An efficient approach for the development of locus specific primers in wheat and its application in a candidate gene based association study on frost tolerance in wheat (*Triticum aestivum* L.)

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. agr.)

der

Naturwissenschaftlichen Fakultät III

– Agrar- und Ernährungswissenschaften, Geowissenschaften und Informatik –

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

M. Sc. **Steve Babben**

Geb. am 08.05.1984 in Finsterwalde

1. Gutachter: Prof. Dr. Frank Ordon und Prof. Dr. Klaus Pillen

2. Gutachter: PD Dr. Ali Ahmad Naz

Verteidigung am 04.Februar 2019

Abstract

Understanding the genetic structure of hexaploid wheat (*Triticum aestivum* L.) is important to develop useful and effective methods for genetic analyses. Particularly frost tolerance (FT) in wheat is essential for preventing yield losses caused by frost due to cellular damage, dehydration and reduced metabolism. For such a complex trait (regulated by various genes and several gene families) in an accordingly complex species (controlled by three genomes) the availability of wheat genomic sequences and their usage for exploring candidate gene diversity are of special interest. Therefore, the objectives of this study were development of genome specific primers of FT candidate genes for re-sequencing, polymorphism detection and candidate gene based association analysis. A set of 39 specific primer pairs corresponding to 19 candidate genes were developed with a success rate of >50%. The genes corresponding to *C-REPEAT BINDING FACTORS* (*CBFs*), i.e. *CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *CBF-A18*, as well as the *VERNALISATION RESPONSE GENES VRN-A1*, *VRN-B3*, and the *PHOTOPERIOD RESPONSE GENES PPD-B1* and *PPD-D1* revealed associations to FT in 235 wheat cultivars. Six of these genes exhibit amino acid (AA) substitution polymorphisms in important protein domains. The effect on FT based on AA substitution in *VRN-A1* is also described in related literature, but the AA substitutions in *CBF-A3*, *CBF-A5*, *VRN-B3*, *PPD-B1* and *PPD-D1* which are located in highly conserved sites were yet unknown. Based on the development of an effective technique to design genome and locus specific primers in hexaploid wheat AA substitutions were identified in six FT genes suited for the application in marker assisted breeding for frost tolerant varieties of winter wheat.

Index

* from Babben et al. (2015); ** from Babben et al. (2018)

* from Babben et al. (2015); ** from Babben et al. (2018)

1 Introduction

Wheat (*Triticum aestivum* L.) is the cereal with the largest acreage worldwide and essential for human nutrition and animal feed. The worldwide production of wheat was 749.5 million tonnes (mt) in 2016 and wheat globally ranks as the second most important crop, excelled only by maize (*Zea mays* L.). Production of wheat in Europe amounted up to 250.13 mt in 2016 with an average yield of 4 t/ha. At the same time 93.35 mt of wheat were harvested in North America with an average yield of 3.45 t/ha, whilst the average yield of 3.25 t/ha in Asia resulted in a production of 326.75 mt (FAO, 2016). In North America, Northern and Eastern Europe and Russia low temperature in wheat crops occurs frequently and subsequent frost damage reduces the productivity. There are two different kinds of wheat cultivars termed as winter and spring wheat. Classification criteria are time of sowing, vernalisation requirement and frost hardiness. Winter wheat varieties require a long cold temperature period of typically 4 to 8 weeks at less than 4 °C to shift from the vegetative to the generative phase (vernalisation) in contrast to varieties of spring wheat (Koemel et al., 2004). Vernalisation requirements extend the vegetative phase of these plants and thereby preserve the plants from frost damage and following high yield losses due to low temperatures during winter.

1.1 Wheat genome

Bread wheat (henceforth referred to as wheat) belongs to the family of *Poaceae* and is an allohexaploid species (2n = 6× = 42) with an AABBDD genome derived from two independent hybridisations (Dvorak and Akhunov, 2005, Huang et al., 2002, Peng et al., 2011). The complex genome of about 17 Giga-base pairs (Gbp) possesses a repeat content of approximately 80% which consists primarily of retroelements (Smith and Flavell, 1975). The gene density lies within a range of one gene per 87 kilo-base pairs (kbp) to one gene per 184 kbp (Choulet et al., 2010). During evolution wheat turned into an allohexaploid organism (2n = 6x = 42) with the genomes called A, B and D. In brief, 300.000–500.000 years ago the first hybridisation between the diploid wild wheat

(*Triticum urartu*, 2n = 2x = 14, genome A^uA^u) and an ancestor closest related to goat grass (*Aegilops speltoides*, 2n = 2x = 14, genome SS, where S is closely related to B) took place (Huang et al., 2002, Dvorak and Akhunov, 2005) leading to the generation of wild emmer wheat (*Triticum dicoccoides*, 2n = 4x = 28, genome A^uA^uBB) (Feldman and Kislev, 2007). Tribal communities formerly making a living of gathering and hunting began to cultivate the wild emmer about 10,000 years ago. Following human selection led to cultivated emmer (*Triticum dicoccum*). A spontaneous hybridisation of cultivated emmer with another goat grass (*Aegilops tauschii* 2n = 2x = 14, genome DD) in combination with a spontaneous mutation (free-threshing ears), is assumed to have generated bread wheat (*Triticum aestivum*, 2n = 6x = 42, genome AABBDD) (Peng et al., 2011) (Figure 1). Due to the hexaploid genome and a very high homology within the three sub-genomes in wheat, the exact genome sequence has an inestimable value for molecular breeding, comparative genomics and association studies.

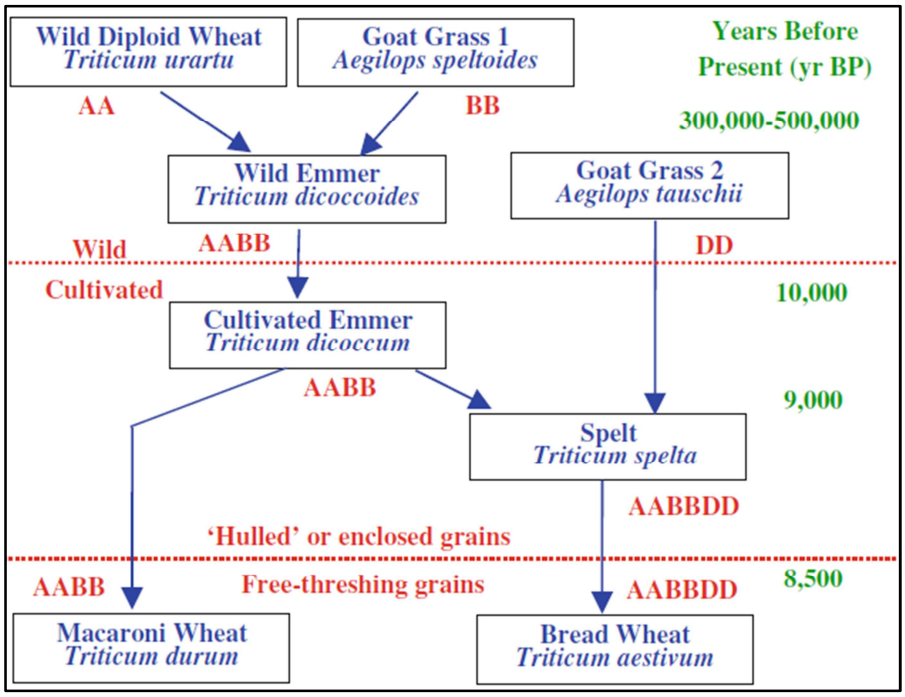


Figure 1: The evolution of *Triticum durum* and *Triticum aestivum* since the Paleolithic (Peng et al., 2011).

To enable efficient exploitation of the wheat genome sequence for molecular breeding, access to and curation of the genome sequence and its derived applications is essential. This is enabled by a variety of publicly accessible databases. The National Center for

Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov/) is a library of genomic, transcriptional and protein sequence data for more than 33,000 organisms (Pruitt et al., 2012). NCBI serves as a web-platform for the identification of target gene sequences in organisms of interest, e.g. *Triticum aestivum*, *Triticum monococcum*, *Hordeum vulgare* etc. The CerealsDB web page as an additional wheat database is provided by members of the Functional Genomics Group at the University of Bristol (http://www.cerealsdb.uk.net) and includes online resources of genomic information, i.e. varietal SNPs (single nucleotide polymorphism), DArT (diversity arrays technology) markers, and EST (expressed sequence tag) sequences all linked to a draft genome sequence of the cultivar Chinese Spring (Wilkinson et al., 2012). Another web based portal is URGI, which includes datasets such as chromosome survey sequences, reference sequences, physical maps, genetic maps, polymorphisms, genetic resources, many phenotypic data and various genomic arrays (http://wheat-urgi.versailles.inra.fr). The efforts of the International Wheat Genome Sequencing Consortium (IWGSC) in the development of a physical map and a reference sequence facilitate many downstream applications, i.e. development of high throughput genotyping platforms (Wang et al., 2014a), efficient development of genome specific primers, and exome sequencing in large gene bank collections (Jordan et al., 2015). Wheat reference sequence, survey sequence and physical map are available at many public databases. The annotated chromosomal sequence is available at the IWGSC. All mentioned databases are suitable for the identification of homologous chromosome sequences in bread wheat. In addition to these resources, an important tool for the examined species is the upcoming Genome Zipper of wheat (http://wheat-urgi.versailles.inra.fr). Summarising, a lot of sequence information of sorted bread wheat chromosome arms (Brenchley et al., 2012, Raats et al., 2013, Appels et al., 2018, IWGSC, 2014), *Triticum urartu* (Ling et al., 2013) and *Aegilops tauschii* (Jia et al., 2013) has been published during the past few years and is integrated in the public databases mentioned above. Today, the current version of the Chinese Spring IWGSC RefSeq v1.0 (Appels et al., 2018) facilitates studies on wheat in fundamental research as well as in molecular breeding.

1.2 Freezing damage in plants

Abiotic stress in winter has to be distinguished between chilling temperatures (low but positive cold temperatures) and freezing temperatures (temperatures below 0°C) (Ruelland et al., 2009). Some crop species, e.g., corn (*Zea mays*), rice (*Oryza sativa*) and potato (*Solanum tuberosum*) are irreversibly damaged even by chilling temperatures (McKersie and Leshem, 1994). Other species, such as spinach (*Spinacia oleracea*), winter wheat (*Triticum aestivum*) and canola (*Brassica napus*) tolerate chilling and are able to survive freezing temperatures (Ruelland et al., 2009). Since winter wheat tolerates chilling temperatures this chapter is focused on freezing temperatures. Under such conditions freeze of plants is signified by freezing of extracellular water. Over time, the extracellular ice crystals grow and cellular water effuses in response to the change of water potential causing dehydration, shrinkage and osmotic contraction of the cell (Dowgert and Steponkus, 1984, Ruelland et al., 2009). The lower the temperature, the higher the degree of dehydration of plant cells (Gusta et al., 2004) causing a deterioration of the intracellular structures and cell death (Figure 2). Dehydrated cells reduce their size. This can happen in a reversible (exocytic extrusions) or an irreversible (endocytotic vesicles) manner. The reversible process enables an expansion of the cell to its former state, whereas the irreversible process leads to lysis during thawing (Ruelland et al., 2009).

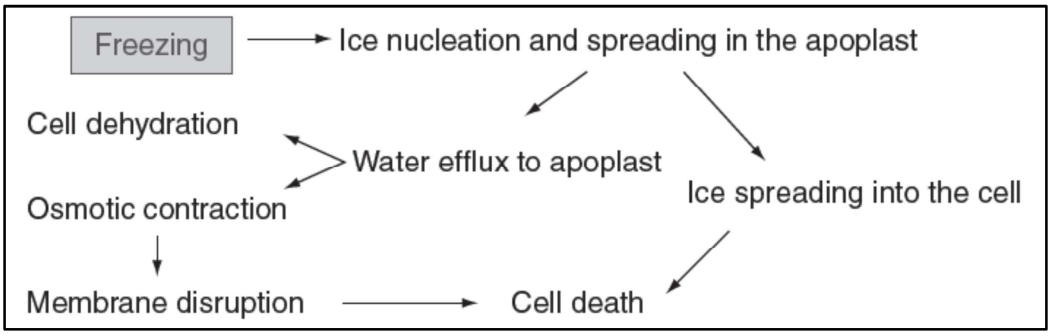


Figure 2: Freezing effects on plant cells (Ruelland et al., 2009).

1.3 Cold stress signalling and transcriptional regulation

Low temperature is one of the most important limiting factors of wheat cultivation in North America and Eastern Europe. To ensure high yields in these areas, introduction of efficient frost tolerance alleles into elite cultivars is a prerequisite. Frost tolerance (FT) is a

complex biological process involving at least two main pathways and many additional processes encompassing a large number of genes. The main pathway is frost response, whereas flowering as alternative pathway involves vernalisation. To avoid yield losses, wheat needs to acclimate to low temperatures, which prevents premature transition to the reproductive phase. This has to take place before the threat of freezing stress during winter has passed (Chinnusamy et al., 2007).

However, plant cells can perceive cold stress in different ways. Light may be of considerable importance in temperature sensing because photosynthetic processes are affected by temperature very fast (Ensminger et al., 2006, Kočová et al., 2009).Furthermore, the DNA-nucleosome composition of H2A.Z plays a key role in perception of temperature as it affects the expression of temperature-sensitive genes (Kumar and Wigge, 2010). Membrane rigidification is involved in temperature sensing as well (Orvar et al., 2000). In consequence, a Ca^{2+} influx into the cytosol arises and these characteristic Ca^{2+} signatures are detected by CALCIUM BINDING PROTEIN (CBPs) (Sangwan et al., 2001, Kaplan et al., 2006). *CALCIUM BINDING* (*CAB*) or *CAM-LIKE* (*CML*) genes, encode proteins composed mostly of EF-hand Ca^{2+}-binding motifs (McCormack and Braam, 2003). As a consequence of cold-induced calcium signalling, INDUCER OF CBF EXPRESSION (ICE) genes are activated. This gene family is regulated by two wheat specific *ICE* genes under cold conditions (Keddie et al., 1998, Chinnusamy et al., 2003). They function as MYC-type basic helix–loop–helix transcription factors which bind to MYC recognition sites of *C-REPEAT BINDING FACTOR* (*CBF*) promoters and consequently activate the expression of these genes (Chinnusamy et al., 2003). The *CBF* transcription factors are a complex gene family and consist of 27 paralogues with 1–3 homologous copies per sub-genome. As such, the gene family contains at least 65 members in total (Mohseni et al., 2012). They are members of the APETALA2/ETHYLENE RESPONSE ELEMENT BINDING PROTEIN (AP2/EREBP) family of DNA-binding proteins (Liu et al., 1998, Stockinger et al., 1997). The AP2/EREBP DNA-binding protein domain comprises a structure with three β-strands and one α-helix (Allen et al., 1998, Dietz et al., 2010, Peng et al., 2013). Furthermore, PKK/RPAGRxKFxETRHP and DSAWR motifs are present, which are typical features of CBF proteins (Jaglo et al., 2001). In addition, it is reported that the lack of introns is widespread in the sequence of *CBF* genes in *Triticeae* (Skinner et al.,

2005). The CBF transcription factors bind to the C-repeat/dehydration-responsive element (CRT/DRE) and induce the expression of *COLD-RESPONSIVE/LATE EMBRYOGENESIS-ABUNDANT* (*COR*/*LEA*) genes or simply *COR* genes (Allagulova et al., 2003, Close, 1997, Winfield et al., 2010). The CRT/DRE element is a highly conserved CCGAC sequence in the promoter of cold- and dehydration-responsive genes (Gilmour et al., 1998). Several of the *COR* genes are *DEHYDRINS* (*DHN*), which are a distinct biochemical group of LEA proteins (Winfield et al., 2010, Close, 1997, Allagulova et al., 2003) for which 54 different unigenes are described, 23 of which are involved in frost tolerance (Wang et al., 2014b). Further *CORs* are *RESPONSIVE TO ABSCISIC ACID* (*RAB*) and LOW TEMPERATURE-RESPONSIVE (*LT*) genes (Winfield et al., 2010). A functionally described member of *LT* genes is *TRITICUM AESTIVUM COLD-REGULATED 7* (*TACR7*) (Gana et al., 1997) whose expression level is altered during cold hardening in wheat (Kocsy et al., 2010).

Further COR/LEA effectors are sugars, proline, chaperones, photosynthetic enzymes and much more (Winfield et al., 2010). Interesting in the case of FT are effectors like the antifreeze proteins (AFPs) like dehydrins and the chitinases, glucanases and thaumatin-like proteins which inhibit both the ice formation and growth (Yeh et al., 2000, Griffith and Yaish, 2004). Also low molecular sugars which reduce plant cell dehydration (Winfield et al., 2010) as well as ice recrystallisation inhibition proteins (IRI), which are specific for *Pooideae* (Tremblay et al., 2005) play a role in FT. The latter stabilise small ice crystals, thereby preventing the formation of large, cell disrupting units (Sandve et al., 2008). Additionally, the WRKY transcription factors are involved in response to abiotic stimuli (Marè et al., 2004) and may be involved in cold hardening in wheat (Talanova et al., 2009). It is reported that the WRKY transcription factors are co-regulators of the glucanases, chitinases and thaumatin-like proteins (Winfield et al., 2010). A general schematic pathway of cold response in plant cells is presented in Figure 3.

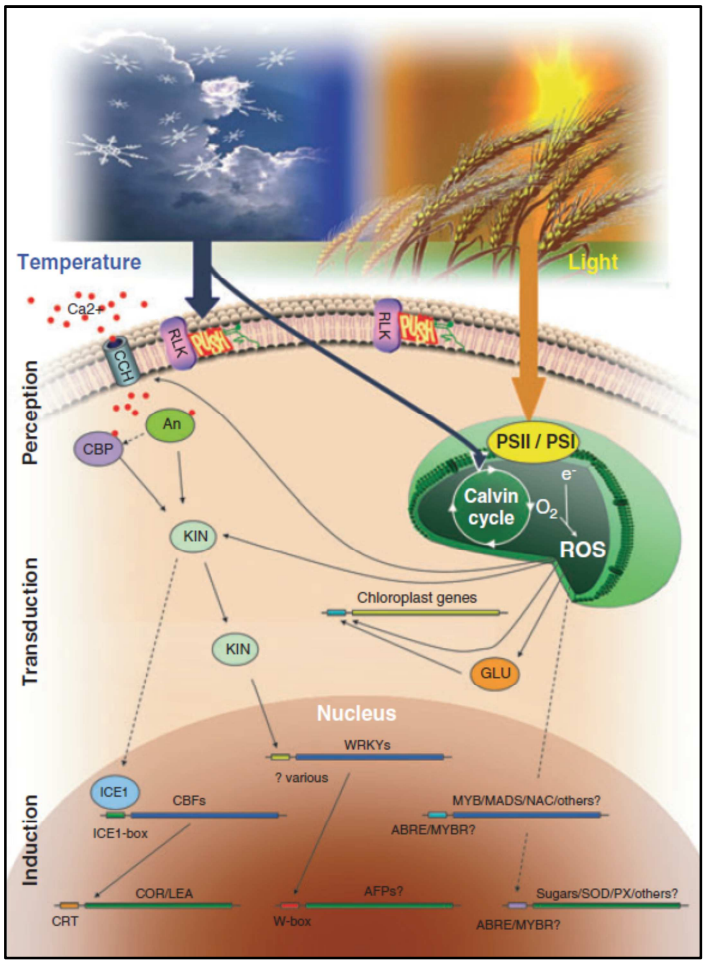


Figure 3: Schematic pathway of the cold response in plant cells. The blue rods in the nucleus represent transcription factor genes, the green rods represent response genes. Each gene has at least one *cis*-acting promotor region, but several may be present. The red dots depict calcium ions; An, annexins; CBP, calcium-binding protein; ABRE, ABA response element; CCH, calcium channel; GLU, glutathione; KIN, kinases and phosphatases; ICE1, inducer of CBF expression 1; RLK, receptor-like kinase, ROS, reactive oxygen species (Winfield et al., 2010).

The flowering pathway is also involved in FT because of genes involved in vernalisation (*VRN*) and photoperiod response (*PPD*) that contribute to low temperature acclimation (Dhillon et al., 2010, Kato and Yamagata, 1988). This pathway is regulated by five major *VRN* genes (*VRN1-5*) and two *PPD* genes (*PPD1/2*) (McIntosh et al., 2013). PPD1 is a member of the PSEUDO RESPONSE REGULATOR (PRR) protein family and interacts with CONSTANS (CO) (Turner et al., 2005). This family possesses a pseudo-receiver domain and a CO motif (Matsushika et al., 2000, Strayer et al., 2000). *VRN1-3* act downstream of *PPDs*. *VRN1* encodes a MADS-box transcription factor, *VRN*2 is similar to a putative zinc finger and a CCT domain and *VRN3* encodes a HEADING DATE 3a (HD3a) protein (Danyluk et al., 2003, Trevaskis et al., 2003, Yan et al., 2003, Yan et al., 2004b, Yan et al., 2006). Both pathways, the flowering and the cold response pathway, are connected by the

interaction of *VRN1* and *CBF* genes. For example, VRN1 is able to reduce the transcript levels of *CBFs* and *COR* genes under long day conditions (Dhillon et al., 2010). Figure 4 displays components of the flowering and FT response pathways that are in the focus of this thesis.

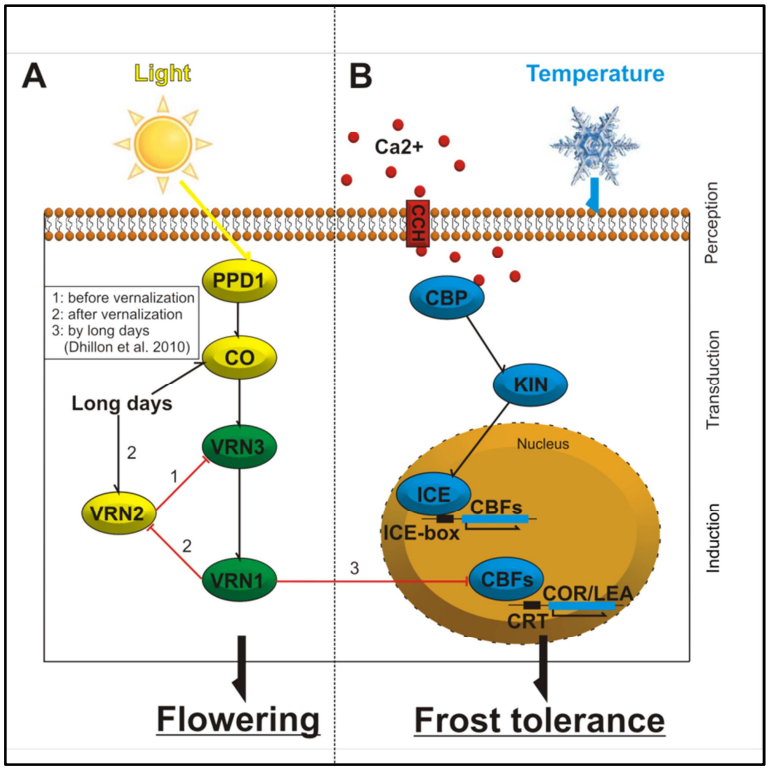


Figure 4: Combined model of flowering and frost tolerance. A Schematic presentation of flowering. The *VRN3* gene interacts with photoperiod (*PPD1-CO* [*CONSTANS*]), vernalisation (*VRN1-VRN2*) and other environmental signals (e.g. stress). Vernalisation induces *VRN1* under short and long days. The interaction among the three vernalisation genes leads to an irreversible induction of flowering(modified from Dhillon et al. (2010) and Distelfeld et al. (2009)). **B** Schematic presentation of cold response in plant cells. The blue rods in the nucleus represent the coding regions of genes and the black rods are *cis*-acting promotor regions. The red dots depict calcium ions; CBP, calcium-binding protein; calcium channel, CCH; kinases and phosphatases, KIN; cold-responsive genes, CORs; late embryogenesis-abundant genes, LEAs (modified from Winfield et al. (2010)).

