromosome genomics. *Biotechnology Advances,* 32**,** 122-136.

Downie, H. F., Adu, M. O., Schmidt, S., Otten, W., Dupuy, L. X., White, P. J. & Valentine, T. A. 2015. Challenges and opportunities for quantifying roots and rhizosphere interactions through imaging and image analysis. *Plant Cell and Environment,* 38**,** 1213-1232.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. & Su, Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research,* 38**,** W64-W70.

Duan, J., Xia, C., Zhao, G., Jia, J. & Kong, X. 2012. Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. *Bmc Genomics,* 13.

Dubcovsky, J. & Dvorak, J. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science,* 316**,** 1862-1866.

Duran, C., Edwards, D. & Batley, J. 2009. Molecular Marker Discovery and Genetic Map Visualisation. *In:* DAVID EDWARDS, J. S., DAVID HANSEN (ed.) *Bioinformatics: Tools and Applications.* 2009 ed.: Springer.

Dvorak, J., Akhunov, E. D., Akhunov, A. R., Deal, K. R. & Luo, M. C. 2006. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Molecular Biology and Evolution,* 23**,** 1386-1396.

Dvorak, J., Luo, M. C., Yang, Z. L. & Zhang, H. B. 1998. The structure of the Aegilops tauschii genepool and the evolution of hexaploid wheat. *Theoretical and Applied Genetics,* 97**,** 657-670.

Ellis, M. H., Bonnett, D. G. & Rebetzke, G. J. 2007. A 192bp allele at the Xgwm261 locus is not always associated with the Rht8 dwarfing gene in wheat (Triticum aestivum L.). *Euphytica,* 157**,** 209-214.

Ellis, M. H., Rebetzke, G. J., Chandler, P., Bonnett, D., Spielmeyer, W. & Richards, R. A. 2004. The effect of different height reducing genes on the early growth of wheat. *Functional Plant Biology,* 31**,** 583-589.

Ensemblplants. 2015. Ensembl Plants. Available: http://plants.ensembl.org/index.html.

Evans, L. T. 1998. Feeding the Ten Billion. Plant and Population Growth.

Food and Agriculture Organisation of the United Nations (FAO) 2009. FAO's Director-General on How to Feed the World in 2050. *Population and Development Review,* 35**,** 837-839.

Food and Agriculture Organisation of the United Nations (FAO) 2012. *http://www.fao.org/corp/statistics/en/* [Online].

Faris, J. D., Fellers, J. P., Brooks, S. A. & Gill, B. S. 2003. A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. *Genetics,* 164**,** 311-321.

Faris, J. D. & Gill, B. S. 2002. Genomic targeting and high-resolution mapping of the domestication gene Q in wheat. *Genome,* 45**,** 706-718.

Faris, J. D., Zhang, Q. J., Chao, S. M., Zhang, Z. C. & Xu, S. S. 2014a. Analysis of agronomic and domestication traits in a durum x cultivated emmer wheat population using a high-density single nucleotide polymorphism-based linkage map. *Theoretical and Applied Genetics,* 127**,** 2333-2348.

Faris, J. D., Zhang, Z., Lu, H., Lu, S., Reddy, L., Cloutier, S., Fellers, J. P., Meinhardt, S. W., Rasmussen, J. B., Xu, S. S., Oliver, R. P., Simons, K. J. & Friesen, T. L. 2010. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 13544-13549.

Faris, J. D., Zhang, Z. C. & Chao, S. M. 2014b. Map-based analysis of the tenacious glume gene Tg-B1 of wild emmer and its role in wheat domestication. *Gene,* 542**,** 198-208.

Faris, J. D., Zhang, Z. C., Garvin, D. F. & Xu, S. S. 2014c. Molecular and comparative mapping of genes governing spike compactness from wild emmer wheat. *Molecular Genetics and Genomics,* 289**,** 641-651.

Feldman, M. 2001. Origin of culitvated wheat. *In:* BONJEAN A.P. & J., A. W. (eds.) *The world wheat book: a history of wheat breeding.* Paris, France: Lavoisier Publishing.

Ferrante, A., Savin, R. & Slafer, G. A. 2013. Floret development and grain setting differences between modern durum wheats under contrasting nitrogen availability. *Journal of Experimental Botany,* 64**,** 169-184.

Feuillet, C., Travella, S., Stein, N., Albar, L., Nublat, A. & Keller, B. 2003. Map-based isolation of the leaf rust disease resistance gene Lr10 from the hexaploid wheat (Triticum aestivum L.) genome. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 15253-15258.

Fischer, R. A. & Quail, K. J. 1990. The effect of major dwarfing genes on yield potential in spring wheats. *Euphytica,* 46**,** 51-56.

Fischer, R. A. & Stapper, M. 1987. Lodging effects on high-yielding crops of irrigated semidwarf wheat. *Field Crops Research,* 17**,** 245-258.

Fleet, C. M. & Sun, T. P. 2005. A DELLAcate balance: the role of gibberellin in plant morphogenesis. *Current Opinion in Plant Biology,* 8**,** 77-85.

Flintham, J. E., Borner, A., Worland, A. J. & Gale, M. D. 1997. Optimizing wheat grain yield: Effects of Rht (gibberellin-insensitive) dwarfing genes. *Journal of Agricultural Science,* 128**,** 11-25.

Foulkes, M. J., Hawkesford, M. J., Barraclough, P. B., Holdsworth, M. J., Kerr, S., Kightley, S. & Shewry, P. R. 2009. Identifying traits to improve the nitrogen economy of wheat: Recent advances and future prospects. *Field Crops Research,* 114**,** 329-342.

Fox, S. E., Geniza, M., Hanumappa, M., Naithani, S., Sullivan, C., Preece, J., Tiwari, V. K., Elser, J., Leonard, J. M., Sage, A., Gresham, C., Kerhornou, A., Bolser, D., Mccarthy, F., Kersey, P., Lazo, G. R. & Jaiswal, P. 2014. De Novo Transcriptome Assembly and Analyses of

Gene Expression during Photomorphogenesis in Diploid Wheat Triticum monococcum. *Plos One,* 9.

Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., Sela, H., Fahima, T. & Dubcovsky, J. 2009. A Kinase-START Gene Confers Temperature-Dependent Resistance to Wheat Stripe Rust. *Science,* 323**,** 1357-1360.

Gale, M. D. & Devos, K. M. 1998. Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 1971-1974.

Gale, M. D. & Marshall, G. A. 1976. Chromosomal location of gai-1 and rht-1, genes for gibberellin insensitivity and semi-dwarfism, in a derivative of Norin-10 wheat. *Heredity,* 37**,** 283-289.

Gale, M. D. & Youssefian, S. 1985. *Dwarfing genes in wheat,* London, Butterworth Co.

Ganeva, G., Korzun, V., Landjeva, S., Tsenov, N. & Atanasova, M. 2005. Identification, distribution and effects on agronomic traits of the semi-dwarfing Rht alleles in Bulgarian common wheat cultivars. *Euphytica,* 145**,** 305-315.

Gasperini, D. 2010. *Genetic and Physiological Characterisation of Rht8 in Hexaploid Wheat.*

Gasperini, D., Greenland, A., Hedden, P., Dreos, R., Harwood, W. & Griffiths, S. 2012. Genetic and physiologcial analysis of Rht8 in bread wheat: an alternative source of semi-dwarfism with a reduced sensitivty to brassinosteroids. *Journal of experimental botany,* 63**,** 4419-4436.

Gegas, V. C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L., Orford, S., Sayers, L., Doonan, J. H. & Snape, J. W. 2010. A Genetic Framework for Grain Size and Shape Variation in Wheat. *Plant Cell,* 22**,** 1046-1056.

Gilissen, L. J. W. J., Van Der Meer, I. M. & Smulders, M. J. M. 2014. Reducing the incidence of allergy and intolerance to cereals. *Journal of Cereal Science,* 59**,** 337-353.

Gilsinger, J., Kong, L., Shen, X. & Ohm, H. 2005. DNA markers associated with low Fusarium head blight incidence and narrow flower opening in wheat. *Theoretical and Applied Genetics,* 110**,** 1218-1225.