1.4 Current knowledge about frost tolerance of wheat

Two major FT loci, *FROST RESISTANCE 1* (*FR1*) and *FROST RESISTANCE 2* (*FR2*), were identified on the long arm of chromosome 5A of wheat (Francia et al., 2004, Vagujfalvi et al., 2003). Zhao et al. (2013) described an additional FT Quantitative Trait Locus (QTL) on chromosome 5B in wheat germplasm from central Europe. Due to the importance of cold acclimation in winter and spring wheat, the locus *FR1* was physically and genetically mapped (Galiba et al., 1995, Sutka et al., 1999). However, it is not clear whether *FR1* is an

independent gene or whether it is triggered by a pleiotropic effect of *VRN1* (Dhillon et al., 2010, Stockinger et al., 2007). As a consequence of the presence of the A, B and D genome, there are three homologous *VRN1* genes (*VRN-A1*, *VRN-B1* and *VRN-D1*) differing in their impact on vernalisation. Wheat plants with a dominant *VRN-A1* allele are spring type and do not need vernalisation for flowering, whilst the dominant *VRN-B1* and *VRN-D1* alleles also result in spring habit, however weaker compared to *VRN-A1* (Santra et al., 2009). The difference between the spring (dominant *VRN-A1* alleles) and winter varieties (recessive *vrn-A1* alleles) is based on a C/T SNP in the fourth exon of the *VRN-A1* gene. The winter varieties contain a T whereas the spring varieties carry a C. An ambiguous nucleotide Y (C/T) on this position also results in winter varieties. This ambiguous nucleotide leads to a higher frost tolerance compared to genotypes carrying the *VRN-B1* or *VRN-D1* and spring varieties carrying a C (Koemel et al., 2004, Chen et al., 2009, Chen et al., 2010, Eagles et al., 2011, Reddy et al., 2006). Zhu et al. (2014) reported increased *VRN-A1* copy numbers are associated with improved frost tolerance. Plenty of studies identified the *FR-A2* locus on chromosome 5A as the most important locus involved in FT in wheat (Baga et al., 2007, Motomura et al., 2013, Vagujfalvi et al., 2005). The *FR-A2* locus includes at least 11 *CBF* genes and is located approximately 30 cM proximal to *VRN1* (Francia et al., 2004, Vagujfalvi et al., 2003, Miller et al., 2006). Two independent studies illustrate that *CBF-A3*, which is located in the *FR-A2* locus, plays an important role in wheat FT (Vagujfalvi et al., 2003, Sutton et al., 2009). Knox et al. (2008) analysed the *FR-A^{m}2* locus of diploid *Triticum monococcum* (A^m genome is very similar to the A genome of hexaploid wheat) and identified three *CBFs* (*CBF12*, *CBF14* and *CBF15*) highly associated with FT. Also Vagujfalvi et al. (2005) identified *CBF14* and *CBF15* as FT associated in *Triticum monococcum* and Soltesz et al. (2013) confirmed this for *Triticum aestivum*. Zhu et al. (2014) reported copy number variation (CNV) for all three genes in the *FR-A2* locus, i.e. *CBF-A12*, *CBF-A14* and *CBF-A15*, to have an influence in winter survival. Likewise, Würschum et al. (2017) described that CNV of *CBF* genes at the *FR-A2* locus is the essential component for winter survival, with CBF-A14 CNV being the most likely causal polymorphism. Additionally, Kocsy et al. (2010) identified three genes, i.e. *TACR7*, *CAB* and *DEM* (*DEFECTIVE EMBRYO AND MERISTEMS*) being differentially expressed during cold hardening in wheat. In addition, on the transcriptome level FT signalling is much more complex. Hundreds to thousands of wheat genes were identified

to be significantly up- or downregulated under low temperature (Winfield et al., 2010, Gulick et al., 2005, Monroy et al., 2007, Laudencia-Chingcuanco et al., 2011, Ganeshan et al., 2011, Winfield et al., 2009).

1.5 The wheat genome sequence and development of genomic markers

For breeding of frost tolerant varieties, fundamental knowledge of genes involved in the regulation of FT (as described above) can be the starting point for the development of diagnostic or gene-derived functional molecular markers (Bagge et al., 2007). Today, marker-assisted selection (MAS) is a basic tool in plant breeding. Furthermore, re-sequencing of target genes allows efficient allele mining (Kaur et al., 2008). However, the development of gene specific primers in wheat is hampered by the large genome size of 17 Gbp, the high repeat content of about 80% (Smith and Flavell, 1975, Choulet et al., 2010), by the close homology of the three genomes (A, B and D), and by the high rate of similarity within genes and gene family members (Brenchley et al., 2012). Comparative analysis of wheat sub-genomes shows high sequence homology and structural conservation, and no significant differences in the rate of duplications between the sub-genomes are observed (IWGSC, 2014). Recent efforts of the scientific community and the IWGSC in sequencing of the three donor genomes as well as of hexaploid wheat offer a solution in deciphering the intron-exon-structure of genes (Brenchley et al., 2012, Raats et al., 2013, Appels et al., 2018, IWGSC, 2014, Ling et al., 2013, Jia et al., 2013). By using intron sequence variation among the homologous and paralogous copies of the various genes, it is possible to reconstruct the gene structure and identify differences between homologues. Continuous improvements of BLAST algorithms enhance the use of the above mentioned wheat genomic resources of the Chinese Spring IWGSC RefSeq v1.0 (Appels et al., 2018), facilitating efficient primer development. Furthermore, specific primers are the basis for the development of molecular marker assays based on SNPs, i.e. cleaved amplified polymorphic sequence (CAPS) (Perovic et al., 2013), pyrosequencing (Silvar et al., 2011), or competitive allele-specific polymerase chain reaction (KASP) (Allen et al., 2011). These types of genomic markers are not only used for MAS, they also enable anchoring physical and sequence contigs (Raats et al., 2013, Appels et al., 2018) as well as germplasm characterisation (Kilian and Graner, 2012).

1.6 Association analysis

Genome-wide association studies (GWAS) are a powerful tool to identify genomic regions and candidate genes involved in FT. High density genotyping arrays i.e. the Illumina 90 K chip (Wang et al., 2014a), the Affymetrix 820 K (Winfield et al., 2016) and the breeding 35 K axiom (Allen et al., 2016) arrays enable the determination of the genetic structure of complex traits by GWAS. However, they are limited in the identification of new alleles involved in FT. One approach that allows mining of novel alleles is re-sequencing of candidate genes followed by a candidate gene association genetics study (Li et al., 2011a). Obviously, this requires successful development of gene-derived functional markers. As described above, in wheat these markers must be able to distinguish homologous sequences on the three sub-genomes. Given these prerequisites, their diagnostic potential can be a valuable addition to classic GWAS.

1.7 Objectives

The aims of this thesis were to (i) develop an efficient genome specific primer design approach for genes involved in frost tolerance (FT), (ii) to sequence candidate genes for FT and identify polymorphisms in these genes, and (iii) to conduct a candidate gene based association genetics approach to get information on the effect of respective SNPs and indels on FT.

Based on the complexity of the wheat genome it is difficult to design gene and genome specific primers for target candidate gene re-sequencing. The reference wheat sequence data enables the development of an efficient approach for specific primer design for FT candidate genes facilitating re-sequencing of candidate gene amplicons and the identification of SNPs and indels. To date no candidate gene based association study has been conducted in wheat for FT. With the aid of the approach developed for primer design FT candidate genes in 235 wheat varieties will be re-sequenced and an association study based on these genotypic data and phenotypic data obtained at five locations in two years will be conducted.

2 Material and Methods

2.1 Specific primer design in hexaploid wheat (*Triticum aestivum* L.)*

2.1.1 Plant material and DNA extraction

In this thesis three cultivars (`Chinese Spring`, `Moskovskaya 39` and `VAKKA`) were used for initial testing of designed primer pairs, while a set of 24 genotypes, comprising two spring and 22 winter wheat cultivars, was used for re-sequencing of amplicons of FT genes (Table 1). For the physical assignment to chromosomes and chromosome segments, 21 nulli-tetrasomic (NT)-lines (Sears, 1966) and 46 deletion-lines (Endo and Gill, 1996) were used (Supplemental Table 1) having the genetic background of 'Chinese Spring'. The DNA was extracted at the three leaf stage according to Stein et al. (2001).

Table 1: Plant material for PCR amplification and re-sequencing.

No.	Genotype	Country	Variety
1	Chinese Spring*	China	spring
2	Zentos	Germany	winter
3	Simila	Czech Republic	winter
4	Roughrider	USA	winter
5	Norstar	USA	winter
6	Moskovskaya 39*	Russia	winter
7	Bezenchukskaja 380	Russia	winter
8	Cheyenne	USA	winter
9	ÄRING II	Sweden	winter
10	VAKKA*	Finland	winter
11	Bezostaja 1	Russia	winter
12	Capelle Desprez	France	winter
13	Centurk	USA	winter
14	Mironovska 808	Ukraine	winter
15	Pobeda	Serbia	winter
16	Renesansa	Serbia	winter
17	Sava	Serbia	winter
18	Triple Dirk B (GK 775)	Australia	winter

* from Babben et al. (2015)

Table 1: (Continued)

No.	Genotype	Country	Variety
19	Triple Dirk S	Australia	spring
20	ISENGRAIN	France	winter
21	APACHE	France	winter
22	SKAGEN	Germany	winter
23	JULIUS	Germany	winter
24	Biryuza	Russia	winter
25	Moskovskaya 40	Russia	winter

Complete set of 24 genotypes (without `Chinese Spring`) were used for sequencing. No.: number; * Genotypes for primer testing.

2.1.2 Sequence retrieval of genes involved in frost tolerance

As a starting point a set of 27 genes involved in FT was selected. Nine *Triticum aestivum* sequences together with nine sequences from *Triticum monococcum* and nine from *Hordeum vulgare*, known to be involved in FT from previous studies, served as a back bone for the identification of bread wheat FT candidate gene sequences (Table 2). If only the coding regions (mRNA-, EST- or protein-sequences) were available, the databases of the International Wheat Genome Sequencing Consortium (IWGSC, http://www.wheatgenome.org/) and/or the Bristol Wheat Genomics (http://www.cerealsdb.uk.net/) were used for the identification of the full genomic sequence and subsequent reconstruction of the gene structure. The BLAST algorithm parameters were set as default.

Table 2: List of identified frost tolerance candidate gene sequences.

Candidate gene	Species	Cultivar	Accession	Type	Citation
CBF1	*Triticum aestivum*	Winoka	AF376136	Gene/CDS	(Jaglo et al., 2001)
CBF4	*Triticum monococcum*	n.a.	AY951945	Gene/CDS	(Miller et al., 2006)
CBF5	*Triticum monococcum*	n.a.	AY951947	Gene/CDS	(Miller et al., 2006)
CBF7	*Triticum monococcum*	DV92	AY785904	Gene/CDS	(Skinner et al., 2005)
CBF8	*Hordeum vulgare*	Tremois	DQ445252	Gene/CDS	(Knox et al., 2010)
CBF10	*Triticum monococcum*	n.a.	AY951950	Gene/CDS	(Miller et al., 2006)
CBF13	*Triticum monococcum*	n.a.	AY951951	Gene/CDS	(Miller et al., 2006)
CBF14	*Triticum monococcum*	n.a.	AY951948	Gene/CDS	(Miller et al., 2006)
CBF15	*Triticum aestivum*	Norstar	EF028765	Gene/CDS	(Badawi et al., 2007)
CBF16	*Triticum monococcum*	G3116	EU076384	Gene/CDS	(Knox et al., 2008)
CBF17	*Triticum monococcum*	n.a.	AY951945	Gene/CDS	(Miller et al., 2006)
CBF18	*Triticum monococcum*	n.a.	AY951946	Gene/CDS	(Miller et al., 2006)

Table 2: (Continued)

Candidate gene	Species	Cultivar	Accession	Type	Citation
DHN1	*Hordeum vulgare*	Dicktoo	AF043087	Gene/CDS	(Choi et al., 1999)
DHN3	*Hordeum vulgare*	Dicktoo	AF043089	Gene/CDS	(Choi et al., 1999)
DHN4	*Hordeum vulgare*	Barke	BQ466915	EST	(Zhang et al., 2004)
ICE2	*Hordeum vulgare*	Morex	DQ113909	Gene/CDS	(Skinner et al., 2006)
VRN-A1	*Triticum aestivum*	Triple Dirk C	AY747600	Gene/CDS	(Fu et al., 2005)
VRN-B1	*Triticum aestivum*	Triple Dirk B	AY747603	Gene/CDS	(Fu et al., 2005)
VRN-D1	*Triticum aestivum*	Triple Dirk C	AY747606	Gene/CDS	(Fu et al., 2005)
VRN2	*Hordeum vulgare*	Dairokkaku	AY485977	partial CDS	(Yan et al., 2004b)
VRN3	*Triticum aestivum*	Chinese Spring	DQ890162	Gene/CDS	(Yan et al., 2006)
CAB	*Hordeum vulgare*	Barke	BQ465487	EST	(Zhang et al., 2004)
DEM	*Hordeum vulgare*	Barke	AL504294	EST	(Michalek et al., 2002)
TACR7	*Hordeum vulgare*	Golden Promise	BQ659345	EST	(Zhang et al., 2004)
PPD-A1	*Triticum aestivum*	Chinese Spring	DQ885753	Gene/CDS	(Beales et al., 2007)
PPD-B1	*Triticum aestivum*	Chinese Spring	DQ885757	Gene/CDS	(Beales et al., 2007)
PPD-D1	*Triticum aestivum*	Chinese Spring	DQ885766	Gene/CDS	(Beales et al., 2007)

n.a.: not available

2.1.3 Reconstruction of intron-exon-structure and gene specific primer development

The reconstruction of the gene intron-exon-structure was performed using the internet platform 'Spidey' (http://www.ncbi.nlm.nih.gov/spidey/spideyweb.cgi) from NCBI, which allows alignment of mRNA to genomic sequences. The intron/UTR regions sequences were used for primer development. The next step was the identification of the best hits to the three different wheat genomes on the IWGSC and/or the Bristol Wheat Genomics website via BLASTn. After collecting three homologous sequences of each targeted gene, the gene structure was reconstructed for each one separately and then used for multiple alignments. Multiple alignments were constructed by using Sequencer 5.1 (Gene Codes Corporation, Ann Arbor, USA) and CLC Main Workbench 7.6 (CLC Bio, Aarhus, Denmark) software and visually inspected for unique stretches among three homologues. The polymorphisms between the three homologous genomes of each gene were detected and used for specific primer development. The primers were developed by using 'Primer3' (v. 0.4.0) (Koressaar and Remm, 2007, Untergasser et al., 2012). Parameters utilised for primer development were set to a maximal 3`stability of 50, primer size between 19 and 28 bp and primer melting temperature between 57° and 63° Celsius. The

maximal fragment length was set to 1200 bp, while optimal fragment length was 900 bp. Other parameters remained as default. Specificity of primers was based on two nucleotide differences within the primer binding site or one difference within the last seven nucleotides at the 3` end of the primer based on the analyses of the three homologous target sequences (Wu et al., 2009). All primers were designed to bind locus specific sequences within the introns/UTR regions of selected genes. At least one primer of a primer pair had to be locus specific for single band amplification.

2.1.4 PCR amplification and fragment analysis

Newly designed PCR primers were amplified in two different reaction volumes i.e. firstly, in a volume of 10 µl for functionality testing and chromosomal assignment, and secondly in a 20 µl reaction volume for re-sequencing. The PCR reactions comprised two different polymerases, FIREPol® DNA polymerase (Solis BioDyne, Tartu, Estonia), in a first round of testing, and MyTaq™ DNA polymerase (BIOLINE, Luckenwalde, Germany), in a second round of testing in case the FIREPol product was very weak, with 50 ng of genomic DNA. The master mix for one PCR reaction comprised 0.4 U FIREPol® DNA Polymerase, 1 x Buffer B, 2.5 mM $MgCl_2$ (Solis BioDyne, Tartu, Estonia), 0.2 mM dNTPs (Fermentas, St. Leon-Rot, Germany) and 0.25 pmol primers (Microsynth, Balgach, Switzerland) or 0.4 U MyTaq™ DNA Polymerase, 1 x My Taq Reaction Buffer B (that comprised 1 mM dNTPs and 3 mM $MgCl_2$) (BIOLINE, Luckenwalde, Germany) and 0.25 pmol primers. The fragment amplification was conducted in a thermal cycler GeneAmp® PCR System 9700 (Applied Biosystems, Darmstadt, Germany) under various PCR profiles (Supplemental Table 2).PCR fragments were separated by using agarose gel electrophoreses and analysed using the imaging system Gel Doc™ XR and the Quantity One® 1-D analysis software (4.6.2) (Bio-Rad, Hercules, USA).

2.1.5 PCR fragment mapping by using NT- and deletion lines

All specific and single banded PCR fragments were assigned to chromosomes by using 21 nulli-tetrasomic (NT) lines (Sears, 1966) and by a set of 46 deletion-lines (Endo and Gill,

1996). The information about chromosomal localisation of these gene specific amplicons was compared to published results. The map of specific PCR fragments was printed via LaTeX 4.4.1 software.

2.1.6 *In silico* analysis of primer sub-genome specificity

A set of 98 primers used for the amplification of 65 PCR fragments with correct chromosomal localisation were *in silico* validated for sub-genome specificity by aligning to the draft sequence of wheat. The primers were aligned via Multiple Alignment using Fast Fourier Transform (MAFFT, http://www.ebi.ac.uk/Tools/msa/mafft/), CLC and Sequencher. Parameters for the Sequencher based alignment were as follows: clean data with minimum overlap of 19 nucleotides and minimum match percentage of 90%, while CLC and MAFFT parameters were as default. The differences between the sub-genome sequences and designed primers were manually inspected. Primers with sub-genome specificity were those having two or more differences in binding site or at least one difference at the last seven nucleotide bases at 3` end of primer.

2.1.7 Re-sequencing of frost tolerance candidate genes and BLAST verification

Sequencing of PCR fragments was performed by Microsynth AG (Balgach, Switzerland) using the Sanger sequencing method (Sanger et al., 1977). First sequencing reactions were performed with primers used for amplification and if quality was lower than 70% an optimisation with redesigned oligos was conducted. Subsequently all fragment sequences were compared to reference sequences and/or candidate genes of related species by using the NCBI MegaBlast function (Tan et al., 2006). The results were limited to five hits, minimum expect threshold of e^{-100} and minimum identity of 85%. All other parameters remained as default. The haplotype diversity (Hd), the nucleotide diversity and the average number of nucleotide diversity in a set of 24 analysed wheat cultivars were calculated using the DnaSP 5.1 freeware software (Rozas et al., 2003, Librado and Rozas, 2009).

2.2 Candidate gene association studies on frost tolerance in wheat (*Triticum aestivum* L.)**

2.2.1 Plant material, DNA extraction and re-sequencing

For further analysis a set of additionally 211 bread wheat genotypes was used for PCR amplification, amplicon sequencing and association genetics studies. In total 235 bread wheat genotypes, comprising 179 cultivars, 48 lines and 8 doubled haploid (DH) lines originating from 28 countries and five continents (Supplemental Table 3) were used. The association panel was selected based on pre-existing knowledge regarding the reaction to growing conditions during winter time, i.e. high latitude and continental European winter wheat collections as well as Russian and North American cultivars. Furthermore, the core collection of the Institute of Field and Vegetable Crops (IFVCNS), Novi Sad, Serbia (Neumann et al., 2011) and parental lines of Western European hybrid breeding programs were included. The method of DNA extraction and fragment re-sequencing are described in chapter 2.1.1 and 2.1.7.

2.2.2 Field experiments and phenotypic data analysis

The field experiments were performed in five environments during 2012 (Gatersleben, Germany; Ranzin, Germany; Puskin, Russia; Roshchinskiy, Russia; Novosibirsk, Russia) and 2013 (Gatersleben, Germany; Ranzin, Germany; Puskin, Russia; Roshchinskiy, Russia) and one in 2014 (Novosibirsk, Russia). All 235 genotypes were tested in Gatersleben, Ranzin, Pushkin, and Novosibirsk in a random design in double rows and two replications per genotype. The trial in Roshchinskiy was conducted as a miniplot (2.5 m^2) trial with one replication. FT was evaluated as winter survival in per cent (%), i.e. the survival of plants per genotype and plot was measured as a quantitative trait (%) ranging from 0% (all dead) to 100% (all alive) after winter.

To take into account the diversity of the environments with respect to climatic conditions, a co-variable comprising the number of days under -15 °C in the period from December 1st to April 30th of each year was calculated. The correlation coefficient r (Pearson, 1914)

** from Babben et al. (2018)

was calculated between the co-variable and FT. Principal component analysis (PCA) was used to get information on the influence of the environment on FT. Correlation coefficient r and PCA were calculated with the JMP® Genomics 5.1 software (SAS, Cary, USA). Data measured at the same location in different years were treated as independent. All data sets that exhibit a deviation described by the second component of PCA analysis indicate an atypical trait value putatively caused by secondary environmental factors and were discarded. Furthermore, the coefficient of variation (CV; standard deviation divided by arithmetic average) was calculated as well as the variance of each environment and year. The data sets with a very low CV (<0.15) and/or variance (<150), were classified as non-representative and were discarded. After editing of the field data, the significance of the influence by environment and genotype was tested by using the analysis of covariance (ANCOVA) and a general linear model (GLM) procedure. Based on this, the Least-Squares means (LS means) were calculated. For all of these analyses the SAS® 9.4 software (SAS, Cary, USA) was used.

2.2.3 *In silico* primer assignment and candidate gene structure re-construction

The primer assignment was verified based on BLASTs of mRNA and genomic regions of close relatives against the Chinese Spring reference assembly v1.0 using NCBI MegaBlast. Subsequently, 5 kb upstream and downstream regions were extracted based on the BLAST results and the analysis of the NT-lines. Primers were aligned to these genomic regions using a free shift alignment with affine gap costs (gap opening = 5, gap elongation = 0.01, match = −1). If ambiguities were detected at the beginning or the end of exons, primers were manually modified to match the consensus dinucleotides of splice sites, GT and AG.

2.2.4 Detection of polymorphisms (SNPs/indels) and haplotypes

The sequencing raw data were edited using Sequencer 5.1 (Gene Codes Corporation, Ann Arbor, USA). Next, the adjusted sequences of each primer pair were aligned by using the Multiple Alignment tool Fast Fourier Transform (MAFFT) (Katoh et al., 2002). MAFFT

parameters were set as default. The polymorphisms between the 235 genotypes were detected automatically via a small multiple sequence alignment (MSA) script (Supplemental Data 1) in the free software Java™. Parameters for the polymorphism detection were as follows: polymorphisms between defined bases (A, T, C or G) and ambiguous nucleotides (N) were ignored. The detected polymorphisms were used for the identification of haplotypes and components of haplotypes for each candidate gene according to the position in promoter, exon, intron or in the 3` untranslated region (UTR).

2.2.5 Population structure and kinship calculation

In order to account for population structure effects in association studies, the population structure was estimated based on 249 SNPs, chosen according to the map location and even distribution along the 21 wheat chromosomes (Wang et al., 2014a, Li et al., 2011b). Population structure of wheat accessions was assessed using STRUCTURE v 2.3.3, which is based on a Bayesian model-based clustering algorithm that incorporates admixture and allele correlation models to account for genetic material exchange in populations, resulting in shared ancestry (Pritchard et al., 2000). Five independent runs were performed setting the number of populations (k) from 1 to 10, burn in time and Markov Chain Monte Carlo (MCMC) replication number both to 100,000. The k-value was determined by ln P(D) in STRUCTURE output and an ad hoc statistic Δk based on the rate of change in ln P(D) between successive k-values (Evanno et al., 2005). Wheat lines with probabilities ≥0.5 were assigned to corresponding clusters. Lines with probabilities <0.5 were assigned to a mixed group. The population structure plot was constructed by using STRUCTURE PLOT (Ramasamy et al., 2014) and the Principal coordinate analysis (PCoA) by using the software package DARwin (http://darwin.cirad.fr). The kinship (K) matrix was calculated on a modified Roger's distance (Goodman and Stuber, 1983, Reif et al., 2005, Wright, 1978) by using R version 3.2.1 free software (https://www.r-project.org). The Roger's distance was calculated as follows:

$$\boldsymbol{d_w = \frac{1}{\sqrt{2m}} \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{nj} (p_{ij} - q_{ij})^2}} \quad (1)$$

where pij and qij are allele frequencies of the jth allele at the ith locus, n number of alleles at the ith locus and m number of loci.

2.2.6 Association genetics analysis

SNP and indel association genetics analysis

The SNP/indel association analysis was performed with a set of 235 genotypes by using TASSEL 5.0.9 (Bradbury et al., 2007) (http://www.maizegenetics.net/tasse). Only the SNPs/indels with minor allele frequencies (MAF) >5% were taken into consideration for analysis. Furthermore, population structure, kinship matrix and phenotypic LS means were included in association studies applying the mixed linear model (MLM) algorithm. The threshold of statistically significant effects was set to 1.30 –log10 of *P* (*P*-value of 0.05) according to Li et al. (2011a, 2011b) who used this threshold for the analysis of candidate genes at chromosomal regions with high linkage disequilibrium (LD). The LD was calculated via TASSEL 5.0.9 by using the full matrix LD type method with 8064 comparisons after MAF selection.

Haplotype association genetics analysis

The haplotype association genetics study was performed by using TASSEL 4.1.20 (Bradbury et al., 2007) (http://www.maizegenetics.net/tasse). The parameters, calculation and significance threshold were the same as used for the SNP and indel association analysis. The same applies to the LD calculation for the genotype groups from Europe, North America and Asia including Australia.

2.2.7 Sequence analysis

The translation of associated gene sequences into amino acid (AA) sequences, identification of homologous AA sequences, AA alignments, identification of protein domains and motifs as well as prediction of secondary protein structures were performed using the following software: CLC Main Workbench 7.6 (CLC Bio, Aarhus, Denmark),

MAFFT, free software Jalview (Waterhouse et al., 2009), NCBI (National Center for Biotechnology) protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cg) and RaptorX (Källberg et al., 2012, Peng and Xu, 2011) (http://raptorx.uchicago.edu/StructurePrediction/predict/). The nucleotide sequences were translated into AA sequences via CLC following BLASTn against the NCBI protein database for homologous AA sequence identification. Parameters were set as default. The alignment of initial and homologous AA sequences was performed using CLC and MAFFT. The last step was the identification of domains and motifs and the prediction of the protein structure via RaptorX. For the quality check of the RaptorX results, the uGDT-value (unnormalised Global Distance Test), uSeqID-value (number of identical residues in the alignment) and *P*-value were taken into account. A uGDT >50, uSeqID >30 and a *P*-value less than $1*10^{-3}$ are indicators for good quality modelling. Nucleotide divergence rates (dN/dS) between the identified haplotypes and reference sequences of *Triticum aestivum* and related species were analysed using a web-based HyPhy application (Delport et al., 2010, Pond et al., 2005). Haplotype and reference sequences were used to generate sequence alignments by applying the L-INS-i option in MAFFT (Katoh et al., 2005). To obtain the best fitting substitution model, the model test application in MEGA-CC was used (Kumar et al., 2012). The reconstruction of the phylogenetic tree was done with maximum likelihood algorithm and 500 bootstraps in MEGA-CC (using the corresponding model). The resulting protein alignment and the corresponding nucleotide sequences were used to compute codon alignments with Pal2Nal (Suyama et al., 2006). The codon alignments and the phylogenetic tree were used to compute dN/dS for each variable site using the SLAC method in HyPhy (Pond et al., 2005).

2.2.8 *In silico* promotor analysis

Identification of promoter regions and regulatory sites was performed using the internet databases SOGO from the National Institute of Agrobiological Sciences (Higo et al., 1999) (https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?lang=en) and Softberry (Solovyev et al., 2010) (http://www.softberry.com/berry.phtml?topic=ann2_ann3&no_menu=on).

3 Results

3.1 Specific primer design in hexaploid wheat (*Triticum aestivum* L.)*

Alignment of candidate gene sequences with corresponding genomic sequences retrieved from the International Wheat Genome Sequencing Consortium, the Bristol Wheat Genomics and NCBI allowed the identification of exon-intron splicing positions, and the identification of coding and non-coding regions. Therefore, reconstruction of the intron-exon structure by using newly available genomic sequences is the basic step towards the development of gene specific primers in polyploid plants such as hexaploid wheat.

3.1.1 Reconstruction of intron-exon-structure and development of gene specific primers

The workflow for the development of gene specific primers and validation regarding PCR specificity, chromosomal localisation and sequence homology contains four steps (Figure 5). In brief, the procedure starts with collecting sequences of candidate genes (described in chapter 2.1.2), followed by the reconstruction of intron and exon structures and sub-genome sequence identification, eventually primer development (described in chapter 2.1.3) and PCR fragment testing (described in chapter 2.1.4). Functionality and validity of PCR fragments were assessed by nullisomic-tetrasomic (NT) mapping (described in chapter 2.1.5), sequencing and BLASTing (described in chapters 2.1.6 and 2.1.7) by using three databases, six tools (‘Spidey’, ‘Primer3’, BLASTn, BLASTx, CLC Main Workbench and Sequencer) and two cytological stocks of wheat.

* from Babben et al. (2015)

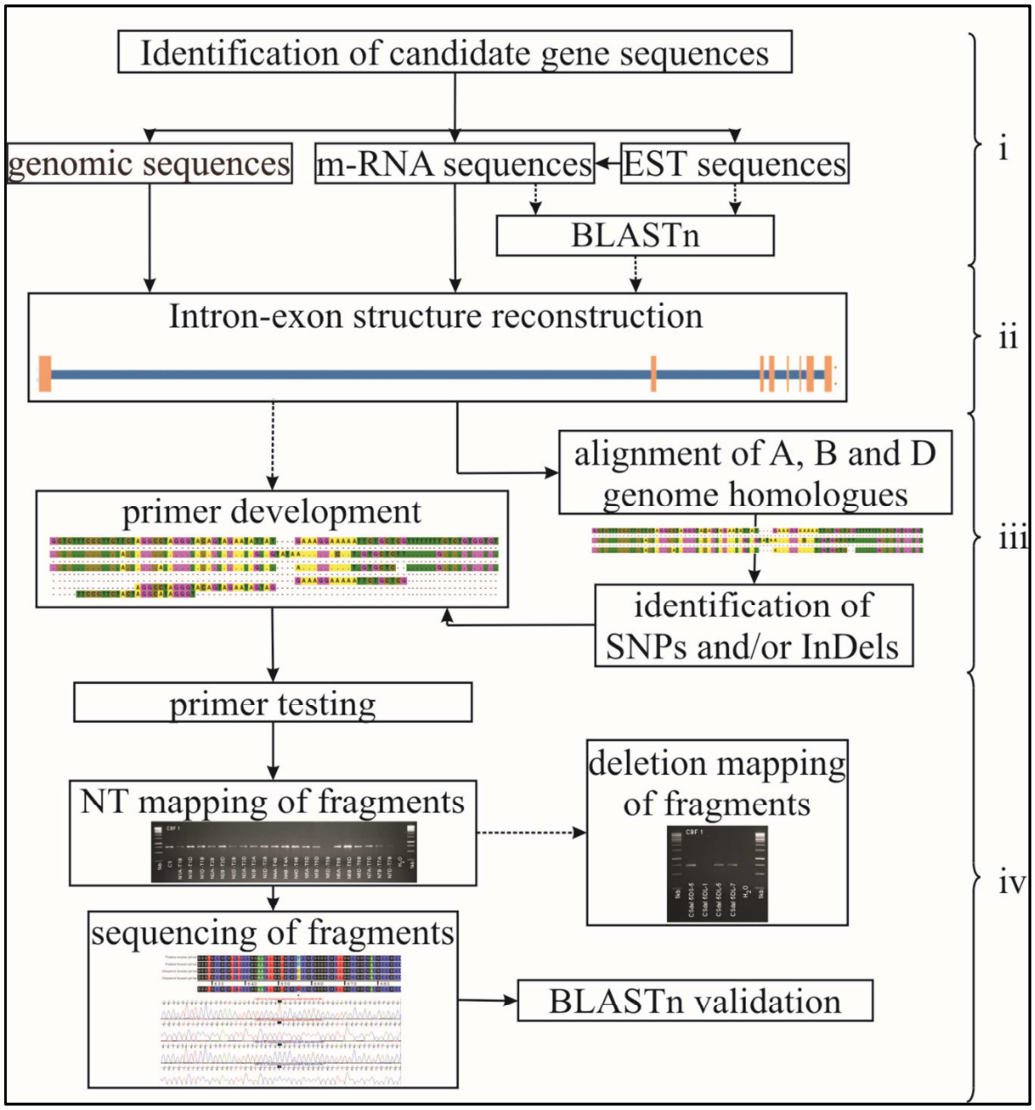


Figure 5: Workflow of development gene specific primers and PCR fragments in wheat. The method comprises four steps, i.e. (i) identification of genomic and coding sequences (CDS) of candidate genes, (ii) intron- and exon-structure reconstruction, (iii) identification of wheat A, B and D sub-genome sequences and primer development based on sequence differences between the three sub-genomes, and (iv); primer and PCR fragment testing for functionality, correct size and localisation. The dashed lines show optional applications (Babben et al., 2015).

For all of the 27 candidates the gene structure or at least a part of it (Table 2) was re-constructed. Fundament for the selection of candidate genes was the description of cold response and sequence data availability in literature and databases, respectively. A set of 119 PCR products was obtained from 157 primers pairs designed in this study, of which 13 have been published in Keilwagen et al. (2014). Additional 12 primers from literature were used for the amplification of target genes. By combining the primers from this study and the 12 primers from the literature a total of 169 primers were analysed. As an example the reconstruction of the three copies of the *VRN1* gene structure, primer positions, intron length differences and exon SNPs are shown in Figure 6.

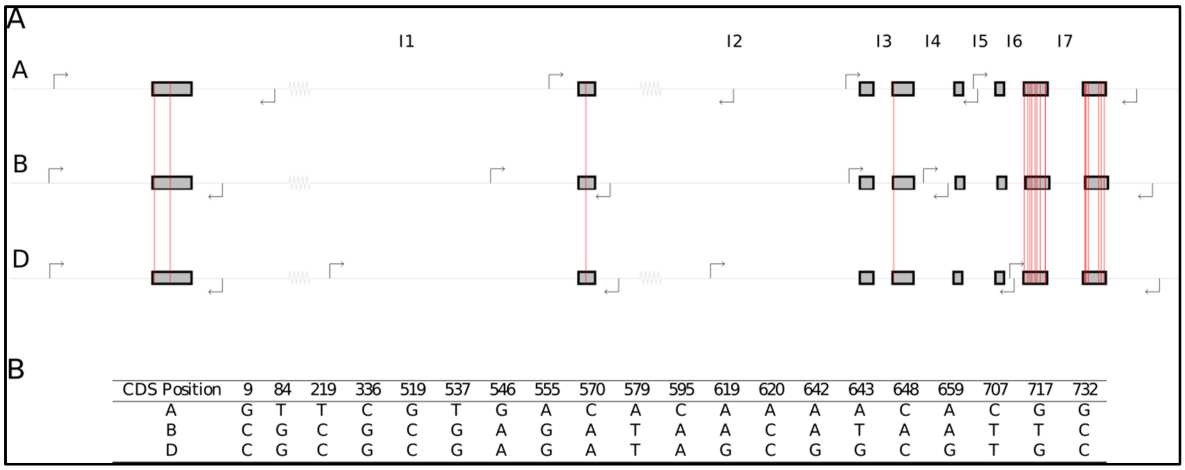


Figure 6: Example for an intron-exon structure, intron length difference, exon SNPs and primer position of the three copies of the *VRN1* gene. A A, B and D on the left border represent the three different wheat sub-genomes, the seven introns of *Vrn1* are labeled I1 to I7. The arrows which bend to the right depict forward primers and the arrows which bend to the left depict reverse primers. The red vertical lines show SNPs between the three gene copies. A I1 has a length of 8518 bp, B I1 2821 bp, D I1 8625bp, A I2 1475 bp, B I2 1246 bp, D I2 1504 bp, A I3 90 bp, B I3 92 bp, D I3 90 bp, A I4 192 bp, B I4 196 bp, D I4 188 bp, A I5 152 bp, B I5 156 bp, D I5 156 bp, A I6 93 bp, B I6 91 bp, D I6 91 bp, A I7 166 bp, B I7 168 bp and D I7 168 bp. **B** This figure shows the SNPs between the three sub-genomes and their coding sequences (CDS) position (Babben et al., 2015).

3.1.2 Testing primers for specificity and chromosomal assignment of PCR products

In total, a set of 169 primers representing 119 PCR products from 27 candidate genes was tested for functionality and specificity. A set of 86 primer combinations from 23 candidate genes showed single band amplification (72.3%).

Chromosomal localisation via NT-lines of Chinese Spring (Sears, 1966) of the 86 single band PCR amplicons revealed that 65 fragments were located on expected chromosomes

according to literature (54.6% success rate). Out of these 65 fragments, six were products of a combination of already published and newly designed primers. The remaining 21 fragments showed an incorrect localisation (literature *vs*. NT-lines) or no localisation was possible as all NT-lines showed a fragment. Figure 7 shows an example of the *CBF-D1* amplicon localisation via NT- and deletion-lines. Correctly assigned amplicons originated from 19 genes and were located on 11 wheat chromosomes (Table 3, Figure 8). The described 65 amplicons are gene specific and represent 19 frost tolerance (FT) genes. They were, therefore, selected for further studies (Table 4, Supplemental Table 2).

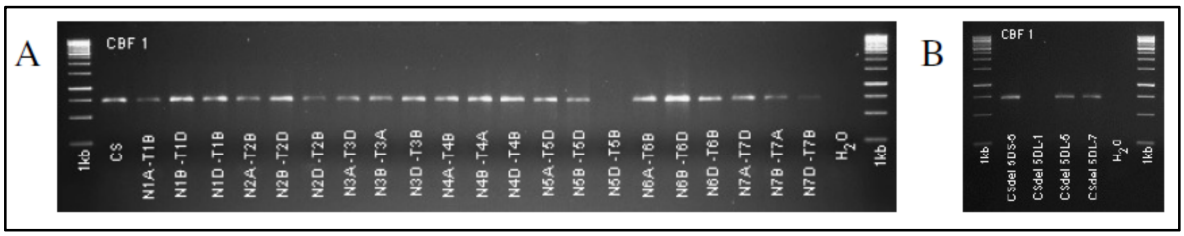


Figure 7: Example of fragment localisation from *CBF-D1* via NT- and deletion-lines. A The missing PCR fragment on NT-line N5D-T5B indicated the location on wheat chromosome 5D. **B** The missing PCR fragment on Csdel 5DL-1 indicated the location on the long arm of chromosome 5D between the deletion segments 1 and 5 (Babben et al., 2015).

Furthermore, 40 (out of 65) amplicons were physically assigned using a set of 46 available deletion-lines (Endo and Gill, 1996) (Figure 8, Table 3). All six genes, which are localised on chromosome 5A via NT-lines, were mapped to a large cluster between sector AL-12 and AL-17 on the long arm of chromosome 5.

Table 3: Overview of chromosome localisation of candidate genes (PCR fragment) via NT-lines, deletion-lines and literature.

Gene	NT-lines	PCR signal present via deletion-lines	Deletion-line localisation section	Literature location	Reference
CBF1	5D	5DS-5; 5DL-5,-7	proximal from 5DL-5 and distal from 5DL-1	n.a	
CBF5	7A	7AL-1	distal on short arm from 7AS-1	$7A^m$	(Miller et al., 2006)
CBF7	2B	2BS-1,-3,-4	distal on long arm from 2BL-6	n.a	
CBF10	5A	5AS-3,-10; 5AL-17,-23	proximal from 5AL-12 and 5AL-17	$5A^m$	(Miller et al., 2006)
CBF13	5A	5AS-3,-10; 5AL-17,-23	proximal from 5AL-12 and 5AL-17	$5A^m$	(Miller et al., 2006)
CBF14	5A	5AS-3,-10; 5AL-17,-23	proximal from 5AL-12 and 5AL-17	$5A^m$	(Miller et al., 2006)
CBF15	5A	5AS-3,-10; 5AL-17,-23	proximal from 5AL-12 and 5AL-17	$5A^m$	(Miller et al., 2006)
CBF18	6A	6AS-1	distal on long arm from 6AL-8	$6A^m$	(Miller et al., 2006)
DHN1	5D	5DS-5; 5DL-5	proximal from 5DL-5 and distal from 5DL-1	5H	(Kosova et al., 2008)
VRN-A1	5A	5AS-3,-10; 5AL-17,-23	proximal from 5AL-12 and 5AL-17	5A	(Yan et al., 2004a)
VRN-B1	5B	5BS-4,-5,-6,-8; 5BL-16	proxiaml from 5BL-16 and distal from 5BL-9	5B	(Yan et al., 2004a)
VRN-D1	5D	5DS-5; 5DL-7	distal on long arm from 5DL-5	5D	(Yan et al., 2004a)
VRN2	4D	everywhere (4DS-1,-2;-3; 4DL-9,-13)	proximal from 4DS-1 and 4DL-9	$5A^mL$	(Dubcovsky et al., 1998)
VRN3	7B	7BL-7;-6;-10	distal on short arm from 7BS-1	7BS	(Yan et al., 2006)
CAB	5A	5AS-3,-10; 5AL-17,-23	proximal from 5AL-12 and 5AL-17	5HL	(Kocsy et al., 2010)
DEM	6B/6D	6BS-2; 6BL-3,-5,-6; 6DS-4,-6; 6-DL 10	distal on short arm from 6BS-3 and proximal from 6DL-11 and distal from 6DL-12	6HL	(Kocsy et al., 2010)
TACR7	2B	everywhere (2BS-1,-3,-4; 2BL-6)	proximal from 2BS-1 and 2BL-6	2HL	(Kocsy et al., 2010)
PPD-B1	2B	2BS-3; 2BL-6	proximal from 2BS-1 and distal from 2BS-4	2B	(Scarth and Law, 1983, Law et al., 1978, Welsh et al., 1973)
PPD-D1	2D	2DL-9	distal on short arm from 2DS-5	2D	(Law et al., 1978, Welsh et al., 1973)

The table shows the analysed frost tolerance candidate gene, their chromosomal localisation and fine mapping via NT and deletion-lines. The column deletion-line localisation section shows the approximate chromosomal position of respective genes based on deletion break points. n.a.: not available

Table 4: Primer sequences used for amplification of candidate genes.

Fragment	Forward primer name	Forward primer sequence (5` - 3`)	Reverse primer name	Reverse primer sequence (5` - 3`)
CBF1	AF376136_s1	TTTTTGACGCTGCAACTGAT	AF376136_as709	TTTACCGAGGGAGTAGTTTCCA
CBF5	TmCBF5_F [*2]	CGATGCAAAGTGTGCAATTC	AY951947_as1691	ACTAGCTCATGCGAATATGGTGT
CBF7	AY785904_s4	TTCTAGTCCACCTAGCTACAGGC	AY785904_as926	CACTAGCAAAGCAATTCATGAGC
CBF10	AY951950_s1522	ACATCTCACACACTCCACAGATG	Cbf4B_R [*3]	GCAGAATCGGCTACAAGCTCCAG
CBF13	Cbf5_F [*3]	CAGAGCAGAATCAGATGGGGAATC	AY951951_as1964	GCTAAGCTCACACTCCTCGATAA
CBF14	AY951948_s_565	TAAACTCGCTGCTTAATTACCCC	AY951948_as_1312	ATATTTGGTGGAACAGAAGCAGA
	AY951948_s_528	CAGCATCCATCTCTCTCAAATCT	AY951948_as_1299	CAGAAGCAGAGAAACCGTCTAAA
	AY951948_s_565	TAAACTCGCTGCTTAATTACCCC	AY951948_as_1299	CAGAAGCAGAGAAACCGTCTAAA
	AY951948_s_528	CAGCATCCATCTCTCTCAAATCT	AY951948_as_1312	ATATTTGGTGGAACAGAAGCAGA
CBF15	EF028765_s_90	ACCGACCACCTGCAGTACC	EF028765_as_875	TTGTTCCATGCATAGAGTCAAAG
CBF18	AY951946_s400	CGTATAAATACGCACACGCACTA	AY951946_as1445	ACATGGTGGAGGGATCTTTTTAT
Dhn1	ScDhn1_F [*1]	CCACGTAGCACGCACGCTGT	AF043087_as1808	TCGGAACATAGAGAAGACACACA
VRN-A1b	VRN1-A_F [*4]	GAAAGGAAAAATTCTGCTCG	AY747600_as1083	GATTACCGTCTTAACCCTTCCAC
VRN-A1c	AY747600_s9072	CATGAAACAACGCATTACAGAAA	AY747600_as10169	CAGATAGAACTGGTTGGATCCCT
VRN-A1d	AY747600_s_10698	TTTCTGTCATTGTTCCTTCCTGT	AY747600_as_11318	CAAGCTAAGGCTTCATGACAAGT
	AY747600_s_10718	TGTCCCACCCAAAGTTAGTAATG	AY747600_as_11390	AACGATGTAATGAGGTTACGTGC
	AY747600_s_10698	TTTCTGTCATTGTTCCTTCCTGT	AY747600_as_11390	AACGATGTAATGAGGTTACGTGC
	AY747600_s_10718	TGTCCCACCCAAAGTTAGTAATG	AY747600_as_11318	CAAGCTAAGGCTTCATGACAAGT
VRN-A1e	AY747600_s_11297	CTTGTCATGAAGCCTTAGCTTGT	AY747600_as_12066	GCTGCAGCTTGCTACTTTACTCT
	AY747600_s_11297	CTTGTCATGAAGCCTTAGCTTGT	AY747600_as_12099	AAACTGAGGTGGACAAAGTGAAA
VRN-B1b	AY747603_s18	AGGCCTAGGGTACAGTAGAATAGTAG	AY747606_as820	CAAACGGAATCAACCAAACAG
VRN-B1c	AY747603_s3097	TCTGAGCAGAATTATACTTACCTTGC	AY747606_as9488	AGATCATCTGATATCGGCAAAAA
VRN-B1d	AY747603_s_4783	CCTTCCTGTTCCACTCAAAGTTA	AY747603_as_5249	TTTTTAACTGTGAAGAGCATATGACTAA
VRN-B1e	AY747603_s5134	AAACAAGAAAAACACTTGCAGAGA	AY747603_as6211	ATTACATGGTAAATTGAGCCCAG
VRN-D1b	AY747606_s6	TTCCCTTCTACTAGGCATAGGGT	AY747606_as820	CAAACGGAATCAACCAAACAG
VRN-D1c	AY747606_s8129	GTGTTGGTAGAAGGCTAGAAGCA	AY747606_as9488	AGATCATCTGATATCGGCAAAAA

Table 4: (Continued)

Fragment	Forward primer name	Forward primer sequence (5` - 3`)	Reverse primer name	Reverse primer sequence (5` - 3`)
VRN-D1d	AY747606_s10179	GACCTCACGCCAATTTTTGT	AY747606_as11608	TACGAAACAATTTAGACCGGTTG
VRN-D1e	AY747606_s11586	CAACCGGTCTAAATTGTTTCGTA	AY747606_as12291	TTAATTCACATAAACAACATCCCACTA
VRN2a	AY485977_s306	AAACAAGCAAACGTTGGAGTTAG	AY485977_as1282	AATAAGCAATTTCCTGATGCAAA
VRN2b	AY485977_s_1542	CAACACTGAATGAAAATGGATCA	AY485977_as_1985	GAACCATCCGAGGTGAAGTTTA
	AY485977_s_1542	CAACACTGAATGAAAATGGATCA	AY485977_as_1972	TGAAGTTTACTAGGATCATGGGG
	AY485977_s_1439	CCATAGAGCAATTGAGTTTGGAC	AY485977_as_1972	TGAAGTTTACTAGGATCATGGGG
VRN2a/b	AY485977_s306	AAACAAGCAAACGTTGGAGTTAG	AY485977_as_1972	TGAAGTTTACTAGGATCATGGGG
VRN3a	DQ890162_s_1430	AAGGAGTACTAGAGCGGCGAG	DQ890162_as_1915	TGTGGTGAGCACTTTCAGAGATA
	DQ890162_s_1552	TTCCTCAATTCACAGCTTACTCC	DQ890162_as_1915	TGTGGTGAGCACTTTCAGAGATA
VRN3b	DQ890162_s2159	TCTTAAATACTCTCTCCGTCCGA	DQ890162_as3153	AAGCCATTGATCTAGGGTTCAC
	DQ890162_s2396	GAAGTACACTTATTCGTGGACGG	DQ890162_as3153	AAGCCATTGATCTAGGGTTCAC
VRN3a/b	DQ890162_s_1552	TTCCTCAATTCACAGCTTACTCC	DQ890162_as3153	AAGCCATTGATCTAGGGTTCAC
Cab b	contig22616_s209	TTTTGCGAAAGCACACTTATACA	contig22616_as938	GAAGCATCGCCAGCTATAAATAC
Cab d	contig22616_s209	TTTTGCGAAAGCACACTTATACA	contig22616_as828	CAGTTGCAGCAGAGAGATTCTT
Dem	CD937801_s29	ATACCATCGGCAACTCCTCTG	contig1013618_as520	CCATTATGGATAGCGAAATTTGA
Tacr7 b	contig4120743_s26	CAACCAAAACTCGCCTATAAAAG	contig2688312_as455	AATCGGAGAGGAAGCTCTCTTTA
Tacr7 c	contig4120743_s271	CGAGGAGAAGGTTTGGGGTT	contig2688312_as455	AATCGGAGAGGAAGCTCTCTTTA
Tacr7 d	BJ246882_s196	GTCGGCGAGGAGAAGGTTTT	contig2688312_as455	AATCGGAGAGGAAGCTCTCTTTA
PPD-B1c	DQ885757_s11028	TCCTTCCAGCTTACTAGTGCATC	DQ885757_as11954	ATCACCTGGAAAACATATTGGAA
PPD-B1d	DQ885757_s11883	AACTGAACCAAAAGCCTGCTACT	DQ885757_as12453	GTACCTTGCAAAGAATGAAAACG
PPD-B1e	DQ885757_s_12390	CCTTTGTGAATCCTTAAATCATCC	DQ885757_as_13162	AACAGAGAACAAACGAAATCGG
PPD-B1f	DQ885757_s_13184	GGGCTTATTCATGATAGCTGATG	DQ885757_as_13562	ATCGACTCCGCACTTCTACTATG
	DQ885757_s_13148	CGTTTGTTCTCTGTTCTTCGTTT	DQ885757_as_13625	ACCGTTACACAGGTTCAGACATT
	DQ885757_s_13184	GGGCTTATTCATGATAGCTGATG	DQ885757_as_13625	ACCGTTACACAGGTTCAGACATT
	DQ885757_s_13148	CGTTTGTTCTCTGTTCTTCGTTT	DQ885757_as_13562	ATCGACTCCGCACTTCTACTATG
PPD-D1 Prom	DQ885766_s3601 [†6]	CTTGTCCAACTCCCAATCTAGTG	DQ885766_as4689 [†6]	TCCTCCCCTGTTTCTTTTTACTC
	DQ885766_s4578	TCGTCCATCCAAAGATACTGATT	DQ885766_as5712 [†6]	AGTACGCTGCCGTGAGTAATAAT

Table 4: (Continued)