Giorgi, D., Farina, A., Grosso, V., Gennaro, A., Ceoloni, C. & Lucretti, S. 2013. FISHIS: Fluorescence In Situ Hybridization in Suspension and Chromosome Flow Sorting Made Easy. *Plos One,* 8.

Girin, T., David, L. C., Chardin, C., Sibout, R., Krapp, A., Ferrario-Mery, S. & Daniel-Vedele, F. 2014. Brachypodium: a promising hub between model species and cereals. *Journal of Experimental Botany,* 65**,** 5683-5696.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. & Toulmin, C. 2010. Food Security: The Challenge of Feeding 9 Billion People. *Science,* 327**,** 812-818.

Goncharov, N. P. & Gaidalenok, R. F. 2005. Localization of genes controlling spherical grain and compact ear in Triticum antiquorum Heer ex Udacz. *Russian Journal of Genetics,* 41**,** 1262-1267.

Gooding, M. J., Addisu, M., Uppal, R. K., Snape, J. W. & Jones, H. E. 2012. Effect of wheat dwarfing genes on nitrogen-use efficiency. *Journal of Agricultural Science,* 150**,** 3-22.

Gower, B. A. & Goss, A. M. 2015. A Lower-Carbohydrate, Higher-Fat Diet Reduces Abdominal and Intermuscular Fat and Increases Insulin Sensitivity in Adults at Risk of Type 2 Diabetes. *Journal of Nutrition,* 145**,** 177-183.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology,* 29**,** 644-U130.

Griffiths, S., Sharp, R., Foote, T. N., Bertin, I., Wanous, M., Reader, S., Colas, I. & Moore, G. 2006. Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature,* 439**,** 749-752.

Griffiths, S., Simmonds, J., Leverington, M., Wang, Y., Fish, L., Sayers, L., Alibert, L., Orford, S., Wingen, L. & Snape, J. 2012. Meta-QTL analysis of the genetic control of crop height in elite European winter wheat germplasm. *Molecular Breeding,* 29**,** 159-171.

Griffiths, S., Wingen, L., Pietragalla, J., Garcia, G., Hasan, A., Miralles, D., Calderini, D. F., Ankleshwaria, J. B., Waite, M. L., Simmonds, J., Snape, J. & Reynolds, M. 2015. Genetic Dissection of Grain Size and Grain Number Trade-Offs in CIMMYT Wheat Germplasm. *Plos One,* 10.

GRIS. 2015. Available: http://wheatpedigree.net/ [Accessed 2015.

Gul, A. & Allan, R. E. 1972. Relation of club gene with yield and yield components of near-isogenic wheat lines. *Crop Science,* 12**,** 297-&.

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

Hadjichristodoulou, A., Della, A. & Photiades, J. 1977. Effect of sowing depth on plant establishment, tillering capacity and other agronomic characters of cereals. *Journal of Agricultural Science,* 89**,** 161-167.

Harper, A., Trick, M., He, Z., Clissold, L., Fellgett, A., Griffiths, S. & Bancroft, I. 2015. Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat. *Plant Biotechnology Journal,* Accepted.

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

Hawkesford, M. J. 2014. Reducing the reliance on nitrogen fertilizer for wheat production. *Journal of Cereal Science,* 59**,** 276-283.

He, C., Holme, J. & Anthony, J. 2014. SNP genotyping: the KASP assay. *Methods in molecular biology (Clifton, N.J.),* 1145**,** 75-86.

Hedden, P. 2003. The genes of the Green Revolution. *Trends in genetics : TIG,* 19**,** 5-9.

Hedden, P. & Kamiya, Y. 1997. Gibberellin biosynthesis: Enzymes, genes and their regulation. *Annual Review of Plant Physiology and Plant Molecular Biology,* 48**,** 431-460.

Helguera, M., Rivarola, M., Clavijo, B., Martis, M. M., Vanzetti, L. S., Gonzalez, S., Garbus, I., Leroy, P., Simkova, H., Valarik, M., Caccamo, M., Dolezel, J., Mayer, K. F. X., Feuillet, C., Tranquilli, G., Paniego, N. & Echenique, V. 2015. New insights into the wheat chromosome 4D structure and virtual gene order, revealed by survey pyrosequencing. *Plant Science,* 233**,** 200-212.

Herrera, J. M., Verhulst, N. & Govaerts, B. 2012. Strategies to identify genetic diversity in root traits. *In:* REYNOLDS, M., PASK, AJD. AND MULLAN DM. (ed.) *Physiological Breeding I: Interdisciplinary Approaches to Improve Crop Adaptation.* Mexico: D.F.: CIMMYT.

Home-Grown Cereals Authority (HGCA) 2008. The wheat growth guide.

Houston, K., Mckim, S. M., Comadran, J., Bonar, N., Druka, I., Uzrek, N., Cirillo, E., Guzy-Wrobelska, J., Collins, N. C., Halpin, C., Hansson, M., Dockter, C., Druka, A. & Waugh, R. 2013. Variation in the interaction between alleles of HvAPETALA2 and microRNA172 determines the density of grains on the barley inflorescence. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 16675-16680.

Http://Aegilops.Wheat.Ucdavis.Edu/Atgsp/. 2015. *"Sequencing the Aegilops tauschii Genome" Project* [Online]. Available: http://aegilops.wheat.ucdavis.edu/ATGSP/data.php [Accessed July 2015.

Hu, Z., Han, Z., Song, N., Chai, L., Yao, Y., Peng, H., Ni, Z. & Sun, Q. 2013. Epigenetic modification contributes to the expression divergence of three TaEXPA1 homoeologs in hexaploid wheat (Triticum aestivum). *New Phytologist,* 197**,** 1344-1352.

Huang, L., Brooks, S. A., Li, W. L., Fellers, J. P., Trick, H. N. & Gill, B. S. 2003. Map-based cloning of leaf rust resistance gene Lr21 from the large and polyploid genome of bread wheat. *Genetics,* 164**,** 655-664.

Husar, S., Berthiller, F., Fujioka, S., Rozhon, W., Khan, M., Kalaivanan, F., Elias, L., Higgins, G. S., Li, Y., Schuhmacher, R., Krska, R., Seto, H., Vaistij, F. E., Bowles, D. & Poppenberger,

B. 2011. Overexpression of the UGT73C6 alters brassinosteroid glucoside formation in Arabidopsis thaliana. *Bmc Plant Biology,* 11.

Hwang, C., Ross, V. & Mahadevan, U. 2014. Popular Exclusionary Diets for Inflammatory Bowel Disease: The Search for a Dietary Culprit. *Inflammatory Bowel Diseases,* 20**,** 732-741.

International Brachypodium Initiative (IBI), **Vogel, J. P., Garvin, D. F., Mockler, T. C., Schmutz, J., Rokhsar, D., Bevan, M. W., Barry, K., Lucas, S., Harmon-Smoth, M., Lail, K., Tice, H., Grimwood, J., Mckenzie, N., Huo, N., Gu, Y. Q., Lazo, G. R., Anderson, O. D., You, F. M., Luo, M.-C., Dvorak, J., Wright, J., Febrer, M., Idziak, D., Hasterok, R., Lindquist, E., Wang, M., Fox, S. E., Priest, H. D., Filichkin, S. A., Givan, S. A., Bryant, D. W., Chang, J. H., Wu, H., Wu, W., Hsia, A.-P., Schnable, P. S., Kalyanaraman, A., Baarbazuk, B., Michael, T. P., Hazen, S. P., Bragg, J. N., Laudencia-Chingcuanco, D., Weng, Y., Haberer, G., Spannagl, M., Mayer, K., Rattei, T., Mitros, T., Lee, S.-J., Rose, J. K. C., Mueller, L. A., York, T. L., Wicker, T., Buchmann, J. P., Tanskanen, J., Schulman, A. H., Gundlach, H., De Oliveira, A. C., Maia, L. D. C., Belknap, W., Jiang, N., Lai, J., Zhu, L., Ma, J., Sun, C., Pritham, E., Salse, J., Murat, F., Abrouk, M., Bruggmann, R., Messing, J., Fahlgren, N., Sullivan, C. M., Carrington, J. C., Chapman, E. J., May, G. D., Zhai, J., Ganssmann, M., Gurazada, S. G. R., German, M., Meyers, B. C., Green, P. J., Tyler, L., Wu, J., Thomson, J., Chen, S., Scheller, H. V., Harholt, J., Ulvskov, P., Kimbrel, J. A., Bartley, L. E., Cao, P., Jung, K.-H., Sharma, M. K., Vega-Sanchez, M., Ronald, P., Dardick, C. D., De Bodt, S., Verelst, W., Inze, D., et al.** 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature,* 463**,** 763-768.