Fragment	Forward primer name	Forward primer sequence (5` - 3`)	Reverse primer name	Reverse primer sequence (5` - 3`)
PPD-D1 Prom	DQ885766_s4450 [†6]	CATACTCCCTCCGTTTCTTCTTT	DQ885766_as5712 [†6]	AGTACGCTGCCGTGAGTAATAAT
	DQ885766_s4578	TCGTCCATCCAAAGATACTGATT	DQ885766_as5700	TGAGTAATAATCGAACCTCGGTC
PPD-D1a	DQ885766_s5689	ATTATTACTCACGGCAGCGTACT	DQ885766_as6299 [†6]	TACTGAAACATTTTAGGGCCAAG
	DQ885766_s5677 [†6]	GACCGAGGTTCGATTATTACTCA	DQ885766_as6299 [†6]	TACTGAAACATTTTAGGGCCAAG
PPD-D1a2	DQ885766_s5766 [†6]	CAACATGTTTCCTCTTGGAGC	DQ885766_as6535 [†6]	GAACAGAGTCAAACACCATCAGA
PPD-D1b	DQ885766_s6298	TATCAGGTTCATTTGCTTCAGTG	DQ885766_as7002 [†6]	ATGGACAAATTGACCTCTAGTGC
	DQ885766_s6277 [†6]	CTTGGCCCTAAAATGTTTCAGTA	DQ885766_as7002 [†6]	ATGGACAAATTGACCTCTAGTGC
	DQ885766_s6277 [†6]	CTTGGCCCTAAAATGTTTCAGTA	DQ885766_as6963	GCCATTCAGTTTTATCTAGCTTCC
PPD-D1c	DQ885766_s7244	TGACAAGTATCTGCATCTGAACC	DQ885766_as8033 [†6]	GATTCGCAAAGGACACTGATATT
PPD-D1d	DQ885766_s6939 [†6]	GGAAGCTAGATAAAACTGAATGGC	DQ885766_as8033 [†6]	GATTCGCAAAGGACACTGATATT
PPD-D1e	DQ885766_s8011 [†6]	AATATCAGTGTCCTTTGCGAATC	Ppd-D1exon8_R1 [*5]	gtctaaatagtaggtactagg
PPD-D1 3`UTR	DQ885766_s8771	CTGCTCTCTGTTCTTGGTTTCAT	DQ885766_as9720	ACCTCCCTGACGAAAAGCTC

Primer names with [†] are developed in course of this work but published from Keilwagen et al. (2014). Primer names with [*] as already published were used in combination with primers with [†] and without labels. [1] (Li et al., 2011b); [2][(Miller et al., 2006); [3] (Vagujfalvi et al., 2005); [4] (Yan et al., 2004a); [5] (Beales et al., 2007); [6] (Keilwagen et al., 2014)

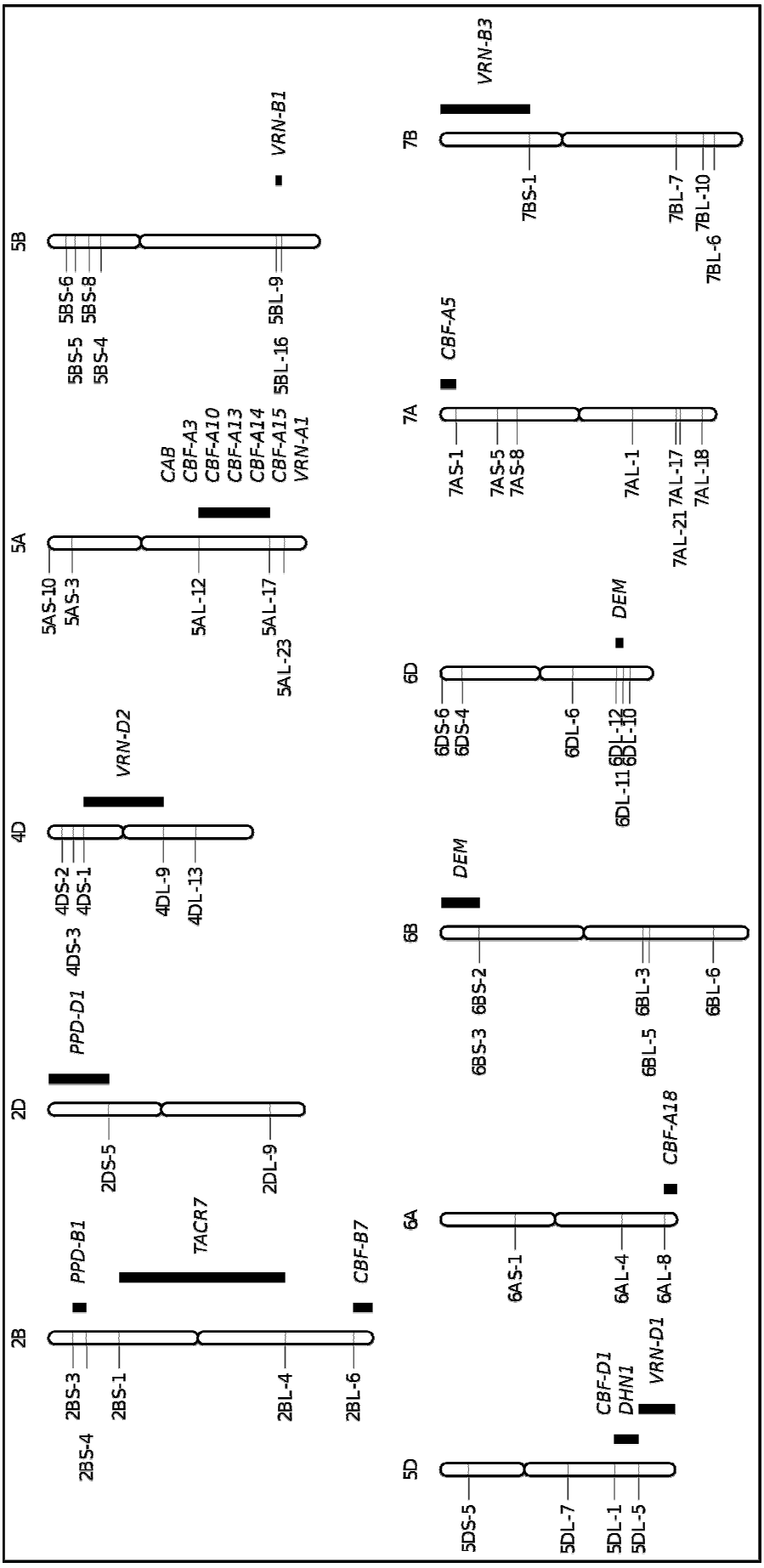


Figure 8: Map of gene specific PCR fragments by using wheat NT- and deletion lines. In this figure only wheat chromosomes harboring mapped PCR fragments are shown. The white bar depicts the chromosome, the constriction symbolises the centromere, to the left of the chromosomes deletion break points are listed and the black bars correspond to the regions of mapped PCR fragments with appended candidate gene (Babben et al., 2015).

3.1.3 *In silico* analysis of primer sub-genome specificity

The draft sequences of wheat and related species allow detailed *in silico* analysis of oligos used in this thesis by conducting a simple BLAST comparison. Out of 98 oligos that were used for the amplification of 65 PCR fragments, 54 turned out to be specific to one sub-genome, 21 specific to two sub-genomes, and 14 were non-specific. For nine oligos the comparison could not be performed due to the non-availability of sub-genome sequences (Supplemental Table 2). 57 out of 65 amplicons comprise at least one sub-genome specific primer. For five PCR fragments (CBF5, Dhn1, Cab b, Cab d and Dem) no wheat sub-genome sequences could be identified. Both primers of PCR fragments CBF7, PPD-B1f and PPD-D1b showed no specificity for any sub-genome in reference to Wu et al. (2009). Nevertheless, all three fragments showed single bands and correct chromosome localisation via NT-lines (Supplemental Figure 1). The primer sequences of PPD-B1f and PPD-D1b were derived from a specific sub-genome. At least one of the primers showed one or more differences to corresponding regions on the chromosomes in alignments with the other two sub-genomes. Special cases are the primers of fragment CBF7. The forward primer has no sub-genome specificity and the reverse primer is specific to sub-genomes A and B (Supplemental Table 4).

3.1.4 Re-sequencing of genes involved in frost tolerance and homology validation via BLAST

For the re-sequencing of candidate genes 24 wheat varieties comprising two spring and 22 winter cultivars (Table 1) were used. Five out of 40 amplicons revealed a presence/absence polymorphisms (dominant) and were therefore not sequenced. These five dominant markers were directly used for genotyping of a *PPD-D1* deletion in the promoter and a transposable element (TE) in intron1 (Guo et al., 2010). One PCR fragment (*CBF7*) could not be sequenced due to very low quality. Finally, 34 amplicons, representing 18 FT genes, were selected for sequencing and all 34 obtained sequences were compared to retrieved gene models via MegaBlast. In case of *TACR7,* Kocsy et al. (2010) identified BQ659345 of *Hordeum vulgare* as reference, but the fragments of 24 sequenced genotypes of *TACR7* did not exceed an identity of 84% to the published wheat

reference sequence L28093 for *TACR7* (Gana et al., 1997). Out of the 24 sequences the initial barley sequence BQ659345 showed highest homology with an identity of 92.3%, whereas the second best barley BLAST hit of 91.8% identity was to X97916, the barley low temperature gene 14.1 abbreviated as *Blt14.1*. The rest of 33 BLAST results showed very high identities from 88.8% to 100% to initial gene sequences (Table 5).

To identify polymorphisms of the FT candidate genes a SNP/indel detection was performed. Within 12 out of the 18 sequenced candidate genes, differences between the 24 genotypes were determined revealing a high level of polymorphism of 66.7%. The remaining six candidate genes were monomorphic. The number of polymorphic sites ranged from 1 to 37, the number of haplotypes (h) from two to three, the haplotype diversity (Hd) from 0.08 to 0.61 and the nucleotide diversity (π) from 0.00008 to 0.00757 (Table 6). This numbers derive from a set of up to 24 accessions.

Taken together, the results of the workflow for locus specific primer development presented in this study were very promising. The main workflow step is the identification of sub-genome sequences and the design of primers based on sub-genome sequence differences. This is the essential step of this workflow and is crucial for the success of this approach. Primer amplification tests for single bands and fragment mapping via NT-lines are simple ways to verify locus specificity. Sequencing of selected locus specific amplicons and BLAST analysis of these fragment sequences versus initial data bases is the last step of safe-guarding the correct amplification. The results of this BLAST search showed no critical differences to the initially selected sequences. The developed primer pairs are ideally suited for the candidate gene based association approach, whose results I will present in the following.

Table 5: BLASTn results of sequenced PCR fragments versus NCBI nucleotide collection (nr/nt) and NCBI candidate gene reference EST.

Gene	PCR fragment	Database	Subject Seq-id (ID of the database hit)	Percentage of identical matches	Expect value (e-value)	Bit score	Subject discribtion
CAB	Cab b	EST	gi|21273269|gb|BQ465487.1|	93.71	0	643	HU03M14r HU *Hordeum vulgare* subsp. *vulgare* cDNA clone HU03M14 5-PRIME, mRNA sequence.
CBF-D1	CBF1	nucleotide	gi|17226800|gb|AF376136.1|AF376136	100	0	1279	*Triticum aestivum* putative *CRT/DRE-binding factor* (*CBF1*) mRNA, complete cds
CBF-A5	CBF5	nucleotide	gi|404415276|gb|JN987194.1|	100	0	1519	*Triticum aestivum AP2 domain CBF protein* (*CBFII*) mRNA, *CBFII-5.4* allele, complete cds
CBF-A10	CBF10	nucleotide	gi|404415286|gb|JN987199.1|	99.53	0	1548	*Triticum aestivum AP2 domain CBF protein* (*CBFIIIc*) mRNA, *CBFIIIc-10.1* allele, complete cds
CBF-A13	CBF13	nucleotide	gi|404415320|gb|JN987217.1|	100	0	1493	*Triticum aestivum AP2 domain CBF protein* (*CBFIIIc*) pseudogene, *CBFIIIc-13.1c* allele, complete sequence
CBF-A14	CBF14	nucleotide	gi|158999375|gb|EU076382.1|	99.01	0	902	*Triticum monococcum CBF14* gene, complete cds
CBF-A15	CBF15	nucleotide	gi|404415321|gb|JN987218.1|	100	0	1325	*Triticum aestivum AP2 domain CBF protein* (*CBFIIId*) gene, *CBFIIId-15.2b* allele, complete cds
CBF-A18	CBF18	nucleotide	gi|63098599|gb|AY951946.1|	94.92	0	1559	*Triticum monococcum CRT/DRE binding factor 18* (*CBF18*) gene, complete cds
DEM	Dem b	nucleotide	gi|241986478|dbj|AK333739.1|	97.37	0	713	*Triticum aestivum* cDNA, clone: WT008_O03, cultivar: Chinese Spring
		EST	gi|12030509|emb|AL504294.1|	90.37	$7.00E^{-128}$	466	AL504294 *Hordeum vulgare* Barke roots *Hordeum vulgare* subsp. *vulgare* cDNA clone HW04N07 5', mRNA sequence.
DHN1	Dhn1	nucleotide	gi|59894280|gb|AY895879.1|	88.79	$3.00E^{-176}$	625	*Hordeum vulgare* subsp. *spontaneum* voucher NPGS PI 559556 *dehydrin 1* (*Dhn1*) gene, partial cds
PPD-B1	PPD-B1c	nucleotide	gi|456359289|dbj|AB745620.1|	100	0	1679	*Triticum turgidum* subsp. *pyramidale Ppd-B1* gene for *pseudo-response regulator*, complete cds, strain: KU-9882
	PPD-B1d	nucleotide	gi|456359289|dbj|AB745620.1|	99.82	0	1009	*Triticum turgidum* subsp. *pyramidale Ppd-B1* gene for *pseudo-response regulator*, complete cds, strain: KU-9882
	PPD-B1e	nucleotide	gi|456359289|dbj|AB745620.1|	99.3	0	1297	*Triticum turgidum* subsp. *pyramidale Ppd-B1* gene for *pseudo-response regulator*, complete cds, strain: KU-9882
	PPD-B1f	nucleotide	gi|383215299|gb|JF946486.1|	99.74	0	693	*Triticum aestivum* transposon TREP 3040_Harbinger, complete sequence; *pseudo-response regulator* (*Ppd-B1*) gene, *Ppd-B1a* allele, complete cds; and retrotransposon Gypsy TREP 3457_Danae, complete sequence

Table 5: (Continued)

Gene	PCR fragment	Database	Subject Seq-id (ID of the database hit)	Percentage of identical matches	Expect value (e-value)	Bit score	Subject discribtion
PPD-D1	PPD-D1d	nucleotide	gi|395759126|dbj|AB646977.1|	99.91	0	1965	*Triticum aestivum PRR* gene for *pseudo-response regulator*, complete cds, allele: *Ppd-D1b.2*
	PPD-D1e	nucleotide	gi|395759124|dbj|AB646976.1|	100	0	1731	*Triticum aestivum PRR* gene for *pseudo-response regulator*, complete cds, allele: *Ppd-D1a.1*
	PPD1 3`UTR	nucleotide	gi|118638641|gb|DQ885766.1|	100	0	1629	*Triticum aestivum* cultivar Chinese Spring chromosome 2D *pseudo-response regulator* (*PRR)* gene, complete cds
TACR7	Tacr7 b	nucleotide	gi|1418967|emb|X97916.1|	91.75	0	778	*H.vulgare blt14.1* gene
		EST	gi|21800478|gb|BQ659345.1|	92.34	$5.00E^{-180}$	640	HD01A06w HD *Hordeum vulgare* cDNA clone HD01A06 3-PRIME, mRNA sequence.
VRN-D2	VRN2a	nucleotide	gi|211593611|gb|FJ173824.1|	91.07	$6.00E^{-149}$	534	*Triticum turgidum* retrotransposon Wilma, partial sequence; and *ZCCT2-B2b* (*VRN2*) gene, complete cds
	VRN2b	nucleotide	gi|45390737|gb|AY485977.1|	92.57	0	1120	*Hordeum vulgare* cultivar Dairokkaku *ZCCT-Ha* (*VRN2*) gene, partial cds
VRN-B3	VRN3a	nucleotide	gi|117168399|gb|DQ890162.1|	100	0	1825	*Triticum aestivum* cultivar Chinese Spring *VRN3* (*vrn-B3*) gene, complete cds
	VRN3b	nucleotide	gi|117168399|gb|DQ890162.1|	100	0	1801	*Triticum aestivum* cultivar Chinese Spring *VRN3* (*vrn-B3*) gene, complete cds
VRN-A1	VRN-A1b	nucleotide	gi|383215288|gb|JF965395.1|	100	0	1829	*Triticum aestivum* cultivar Claire *VRN-A1* (*VRN-A1*) gene, complete cds
	VRN-A1c	nucleotide	gi|383215290|gb|JF965396.1|	100	0	1995	*Triticum aestivum* cultivar Malacca *VRN-A1* (*VRN-A1*) gene, complete cds
	VRN-A1d	nucleotide	gi|383215290|gb|JF965396.1|	100	0	1109	*Triticum aestivum* cultivar Malacca *VRN-A1* (*VRN-A1*) gene, complete cds
	VRN-A1e	nucleotide	gi|383215292|gb|JF965397.1|	100	0	1279	*Triticum aestivum* cultivar Hereward *VRN-A1* (*VRN-A1*) gene, complete cds
VRN-B1	VRN-B1b	nucleotide	gi|384371844|gb|HQ130483.2|	100	0	1459	*Triticum aestivum* cultivar Diamant2 *Vrn-B1* (*Vrn-B1*) gene, *Vrn-B1-a* allele, promoter region and complete cds
	VRN-B1c	nucleotide	gi|384371844|gb|HQ130483.2|	99.82	0	1016	*Triticum aestivum* cultivar Diamant2 *Vrn-B1* (*Vrn-B1*) gene, *Vrn-B1-a* allele, promoter region and complete cds

Table 5: (Continued)

Gene	PCR fragment	Database	Subject Seq-id (ID of the database hit)	Percentage of identical matches	Expect value (e-value)	Bit score	Subject discribtion
	VRN-B1d	nucleotide	gi|384371844|gb|HQ130483.2|	100	0	867	*Triticum aestivum* cultivar Diamant2 *Vrn-B1* (*Vrn-B1*) gene, *Vrn-B1-a* allele, promoter region and complete cds
	VRN-B1e	nucleotide	gi|58423007|gb|AY747604.1|	100	0	1969	*Triticum aestivum* cultivar Triple Dirk C line *VRN-B1* (*VRN-B1*) gene, complete cds
VRN-D1	VRN-D1b	nucleotide	gi|58423011|gb|AY747606.1|	100	0	1328	*Triticum aestivum* cultivar Triple Dirk C line *VRN-D1* (*VRN-D1*) gene, complete cds
	VRN-D1c	nucleotide	gi|58423011|gb|AY747606.1|	100	0	1701	*Triticum aestivum* cultivar Triple Dirk C line *VRN-D1* (*VRN-D1*) gene, complete cds
	VRN-D1d	nucleotide	gi|58423011|gb|AY747606.1|	99.16	0	1714	*Triticum aestivum* cultivar Triple Dirk C line *VRN-D1* (*VRN-D1*) gene, complete cds
	VRN-D1e	nucleotide	gi|58423011|gb|AY747606.1|	100	0	1327	*Triticum aestivum* cultivar Triple Dirk C line *VRN-D1* (*VRN-D1*) gene, complete cds

Table 6: Nucleotide polymorphisms of coding and non-coding candidate gene regions.

Gene	No. accessions	No. of bp	No. of polymorphic sites	Percentage poly-morphism	h	Hd	k	π	k (i)	π (i)
CBF-A5	23	824	2	0.24	3	0.49	0.95	0.00115	0.44	0.00054
CBF-A10	23	776	2	0.26	2	0.47	0.95	0.0012	n/a	n/a
CBF-A13	23	773	4	0.52	2	0.47	1.42	0.00193	0.95	0.00123
CBF-A14	22	1184	6	0.51	3	0.48	1.91	0.00163	0.46	0.00038
CBF-A15	24	755	7	0.93	2	0.49	2.94	0.00395	0.49	0.00065
CBF-A18	22	951	37	3.89	2	0.09	3.09	0.00328	0.27	0.00029
VRN-A1	24	2954	9	0.30	2	0.08	0.42	0.00014	0.33	0.00011
VRN-D1	23	3093	1	0.03	2	0.24	0.24	0.00008	n/a	n/a
VRN-B3	24	1566	1	0.06	2	0.52	0.52	0.00033	n/a	n/a
CAB	24	707	13	1.84	3	0.61	5.18	0.00757	1.26	0.00179
PPD-B1	24	3971	2	0.05	3	0.36	0.37	0.00009	n/a	n/a
PPD-D1	24	2642	1	0.04	2	0.23	n/a	n/a	0.23	0.00009

No.: number; h: haplotypes; Hd: haplotype diversity; k: average number of nucleotide differences; π: nucleotide diversity; (i): Indel; n.a.: not available

3.2 Candidate gene association studies for frost tolerance in wheat (*Triticum aestivum* L.)**

Wheat specific primer pairs for 18 FT candidate genes were developed, which were used in a candidate gene association study on a diverse panel of 235 genotypes. The winter survival rates were scored on field locations in Germany and Russia. Based on the sequence and phenotypic data an association study was performed.

3.2.1 Phenotypic data analysis

All 235 wheat genotypes, originating from 28 countries and five continents, were tested in five environments (Gatersleben, Germany; Ranzin, Germany; Puskin, Russia; Roshchinskiy, Russia; Novosibirsk, Russia) in two years. The genotypes were selected based on pre-existing knowledge regarding the reaction to growing conditions during winter time (Supplemental Table 3). Phenotypic data were collected as winter survival in % (0% all dead to 100% all alive) after winter (Figure 9 and Supplemental Figure 2).

** from Babben et al. (2018)

Furthermore, to take the diversity of the environments with respect to climatic conditions into account a co-variable was calculated (number of days under −15 °C). The phenotypic data sets and the established co-variable are only weakly correlated (r = 0.24), indicating that all phenotypic data are not suitable to be used for FT analysis. All the FT scores were transformed using arcsine and Cox-Box (data not shown), but transformations did not result in any improvement. A scatter plot illustrates that the phenotypic data obtained in Pushkin_2013 and Novosibirsk_2014 are not due to FT (Supplemental Figure 2). A reason for this may be the very few days with temperature below −15 °C at Pushkin in 2013 (10 days) and a rather low winter hardiness (mean winter survival of 16.2%) due to missing snow coverage; while in Novosibirsk in 2014 a very high number of days with temperature below −15 °C (41 days) but less frost damage was observed due to continuous snow coverage (mean winter survival of 82.6%). The PCA calculation shows that the environments Ranzin_2013, Gatersleben_2013 and Novosibirsk_2014 are separated from the other environments described by the second component (Figure 10). This indicates that at these locations, in the respective years, phenotypic data are influenced by additional factors and not entirely due to differences in FT. An additional characteristic of phenotypic data is the coefficient of variation (CV) and variance calculation (in %) of each environment per year. High CVs (>0.15) and/or variances (>150) represent a good phenotypic distribution of all 235 genotypes within the field data sets. Ranzin_2013, Gatersleben_2013 and Novosibirsk_2014 show very low CVs 0.08 and 0.15 and variances between 54.3 and 149.4. The environment Pushkin_2013 shows a small variance with 51 but a high CV value of 0.44. These values resulted from a very slight winter survival with a small standard deviation (Table 7). In conclusion of the phenotypic data analysis, environments Ranzin_2013, Gatersleben_2013, Pushkin_2013 and Novosibirsk_2014 were excluded from further analysis because the results are not coherent with FT effects. Out of the data sets Ranzin_2012, Gatersleben_2012, Pushkin_2012, Novosibirsk_2012, Roshchinskiy_2012 and Roshchinskiy_2013 the LS means were calculated (Supplemental Table 3).