International Barley Genome Sequencing Consortium (IBGSC), **Mayer, K. F. X., Waugh, R., Langridge, P., Close, T. J., Wise, R. P., Graner, A., Matsumoto, T., Sato, K., Schulman, A., Muehlbauer, G. J., Stein, N., Ariyadasa, R., Schulte, D., Poursarebani, N., Zhou, R., Steuernagel, B., Mascher, M., Scholz, U., Shi, B., Langridge, P., Madishetty, K., Svensson, J. T., Bhat, P., Moscou, M., Resnik, J., Close, T. J., Muehlbauer, G. J., Hedley, P., Liu, H., Morris, J., Waugh, R., Frenkel, Z., Korol, A., Berges, H., Graner, A., Stein, N., Steuernagel, B., Taudien, S., Groth, M., Felder, M., Platzer, M., Brown, J. W. S., Schulman, A., Platzer, M., Fincher, G. B., Muehlbauer, G. J., Sato, K., Taudien, S., Sampath, D., Swarbreck, D., Scalabrin, S., Zuccolo, A., Vendramin, V., Morgante, M., Mayer, K. F. X. & Schulman, A.** 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature,* 491**,** 711.

International Wheat Genome Ssequencing Consortium (IWGSC), **Mayer, K. F. X., Rogers, J., Dolezel, J., Pozniak, C., Eversole, K., Feuillet, C., Gill, B., Friebe, B., Lukaszewski, A. J., Sourdille, P., Endo, T. R., Dolezel, J., Kubalakova, M., Cihalikova, J., Dubska, Z., Vrana, J., Sperkova, R., Simkova, H., Rogers, J., Febrer, M., Clissold, L., Mclay, K., Singh, K., Chhuneja, P., Singh, N. K., Khurana, J., Akhunov, E., Choulet, F., Sourdille, P., Feuillet, C., Alberti, A., Barbe, V., Wincker, P., Kanamori, H., Kobayashi, F., Itoh, T., Matsumoto, T., Sakai, H., Tanaka, T., Wu, J., Ogihara, Y., Handa, H., Pozniak, C., Maclachlan, P. R., Sharpe, A., Klassen, D., Edwards, D., Batley, J., Olsen, O.-A., Sandve, S. R., Lien, S., Steuernagel, B., Wulff, B., Caccamo, M., Ayling, S., Ramirez-Gonzalez, R. H., Clavijo, B. J., Steuernagel, B., Wright, J., Pfeifer, M., Spannagl, M., Mayer, K. F. X., Martis, M. M., Akhunov, E., Choulet, F., Mayer, K. F. X., Mascher, M., Chapman, J., Poland, J. A., Scholz, U., Barry, K., Waugh, R., Rokhsar, D. S., Muehlbauer, G. J., Stein, N., Gundlach, H., Zytnicki, M., Jamilloux, V., Quesneville, H., Wicker, T., Mayer, K. F. X., Faccioli, P., Colaiacovo, M., Pfeifer, M., Stanca, A. M., Budak, H., Cattivelli, L., Glover, N., Martis, M. M., Choulet, F., Feuillet, C., Mayer, K. F. X., Pfeifer, M., Pingault, L., Mayer, K. F. X., Paux, E., Spannagl, M., Sharma, S., Mayer, K. F. X., Pozniak, C., et al.** 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science,* 345.

Ishikawa, G., Nakamura, K., Ito, H., Saito, M., Sato, M., Jinno, H., Yoshimura, Y., Nishimura, T., Maejima, H., Uehara, Y., Kobayashi, F. & Nakamura, T. 2014. Association Mapping and

Validation of QTLs for Flour Yield in the Soft Winter Wheat Variety Kitahonami. *Plos One,* 9.

James, G. V., Patel, V., Nordstroem, K. J. V., Klasen, J. R., Salome, P. A., Weigel, D. & Schneeberger, K. 2013. User guide for mapping-by-sequencing in Arabidopsis. *Genome Biology,* 14.

Jantasuriyarat, C., Vales, M. I., Watson, C. J. W. & Riera-Lizarazu, O. 2004. Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (Triticum aestivum L.). *Theoretical and Applied Genetics,* 108**,** 261-273.

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

Jiang, Y., Zhao, Y., Rodemann, B., Plieske, J., Kollers, S., Korzun, V., Ebmeyer, E., Argillier, O., Hinze, M., Ling, J., Roder, M. S., Ganal, M. W., Mette, M. F. & Reif, J. C. 2015a. Potential and limits to unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (Triticum aestivum L.). *Heredity,* 114**,** 318-326.

Jiang, Y., Zhao, Y., Rodemann, B., Plieske, J., Kollers, S., Korzun, V., Ebmeyer, E., Argillier, O., Hinze, M., Ling, J., Roeder, M. S., Ganal, M. W., Mette, M. F. & Reif, J. C. 2015b. Potential and limits to unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (Triticum aestivum L.). *Heredity,* 114**,** 318-326.

Johnson, E. B., Nalam, V. J., Zemetra, R. S. & Riera-Lizarazu, O. 2008. Mapping the compactum locus in wheat (Triticum aestivum L.) and its relationship to other spike morphology genes of the Triticeae. *Euphytica,* 163**,** 193-201.

Jones, H., Gosman, N., Horsnell, R., Rose, G. A., Everest, L. A., Bentley, A. R., Tha, S., Uauy, C., Kowalski, A., Novoselovic, D., Simek, R., Kobiljski, B., Kondic-Spika, A., Brbaklic, L., Mitrofanova, O., Chesnokov, Y., Bonnett, D. & Greenland, A. 2013. Strategy for exploiting exotic germplasm using genetic, morphological, and environmental diversity: the Aegilops tauschii Coss. example. *Theoretical and Applied Genetics,* 126**,** 1793-1808.

Jones, S. S. & Cadle, M. M. 1997. Effect of variation at Glu-D1 on club wheat end-use quality. *Plant Breeding,* 116**,** 69-72.

Kasperbauer, M. J. & Karlen, D. L. 1986. Light-mediated bioregulation of tillering and photosynthate partitioning in wheat. *Physiologia Plantarum,* 66**,** 159-163.

Kasprzyk, A. 2011. BioMart: driving a paradigm change in biological data management. *Database,* 2011.

Kato, K. & Yokoyama, H. 1992. Geographical variation in heading characters among wheat landraces, Triticum-aestivum L, and its implication for their adaptability. *Theoretical and Applied Genetics,* 84**,** 259-265.

Kawahara, Y., De La Bastide, M., Hamilton, J. P., Kanamori, H., Mccombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J., Zhou, S., Childs, K. L., Davidson, R. M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., Itoh, T., Buell, C. R. & Matsumoto, T. 2013. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice,* 6.

Kerber, E. R. & Rowland, G. G. 1974. Origin of free threshing character in hexaploid wheat. *Canadian Journal of Genetics and Cytology,* 16**,** 145-154.

Kjaer, K. H. & Ottosen, C.-O. 2015. 3D Laser Triangulation for Plant Phenotyping in Challenging Environments. *Sensors,* 15**,** 13533-13547.

Klement, R. J. & Kaemmerer, U. 2011. Is there a role for carbohydrate restriction in the treatment and prevention of cancer? *Nutrition & Metabolism,* 8.