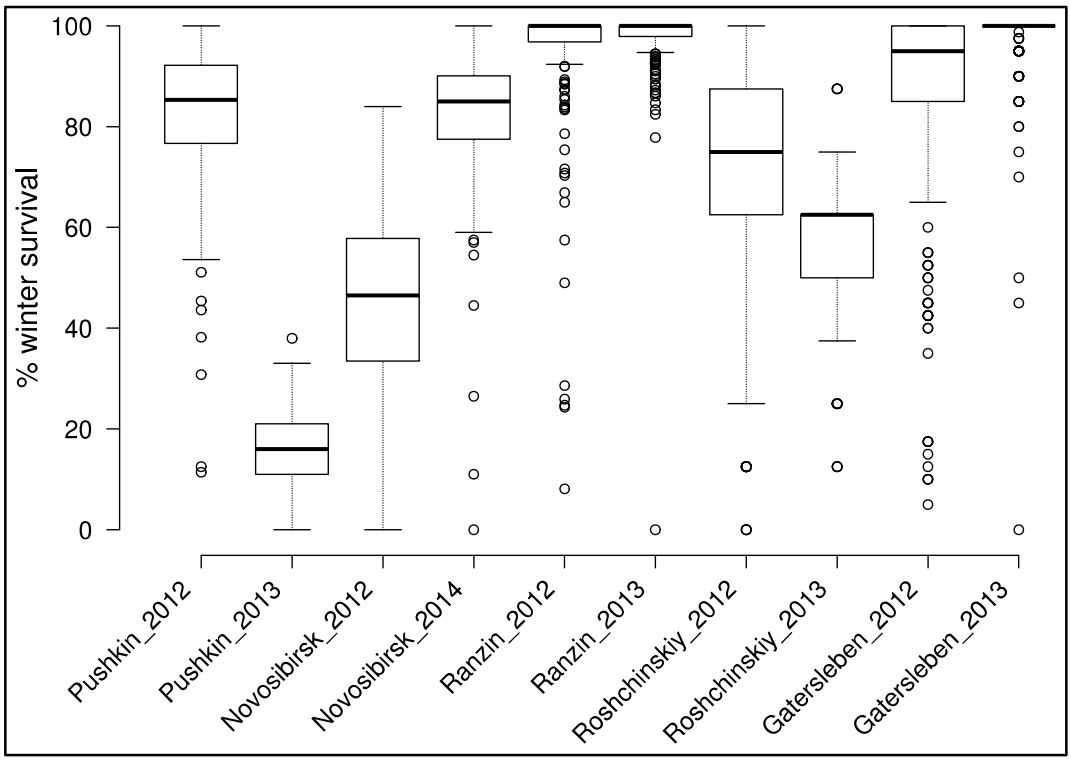


Figure 9: Phenotypic variation at five field locations during two experimental years. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend to 5th and 95th percentiles, outliers are represented by dots. n > 203 (Babben et al., 2018).

An efficient approach for the development of genome-specific markers in allohexaploid wheat (Triticum aestivum L.) and its application in the construction of high-density linkage maps of the D genome

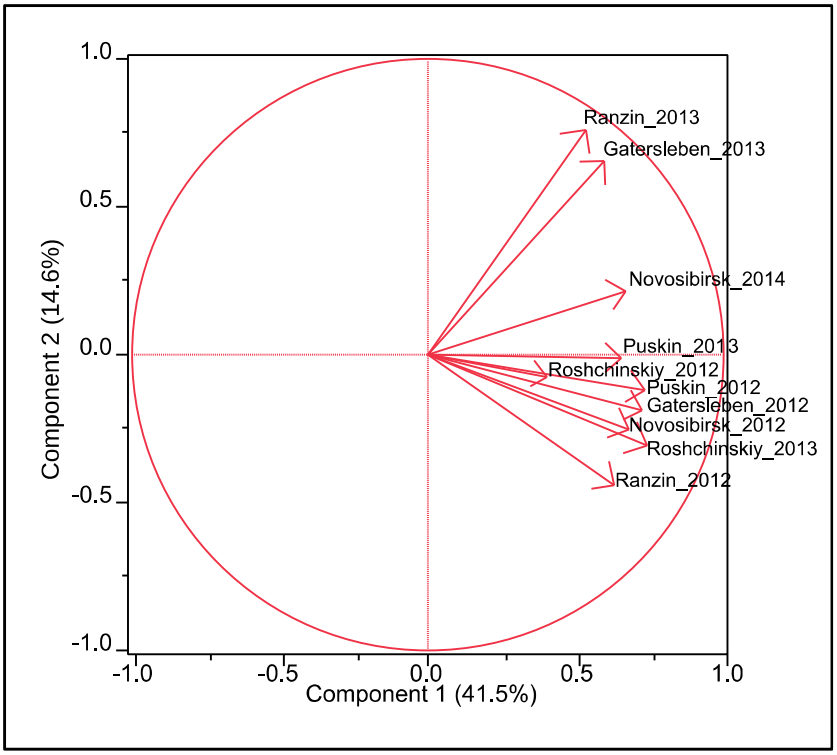


Figure 10: PCA plot of mean winter survival and number of days below −15 °C from ten environments (Babben et al., 2018).

Table 7: Variance and coefficient of variation of frost tolerance data.

Location_Year	Size	Missing data	Mean FT in %	Standard deviation	Variance	Coefficient of variation
Pushkin_2012	235	0	82.40	13.88	192.52	0.17
Pushkin_2013	235	0	16.16	7.14	51.01	0.44
Novosibirsk_2012	235	0	45.60	16.46	271.05	0.36
Novosibirsk_2014	235	0	82.58	12.22	149.42	0.15
Ranzin_2012	235	0	95.20	13.31	166.12	0.14
Ranzin_2013	235	0	97.67	8.19	54.27	0.08
Roshchinskyiy_2012	235	32	69.33	22.64	512.53	0.33
Roshchinskiy_2013	235	4	56.55	13.68	187.24	0.24
Gatersleben_2012	235	0	85.89	22.14	490.01	0.26
Gatersleben_2013	235	0	97.18	9.12	83.21	0.09

FT: frost tolerance

3.2.2 Candidate gene polymorphisms

A set of 65 specific primer pairs which are described in chapters 3.1.2, 3.1.3 and 3.1.4 was BLASTed against the IWGSC RefSeq, allowing the identification of 39 primer combinations that cover the full length of 19 genes. An optimised set of 39 primer pairs was used for PCR amplification, amplicon sequencing (34 primer pairs; described in chapter 3.1.4) (Supplemental Table 5) and association genetics studies. No exact position could be determined for the first forward primer of *PPD-B1* and the third forward primer of *VRN-D2*, since BLASTn revealed seven and five miss matches, respectively. For all other primers an exact position was determined. In addition, an unassigned scaffold was identified for *PPD-B1* located on chromosome 2B (Supplemental Figure 3). The total re-sequenced length of candidate genes was 13.3 kbp coding DNA sequence (CDS) and 43.76 kbp genomic lengths, with a CDS/genomic length ratio of 0.3. In total, a sequence length of 33.5 kbp was analysed. In our set of 235 genotypes, the 15 candidate genes *CAB*, *CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *CBF-A18*, *TACR7*, *VRN-B3*, *VRN-A1*, *VRN-B1*, *VRN-D1*, *PPD-B1* and *PPD-D1* were polymorphic whilst the four genes *CBF-D1*, *DHN1*, *VRN2* and *DEM* (Table 8) were monomorphic. 254 polymorphic sites, i.e. 221 SNPs and 33 indels, were identified in total. The SNP number per gene ranged from 0 to 97 and the indel number from 0 to 12. Summarising all genes, 42 polymorphic sites in promoter regions, 64 in introns, 25 in 3` UTRs and 123 in exons were identified. Out of the 254

polymorphic sites, 131 were located in non-coding regions, and 54 synonymous and 69 non-synonymous polymorphic sites were identified (Figure 11). The number of haplotypes ranged between two and six and the haplotype diversity (Hd) between 0.07 and 0.68 (Table 9). This numbers derive from a set of up to 235 accessions, therefore the results differ from chapter 3.1.4.

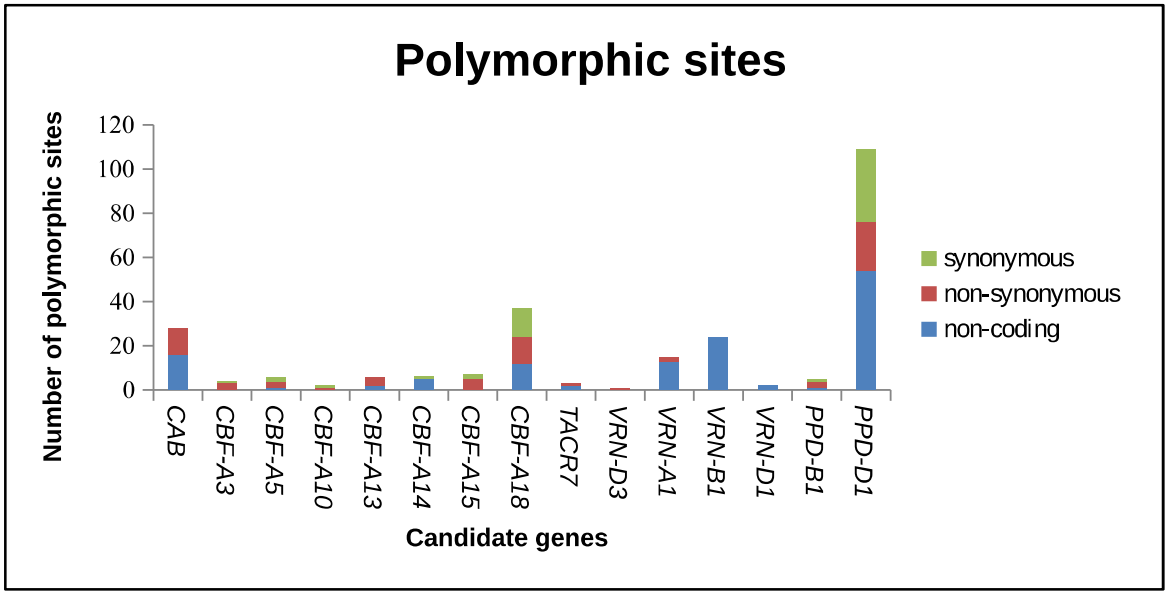


Figure 11: Number of synonymous, non-synonymous and non-coding mutations per candidate gene (Babben et al., 2018).

Table 8: List of analysed FT candidate genes, sequence length, number of specific PCR fragments and detected mutations.

Gene	Chr. position	No. of specific PCR fragments	Sequence length in bp	CDS length in bp	Gene length in bp	CDS/gene length ratio	Detected mutations	Number of polymorphic sites
CBF-D1	5D	1	709	639	639	1.00	no	0
CBF-A3	5A	1	790	741	741	1.00	yes	4
CBF-A5	7A	1	1027	633	633	1.00	yes	6
CBF-A10	5A	1	867	720	720	1.00	yes	2
CBF-A13	5A	1	855	720	720	1.00	yes	6
CBF-A14	5A	2	610/748	639	639	1.00	yes	6
CBF-A15	5A	1	786	726	726	1.00	yes	7
CBF-A18	6A	1	1045	738	738	1.00	yes	37
DHN1	5D	1	936	420	638	0.66	no	0
VRN-A1	5A	4	1039/1097/621/770	735	11414	0.06	yes	15
VRN-B1	5B	5	600/817/662/467/1077	735	5498	0.13	yes	24
VRN-D1	5D	4	814/1359/1429/705	735	11550	0.06	yes	2
VRN-D2	4D	2	976/534	639	1650	0.39	no	0
VRN-B3	7B	2	1602/994	534	1258	0.42	yes	1
CAB	5A	1	728	n.a.	n.a.	n.a.	yes	28
DEM	6B/6D	1	616	n.a.	n.a.	n.a.	no	0
TACR7	2B	1	765	n.a.	n.a.	n.a.	yes	3
PPD-B1	2B	6	1600/954/927/571/773/378	1995	3053	0.65	yes	5
PPD-D1	2D	3	726/1094/966	1983	3141	0.63	yes	109
total		39	34034	13332	43758	0.30		255

Chr.: chromosome; No.: number; bp: base pairs; CDS: coding DNA sequence; n.a.: not available

Table 9: Polymorphic candidate genes, location of polymorphisms, amino acid change and haplotype diversity.

Gene	Chr.	Accessions	Polymorphic sites	SNPs	indels	Promotor	Intron	Exon	3` UTR	Non-coding	Non-synonymous	Synonymous	Haplotypes	Hd
CBF-A3	5A	235	4	4	0	0	0	4	0	0	3	1	2	0.30
CBF-A5	7A	235	6	5	1	1	0	5	0	1	3	2	4	0.54
CBF-A10	5A	235	2	2	0	0	0	2	0	0	1	1	2	0.30
CBF-A13	5A	235	6	4	2	1	0	5	0	2	4	0	3	0.30
CBF-A14	5A	235	6	5	1	5	0	1	0	5	0	1	3	0.30
CBF-A15	5A	235	7	6	1	0	0	7	0	0	5	2	2	0.30
CBF-A18	6A	235	37	34	3	1	0	25	11	12	12	13	3	0.20
VRN-A1	5A	235	15	10	5	7	6	2	0	13	2	0	6	0.19
VRN-B1	5B	235	24	21	3	20	4	0	0	24	0	0	5	0.07
VRN-D1	5D	235	2	2	0	0	1	0	1	2	0	0	3	0.14
VRN-B3	7B	235	1	1	0	0	0	1	0	0	1	0	3	0.50
CAB	5A	235	28	24	4	5	0	12	11	16	12	0	6	0.64
TACR7	2B	235	3	2	1	0	0	1	2	2	1	0	3	0.40
PPD-B1	2B	235	5	5	0	1	0	4	0	1	3	1	6	0.37
PPD-D1	2D	235	109	97	12	1	53	55	0	54	22	33	6	0.68
total			255	222	33	42	64	124	25	132	69	54	58	

Chr.: chromosome; SNP: single nucleotide polymorphism; indel: insertion-deletion; UTR: untranslated region; Hd: Haplotype Diversity

3.2.3 Population structure and kinship

Calculations of the population structure and kinship matrix of the 235 genotypes were carried as a basis to correct for false positives in association studies. Genetic profiles were obtained from the ILLUMINA Infinium iSelect 90 k wheat chip. These data were kindly provided by the Dr. A. Börner from IPK Gatersleben. As an outcome of the STRUCTURE analysis based on 249 SNPs covering the whole genome, k = 3 was the most probable number of sub-populations mostly corresponding to the origin of genotypes. The neighbour-joining tree revealed three sub-populations i.e. North American, Russian, and North and Central European genotypes. This population structure resembled a tight association to the origin of the genotypes analysed. In the first sub-population, accessions from European countries, subdivided into two groups, i.e. North and Central European genotypes, are included. Accessions from Canada, Mexico and USA were predominantly in the second sub-population, and the third sub-population contained predominantly accessions from Russia and Kazakhstan. The population structure is shown in Figure 12 and Supplemental Figure 4. The STRUCTURE membership coefficients and the modified Roger's distance revealed a high degree of admixture in a large number of accessions. Therefore, many accessions could not be classified to main groups, because their genomes represent a mixture of the main groups. This may be due to germplasm exchange and its use in breeding programs.

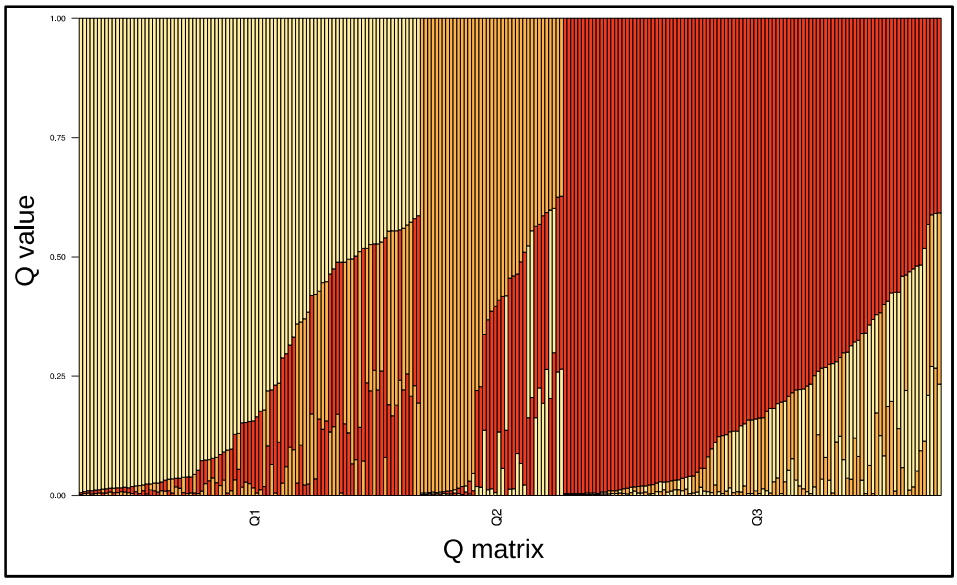


Figure 12: Population structure of 235 wheat cultivars based on 249 SNPs. Each individual is represented by a single vertical line that is partitioned into Q colored segments (Q = 3) in the x-axis. The y-axis illustrated the Q-value (Babben et al., 2018).

3.2.4 SNP/indel association analysis

The association analysis was performed using 255 polymorphic sites identified in 15 candidate genes. After minor allele frequency (MAF) selection, 58 polymorphic sites, i.e. 46 SNPs and 12 indels, from 13 candidate genes (*CAB*, *CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *TACR7*, *VRN-B3*, *VRN-A1*, *VRN-D1*, *PPD-B1* and *PPD-D1*) were included in further analysis. Additionally, the Least-Squares means (LS means), population structure and kinship matrix were included in both association analyses. In the SNP/indel association, 27 statistically significant ($P < 0.05$) polymorphic sites (21 SNPs and six indels) in nine candidate genes were identified (*CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *VRN-B3*, *VRN-A1* and *PPD-D1*) (Table 10). Four SNPs and two indels were located in the promoter region of *CBF-A5*, *CBF-A13* and *CBF-A14*. The remaining 17 SNPs and four indels were located in exon regions of *CBF*s, *PPD-D1*, *VRN-A1* and *VRN-B3*. 11 of these SNPs are non-synonymous. Six associated genes were located on wheat chromosome 5A and one each on chromosome 2D, 7A and 7B. Allelic effects of significantly associated polymorphic sites on FT ranged from 0.1% to 15% (Figure 13, Table 10, Supplemental Table 6). With an LD of $r^2 = 0.92$ to 1, the statistical significance of associated SNPs and indels from the five *CBF* genes on chromosome 5A and *FR-A2* locus (*CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15*) is very high. The LD plot combining all used SNPs/indels for association calculation is shown in Figure 14.

Table 10: Statistics of significantly associated SNPs and indels.

Gene/ Polymorphism name	CDS and promotor site	*P*-value	-Log of *P*	Polymorphism		Effect in %		Observations*	
CBF-A3_SNP1	222	$6.2823E^{-10}$	9.20	A	C	15.01	0.00	193	42
CBF-A3_SNP2	263	$6.2823E^{-10}$	9.20	C	G	15.01	0.00	193	42
CBF-A3_SNP3	367	$6.2823E^{-10}$	9.20	A	G	15.01	0.00	193	42
CBF-A3_SNP4	407	$6.2823E^{-10}$	9.20	C	T	15.01	0.00	193	42
CBF-A5_InDel1	-83	0.006	2.22	C	-	-4.52	0.00	141	84
CBF-A10_SNP1	471	$6.2823E^{-10}$	9.20	C	T	15.01	0.00	193	42
CBF-A10_SNP2	518	$6.2823E^{-10}$	9.20	G	C	15.01	0.00	193	42
CBF-A13_SNP1	-11	$6.2823E^{-10}$	9.20	T	C	15.01	0.00	193	42
CBF-A13_InDel1	19	$6.2823E^{-10}$	9.20	T	-	15.01	0.00	193	42
CBF-A13_SNP2	92	$6.2823E^{-10}$	9.20	G	A	15.01	0.00	193	42
CBF-A13_InDel2	133	$6.2823E^{-10}$	9.20	-	C	15.01	0.00	193	42
	134	$6.2823E^{-10}$	9.20	-	G	15.01	0.00	193	42

Table 10: (Continued)

Gene/ Polymorphism name	CDS and promotor site	*P*-value	-Log of *P*	Polymorphism		Effect in %			Observations*		
CBF-A13_InDel2	135	$6.2823E^{-10}$	9.20	-	T	15.01	0.00		193	42	
	136	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	137	$6.2823E^{-10}$	9.20	-	C	15.01	0.00		193	42	
	138	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	139	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	140	$6.2823E^{-10}$	9.20	-	C	A	15.22	0.00	6.75	193	41	1
	141	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	142	$6.2823E^{-10}$	9.20	-	C	15.01	0.00		193	42	
	143	$6.2823E^{-10}$	9.20	-	A	15.01	0.00		193	42	
	144	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	145	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	146	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	147	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	148	$6.2823E^{-10}$	9.20	-	C	15.01	0.00		193	42	
	149	$6.2823E^{-10}$	9.20	-	A	15.01	0.00		193	42	
	150	$6.2823E^{-10}$	9.20	-	A	15.01	0.00		193	42	
	151	$6.2823E^{-10}$	9.20	-	C	15.01	0.00		193	42	
	152	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	153	$6.2823E^{-10}$	9.20	-	C	15.01	0.00		193	42	
	154	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	155	$4.00E^{-10}$	9.40	-	G	15.01	0.00		193	42	
	156	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	157	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	158	$6.2823E^{-10}$	9.20	-	C	15.01	0.00		193	42	
	159	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	160	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	161	$6.2823E^{-10}$	9.20	-	T	15.01	0.00		193	42	
	162	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	163	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
	164	$6.2823E^{-10}$	9.20	-	G	15.01	0.00		193	42	
CBF-A13_SNP3	294	$6.2823E^{-10}$	9.20	C	A	15.01	0.00		193	42	
CBF-A14_InDel1	-506	$6.2823E^{-10}$	9.20	G	-	15.01	0.00		193	42	
	-505	$6.2823E^{-10}$	9.20	T	-	15.01	0.00		193	42	
	-504	$6.2823E^{-10}$	9.20	G	-	15.01	0.00		193	42	
	-503	$6.2823E^{-10}$	9.20	A	-	15.01	0.00		193	42	
	-502	$6.2823E^{-10}$	9.20	G	-	15.01	0.00		193	42	
	-501	$6.2823E^{-10}$	9.20	T	-	15.01	0.00		193	42	
	-500	$6.2823E^{-10}$	9.20	G	-	15.01	0.00		193	42	
	-499	$6.2823E^{-10}$	9.20	T	-	15.01	0.00		193	42	
	-498	$6.2823E^{-10}$	9.20	G	-	15.01	0.00		193	42	
	-497	$6.2823E^{-10}$	9.20	A	-	15.01	0.00		193	42	
	-496	$6.2823E^{-10}$	9.20	G	-	15.01	0.00		193	42	
	-495	$6.2823E^{-10}$	9.20	T	-	15.01	0.00		193	42	
CBF-A14_SNP1	-449	$6.2823E^{-10}$	9.20	T	C	15.01	0.00		193	42	

Table 10: (Continued)

Gene/ Polymorphism name	CDS and promotor site	*P*-value	-Log of *P*	Polymorphism			Effect in %			Observations*		
CBF-A14_SNP2	-158	$6.2823E^{-10}$	9.20	T	C		15.01	0.00		193	42	
CBF-A14_SNP3	-53	$6.57E^{-10}$	9.18	C	G		15.01	0.00		193	42	
CBF-A14_SNP4	576	$6.2823E^{-10}$	9.20	G	A		15.01	0.00		193	42	
CBF-A15_SNP1	84	$6.2823E^{-10}$	9.20	C	T		15.01	0.00		193	42	
CBF-A15_SNP2	243	$6.2823E^{-10}$	9.20	T	C		15.01	0.00		193	42	
CBF-A15_SNP3	293	$6.2823E^{-10}$	9.20	C	T		15.01	0.00		193	42	
CBF-A15_SNP4	397	$6.2823E^{-10}$	9.20	G	A		15.01	0.00		193	42	
CBF-A15_InDel1	503	$6.2823E^{-10}$	9.20	T	-		15.01	0.00		193	42	
	504	$6.2823E^{-10}$	9.20	C	-		15.01	0.00		193	42	
	505	$6.2823E^{-10}$	9.20	G	-		15.01	0.00		193	42	
	506	$6.2823E^{-10}$	9.20	T	-		15.01	0.00		193	42	
	507	$6.2823E^{-10}$	9.20	C	-		15.01	0.00		193	42	
	508	$6.2823E^{-10}$	9.20	G	-		15.01	0.00		193	42	
	509	$6.2823E^{-10}$	9.20	T	-		15.01	0.00		193	42	
CBF-A15_InDel1	510	$6.2823E^{-10}$	9.20	C	-		15.01	0.00		193	42	
	511	$6.2823E^{-10}$	9.20	G	-		15.01	0.00		193	42	
CBF-A15_SNP5	694	$6.2823E^{-10}$	9.20	A	G		15.01	0.00		193	42	
CBF-A15_SNP6	716	$6.2823E^{-10}$	9.20	C	G		15.01	0.00		193	42	
PPD-D1_Indel1	1266	0.03771	1.42	-	C		3.96	0.00		68	166	
	1267	0.03771	1.42	-	G		3.96	0.00		68	166	
	1268	0.03771	1.42	-	T		3.96	0.00		68	166	
	1269	0.03771	1.42	-	C		3.96	0.00		68	166	
	1270	0.03771	1.42	-	G		3.96	0.00		68	166	
VRN-A1_SNP1	349	$3.03E^{-04}$	3.52	C	T	Y	0.00	1.35	0.11	24	2	209
VRN-B3_SNP1	19	0.03758	1.43	C	G		3.05	0.00		105	128	

CDS: coding DNA sequence; *P*: probability; Log: logarithm; * number of genotypes per allele

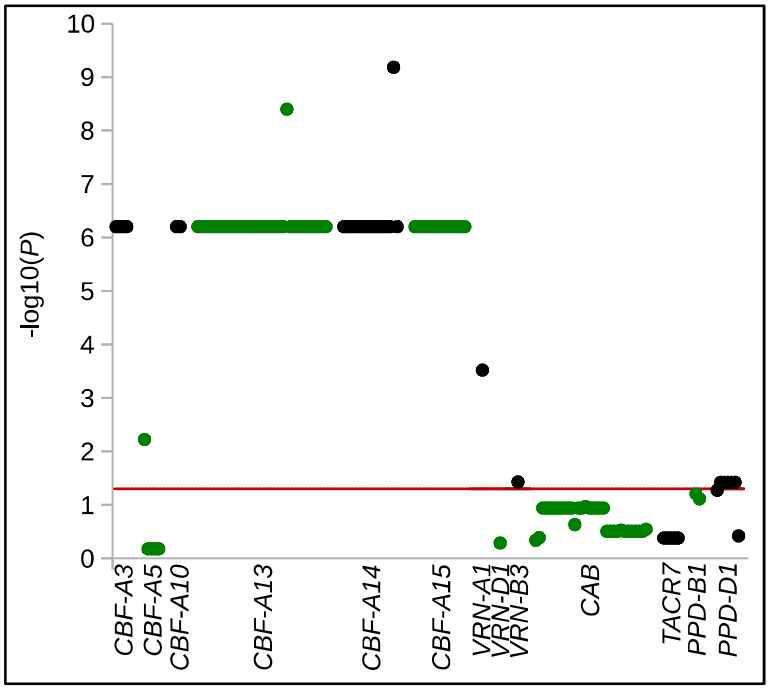


Figure 13: Manhattan plot of SNPs/indels in candidate genes for FT. The-log10 (*P*-values) from the association analysis are plotted against the respective candidate gene. The red horizontal line indicates the significance threshold at $P < 0.05$. For better visualisation, the successive genes are shown in alternating black and green colors (Babben et al., 2018).