Klonoff, D. C. 2009. The beneficial effects of a Paleolithic diet on type 2 diabetes and other risk factors for cardiovascular disease. *Journal of diabetes science and technology,* 3**,** 1229-32.

Knight, E., Binnie, A., Draeger, T., Moscou, M., Rey, M. D., Sucher, J., Mehra, S., King, I. & Moore, G. 2015. Mapping the 'breaker' element of the gametocidal locus proximal to a block of sub-telomeric heterochromatin on the long arm of chromosome 4S of Aegilops sharonensis. *Theoretical and Applied Genetics,* 128, 1049-1059.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., Mclellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L. & Wilson, R. K. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research,* 22**,** 568-576.

Kohorn, B. D. & Kohorn, S. L. 2012. The cell wall-associated kinases, WAKs, as pectin receptors. *Front Plant Sci,* 3**,** 88.

Korzun, V., Roder, M. S., Ganal, M. W., Worland, A. J. & Law, C. N. 1998. Genetic analysis of the dwarfing gene (Rht8) in wheat. Part I. Molecular mapping of Rht8 on the short arm of chromosome 2D of bread wheat (Triticum aestivum L.). *Theoretical and Applied Genetics,* 96**,** 1104-1109.

Kosuge, K., Watanabe, N., Melnik, V. M., Laikova, L. I. & Goncharov, N. P. 2012. New sources of compact spike morphology determined by the genes on chromosome 5A in hexaploid wheat. *Genetic Resources and Crop Evolution,* 59**,** 1115-1124.

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

Krattinger, S., Wicker, T. & Keller, B. 2009a. Map-Based Cloning of Genes in Triticeae (Wheat and Barley). *In:* MUEHLBAUER, G. J. & FEUILLET, C. (eds.) *Genetics and Genomics of the Triticeae.* Springer US.

Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., Mcfadden, H., Bossolini, E., Selter, L. L. & Keller, B. 2009b. A Putative ABC Transporter Confers Durable Resistance to Multiple Fungal Pathogens in Wheat. *Science,* 323**,** 1360-1363.

Lamport, D. T. A., Kieliszewski, M. J., Chen, Y. & Cannon, M. C. 2011. Role of the Extensin Superfamily in Primary Cell Wall Architecture. *Plant Physiology,* 156**,** 11-19.

Langmead, B. & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9**,** 357-U54.

Lanning, S. P., Martin, J. M., Stougaard, R. N., Guillen-Portal, F. R., Blake, N. K., Sherman, J. D., Robbins, A. M., Kephart, K. D., Lamb, P., Carlson, G. R., Pumphrey, M. & Talbert, L. E. 2012. Evaluation of Near-Isogenic Lines for Three Height-Reducing Genes in Hard Red Spring Wheat. *Crop Science,* 52**,** 1145.

Laperche, A., Le Gouis, J., Hanocq, E. & Brancourt-Hulmel, M. 2008. Modelling nitrogen stress with probe genotypes to assess genetic parameters and genetic determinism of winter wheat tolerance to nitrogen constraint. *Euphytica,* 161**,** 259-271.

Lassmann, T., Hayashizaki, Y. & Daub, C. O. 2011. SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics,* 27**,** 130-131.

Law, C. C. & Worland, A. J. 1973. Aneuploidy in wheat and its uses in genetic analysis. *UK, Plant Breeding Institute: Annual report 1972.***,** 25-65.

Law, C. N. 1966. Location of genetic factors affecting a quantitative character in wheat. *Genetics,* 53**,** 487.

Law, C. N. 1967. Location of genetic factors controlling a number of quantitative characters in wheat. *Genetics,* 56**,** 445.

Law, C. N., Snape, J. W. & Worland, A. J. 1978. Genetic relationship between height and yield in wheat. *Heredity,* 40**,** 133-151.

Law, C. N. & Worland, A. J. 1997. The control of adult-plant resistance to yellow rust by the translocated chromosome 5BS-7BS of bread wheat. *Plant Breeding,* 116**,** 59-63.

Leach, L. J., Belfield, E. J., Jiang, C., Brown, C., Mithani, A. & Harberd, N. P. 2014. Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. *BMC Genomics,* 15.

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., Yan, J., Jiang, X., Zhang, L., Wu, J., Yin, L., Zhang, R., Wu, L., Zheng, Y. & Mao, L. 2014. mRNA and Small RNA Transcriptomes Reveal Insights into Dynamic Homoeolog Regulation of Allopolyploid Heterosis in Nascent Hexaploid Wheat. *Plant Cell,* 26**,** 1878-1900.

Li, A., Yang, W., Lou, X., Liu, D., Sun, J., Guo, X., Wang, J., Li, Y., Zhan, K., Ling, H.-Q. & Zhang, A. 2013. Novel Natural Allelic Variations at the Rht-1 Loci in Wheat. *Journal of Integrative Plant Biology,* 55**,** 1026-1037.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data, P. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

Li, H., Ruan, J. & Durbin, R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research,* 18**,** 1851-1858.

Li, W. & Gill, B. S. 2006. Multiple genetic pathways for seed shattering in the grasses. *Functional & Integrative Genomics,* 6**,** 300-309.

Liang, Y., Zhang, D.-Y., Ouyang, S., Xie, J., Wu, Q., Wang, Z., Cui, Y., Lu, P., Zhang, D., Liu, Z.-J., Zhu, J., Chen, Y.-X., Zhang, Y., Luo, M.-C., Dvorak, J., Huo, N., Sun, Q., Gu, Y.-Q. & Liu, Z. 2015. Dynamic evolution of resistance gene analogs in the orthologous genomic regions of powdery mildew resistance gene MlIW170 in Triticum dicoccoides and Aegilops tauschii. *Theoretical and Applied Genetics,* 128**,** 1617-1629.

Lim, E. K., Doucet, C. J., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Ross, J. & Bowles, D. J. 2002. The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *Journal of Biological Chemistry,* 277**,** 586-592.

Lin, P. Y. & Czuchajowska, Z. 1997. General characteristics and milling performance of club wheat vs. soft white winter wheat. *Cereal Foods World,* 42**,** 861-867.

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

Liu, C. J., Atkinson, M. D., Chinoy, C. N., Devos, K. M. & Gale, M. D. 1992. Nonhomologous translocations between group-4, group-5 and group-7 chromosomes within wheat and rye. *Theoretical and Applied Genetics,* 83**,** 305-312.

Liu, X., Khajuria, C., Li, J., Trick, H. N., Huang, L., Gill, B. S., Reeck, G. R., Antony, G., White, F. F. & Chen, M.-S. 2013. Wheat Mds-1 encodes a heat-shock protein and governs susceptibility towards the Hessian fly gall midge. *Nature Communications,* 4.

Liu, Y., Liu, D. C., Zhang, H. Y., Wang, J., Sun, J. Z., Guo, X. L. & Zhang, A. M. 2005. Allelic variation, sequence determination and microsatellite screening at the XGWM261 locus in Chinese hexaploid wheat (Triticum aestivum) varieties. *Euphytica,* 145**,** 103-112.

Lorenzetti, R. 2000. *La scienza del grano. L'esperienza scientifica di Nazareno Strampelli e la granicoltura italiana dal periodo giolittiano al secondo dopoguerra,* Rome, Ministero per i beni e le attivita culturali.

Lu, Y., Wu, X., Yao, M., Zhang, J., Liu, W., Yang, X., Li, X., Du, J., Gao, A. & Li, L. 2015. Genetic mapping of a putative Agropyron cristatum-derived powdery mildew resistance gene by a combination of bulked segregant analysis and single nucleotide polymorphism array. *Molecular Breeding,* 35.

Luo, M.-C., Gu, Y. Q., You, F. M., Deal, K. R., Ma, Y., Hu, Y., Huo, N., Wang, Y., Wang, J., Chen, S., Jorgensen, C. M., Zhang, Y., Mcguire, P. E., Pasternak, S., Stein, J. C., Ware, D., Kramer, M., Mccombie, W. R., Kianian, S. F., Martis, M. M., Mayer, K. F. X., Sehgal, S.

K., Li, W., Gill, B. S., Bevan, M. W., Šimková, H., Doležel, J., Weining, S., Lazo, G. R., Anderson, O. D. & Dvorak, J. 2013. A 4-gigabase physical map unlocks the structure and evolution of the complex genome of Aegilops tauschii, the wheat D-genome progenitor. *Proceedings of the National Academy of Sciences,* 110**,** 7940-7945.

Luo, M. C., Yang, Z. L., You, F. M., Kawahara, T., Waines, J. G. & Dvorak, J. 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. *Theoretical and Applied Genetics,* 114**,** 947-959.

Lynch, J. P. 2007. Turner review No. 14: Roots of the Second Green Revolution. *Australian Journal of Botany,* 55**,** 493-512.

Ma, J., Stiller, J., Wei, Y., Zheng, Y.-L., Devos, K. M., Dolezel, J. & Liu, C. 2014. Extensive Pericentric Rearrangements in the Bread Wheat (Triticum aestivum L.) Genotype "Chinese Spring" Revealed from Chromosome Shotgun Sequence Data. *Genome Biology and Evolution,* 6**,** 3039-3048.

Ma, J., Stiller, J., Zheng, Z., Wei, Y., Zheng, Y.-L., Yan, G., Dolezel, J. & Liu, C. 2015a. Putative interchromosomal rearrangements in the hexaploid wheat (Triticum aestivum L.) genotype 'Chinese Spring' revealed by gene locations on homoeologous chromosomes. *BMC Evolutionary Biology,* 15.

Ma, J., Wingen, L. U., Orford, S., Fenwick, P., Wang, J. K. & Griffiths, S. 2015b. Using the UK reference population Avalon 3 Cadenza as a platform to compare breeding strategies in elite Western European bread wheat. *Molecular Breeding,* 35.

Ma, Z., Zhao, D., Zhang, C., Zhang, Z., Xue, S., Lin, F., Kong, Z., Tian, D. & Luo, Q. 2007. Molecular genetic analysis of five spike-related traits in wheat using RIL and immortalized F-2 populations. *Molecular Genetics and Genomics,* 277**,** 31-42.

Mahdi, L., Bell, C. J. & Ryan, J. 1998. Establishment and yield of wheat (Triticum turgidum L.) after early sowing at various depths in a semi-arid Mediterranean environment. *Field Crops Research,* 58**,** 187-196.

Manickavelu, A., Kawaura, K., Imamura, H., Mori, M. & Ogihara, Y. 2011. Molecular mapping of quantitative trait loci for domestication traits and beta-glucan content in a wheat recombinant inbred line population. *Euphytica,* 177**,** 179-190.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z. T., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., Mcdade, K. E., Mckenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P. G., Begley, R. F. & Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-380.

Martins, W. S., Lucas, D. C. S., Neves, K. F. D. S. & Bertioli, D. J. 2009. WebSat--a web software for microsatellite marker development. *Bioinformation,* 3**,** 282-3.

Mascher, M. 2014. *Genetic positions and consensus genotypes of 437,973 scaffolds.* [Online]. //scholz@IPK-GATERSLEBEN.DE/W7984_WGS_assembly_SNP-based_genetic_positions_and_consensus_genotypes_of_scaffolds: IPK. Available: https://doi.ipk-gatersleben.de/DOI/7f5bebea-357c-4515-8f44-eb3885bb764f/a79c04d1-23d5-42b3-9c24-a0c765a42f6f/2 [Accessed 23rd April 2015].

Mascher, M., Muehlbauer, G. J., Rokhsar, D. S., Chapman, J., Schmutz, J., Barry, K., Munoz-Amatriain, M., Close, T. J., Wise, R. P., Schulman, A. H., Himmelbach, A., Mayer, K. F. X., Scholz, U., Poland, J. A., Stein, N. & Waugh, R. 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant Journal,* 76**,** 718-727.

Massa, A. N., Wanjugi, H., Deal, K. R., O'brien, K., You, F. M., Maiti, R., Chan, A. P., Gu, Y. Q., Luo, M. C., Anderson, O. D., Rabinowicz, P. D., Dvorak, J. & Devos, K. M. 2011. Gene Space Dynamics During the Evolution of Aegilops tauschii, Brachypodium distachyon,

Oryza sativa, and Sorghum bicolor Genomes. *Molecular Biology and Evolution,* 28**,** 2537-2547.

Mayer, K. F. X., Martis, M., Hedley, P. E., Simkova, H., Liu, H., Morris, J. A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubalakova, M., Suchankova, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T., Sakai, H., Tanaka, T., Itoh, T., Sato, K., Platzer, M., Matsumoto, T., Scholz, U., Dolezel, J., Waugh, R. & Stein, N. 2011. Unlocking the Barley Genome by Chromosomal and Comparative Genomics. *Plant Cell,* 23**,** 1249-1263.

Mcintosh, R. A., Yamazaki, Y., Dubcovsky, J., Rogers, W. J., Morris, C., Appels, R. & Xia, X. Catalogue of gene symbols for wheat Proceedings of 12th International wheat genetics Symposium, 2013 Yokohama, Japan. 12.

Mcvittie, J. A., Gale, M. D., Marshall, G. A. & Westcott, B. 1978. The intra-chromosomal mapping of the Norin 10 and Tom Thumb dwarfing genes. *Heredity,* 40**,** 67-70.

Michelmore, R. W., Paran, I. & Kesseli, R. V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America,* 88**,** 9828-9832.

Middleton, C. P., Senerchia, N., Stein, N., Akhunov, E. D., Keller, B., Wicker, T. & Kilian, B. 2014. Sequencing of Chloroplast Genomes from Wheat, Barley, Rye and Their Relatives Provides a Detailed Insight into the Evolution of the Triticeae Tribe. *Plos One,* 9.

Millet, E. 1986. Relationships between grain weight and the size of floret cavity in the wheat spike. *Annals of Botany,* 58**,** 417-423.

Miralles, D. J., Katz, S. D., Colloca, A. & Slafer, G. A. 1998. Floret development in near isogenic wheat lines differing in plant height. *Field Crops Research,* 59**,** 21-30.

Mochida, K., Yoshida, T., Sakurai, T., Ogihara, Y. & Shinozaki, K. 2009. TriFLDB: A Database of Clustered Full-Length Coding Sequences from Triticeae with Applications to Comparative Grass Genomics. *Plant Physiology,* 150**,** 1135-1146.

Moore, G., Devos, K. M., Wang, Z. & Gale, M. D. 1995. Cereal genome evolution - grasses, line up and form a circle. *Current Biology,* 5**,** 737-739.

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R. & Schaffer, A. A. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics,* 24**,** 1757-64.

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L. & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods,* 5**,** 621-628.

Muramatsu, M. 1963. Dosage effect of spelta gene Q of hexaploid wheat. *Genetics,* 48**,** 469-&.

Muramatsu, M. 1986. The vulgare super gene, Q - its universality in durum-wheat and its phenotypic effects in tetraploid and hexaploid wheats. *Canadian Journal of Genetics and Cytology,* 28**,** 30-41.

Mustalahti, K., Catassi, C., Reunanen, A., Fabiani, E., Heier, M., Mcmillan, S., Murray, L., Metzger, M.-H., Gasparin, M., Bravi, E. & Maki, M. 2010. The prevalence of celiac disease in Europe: Results of a centralized, international mass screening project. *Annals of Medicine,* 42**,** 587-595.

Nalam, V. J., Vales, M. I., Watson, C. J. W., Kianian, S. F. & Riera-Lizarazu, O. 2006. Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (Triticum turgidum L.). *Theoretical and Applied Genetics,* 112**,** 373-381.

Narayanan, S., Mohan, A., Gill, K. S. & Prasad, P. V. V. 2014. Variability of Root Traits in Spring Wheat Germplasm. *Plos One,* 9.

Nelson, D. R. 2009. The cytochrome p450 homepage. *Human genomics,* 4**,** 59-65.