3.2.5 Linkage disequilibrium (LD) and germplasm origin

Out of 235 studied genotypes, 116 originate from Europe, 38 from North America and 81 from Asia. After MAF selection, 51 polymorphic sites were identified in the genotypes from Europe, 88 in those derived from North America and 25 in Asian genotypes. All three different sub-populations show that the *CBF* genes on chromosome 5A are in a very high LD and are separated from other 5A located genes, i.e. *VRN-A1* and *CAB*. All non-5A candidate genes show a low LD to each other in all sub-populations (Figure 14).

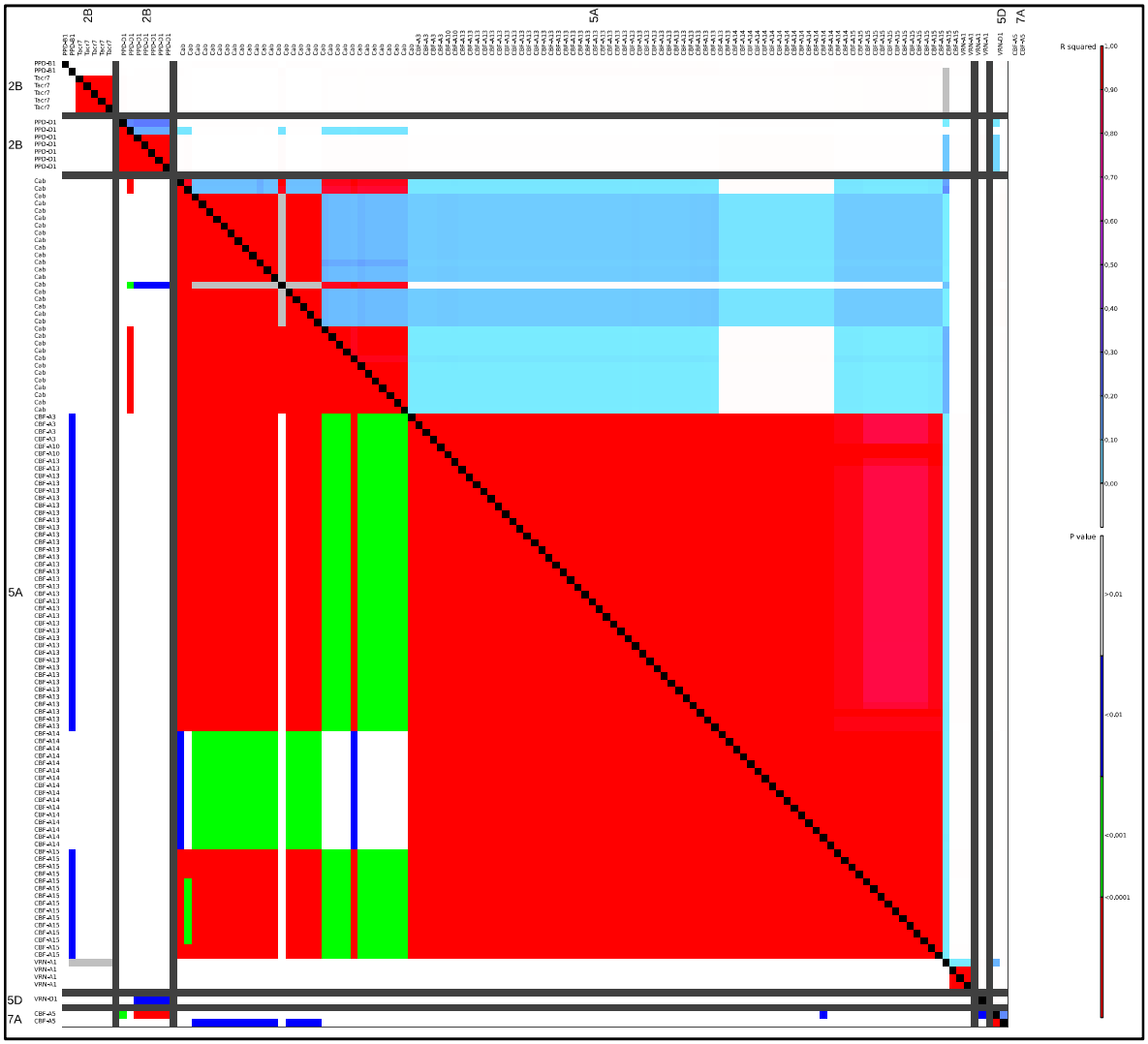


Figure 14: LD plot of all used SNPs and indels for association analysis. On the left side and at the top the genes and chromosomes with polymorphic sites are shown. Each colored square in the below triangle represents the intensity of LD expressed by *P*-values for each pairwise comparison between polymorphic sites. Each colored square in the top triangle represents the intensity of LD expressed by r^2 for each pairwise comparison between polymorphic sites. To the right the legend for r^2 values and *P*-values are shown

3.2.6 Haplotype association analysis

A set of 44 haplotypes and components of haplotypes from 15 candidate genes were identified. After MAF selection, 32 haplotypes (six promoter haplotypes, 12 exon haplotypes, one intron haplotype, three 3` UTR haplotypes and 10 whole gene haplotypes) in 14 candidate genes (*CAB*, *CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *CBF-A18*, *TACR7*, *VRN-B3*, *VRN-A1*, *VRN-D1*, *PPD-B*1 and *PPD-D1*) were used for haplotype analysis. In the haplotype association analysis, 22 haplotypes (five promoter haplotypes, nine exon haplotypes, one intron haplotype and seven whole gene

haplotypes) in ten candidate genes revealed significant associations ($P < 0.05$) (Table 11). *CBF-A18* is the only exon haplotype not significantly associated to FT. The same applies for promoter haplotypes of *CBF-A5*, *CBF-A13*, *CBF-A14*, *PPD-B1* and *PPD-D1*, intron haplotype of *PPD-D1* and the whole gene haplotypes of *CBF-A5*, *CBF-A13*, *CBF-A14*, *CBF-A18*, *PPD-B1*, *PPD-D1* and *VRN-A1*. The associated genes are located on chromosome 2B (*PPD-B1*), 2D (*PPD-D1*), 5A (*CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *VRN-A1*), 6A (*CBF-A18*) and 7A (*CBF-A5*). The allelic effects of associated haplotypes for winter survival ranged from 0.7% to 34%. The haplotypes of the five *CBF* genes on chromosome 5A (*CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15*) showed very high –log of *P*-values. All of these are localised at the *FR-A2* locus and underlie high LD (Figure 15, Table 11, Supplemental Table 6).

FT associated SNPs/indels and haplotypes were identified in 11 genes (*CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *CBF-A18*, *VRN-A1*, *VRN-B3*, *PPD-B1* and *PPD-D1*). The presented analysis merely detects statistically conspicuous genes and their polymorphisms. Therefore, the SNPs and indels must be further analysed for their functional relevance. Since we focused on the genetic aspect of FT as a trait up to this point, we neglect the fact that in growing organisms proteins rather than genes are responsible for vital functions. Therefore, an analysis which changes in the AA sequence are generated by the detected polymorphisms and what effect they have on protein functionality are of great interest.

Table 11: Statistics of haplotypes significantly associated to FT.

Haplotype name	Chr.	*P*-value	-Log of *P*	Haplotype						FT effect of haplotypes in %						Observations*					
CBF-A3 exon haplotype	5A	$2.37E^{-11}$	10.63	1	2					15.01	0.00					193	42				
CBF-A5 promotor haplotype	7A	$7.17E^{-03}$	2.14	1	2					4.44	0.00					82	143				
CBF-A5 exon haplotype	7A	$1.58E^{-02}$	1.80	1	2	3				19.51	18.09	0.00				204	28	3			
CBF-A5 haplotype	7A	$5.16E^{-04}$	3.29	1	2	3	4			24.18	19.29	18.88	0.00			55	27	140	3		
CBF-A10 exon haplotype	5A	$2.37E^{-11}$	10.63	1	2					15.01	0.00					193	42				
CBF-A13 exon haplotype	5A	$1.57E^{-10}$	9.80	1	2	3				15.22	0.00	8.47				193	41	1			
CBF-A13 promotor haplotype	5A	$2.37E-^{11}$	10.63	1	2					15.01	0.00					193	42				
CBF-A13 haplotype	5A	$1.57E-^{10}$	9.80	1	2	3				15.22	0.00	8.47				193	41	1			
CBF-A14 promotor haplotype	5A	$2.37E^{-11}$	10.63	1	2					15.01	0.00					193	42				
CBF-A14 exon haplotype	5A	$1.90E^{-10}$	9.72	1	2	3				17.01	2.11	0.00				193	40	2			
CBF-A14 haplotype	5A	$1.90E^{-10}$	9.72	1	2	3				17.01	2.11	0.00				193	40	2			
CBF-A15 exon haplotype	5A	$2.37E^{-11}$	10.63	1	2					15.01	0.00					193	42				
CBF-A18 haplotype	6A	$6.88E^{-03}$	2.16	1	2	3				9.32	6.66	0.00				209	9	17			
PPD-B1 promotor haplotype	2B	$4.23E^{-02}$	1.37	1	2					5.50	0.00					213	22				
PPD-B1 exon haplotype	2B	$3.69E^{-07}$	6.43	1	2	3	4			18.97	0.00	28.28	33.39			3	5	207	20		
PPD-B1 haplotype	2B	$1.31E^{-07}$	6.88	1	2	3	4	5		19.03	0.00	33.95	29.07	22.63		3	5	20	185	22	
PPD-D1 promotor haplotype	2D	$6.95E^{-03}$	2.16	1	2	3				14.38	10.75	0.00				141	80	4			
PPD-D1 intron haplotype	2D	$3.93E^{-02}$	1.41	1	2	3				18.66	19.56	0.00				162	69	3			
PPD-D1 exon haplotype	2D	$7.05E^{-03}$	2.15	1	2	3				21.80	18.18	0.00				68	163	3			
PPD-D1 haplotype	2D	$3.78E-^{03}$	2.42	1	2	3	4	5	6	21.47	3.77	19.28	16.13	0.00	20.55	68	3	69	90	3	1
VRN-A1 exon haplotype	5A	$2.57E^{-04}$	3.59	1	2	3	4			0.00	0.68	13.12	11.89			9	15	2	209		
VRN-A1 haplotype	5A	$2.87E^{-03}$	2.54	1	2	3	4	5	6	14.13	15.90	2.26	15.30	0.00	8.90	207	2	13	2	6	1

Chr.: Chromosome; *P*: probability; Log: logarithm; FT: frost tolerance; * number of genotypes per allele

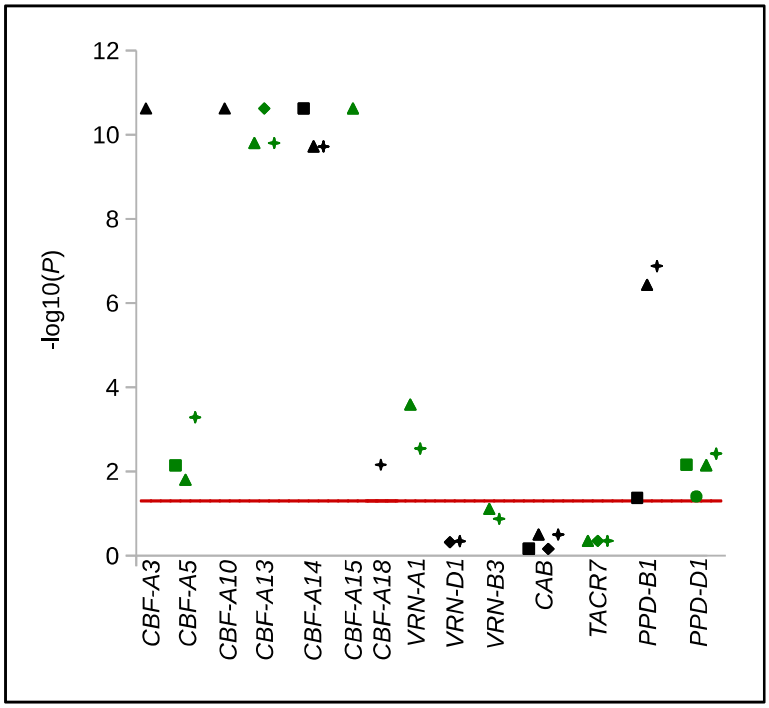


Figure 15: Manhattan plot of haplotypes based on FT candidate genes. The -log10 (*P*-values) from the association analysis are plotted against candidate genes. The red horizontal line indicates the significance threshold at $P < 0.05$. The squares show the promotors, triangle the exons, circles the introns, diamonds the 3' UTRs and stars the whole haplotypes. For better visualisation, the successive genes are shown in alternating black and green colors (Babben et al., 2018).

3.2.7 *In silico* sequence analysis

C-REPEAT BINDING FACTORS (CBFs)

To validate and annotate obtained sequences of candidate genes, an *in silico* analysis was performed. NCBI protein BLASTx of CBF AA sequences results in nine homologues from five related and four unrelated species (Supplemental Table 7). For the CBFs (CBF-A3, CBF-A5, CBF-A10, CBF-A13, CBF-A14, CBF-A15, CBF-A18) that showed a significant association to FT, the AP2/EREBP transcription factor domain was identified, which is predicted to encode a protein structure with three β-strands and one α-helix (2gcc). The results of protein structure modelling are shown in Table 12. The CBF-A13_1 haplotype protein shows an appropriate *P*-value of 4.79×10^{-4}, a high uGDT value of 51 and a low uSeqID value of 15. The protein structure of CBF-A13_2 could not be predicted. In all remaining cases protein structure models were of high quality. Additionally, the

PKK/RPAGRxKFxETRHP and DSAWR motifs were identified, which are typical features of CBF proteins. Comparison of AA sequences of two CBF-A3 haplotypes revealed three AA substitutions. The CBF-A3_SNP2 [C/G] on CDS site 263 leads to an Ala/Gly substitution, CBF-A3_SNP3 [A/G] on CDS site 367 to a Ser/Gly substitution and CBF-A3_SNP4 [C/T] on CDS site 407 to a Ser/Phe substitution. The AA substitution of CBF-A3_SNP2 is located in the α-helix of the AP2 domain. The CBF-A3_SNP3 and CBF-A3_SNP4 is located upstream of the identified domains/motifs. The AA site of CBF-A3_SNP2 is significant for negative selection (Figure 16). The AA haplotypes of CBF-A5 show three AA changes, which are not significantly associated in the SNP/indel analysis. In contrast CBF-A5_SNP2 [C/A] resulted in an Arg/Ser and CBF-A5_SNP4 [A/G] in a His/Arg substitution which are located in the β-strand 3 and the α-helix of the AP2 domain, respectively. Additionally, CBF-A5_SNP4 site underlies negative selection (Supplemental Figure 5). The AA haplotypes of CBF-A10 show an AA substitution of Gly/Ala based on CBF-A10_SNP2 [G/C] on CDS site 518. This substitution is upstream of the identified domains/motifs and at a site not showing positive or negative selection (Supplemental Figure 6). The haplotypes of CBF-A13 show completely different AA sequences. The single bp deletion on CDS site 19 of CBF-A13_indel1 results in a frame shift that replaces every AA from the seventh AA onwards. Furthermore, a premature stop codon at position AA 74 was detected. The haplotype one shows a 32 bp deletion at CDS site 133 to 164 (CBF-A13_indel2). This deletion is localised in the AP2 domain and causes a stop codon after 135 AAs. In summary, haplotype one of CBF-A13 shows a higher similarity to homologous AA sequences of related species and the AP2 domain than haplotype two. Due to the short AA sequences of both CBF-A13 haplotypes and low similarity to homologous AA sequences, no assumptions about positive or negative selection or functionality can be made (Supplemental Figure 7). The AA analysis of two CBF-A15 haplotypes revealed four AA substitutions and one AA indel. A nine-nucleotide insertion (CBF-A15_indel1) within the coding region (sites 503 to 511) resulted in three additional Ser residues in the haplotype CBF-A15_1. The AA substitution Ala/Val caused by CBF-A15_SNP3 [C/T] on CDS site 293 is located in the AP2 domain. The CBF-A15_SNP4 [G/A] on CDS site 397 resulted in an Ala/Thr AA substitution. The last AA substitutions are Ser/Gly caused by CBF-A15_SNP5 [A/G] on CDS site 694 and Ser/Trp caused by CBF-A15_SNP6 [C/G] on CDS site 716. The CBF-A15_SNP4, CBF-A15_SNP5, CBF-

A15_SNP6 and CBF-A15_inde1 are located upstream of the identified domains/motifs. All five AA substitution regions show no signatures of positive or negative selection (Figure 17A, Supplemental Figure 8). The CBF-A14 AA haplotypes do not result in AA substitutions (Supplemental Figure 9). The CBF-A18 haplotypes carry 14 AA substitutions. However, none of these is associated with FT or underlies negative or positive selection (Supplemental Figure 10).

VERNALISATION GENES (VRN-A1 and VRN-B3)

Nine homologous AA sequences of VRN-A1 and VRN-B3 were identified (Supplemental Table 7). The RaptorX analysis of VRN-A1 identified protein structures from MADS-box/MEF2 (myocyte enhancer factor 2) domain (3kov) and MEF2A (1egw), which are located in the same conserved domain. In addition, the keratin-like (K) domain (4ox0) was identified. The complete complex was described by Theissen et al. (1996) and Kaufmann et al. (2005) with a MADS DNA binding (M) domain, a type II transcription factor containing the Intervening (I) domain, a Keratin-like coiled-coil (K) domain and a C-terminal (C) domain (MIKC-type). The parameters of protein structure modelling are shown in Table 12. The VRN-A1_SNP1 [C/T/Y] on CDS site 349 generated an AA substitution of Leu/Phe in the K domain between α-helix three and α-helix four. Furthermore, in the region of this AA substitution, no positive or negative selection was identified (Figure 17B, Supplemental Figure 11).

The association study revealed that VRN-B3 is significantly associated with FT only in the SNP/indel analysis. The protein prediction analysis identified an Hd3A protein (3axy), which is a mobile flowering signal in rice (Table 12). This flowering time family protein contains four α-helices, seven β-strands and one segment B (Taoka et al., 2011). The VRN-B3_SNP1 [C/G] on CDS site 19 generates an AA substitution of His/Asp directly before the start of the α-helix 1. Furthermore, this site is subject to significant negative selection (Figure 17C, Supplemental Figure 12).

PHOTOPERIOD RESPONSE GENES (PPD-B1 and PPD-D1)

The association study revealed that the gene *PPD-B1* is significantly associated with FT only in the haplotype analysis, *PPD-D1* however in both analyses. The NCBI protein BLASTx of the PPD-B1 and PPD-D1 AA haplotype sequences showed nine homologues (Supplemental Table 7). On the basis of the protein BLAST analysis, the Pseudo-Receiver domain and the CONSTANS motif of the pseudo-response regulator (PRR) protein family were identified (Matsushika et al., 2000, Strayer et al., 2000). In the RaptorX analysis the protein models of a putative response regulator domain (3t6k) and response regulator receiver domain (3jte) were identified (Table 12). These domains contain five α-helices and five β-strands. The PPD-B1 AA haplotypes exhibited three AA substitutions. The PPD-B1_SNP2 [A/G] on CDS site 304 generated an AA substitution of Arg/Gly and PPD-B1_SNP3 [A/G] on CDS site 368 generated an AA substitution of Asn/Asp. Both are located in the Pseudo-Receiver domain. The PPD-B1_SNP2 is located in the α-helix 3 and PPD-B1_SNP3 between β-strand 4 and α-helix 4. The third AA substitution from Asp to Asn is caused by PPD-B1_SNP5 [G/A] on CDS site 623. All three AA substitution sites show no significant positive or negative selection (Figure 17D, Supplemental Figure 13). The associated PPD-D1_indel1 on CDS site 1266 to 1270 bp produced a stop codon on AA position 470. Therefore, haplotype one has no CONSTANS (CO) motif. In consequence, the PPD-D1 protein cannot interact with the CO protein. All other AA substitutions between haplotype two and three did not show significant associations in the SNP/indel association study (Supplemental Figure 14).

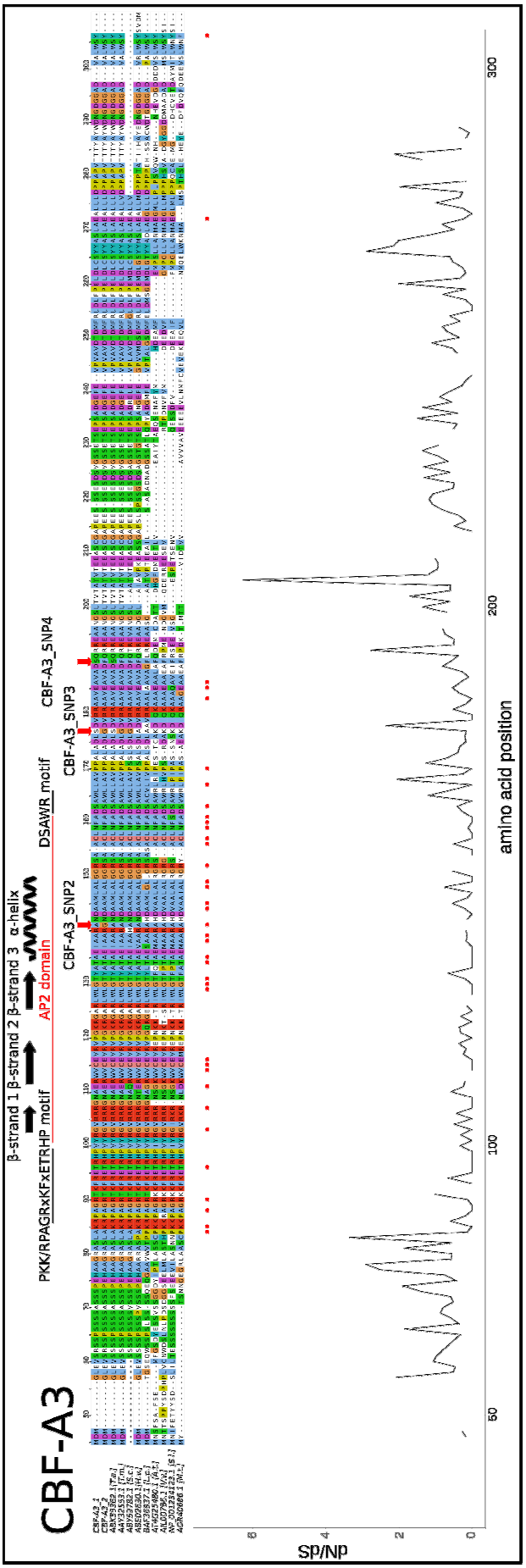


Figure 16: Amino acid alignment and nucleotide divergence rates (dN/dS) of CBF-A3 and homologous amino acid sequences. Illustrated are alignments of two haplotype AA sequences of CBF-A3 and nine homologue plant AA sequences. The numbers above the alignment indicate the sites of AAs. The black line above the alignment illustrates the PKK/RPAGRxKFxETRHP and DSAWR motif, the red line the AP2 domain, the black arrows the β-strands and the black spirals the α-helices. The red arrows label the AA changes based on significantly associated SNPs or indels. The plot below the alignment shows the nucleotide divergence rates (dN/dS). The black line describes the dN/dS ratio. The red dots between alignment and plot indicate sites with significant negative selection ($P < 0.05$) (Babben et al., 2018).

Table 12: Protein structure modelling of associated genes.