Nelson, J. C., Sorrells, M. E., Vandeynze, A. E., Lu, Y. H., Atkinson, M., Bernard, M., Leroy, P., Faris, J. D. & Anderson, J. A. 1995. Molecular mapping of wheat - major genes and rearrangements in homoeologous group-4, group-5, and group-7. *Genetics,* 141**,** 721-731.

Noir, S., Brautigam, A., Colby, T., Schmidt, J. & Panstruga, R. 2005. A reference map of the Arabidopsis thaliana mature pollen proteome. *Biochemical and Biophysical Research Communications,* 337**,** 1257-1266.

O'Donovan, J. T., Blackshaw, R. E., Harker, K. N., Clayton, G. W. & Mckenzie, R. 2005. Variable crop plant establishment contributes to differences in competitiveness with wild oat among cereal varieties. *Canadian Journal of Plant Science,* 85**,** 771-776.

Oono, Y., Kobayashi, F., Kawahara, Y., Yazawa, T., Handa, H., Itoh, T. & Matsumoto, T. 2013. Characterisation of the wheat (triticum aestivum L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. *BMC Genomics,* 14.

Ortiz-Monasterio, J. I., Manske G., Van Ginkel, M. 2012. Nitrogen and phosphorus use efficiency. *In:* REYNOLDS, M., PASK, AJD. AND MULLAN DM. (ed.) *Physiological Breeding I: Interdisciplinary Approaches to Improve Crop Adaptation.* Mexico: D.F.: CIMMYT.

Ostrowski, M. & Jakubowska, A. 2014. Udp-Glycosyltransferases of Plant Hormones. *Advances in Cell Biology.*

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R. L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J. & Buell, C. R. 2007. The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Research,* 35**,** D883-D887.

Paillard, S., Schnurbusch, T., Winzeler, M., Messmer, M., Sourdille, P., Abderhalden, O., Keller, B. & Schachermayr, G. 2003. An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics,* 107**,** 1235-1242.

Pallotta, M. A., Asayama, S., Reinheimer, J. M., Davies, P. A., Barr, A. R., Jefferies, S. P., Chalmers, K. J., Lewis, J., Collins, H. M., Roumeliotis, S., Logue, S. J., Coventry, S. J., Lance, R. C. M., Karakousis, A., Lim, P., Verbyla, A. P. & Eckermann, P. J. 2003. Mapping and QTL analysis of the barley population Amagi Nijo x WI2585. *Australian Journal of Agricultural Research,* 54**,** 1141-1144.

Pastore, R. L., Brooks, J. T. & Carbone, J. W. 2015. Paleolithic nutrition improves plasma lipid concentrations of hypercholesterolemic adults to a greater extent than traditional heart-healthy dietary recommendations. *Nutrition Research,* 35**,** 474-479.

Paterson, T. & Law, A. 2013. ArkMAP: integrating genomic maps across species and data sources. *BMC Bioinformatics,* 14.

Paux, E., Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., Korol, A., Michalak, M., Kianian, S., Spielmeyer, W., Lagudah, E., Somers, D., Kilian, A., Alaux, M., Vautrin, S., Bergès, H., Eversole, K., Appels, R., Safar, J., Simkova, H., Dolezel, J., Bernard, M. & Feuillet, C. 2008. A physical map of the 1-gigabase bread wheat chromosome 3B. *Science (New York, N.Y.),* 322**,** 101-4.

Peel, C. H. 1987. A rising disc apparatus for the measurement of turfgrass sward heights. *Journal of the Sports Turf Research Institute,* 63**,** 153-156.

Peltonen-Sainio, P., Kangas, A., Salo, Y. & Jauhiainen, L. 2007. Grain number dominates grain weight in temperate cereal yield determination: Evidence based on 30 years of multi-location trials. *Field Crops Research,* 100**,** 179-188.

Peng, J. R., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D. & Harberd, N. P. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature,* 400**,** 256-261.

Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., Deal, K., Luo, M., Kong, X., Bariana, H., Mago, R., Mcintosh, R., Dodds, P., Dvorak, J. & Lagudah, E. 2013. The Gene Sr33, an Ortholog of Barley Mla Genes, Encodes Resistance to Wheat Stem Rust Race Ug99. *Science,* 341**,** 786-788.

Pingali, P. L. 2012. Green Revolution: Impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 12302-12308.

Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. 2012. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *Plos One,* 7.

Pysh, L. D., Wysocka-Diller, J. W., Camilleri, C., Bouchez, D. & Benfey, P. N. 1999. The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant Journal,* 18**,** 111-119.

Quarrie, S. A., Lazic-Jancic, V., Kovacevic, D., Steed, A. & Pekic, S. 1999. Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *Journal of Experimental Botany,* 50**,** 1299-1306.

R Development Core Team 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ramirez-Gonzalez, R. H., Segovia, V., Bird, N., Fenwick, P., Holdgate, S., Berry, S., Jack, P., Caccamo, M. & Uauy, C. 2014. RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnology Journal***,** 5, 613-24.

Ramirez-Gonzalez, R. H., Uauy, C. & Caccamo, M. 2015. PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics,* 31**,** 2038-2039.

Rao, M. V. P. 1972. Mapping of the compactum gene C on chromosome 2D of wheat. *Wheat Information Service,* 35**,** 9-9.

Rao, M. V. P. 1977. Mapping of the sphaerococcum gene S on chromosome 3D of wheat. *Cereal Research Communications,* 5**,** 15-17.

Raun, W. R. & Johnson, G. V. 1999. Improving nitrogen use efficiency for cereal production. *Agronomy Journal,* 91**,** 357-363.

Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *Plos One,* 8.

Rebetzke, G. J., Appels, R., Morrison, A. D., Richards, R. A., Mcdonald, G., Ellis, M. H., Spielmeyer, W. & Bonnett, D. G. 2001. Quantitative trait loci on chromosome 4B for coleoptile length and early vigour in wheat (Triticum aestivum L.). *Australian Journal of Agricultural Research,* 52**,** 1221-1234.

Rebetzke, G. J., Bonnett, D. G. & Ellis, M. H. 2012a. Combining gibberellic acid-sensitive and insensitive dwarfing genes in breeding of higher-yielding, sesqui-dwarf wheats. *Field Crops Research,* 127**,** 17-25.

Rebetzke, G. J., Ellis, M. H., Bonnett, D. G., Mickelson, B., Condon, A. G. & Richards, R. A. 2012b. Height reduction and agronomic performance for selected gibberellin-responsive dwarfing genes in bread wheat (Triticum aestivum L.). *Field Crops Research,* 126**,** 87-96.

Rebetzke, G. J. & Richards, R. A. 1999. Genetic improvement of early vigour in wheat. *Australian Journal of Agricultural Research,* 50**,** 291-301.

Rebetzke, G. J. & Richards, R. A. 2000. Gibberellic acid-sensitive dwarfing genes reduce plant height to increase kernel number and grain yield of wheat. *Australian Journal of Agricultural Research,* 51**,** 235-245.

Rebetzke, G. J., Richards, R. A., Fettell, N. A., Long, M., Condon, A. G., Forrester, R. I. & Botwright, T. L. 2007. Genotypic increases in coleoptile length improves stand establishment, vigour and grain yield of deep-sown wheat. *Field Crops Research,* 100**,** 10-23.

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., Van Ginkel, M., Hoisington, D., Bohn, M. & Melchinger, A. E. 2005. Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics,* 110**,** 859-864.

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

Rona, R. J., Keil, T., Summers, C., Gislason, D., Zuidmeer, L., Sodergren, E., Sigurdardottir, S. T., Lindner, T., Goldhahn, K., Dahlstrom, J., Mcbride, D. & Madsen, C. 2007. The prevalence of food allergy: A meta-analysis. *Journal of Allergy and Clinical Immunology,* 120**,** 638-646.

Rubio-Tapia, A., Kyle, R. A., Kaplan, E. L., Johnson, D. R., Page, W., Erdtmann, F., Brantner, T. L., Kim, W. R., Phelps, T. K., Lahr, B. D., Zinsmeister, A. R., Melton, L. J., Iii & Murray, J. A. 2009. Increased Prevalence and Mortality in Undiagnosed Celiac Disease. *Gastroenterology,* 137**,** 88-93.

Ruiz-Nunez, B., Pruimboom, L., Dijck-Brouwer, D. a. J. & Muskiet, F. a. J. 2013. Lifestyle and nutritional imbalances associated with Western diseases: causes and consequences of chronic systemic low-grade inflammation in an evolutionary context. *Journal of Nutritional Biochemistry,* 24**,** 1183-1201.

Safar, J., Bartos, J., Janda, J., Bellec, A., Kubalakova, M., Valarik, M., Pateyron, S., Weiserova, J., Tuskova, R., Cihalikova, J., Vrana, J., Simkova, H., Faivre-Rampant, P., Sourdille, P., Caboche, M., Bernard, M., Dolezel, J. & Chalhoub, B. 2004. Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *Plant Journal,* 39**,** 960-968.

Saintenac, C., Zhang, W., Salcedo, A., Rouse, M. N., Trick, H. N., Akhunov, E. & Dubcovsky, J. 2013. Identification of Wheat Gene Sr35 That Confers Resistance to Ug99 Stem Rust Race Group. *Science,* 341**,** 783-786.

Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R. & Martin, W. 2002. Genetics and geography of wild cereal domestication in the Near East. *Nature Reviews Genetics,* 3**,** 429-441.

Salvi, S., Porfiri, O. & Ceccarelli, S. 2013. Nazareno Strampelli, the 'Prophet' of the green revolution. *Journal of Agricultural Science,* 151**,** 1-5.

Scheible, W.-R., Torjek O & Altmann, T. 2005. From Markers to Cloned Genes: Map-Based Cloning. *In:* LORZ H & G, W. (eds.) *Biotechnology in agriculture and forestry.* Berlin, Germany: Springer.

Schmidt, A. L., Gale, K. R., Ellis, M. H. & Giffard, P. M. 2004. Sequence variation at a microsatellite locus (XGWM261) in hexaploid wheat (Triticum aestivum) varieties. *Euphytica,* 135**,** 239-246.

Schneeberger, K. 2014. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nature Reviews Genetics,* 15**,** 662-676.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., Jorgensen, J.-E., Weigel, D. & Andersen, S. U. 2009. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods,* 6**,** 550-551.

Schneeberger, K. & Weigel, D. 2011. Fast-forward genetics enabled by new sequencing technologies. *Trends in plant science,* 16**,** 282-8.

Sears, E. R. 1947. The sphaerococcum gene in wheat. *Genetics,* 32**,** 102-103.

Sears, E. R. 1952. Misdivision of univalents in common wheat. *Chromosoma,* 4**,** 535-550.

Sears, E. R. 1966. Nullisomic-tetrasomic combinations in hexaploid wheat. *In:* RILEY, R. & LEWIS, K. R. (eds.) *Chromosome Manipulation and Plant Genetics.* Edinburgh: Oliver & Boyd.

Semenov, M. A. & Shewry, P. R. 2011. Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. *Scientific Reports,* 1.

Semenov, M. A., Stratonovitch, P., Alghabari, F. & Gooding, M. J. 2014. Adapting wheat in Europe for climate change. *Journal of Cereal Science,* 59**,** 245-256.

Shewry, P. R. 2009. Wheat. *Journal of Experimental Botany,* 60**,** 1537-1553.

Simkova, H., Svensson, J. T., Condamine, P., Hribova, E., Suchankova, P., Bhat, P. R., Bartos, J., Safar, J., Close, T. J. & Dolezel, J. 2008. Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. *BMC Genomics,* 9.

Simonetti, M. C., Bellomo, M. P., Laghetti, G., Perrino, P., Simeone, R. & Blanco, A. 1999. Quantitative trait loci influencing free-threshing habit in tetraploid wheats. *Genetic Resources and Crop Evolution,* 46**,** 267-271.

Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z. C., Tai, Y. S., Gill, B. S. & Faris, J. D. 2006. Molecular characterization of the major wheat domestication gene Q. *Genetics,* 172**,** 547-555.

Šíp, V., Chrpová, J., Žofajová, A., Pánková, K., Užík, M. & Snape, J. W. 2009. Effects of specific Rht and Ppd alleles on agronomic traits in winter wheat cultivars grown in middle Europe. *Euphytica,* 172**,** 221-233.

Slafer, G. A. 2003. Genetic basis of yield as viewed from a crop physiologist's perspective. *Annals of Applied Biology,* 142**,** 117-128.

Slafer, G. A. 2012. Wheat development: its role in phenotyping and improving crop adaptation. *In:* REYNOLDS, M., PASK, AJD. AND MULLAN DM. (ed.) *Physiological Breeding I: Interdisciplinary Approaches to Improve Crop Adaptation.* Mexico: D.F.: CIMMYT.

Slafer, G. A., Savin, R. & Sadras, V. O. 2014. Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Research,* 157**,** 71-83.

Sorrells, M. E., Gustafson, J. P., Somers, D., Chao, S., Benscher, D., Guedira-Brown, G., Huttner, E., Kilian, A., Mcguire, P. E., Ross, K., Tanaka, J., Wenzl, P., Williams, K. & Qualset, C. O. 2011. Reconstruction of the Synthetic W7984 x Opata M85 wheat reference population. *Genome,* 54**,** 875-882.

Sourdille, P., Tixier, M. H., Charmet, G., Gay, G., Cadalen, T., Bernard, S. & Bernard, M. 2000. Location of genes involved in ear compactness in wheat (Triticum aestivum) by means of molecular markers. *Molecular Breeding,* 6**,** 247-255.

Sparkes, D. L. & King, M. 2008. Disentangling the effects of PAR and R : FR on lodging-associated characters of wheat (Triticum aestivum). *Annals of Applied Biology,* 152**,** 1-9.

Sreenivasulu, N. & Schnurbusch, T. 2012. A genetic playground for enhancing grain number in cereals. *Trends in Plant Science,* 17**,** 91-101.

Sun, H., Guo, Z., Gao, L., Zhao, G., Zhang, W., Zhou, R., Wu, Y., Wang, H., An, H. & Jia, J. 2014. DNA methylation pattern of Photoperiod-B1 is associated with photoperiod insensitivity in wheat (Triticum aestivum). *New Phytologist,* 204**,** 682-692.

Taenzler, B., Esposti, R. F., Vaccino, P., Brandolini, A., Effgen, S., Heun, M., Schafer-Pregl, R., Borghi, B. & Salamini, F. 2002. Molecular linkage map of Einkorn wheat: mapping of storage-protein and soft-glume genes and bread-making quality QTLs. *Genetical Research,* 80**,** 131-143.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. 2011. Differential expression in RNA-seq: A matter of depth. *Genome Research,* 21**,** 2213-2223.

Trethowan, R. M., Singh, R. P., Huerta-Espino, J., Crossa, J. & Van Ginkel, M. 2001. Coleoptile length variation of near-isogenic Rht lines of modern CIMMYT bread and durum wheats. *Field Crops Research,* 70**,** 167-176.

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

Tulpan, D., Leger, S., Tchagang, A. & Pan, Y. 2015. Enrichment of Triticum aestivum gene annotations using ortholog cliques and gene ontologies in other plants. *Bmc Genomics,* 16.

U.S. Department of Agriculture (USDA) 1923. The Club Wheats. 1303 ed.

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.

Uauy, C., Paraiso, F., Colasuonno, P., Tran, R. K., Tsai, H., Berardi, S., Comai, L. & Dubcovsky, J. 2009. A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *Bmc Plant Biology,* 9.

UC Davis Plant Science & USDA 2015. *Aegilops tauschii genome database* [Online]. Available: http://probes.pw.usda.gov/WheatDMarker/downloads/ 2015].

Unrau, J. 1950. The use of monosomes and nullisomes in cytogenetic studies of common wheat. *SCI AGRIC [OTTOWA],* 30**,** 66-89.