Protein haplotypes	*P*-value	uGDT (GDT)	uSeqID (SeqID)	Score	Domain/motif	Literature
CBF-A3_1	$3.03E^{-4}$	71 (62)	30 (26)	51	2gcc (AP2)	(Allen et al., 1998)
CBF-A3_2	$1.52E^{-4}$	71 (61)	29 (25)	51	2gcc (AP2)	(Allen et al., 1998)
CBF-A5_1	$3.22E^{-4}$	69 (67)	30 (29)	51	2gcc (AP2)	(Allen et al., 1998)
CBF-A5_2	$5.75E^{-4}$	56 (54)	30 (29)	54	2gcc (AP2)	(Allen et al., 1998)
CBF-A5_3	$9.71E^{-4}$	57 (56)	29 (28)	56	2gcc (AP2)	(Allen et al., 1998)
CBF-A10_1	$2.78E^{-4}$	70 (66)	31 (29)	51	2gcc (AP2)	(Allen et al., 1998)
CBF-A10_2	$2.88E^{-4}$	71 (67)	31 (29)	50	2gcc (AP2)	(Allen et al., 1998)
CBF-A13_1	$4.79E^{-4}$	51 (38)	15 (11)	36	2gcc (AP2)	(Allen et al., 1998)
CBF-A13_2	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CBF-A14_1	$1.97E^{-4}$	71 (69)	29 (28)	52	2gcc (AP2)	(Allen et al., 1998)
CBF-A14_2	$1.97E^{-4}$	71 (69)	29 (28)	52	2gcc (AP2)	(Allen et al., 1998)
CBF-A15_1	$3.80E^{-4}$	70 (66)	29 (27)	50	2gcc (AP2)	(Allen et al., 1998)
CBF-A15_2	$2.95E^{-4}$	70 (70)	29 (29)	49	2gcc (AP2)	(Allen et al., 1998)
CBF-A18_1	$1.33E^{-3}$	55 (49)	29 (26)	52	2gcc (AP2)	(Allen et al., 1998)
CBF-A18_2	$8.72E^{-4}$	55 (49)	29 (26)	54	2gcc (AP2)	(Allen et al., 1998)
PPD-B1	$1.24E^{-4}$	105 (62)	32 (19)	130	3t6k (putative response regulator domain)	n.a.
	$8.74E^{-5}$	101 (60)	27 (16)	134	3jte (response regulator receiver domain)	(Bachhawat and Stock, 2007)
PPD-B2	$1.21E^{-4}$	104 (61)	32 (19)	129	3t6k (putative response regulator domain)	n.a.
	$8.12E^{-5}$	101 (59)	27 (16)	134	3jte (response regulator receiver domain)	(Bachhawat and Stock, 2007)
PPD-B3	$1.06E^{-4}$	103 (60)	31 (18)	129	3t6k (putative response regulator domain)	n.a.
	$7.21E^{-5}$	102 (59)	27 (16)	134	3jte (response regulator receiver domain)	(Bachhawat and Stock, 2007)
PPD-B4	$1.28E^{-4}$	104 (61)	32 (19)	130	3t6k (putative response regulator domain)	n.a.
	$8.98E^{-5}$	102 (59)	27 (16)	134	3jte (response regulator receiver domain)	(Bachhawat and Stock, 2007)
PPD-D1_1	$1.05E^{-4}$	104 (47)	32 (15)	131	3t6k (putative response regulator domain)	n.a.
	$7.11E^{-5}$	102 (46)	27 (12)	136	3jte (response regulator receiver domain)	(Bachhawat and Stock, 2007)
PPD-D1_2	$1.14E^{-4}$	105 (63)	32 (19)	130	3t6k (putative response regulator domain)	n.a.
PPD-D1_3	$1.16E^{-4}$	104(62)	32 (19)	129	3t6k (putative response regulator domain)	n.a.
	$8.07E^{-5}$	101 (60)	27 (16)	134	3jte (response regulator receiver domain)	(Bachhawat and Stock, 2007)

Table 12: (Continued)

Protein haplotypes	*P*-value	uGDT (GDT)	uSeqID (SeqID)	Score	Domain/motif	Literature
VRN-A1_1	$8.02E^{-3}$	84 (83)	39 (39)	84	4ox0 (K domain)	(Puranik et al., 2014)
	$1.02E^{-4}$	78 (103)	38 (50)	53	3kov (MADS domain/I domain)	(Wu et al., 2010)
	$1.30E^{-4}$	75 (98)	38 (50)	52	1egw (MADS domain/I domain)	(Santelli and Richmond, 2000)
VRN-A1_2	$8.47E^{-3}$	83 (82)	39 (39)	84	4ox0 (K domain)	(Puranik et al., 2014)
	$7.90E^{-5}$	73 (103)	38 (50)	54	3kov (MADS domain/I domain)	(Wu et al., 2010)
	$9.11E^{-5}$	74 (98)	38 (50)	53	1egw (MADS domain/I domain)	(Santelli and Richmond, 2000)
VRN-A1_3	$7.95E^{-3}$	84 (83)	38 (38)	84	4ox0 (K domain)	(Puranik et al., 2014)
	$7.06E^{-5}$	78 (103)	38 (50)	54	3kov (MADS domain/I domain)	(Wu et al., 2010)
	$1.08E^{-4}$	75 (99)	38 (50)	53	1egw (MADS domain/I domain)	(Santelli and Richmond, 2000)
VRN-B3_1	$7.92E-^{13}$	154 (87)	147 (83)	166	3axy (Hd3A protein)	(Taoka et al., 2011)
VRN-B3_2	$1.66E-^{12}$	154 (87)	148 (84)	168	3axy (Hd3A protein)	(Taoka et al., 2011)

P: probability; uGDT: unnormalised GDT (Global Distance Test) score; GDT: calculated as uGDT divided by the protein (or domain) length and multiplied by a 100; uSeqID: number of identical residues in the alignment; SeqID: uSeqID normalised by the protein (or domain) sequence length and multiplied by 100; n.a.: not

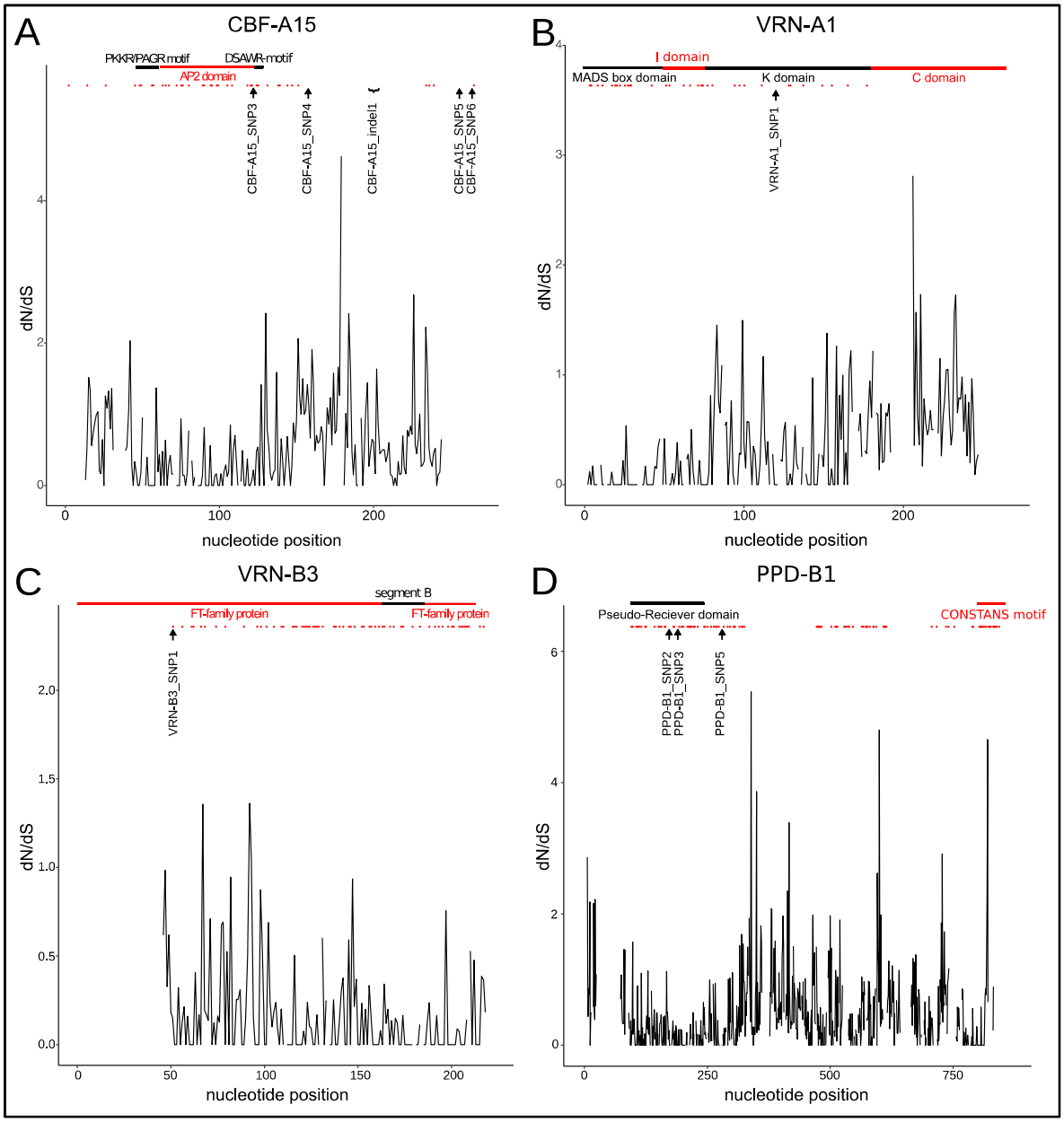


Figure 17: Nucleotide divergence rates (dN/dS) of *CBF-A15*, *VRN-A1*, *VRN-B3* and *PPD-B1* genes and homologue amino acid sequences. A Nucleotide divergence rates (dN/dS) between the identified haplotypes of CBF-A15, reference sequence of *T. aestivum* and eight homologous plant AA sequences. The black line on top indicates the PKK/RPAGRxKFxETRHP motif and DSAWR, the red line the AP2 domain. The black arrows and clamp label the position of AA changes based on significantly associated SNPs and indels. The black line shows the dN/dS ratio. The red dots illustrate sites underlying significant negative selection (*P* < 0.05). **B** Nucleotide divergence rates (dN/dS) between the identified haplotypes of VRN-A1, reference sequence of *T. aestivum* and eight homologous plant AA sequences. The black line on top indicates the MADS box and K domain, the red line the I and C domain. The black arrow labels the position of AA change based on significantly associated SNPs. The black line describes the dN/dS ratio. The red dots indicate sites with significant negative selection (*P* < 0.05). **C** Illustrated are nucleotide divergence rates (dN/dS) between the identified haplotypes of VRN-B3, reference sequence of *T. aestivum* and eight homologous plant AA sequences. The black line on top indicatesthe segment B, the red line the flowering time (FT)-family protein. The black arrow labels the position of AA change based on significant associated SNPs. The black line describes the dN/dS ratio. The red dots illustrate sites with significant negative selection (P < 0.05). **D** Nucleotide divergence rates (dN/dS) between the identified haplotypes of PPD-B1, reference sequence of *T. aestivum* and eight homologous plant AA sequences. The black line on top indicates the Pseudo-Reciver domain, the red line the CONSTANS motif. The black arrows label the position of AA changes significantly associated SNPs. The black line describes the dN/dS ratio. The red dots indicate sites with significant negative selection (P < 0.05) (Babben et al., 2018).

In conclusion, AA substitutions in important protein domains of CBF-A3, CBF-A15, VRN-A1, VRN-B3, PPD-B1 and PPD-D1 were identified. The polymorphic sites CBF-A3_SNP2 and VRN-B3_SNP1 show significant signatures of negative selection and PPD-D1_indel1 removes the CO motif of the PPD-D1 protein of haplotype one.

3.2.8 *In silico* promoter analysis

The significantly associated CBF-A5_inde1 [C/−], which is located 83 bp upstream of the start codon, entailed no changes on the promoter regulatory sites and showed no influence on the transcription of the *CBF-A5* gene in the *in silico* promotor analysis. Also, the associated polymorphic sites CBF-A13_SNP1 of the *CBF-A13* promotor and the sites CBF-A14_indel1, CBF-A14_SNP1, CBF-A14_SNP2 and CBF-A14_SNP3 of the *CBF-A14* promoter exhibited no regulatory site changes which are indicative for a modification of the gene transcription.

4 Discussion

After hybridisation and domestication breeding activities have shaped the genome of bread wheat and reduced the level of genetic diversity which is nowadays the major limiting factor in breeding of cultivars resistant to biotic and abiotic stresses (Tanksley and McCouch, 1997). To exploit the genetic variation, wheat genome sequences along with new bioinformatics platforms and databases containing recent genomic data are a powerful resource for the development of tools for molecular plant breeding. In this thesis, wheat genome sequence information was applied to identify genetic variation in candidate genes for FT in a large genotype collection to rediscover hidden genetic diversity and facilitate the subsequent utilisation in wheat breeding programs.

4.1 Gene specific primer development and chromosomal assignment of specific PCR fragments by using NT- and deletion lines

The rapid progress in sequencing of plant genomes leads to the accumulation of whole genome sequence data, allowing the fast development of locus/genome specific markers in complex plant genomes (e.g. wheat) with a high success rate. Up to now, high homology of the hexaploid wheat genome hampered the success in gene specific primer development. Knowledge of the gene structure is important for marker development, since wheat introns have more sequence differences between the homologous chromosomes than exons (Poczai et al., 2010, Ishikawa et al., 2007). The gene structure reconstruction and comparison of homologous sequences facilitate an improved development of gene specific PCR fragments as well as re-sequencing of target genes. Furthermore, the development of gene specific primers in wheat is depending on their chromosomal location within the genome. Therefore, the application of available wheat NT- and deletion lines is a useful tool to physically map and monitor the locations of the designed PCR fragments in the lab, which allows to verify the specificity of PCR fragments.

4.2 Specificity of developed primers

Specificity of primers is the non-recurring binding in the target genome. This is reflected in a single PCR product and a correct or syntenically localised amplicon. The inspection of primer functionality and single PCR product generation is a standard for the development of primers and therefore the first necessary step of the presented approach. Via the first inspection step eliminated ~28% of studied primer pair combinations were eliminated. Most of these lacked PCR amplification product probably due to non-binding of target sequences. To succeed in amplification of PCR fragments not yet producible, the primer binding sites have to be re-analysed using BLASTn in combination with the recently published wheat reference sequences (RefSeq v1.0) (Appels et al., 2018). Thereby, primer mismatches should be identified and corrected, which would decrease the elimination rate tremendously. The second important step in the assessment of amplicon specificity is mapping of PCR products via NT-lines to get information about the correct amplification on the correct target chromosome template and sub-genome. By using NT-mapping of PCR amplicons ~18% of primer pair combinations were eliminated. One part of the eliminated PCR products showed a chromosomal localisation that differs from the reported ones. In this case, a non-specific binding in the wheat genome is assumed. This can occur if primers are derived from related organisms and not from wheat itself. For seven of eight discarded candidate genes, sequences of related organisms (*Triticum monococcum* and *Hordeum vulgare*) were used for primer development. The other part of eliminated primer pair combinations showed a PCR product on all NT-lines which may be due to the fact that both primers (forward and reverse) bind at least to two sub-genomes. Here the same applies as for the amplification test. Primer binding sites of the PCR fragments which are localised wrong have to be controlled by using BLASTn in combination with the recently published RefSeq v1.0 (Appels et al., 2018). Thereby, additional primer sub-genome annealing positions will be identified enabling sub-genome specific primer design.

By using the wheat chromosome arm sorted sequences (Brenchley et al., 2012, Raats et al., 2013, Appels et al., 2018, IWGSC, 2014) and simple comparative methods gene specific primers in hexaploid wheat were developed with a high success rate of almost 60%. Also an excellent rate of ~55% for specific fragment amplification confirmed the

usefulness of wheat genomic sequence. To my knowledge such an efficiency is not yet described in literature for specific primer/marker development in polyploid plants. An overview of published success rates revealed a variation in microsatellite amplification in wheat between 22.9 and 45% (Roder et al., 1995, Bryan et al., 1997, Nicot et al., 2004, Song et al., 2005). In cotton this rate was 23.3% (Han et al., 2006). Wang et al. (2010) estimated the development of effectively derived primers for sequence tagged sites (STS) with 24.6% and for STS primer combinations with only 3.7% in wheat. Chen et al. (2014) achieved a rate of 27.5% for STS marker development in wheat. In *Brassica oleracea* (which is a paleohexaploid species) a success rate of 29.1% is described in allele specific PCR primer development (Liu et al., 2012). The highest success rate reported in literature is for potato (Poczai et al., 2010). In this study a rate of 51.8% developed intron targeting (IT) markers was achieved. The availability of wheat chromosome arm sorted sequences in combination with a primer pick based on two nucleotide differences within the primer binding site respectively one difference within the last seven nucleotides at the 3` end of the primer based on the analyses of the three homologous target sequences (Wu et al., 2009) permits a gene specific primer development pipeline with such a high efficiency. It has to be mentioned that for this approach inclusion of the intron-exon gene structure is essential to derive primers within the introns. Intron intrinsic primers facilitate higher specificity discriminating different sub-genomes due to a lower conservation rate (Poczai et al., 2010, Ishikawa et al., 2007) and allow sequencing of the entire coding regions.

4.3 Sequencing of frost tolerance candidate genes and BLAST based verification

In this study, 18 out of 19 FT genes (94.7%) were sequenced using the same primers used for PCR amplification. For *CBF-B7*, for which initial sequencing failed, a set of newly designed primers improved the sequencing, therefore optimisation for single band products is recommended as part of the verification procedure. Concerning the gene *TACR7*, Kocsy et al. (2010) claimed BQ659345 of *Hordeum vulgare* as identical to the *TACR7* gene in wheat. However, the analysis of the generated sequences presented in this thesis showed an identity of 84% to the reference sequence L28093 for *TACR7* of wheat

and 92% to BQ659345. In contrast, our sequences reveal an identity of 92% to X97916 of *Hordeum vulgare* which is annotated as the *BARLEY LOW TEMPERATURE 14.1* (*BLT14.1*) gene. *BLT14.1* shows a considerable homology to *WHEAT LOW TEMPERATURE* (*WLT10*), as described by Ohno et al. (2001). Matching BQ659345 against X97916 results in an identity of 99%. Furthermore *TACR7*, *BLT14.1* and *WLT10* are located on chromosome 2 of both, barley and wheat (Kocsy et al., 2010, Ohno et al., 2001, Dunn et al., 1990). PCR fragments derived from *TACR7* were also mapped on chromosome 2B. Furthermore, it was shown recently that the newest sequence of *TACR7* (Kocsy et al., 2010) is very similar to *BLT14.1* and *WLT10*, in contrast to the L28093 sequence (described also as *TACR7* (Gana et al., 1997)). The nucleotide identity of 99% between *BTLT4.1* (X97916) and the initial reference sequence (BQ659345), which is published as *TACR7* (Kocsy et al., 2010), supported this hypothesis. All other PCR fragment sequences have shown a very good sequence identity to the original gene of interest (97.5% or more).

The sequencing of single bands and correct chromosome assigned PCR amplicons followed by BLAST based verification is the last step in the workflow presented in this study. The results of the BLAST based verification demonstrate that the selection of single band PCR products and the assignment to the correct chromosomes of the PCR amplicons is an efficient tool for locus specific primer selection. The combination of sequencing and BLAST based verification using the presented approach leads to very robust results with an error rate near zero.

The identified SNPs at 11 polymorphic candidate genes can be used for developing SNP based markers. Likewise, the indels in eight candidate genes are suited for marker development based on size polymorphisms. Developed PCR amplicons can be employed for re-sequencing of corresponding candidate genes in populations or genotypes not yet analysed, thereby allowing their genetic characterisation for the first time.

Locus specific primers are necessary for locus specific sequencing and detection of gene specific polymorphisms (SNPs and indels) in genotypes of interest. Subsequently sequence information obtained can be used for transcription activator-like effector nucleases (TALENs) (Boch et al., 2009, Moscou and Bogdanove, 2009, Christian et al., 2010) or clustered regularly interspaced short palindromic repeats (CRISPR/Cas) systems

(Upadhyay et al., 2013). Therefore our approach of development of locus specific primers can serve as a basis for various downstream applications including detection of novel polymorphisms, development of new markers, genetic mapping and gene editing in wheat.

4.4 LD and diversity

Application of the developed markers for 18 FT candidate genes to a set of 235 wheat varieties revealed that the group from Asia shows a very small diversity with only 25 polymorphic sites within its 81 genotypes. That implies these genotypes to be closely related to each other based on the analysed genes. The polymorphic sites on the same chromosomes show a high LD among themselves but not between the different chromosomes. Consequently, the polymorphic sites which are in a high LD are inherited together. In contrast, genotypes from North America show a very high diversity with 88 polymorphic sites in only 38 genotypes. That implies these genotypes to be more distantly related to each other. The gene cluster of *CBFs* (*CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15*) on chromosome 5A showed a very high LD of $r^2 = 1$. Consequently, this gene cluster has not been divided by meiotic events but is separated from *CAB* also located on chromosome 5A. An intermediate diversity was detected in genotypes from Europe with 51 polymorphic sites in 116 genotypes. The LD analysis shows a very high LD within three blocks located on chromosome 5A but not between them. Hence the *CBF* gene cluster (*CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15*), *CAB* and *VRN-A1* are inherited independently.

4.5 Association study and AA analysis

C-REPEAT BINDING FACTORS (CBFs)

FT is a highly complex and important trait of winter wheat that is usually studied using QTL and expression profiling approaches (Vagujfalvi et al., 2003, Knox et al., 2010, Zhu et al., 2014). In this study, to the best of my knowledge, the first large scale candidate gene based association analysis of FT in wheat was conducted.

Significantly associated polymorphisms (SNPs/indels) were identified in 11 of the 18 selected genes as well as respective haplotypes. Eight of these were detected in both approaches (*CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *PPD-D1* and *VRN-A1*). Out of the seven *CBF* genes, which are members of a large gene family that was investigated in this study, six revealed FT association using the SNP/indel and haplotype method. Associated polymorphisms concerning the five candidate genes *CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15* are located at the *FR-A2* locus (Francia et al., 2004, Vagujfalvi et al., 2003) on wheat chromosome 5A. The other two *CBF* members, i.e. *CBF-A5* and *CBF-A18*, are located on chromosomes 7A and 6A and do not belong to the *FR-A2* locus. Additionally, all the SNPs identified in *CBFs* on chromosome 5A reveal very high LD (r^2 = 0.92 to 1), indicating localisation at the same chromosomal region without or with very rare recombination events. This result confirms previous knowledge of tight linkage of *CBF* genes and emphasises the importance of the *FR-A2* locus in frost response. On the other hand, results reveal that not all members of the *CBF* family are involved in FT as also reported in several studies (Sutka et al., 1999, Campoli et al., 2009). Based on the occurrence of only two haplotypes of the *CBF* genes and the fact that they are very closely linked, genotypes can be divided into two groups. Group one with 193 genotypes which showed 15% better winter survival in comparison to group two (42 genotypes) (Table 10 and 11). Therefore, breeding efforts in combining two FT haplotypes are highly desirable towards creating elite cultivars exhibiting exceptional FT.

The CBF-A3_SNP2 [C/G] which is significantly associated with FT ($P = 6.28 \times 10^{-10}$) results in an Ala/Gly AA substitution. The affected AA is localised within the α-helix of the AP2 domain and shows a high conservation of Ala in the AA alignment with homologues. Merely AA haplotype two had a Gly which corresponds to a significant reduction in the winter survival rate of 15% in the SNP/indel association study. Haplotype one also shows a 15% better winter survival compared to haplotype two with Gly in the haplotype association study (Tables 10 and 11). Allen et al. (1998) described Ala contributing to the stabilisation of the protein structure by its hydrophobic side chain. It is possible that the appearance of Gly in haplotype two results in a structurally more instable protein or, alternatively, impairs the functionality as a transcription factor. Additionally, the dN/dS

ratio analysis shows that the CBF-A3_SNP2 site underlies negative selection (Figure 16), reflecting the association effects.

The SNP/indel association study of *CBF-A5* exon SNPs revealed no significant association. In contrast, compared to exon haplotype three (Ser on site CBF-A5_SNP2) the exon haplotypes are significantly associated with a reduced winter survival of 19.5% (exon haplotype one) and 18% (exon haplotype two). Due to the fact that exon haplotype three consists of only three genotypes no identification in the SNP/indel association analysis was possible. The results of the haplotype association analysis suggest a high influence of CBF-A5_SNP2 in FT. But in consequence of the very low observation numbers of haplotype three the results must be considered with caution (Table 11, Supplemental Figure 5).

The significantly associated SNPs and indels of *CBF-A13* may disorganise the protein structure very strongly. In detail, the one bp CBF-A13_indel1 of haplotype two modified the complete AA sequences from the seventh AA onwards and a stop codon is present after 74 AAs. Since only six AAs are identical to the reference protein sequence, a complete loss of function may be assumed. The 32 bp deletion (CBF-A13_indel2) of haplotype one also generates a frame shift from AA 45 on and a stop codon after 135 AAs. Hence, the AA haplotype two exhibits no similarity to the reference protein sequence and AA haplotype one shows a very low uSeqID of 15 but an appropriate *P*-value of 4.79×10^{-4} compared to the AP2 domain. Therefore, it is most likely that both proteins are non-functional. However, haplotype one shows about 15.01% higher winter survival in the SNP/indel association study and 15.22% in the haplotype association study compared to haplotype two. Consequently, it is possible that the proteins of AA haplotype one have an increased transcription factor efficiency compared to the proteins of AA haplotype two, which results in higher FT. On the other hand the association may be due to the very high LD within the *FR-A2* locus.