Unité de Recherche Génomique Info (URGI). 2013. Available: https://urgi.versailles.inra.fr/blast/blast.php.

Unité de Recherche Génomique Info (URGI). 2015a. *Genome Zipper 2015* [Online]. Available: http://wheat-urgi.versailles.inra.fr/Seq-Repository/Genes-annotations.

Unité de Recherche Génomique Info (URGI). 2015b. *IWGSC-2 POPSEQ* [Online]. Available: http://wheat-urgi.versailles.inra.fr/About-us/News/3B-survey-seq-POPSEQ-GenomeZipper-data-available.

Van Ooijen, J. W. & Voorrips, R. E. 2001. *JoinMap® 3.0, Software for the calculation of genetic linkage maps* [Online].

Vazquez, F., Legrand, S. & Windels, D. 2010. The biosynthetic pathways and biological scopes of plant small RNAs. *Trends in Plant Science,* 15**,** 337-345.

Villalba, M., Batanero, E., López-Otín, C., Sánchez, L. M., Monsalve, R. I., González De La Peña, M. A., Lahoz, C. & Rodríguez, R. 1993. The amino acid sequence of Ole e I, the major allergen from olive tree (Olea europaea) pollen. *European journal of biochemistry / FEBS,* 216**,** 863-869.

Voorrips, R. E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity,* 93**,** 77-78.

Vrana, J., Simkova, H., Kubalakova, M., Cihalikova, J. & Dolezel, J. 2012. Flow cytometric chromosome sorting in plants: The next generation. *Methods,* 57**,** 331-337.

Wang, M., Wang, S. & Xia, G. 2015a. From genome to gene: a new epoch for wheat research? *Trends in Plant Science,* 20**,** 380-387.

Wang, S. C., Wong, D. B., Forrest, K., Allen, A., Chao, S. M., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G., Cattivelli, L., Mastrangelo, A. M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A. R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M. C., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K. J., Hayden, M., Akhunov, E. & Int Wheat Genome, S. 2014a. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal,* 12**,** 787-796.

Wang, Y., Chen, L., Du, Y., Yang, Z., Condon, A. G. & Hu, Y.-G. 2014b. Genetic effect of dwarfing gene Rht13 compared with Rht-D1b on plant height and some agronomic traits in common wheat (Triticum aestivum L.). *Field Crops Research,* 162**,** 39-47.

Wang, Y., Du, Y., Yang, Z., Chen, L., Condon, A. G. & Hu, Y.-G. 2015b. Comparing the effects of GA-responsive dwarfing genes Rht13 and Rht8 on plant height and some agronomic traits in common wheat. *Field Crops Research,* 179**,** 35-43.

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics,* 10**,** 57-63.

Wang, Z., Wu, X., Ren, Q., Chang, X., Li, R. & Jing, R. 2010. QTL mapping for developmental behavior of plant height in wheat (Triticum aestivum L.). *Euphytica,* 174**,** 447-458.

Warden, C. D., Yuan, Y. & Wu, X. 2013. Optimal Calculation of RNA-Seq Fold-Change Values. *International Journal of Computational ioinformatics and In Silico Modeling,* 2**,** 285-292.

Weichselbaum, E. 2012. Does bread cause bloating? *Nutrition Bulletin,* 37**,** 30-36.

Wilhelm, E. P., Mackay, I. J., Saville, R. J., Korolev, A. V., Balfourier, F., Greenland, A. J., Boulton, M. I. & Powell, W. 2013. Haplotype dictionary for the Rht-1 loci in wheat. *Theoretical and Applied Genetics,* 126**,** 1733-1747.

Wojciechowski, T., Gooding, M. J., Ramsay, L. & Gregory, P. J. 2009. The effects of dwarfing genes on seedling root growth of wheat. *Journal of Experimental Botany,* 60**,** 2565-2573.

Wolfe, M. S., Baresel, J. P., Desclaux, D., Goldringer, I., Hoad, S., Kovacs, G., Loeschenberger, F., Miedaner, T., Ostergard, H. & Van Bueren, E. T. L. 2008. Developments in breeding cereals for organic agriculture. *Euphytica,* 163**,** 323-346.

Worland, A. J. 1999. The importance of Italian wheats to worldwide varietal improvement. *Journal of Genetics and Breeding,* 53**,** 165-173.

Worland, A. J., Borner, A., Korzun, V., Li, W. M., Petrovic, S. & Sayers, E. J. 1998a. The influence of photoperiod genes on the adaptability of European winter wheats (Reprinted from Wheat: Prospects for global improvement, 1998). *Euphytica,* 100**,** 385-394.

Worland, A. J., Korzun, V., Ro, M. S., Ganal, M. W. & Law, C. N. 1998b. 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. 1110-1120.

Worland, A. J. & Law, C. N. 1986. Genetic-analysis of chromosome 2D of wheat .1. The location of genes affecting height, day-length insensitivity, hybrid dwarfism and yellow-rust resistance. *Zeitschrift Fur Pflanzenzuchtung-Journal of Plant Breeding,* 96**,** 331-345.

Worland, A. J., Sayers, E. J. & Korzun, V. 2001. Allelic variation at the dwarfing gene Rht8 locus and its significance in international breeding programmes. *Euphytica,* 119**,** 155-159.

Wu, R. L. & Lin, M. 2006. Opinion - Functional mapping - how to map and study the genetic architecture of dynamic complex traits. *Nature Reviews Genetics,* 7**,** 229-237.

Wuerschum, T., Liu, W., Busemeyer, L., Tucker, M. R., Reif, J. C., Weissmann, E. A., Hahn, V., Ruckelshausen, A. & Maurer, H. P. 2014. Mapping dynamic QTL for plant height in triticale. *BMC Genetics,* 15.

Yahiaoui, N., Srichumpa, P., Dudler, R. & Keller, B. 2004. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene Pm3b from hexaploid wheat. *Plant Journal,* 37**,** 528-538.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Sanchez, A., Valarik, M., Yasuda, S. & Dubcovsky, J. 2006. The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 19581-19586.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T. & Dubcovsky, J. 2003. Positional cloning of the wheat vernalization gene VRN1. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 6263-6268.

Yan, L. L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., Sanmiguel, P., Bennetzen, J. L., Echenique, V. & Dubcovsky, J. 2004. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. *Science,* 303**,** 1640-1644.

Yediay, F. E., Andeden, E. E., Baloch, F. S., Borner, A., Kilian, B. & Ozkan, H. 2011. The allelic state at the major semi-dwarfing genes in a panel of Turkish bread wheat cultivars and landraces. *Plant Genetic Resources-Characterization and Utilization,* 9**,** 423-429.

Yendrek, C. R., Ainsworth, E. A. & Thimmapuram, J. 2012. The bench scientist's guide to statistical analysis of RNA-Seq data. *BMC research notes,* 5**,** 506-506.

Zadoks, J. C., Chang, T. T. & Konzak, C. F. 1974. Decimal code for growth stages of cereals. *Weed Research,* 14**,** 415-421.

Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, K., Ganal, M. W. & Roeder, M. S. 2014. Whole Genome Association Mapping of Plant Height in Winter Wheat (Triticum aestivum L.). *Plos One,* 9.

Zhang, Z., Belcram, H., Gornicki, P., Charles, M., Just, J., Huneau, C., Magdelenat, G., Couloux, A., Samain, S., Gill, B. S., Rasmussen, J. B., Barbe, V., Faris, J. D. & Chalhoub, B. 2011. Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 18737-18742.

Zhang, Z., Zhu, H., Gill, B. S. & Li, W. 2015. Fine mapping of shattering locus Br2 reveals a putative chromosomal inversion polymorphism between the two lineages of Aegilops tauschii. *Theoretical and Applied Genetics,* 128**,** 745-755.

Zwer, P. K., Sombrero, A., Rickman, R. W. & Klepper, B. 1995. Club and common wheat yield component and spike development in the Pacific-Northwest. *Crop Science,* 35**,** 1590-1597.