Regarding the two AA haplotypes of CBF-A15 an AA substitution in the AP2 domain from Ala to Val was identified which matched to the statistically significant associated CBF-A15_SNP3 [C/T] with a *P*-value of 6.28×10^{-10}. Ala is conserved at this position for the homologous AA sequences. Haplotype two which comprises Val in the AA sequence

shows 15% less winter survival in the SNP/indel association study and 15% in the haplotype association study. That indicates a functional loss, although Ala as well as Val possess a hydrophobic side chain and are very similar in molecule size. A functional loss due to this AA substitution is unlikely. On the other hand, no statistically significant positive or negative selection at this site was detected.

Uniquely the complete gene haplotype of *CBF-A18* is significantly associated to FT. Due to the MAF selection, no association study could be performed for all other haplotype components and polymorphic sites of *CBF-A18* (Table 11). The number of rare alleles limits further analysis of this effect. Therefore, an increase of observation numbers would reveal this effect as an artefact or give more details like AA substitutions of polymorphic site associations in important domains/motifs etc.

All other SNPs/indels of the seven *CBFs*, which resulted in AA substitutions, may be involved in FT but are not located in the highly conserved AP2 domain. In contrast, CBF-A3_SNP4, which is located upstream of the AP2 domain, shows conserved AA sites. This may play an important regulatory role in FT. The associated polymorphisms of *CBF-A5* and *CBF-A14* are located within the promoter and the *in silico* promoter analysis of these haplotypes shows no promoter region differences which would indicate modifications in gene transcription. However, only gene expression analyses can answer this question.

In summary, this study revealed significantly associated SNPs/indels and haplotypes in seven *CBF* genes. Out of the associated polymorphisms, two SNPs were identified, which resulted in an AA substitution in the highly conserved AP2 domain of the CBF-A3 and CBF-A15 protein. Both *CBF* genes were identified as important FT genes in *Triticum monococcum* and *Triticum aestivum* (Vagujfalvi et al., 2003, Vagujfalvi et al., 2005, Sutton et al., 2009, Knox et al., 2008, Soltesz et al., 2013). All this leads to the conclusion that CBF-A3_SNP2 is the most interesting *CBF* allele for FT improvement together with CBF-A15_SNP3. Further details remain to be revealed by future investigations such as complementation of promising alleles in spring or winter wheat varieties or protein functionality analysis of these alleles.

VERNALISATION GENES (VRN-A1 and VRN-B3)

Identification of the VRN-A1_SNP1 [C/T/Y] revealed that the genotypes with the base T and the genotypes with the ambiguous nucleotide Y correlate with increased FT by 1.4% respectively 0.1% in comparison to the genotypes with the base C. The exon haplotype association study shows stronger effects. In detail, in comparison to haplotype two (C), haplotype three (T) and haplotype four (Y) increase FT by 12.4% and 11.2%, respectively. As significant effects of these allele had previously also been described by other groups using different populations and experimental sites (Zhu et al., 2014, Diaz et al., 2012, Chen et al., 2009, Eagles et al., 2011), functional relevance of these variants is highly likely. Diaz et al. (2012) and Zhu et al. (2014), for example, described the C allele to be associated with a lower *VRN-A1* copy number and the Y allele with a higher copy number. In addition, Zhu et al. (2014) described an increase of the *VRN-A1* copy number associated with improved FT among the *FR-A2-T* allele of the *CBF12* and *CBF15* genes but not among the *FR-A2-S* which are reflected in our haplotypes one and two, respectively. However, the VRN-A1_SNP1 generates an AA substitution on a Leu conserved site in the K-domain of a MIKC-type transcription factor between α-helix three and α-helix four to a Phe. Puranik et al. (2014) showed that this Leu stabilizes the kink region between α-helix three and α-helix four by extensive intra-molecule hydrophobic interactions of multiple Leu residues. Both AA have hydrophobic side chains but Phe with its benzene ring strongly differs from Leu for its steric requirements. Therefore, it is possible that the Phe residue increases the angle between both α-helices due to its bulkiness and the attachment to the target sequence of the transcription factor is improved (Puranik et al., 2014). Additionally the VRN-A1_SNP1 underlies no negative or positive selection.

The *VRN-B3* gene is significantly associated to FT only in the SNP/indel association study. The polymorphism VRN-B3_SNP1 [C/G] shows a *P*-value of 3.76×10^{-2}. This SNP generates a His/Asp AA substitution. The genotypes with His reveal a higher winter survival of 3.1% compared to those containing Asp (Table 10). Additionally, this site exhibits Asp highly conserved regarding all nine homologous AA sequences and underlies significant negative selection. The high conserved AA Asp contains a negative charged side chain and the AA His a positive charged side chain. Also the AAs differ in molecule size, His being larger than Asp. Both facts suggest that if a His is involved, the protein

structure might be changed and as a consequence also the functionality. Also the fact of negative selection supported the hypothesis of protein functionality modification. If the VRN3 protein functionality is changed, the flowering pathway is affected and it could eventually lead to a reduction of the generation of flowers or no flowers at all. To conclude, the VRN3 protein lies upstream the VRN1 protein along the signalling cascade and enhances VRN1 (Distelfeld et al., 2009). VRN1 can inhibit CBFs and subsequently COR/LEA proteins. If the protein functionality of VRN-B3 is negatively affected the FT could be increased (Dhillon et al., 2010). All this data indicates that *VRN-B3* and respective homologues play a role in FT.

PHOTOPERIOD RESPONSE GENES (PPD-B1 and PPD-D1)

The polymorphisms of the *PPD-B1* gene merely show significance in the haplotype association study. The exon haplotype three displays 5.1% lower winter survival than haplotype four. The difference between both haplotypes is the PPD-B1_SNP2 [A/G] which generates an Arg/Gly substitution in the α-helix 3 of the Pseudo-Receiver domain. The Arg is highly conserved for grasses (Matsushika et al., 2000, Strayer et al., 2000). The AA Arg contains a negative charged side chain and Gly is ambivalent. Also the molecule size is different as Arg is much larger than Gly. All facts suggest a modification of protein functionality. Therefore, it is possible that the Gly from haplotype three has a negative effect on FT. Another AA substitution in the Pseudo-Receiver domain from Asn to Asp is caused by the PPD-B1_SNP3 [A/G]. Asp at this position is highly conserved. Only haplotype one and the homologues AA of *Triticum aestivum* contain Asn resulting in a 14.4% decreased winter survival in comparison to haplotype four. Asp contains a negative charged side chain and Asn contains a polar side chain. The molecule size of both is equivalent. The loss of polarity due to Asn suggests a modification of protein functionality and is followed by lower FT. The haplotype two, which is associated with PPD-B1_SNP5 [G/A], and an Asp/Asn AA substitution reveals 28.3% less winter survival in comparison to haplotype four. The most tolerant haplotype four originates from the Asia group. The position of this AA is 51 residues downstream of the pseudo-receiver domain and is moderately conserved. This AA substitution is comparable to PPD-B1_SNP3. Therefore, the loss of polarity suggests a modification of protein functionality. Associations of the haplotypes one (three observations) and two (five observations) are based on very few

observations and therefore the results should be interpreted with caution. In this case, PPD-B1_SNP2, associated with the most drastic AA substitution, plays the most important role in FT. The dN/dS analysis revealed no significant negative or positive selection for all three AA substitutions (PPD-B1_SNP2, PPD-B1_SNP3 and PPD-B1_SNP5), suggesting either an insufficient number of genotypes that were analysed or, alternatively, questioning the relevance of PPD (Supplemental Figure 13).

The associated indel of *PPD-D1* has an effect of 4% in the SNP/indel association study (Table 10). Haplotype one (with a deletion) exhibits 3.6% better winter survival compared to haplotype two in the haplotype association study (Table 11). As a result of this deletion, a stop codon on AA position 470 occurs and the CONSTANS motif is missing. In consequence, the PPD-D1 protein cannot interact with the CO protein and the flowering control pathway is interrupted (Turner et al., 2005). Furthermore, the interaction between the flowering time and FT pathway is disturbed (Supplemental Figure 14).

This study demonstrated polymorphisms significantly associated with FT and the importance of the AA substitutions of seven *CBF* gene family members and the *VRN-A1*, *VRN-B3*, *PPD-B1* and *PPD-D1* genes (Figure 18). Additionally, investigations on functionality and structure of encrypted proteins would complete our insight on FT from cellular level.

These results ought be used to design wheat cultivars with exceptional frost tolerance via gene engineering or classical breeding. To achieve this, six associated genes (*CBF-A3*, *CBF-A15*, *VRN-A1*, *VRN-B3*, *PPD-B1* and *PPD-D1*) have to be combined, employing the alleles which show the strongest positive effect in FT. In this respect a wheat cultivar with the haplotype one of both *CBF* genes, haplotype three of PPD-B1, haplotype one of PPD-D1, haplotype three of VRN-A1 and VRN-B3_SNP1 of VRN-B3 is suggested to create a genotype with the theoretically highest FT regarding the investigated genes. Additional candidate gene based association genetics studies in the field of FT should focus on the *COR* (cold-regulated) genes and proteins, since they are also part of the signalling cascade and have not yet been investigated.

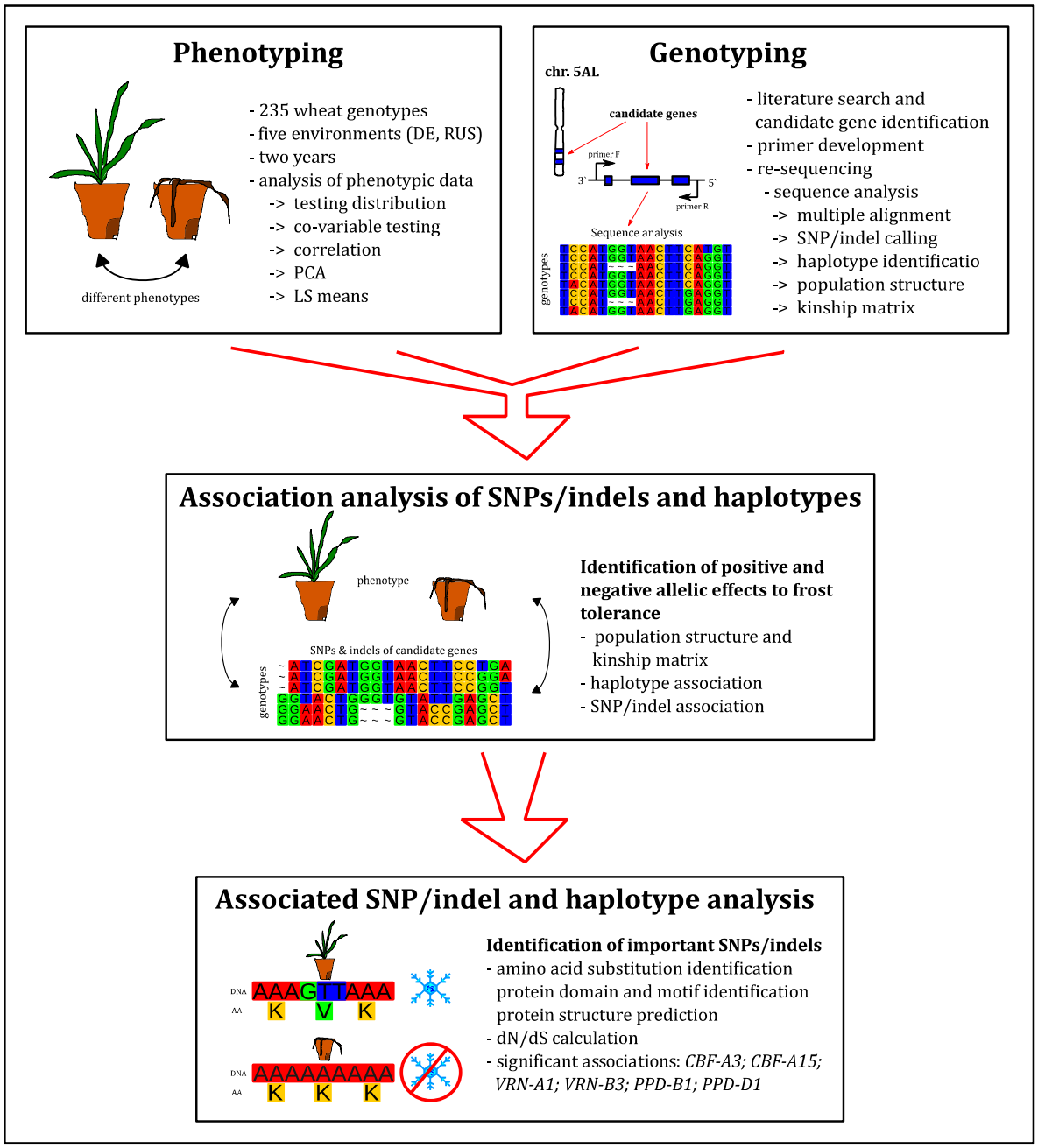


Figure 18: Workflow for identifyimg candidate gene based associations for frost tolerance in wheat (Babben et al., 2018).

5 Conclusion

Triticum aestivum is a genetically very complex species owing to its very large genome size and hexaploidy. Therefore, it is very difficult to design genome and locus specific primers. For MAS for improved FT, it is important to identify genes and alleles involved in freezing response. This thesis presents an efficient approach for FT gene primer development by using sequence data of wheat. By using the presented approach for gene and genome specific primer development, it is possible to sequence and analyse genes of interest in wheat (and other polyploid species) very efficiently by using gene sequences of related plant species as a starting point. Due to the recently published wheat sequence RefSeq v1.0 (reference) and using the approach presented a further increase of the success rate and effectivity is achievable. The designed primer pairs were used for re-sequencing of 18 FT genes and the detection of polymorphisms in a set of 235 wheat cultivars. The following candidate gene association study based on phenotypic data from German and Russian field trials resulted in the identification of 11 genes significantly associated with FT. AA substitutions in important protein domains of some of these genes/proteins suggest functional relevance.

In summary, the results of the candidate gene based association study prove it a very useful method to identify alleles that positively contribute to FT in hexaploid wheat. The identified FT associated SNPs/indels and haplotypes ought to be used for developing diagnostic markers for MAS in wheat breeding. Combination and pyramiding of alleles with positive effects on FT in elite breeding material promise the generation of superior germplasm with regard to FT.

6 References

ALLAGULOVA, C. R., GIMALOV, F. R., SHAKIROVA, F. M. & VAKHITOV, V. A. 2003. The plant dehydrins: Structure and putative functions. *Biochemistry-Moscow,* 68**,** 945-951.

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., WINFIELD, M. O. & BURRIDGE, A. J. 2016. Characterization of a wheat breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnology Journal,* 15.

ALLEN, M. D., YAMASAKI, K., OHME-TAKAGI, M., TATENO, M. & SUZUKI, M. 1998. A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *Embo Journal,* 17**,** 5484-5496.

APPELS, R., EVERSOLE, K., FEUILLET, C., KELLER, B., ROGERS, J., STEIN, N., POZNIAK, C. J., STEIN, N., CHOULET, F., DISTELFELD, A., EVERSOLE, K., POLAND, J., ROGERS, J., RONEN, G., SHARPE, A. G., POZNIAK, C., RONEN, G., STEIN, N., BARAD, O., BARUCH, K., CHOULET, F., KEEBLE-GAGNÈRE, G., MASCHER, M., SHARPE, A. G., BEN-ZVI, G., JOSSELIN, A.-A., STEIN, N., MASCHER, M., HIMMELBACH, A., CHOULET, F., KEEBLE-GAGNÈRE, G., MASCHER, M., ROGERS, J., BALFOURIER, F., GUTIERREZ-GONZALEZ, J., HAYDEN, M., JOSSELIN, A.-A., KOH, C., MUEHLBAUER, G., PASAM, R. K., PAUX, E., POZNIAK, C. J., RIGAULT, P., SHARPE, A. G., TIBBITS, J., TIWARI, V., CHOULET, F., KEEBLE-GAGNÈRE, G., MASCHER, M., JOSSELIN, A.-A., ROGERS, J., SPANNAGL, M., CHOULET, F., LANG, D., GUNDLACH, H., HABERER, G., KEEBLE-GAGNÈRE, G., MAYER, K. F. X., ORMANBEKOVA, D., PAUX, E., PRADE, V., ŠIMKOVÁ, H., WICKER, T., CHOULET, F., SPANNAGL, M., SWARBRECK, D., RIMBERT, H., FELDER, M., GUILHOT, N., GUNDLACH, H., HABERER, G., KAITHAKOTTIL, G., KEILWAGEN, J., LANG, D., LEROY, P., LUX, T., MAYER, K. F. X., TWARDZIOK, S., VENTURINI, L., APPELS, R., RIMBERT, H., CHOULET, F., JUHÁSZ, A., KEEBLE-GAGNÈRE, G., CHOULET, F., SPANNAGL, M., LANG, D., ABROUK, M., HABERER, G., KEEBLE-GAGNÈRE, G., MAYER, K. F. X., WICKER, T., CHOULET, F., WICKER, T., GUNDLACH, H., LANG, D., SPANNAGL, M., LANG, D., SPANNAGL, M., APPELS, R., et al. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science,* 361.

BABBEN, S., PEROVIC, D., KOCH, M. & ORDON, F. 2015. An Efficient Approach for the Development of Locus Specific Primers in Bread Wheat (*Triticum aestivum* L.) and Its Application to Re-Sequencing of Genes Involved in Frost Tolerance. *Plos One,* 10**,** e0142746.

BABBEN, S., SCHLIEPHAKE, E., JANITZA, P., BERNER, T., KEILWAGEN, J., KOCH, M., ARANA-CEBALLOS, F. A., TEMPLER, S. E., CHESNOKOV, Y., PSHENICHNIKOVA, T., SCHONDELMAIER, J., BÖRNER, A., PILLEN, K., ORDON, F. & PEROVIC, D. 2018. Association genetics studies on frost tolerance in wheat (*Triticum aestivum* L.) reveal new highly conserved amino acid substitutions in CBF-A3, CBF-A15, *VRN3* and *PPD1* genes. *BMC Genomics,* 19**,** 409.

BACHHAWAT, P. & STOCK, A. M. 2007. Crystal structures of the receiver domain of the response regulator PhoP from *Escherichia coli* in the absence and presence of the phosphoryl analog Beryllofluoride. *Journal of Bacteriology,* 189.

BADAWI, M., DANYLUK, J., BOUCHO, B., HOUDE, M. & SARHAN, F. 2007. The *CBF* gene family in hexaploid wheat and its relationship to the phylogenetic complexity of cereal *CBFs*. *Molecular Genetics and Genomics,* 277**,** 533-554.

BAGA, M., CHODAPARAMBIL, S. V., LIMIN, A. E., PECAR, M., FOWLER, D. B. & CHIBBAR, R. N. 2007. Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. *Functional & Integrative Genomics,* 7**,** 53-68.

BAGGE, M., XIA, X. & LÜBBERSTEDT, T. 2007. Functional markers in wheat. *Current Opinion in Plant Biology,* 10**,** 211-216.

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.). *Theoretical and Applied Genetics,* 115**,** 721-733.

BOCH, J., SCHOLZE, H., SCHORNACK, S., LANDGRAF, A., HAHN, S., KAY, S., LAHAYE, T., NICKSTADT, A. & BONAS, U. 2009. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science,* 326**,** 1509-1512.

BRADBURY, P. J., ZHANG, Z., KROON, D. E., CASSTEVENS, T. M., RAMDOSS, Y. & BUCKLER, E. S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics,* 23**,** 2633-2635.

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 bread wheat genome using whole-genome shotgun sequencing. *Nature,* 491**,** 705-710.

BRYAN, G. J., COLLINS, A. J., STEPHENSON, P., ORRY, A., SMITH, J. B. & GALE, M. D. 1997. Isolation and characterisation of microsatellites from hexaploid bread wheat. *Theoretical and Applied Genetics,* 94**,** 557-563.

CAMPOLI, C., MATUS-CADIZ, M. A., POZNIAK, C. J., CATTIVELLI, L. & FOWLER, D. B. 2009. Comparative expression of *Cbf* genes in the *Triticeae* under different acclimation induction temperatures. *Molecular Genetics and Genomics,* 282**,** 141-152.

CHEN, J., HU, G., ZHANG, J., CHU, C. & WU, Y. 2014. Mapping of STS Markers Developed from Drought Tolerance Candidate Genes and Preliminary Analysis of their Association with Yield-related Traits in Common Wheat (*Triticum aestivum*). *Cereal Research Communications,* 42**,** 199-208.

CHEN, Y., CARVER, B. F., WANG, S., CAO, S. & YAN, L. 2010. Genetic regulation of developmental phases in winter wheat. *Molecular Breeding,* 26**,** 573-582.

CHEN, Y. H., CARVER, B. F., WANG, S. W., ZHANG, F. Q. & YAN, L. L. 2009. Genetic loci associated with stem elongation and winter dormancy release in wheat. *Theoretical and Applied Genetics,* 118**,** 881-889.

CHINNUSAMY, V., OHTA, M., KANRAR, S., LEE, B.-H., HONG, X., AGARWAL, M. & ZHU, J.-K. 2003. *ICE1*: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes & Development,* 17**,** 1043-1054.

CHINNUSAMY, V., ZHU, J. & ZHU, J. K. 2007. Cold stress regulation of gene expression in plants. *Trends in Plant Science,* 12**,** 444-451.

CHOI, D. W., ZHU, B. & CLOSE, T. J. 1999. The barley (*Hordeum vulgare* L.) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 *Dhn* genes of cv Dicktoo. *Theoretical and Applied Genetics,* 98**,** 1234-1247.

CHOULET, F., WICKER, T., RUSTENHOLZ, C., PAUX, E., SALSE, J., LEROY, P., SCHLUB, S., LE PASLIER, M. C., MAGDELENAT, G., GONTHIER, C., COULOUX, A., BUDAK, H., BREEN, J., PUMPHREY, M., LIU, S. X., KONG, X. Y., JIA, J. Z., GUT, M., BRUNEL, D., ANDERSON, J. A., GILL, B. S., APPELS, R., KELLER, B. & FEUILLET, C. 2010. Megabase Level Sequencing Reveals Contrasted Organization and Evolution Patterns of the Wheat Gene and Transposable Element Spaces. *Plant Cell,* 22**,** 1686-1701.

CHRISTIAN, M., CERMAK, T., DOYLE, E. L., SCHMIDT, C., ZHANG, F., HUMMEL, A., BOGDANOVE, A. J. & VOYTAS, D. F. 2010. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics,* 186**,** 757-U476.

CLOSE, T. J. 1997. Dehydrins: A commonality in the response of plants to dehydration and low temperature. *Physiologia Plantarum,* 100**,** 291-296.

DANYLUK, J., KANE, N. A., BRETON, G., LIMIN, A. E., FOWLER, D. B. & SARHAN, F. 2003. *TaVRT-1*, a putative transcription factor associated with vegetative to reproductive transition in cereals. *Plant Physiology,* 132**,** 1849-1860.

DELPORT, W., POON, A. F. Y., FROST, S. D. W. & KOSAKOVSKY POND, S. L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics,* 26**,** 2455-2457.

DHILLON, T., PEARCE, S. P., STOCKINGER, E. J., DISTELFELD, A., LI, C. X., KNOX, A. K., VASHEGYI, I., VAGUJFALVI, A., GALIBA, G. & DUBCOVSKY, J. 2010. Regulation of Freezing Tolerance and Flowering in Temperate Cereals: The *VRN-1* Connection. *Plant Physiology,* 153**,** 1846-1858.

DIAZ, A., ZIKHALI, M., TURNER, A. S., ISAAC, P. & LAURIE, D. A. 2012. Copy Number Variation Affecting the *Photoperiod-B1* and *Vernalization-A1* Genes Is Associated with Altered Flowering Time in Wheat (*Triticum aestivum*). *Plos One,* 7.

DIETZ, K.-J., VOGEL, M. O. & VIEHHAUSER, A. 2010. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma,* 245**,** 3-14.

DISTELFELD, A., LI, C. & DUBCOVSKY, J. 2009. Regulation of flowering in temperate cereals. *Current Opinion in Plant Biology,* 12**,** 178-184.

DOWGERT, M. F. & STEPONKUS, P. L. 1984. Behavior of the Plasma Membrane of Isolated Protoplasts during a Freeze-Thaw Cycle. *Plant Physiology,* 75**,** 1139-1151.

DUBCOVSKY, J., LIJAVETZKY, D., APPENDINO, L. & TRANQUILLI, G. 1998. Comparative RFLP mapping of *Triticum monococcum* genes controlling vernalization requirement. *Theoretical and Applied Genetics,* 97**,** 968-975.

DUNN, M. A., HUGHES, M. A., PEARCE, R. S. & JACK, P. L. 1990. MOLECULAR CHARACTERIZATION OF A BARLEY GENE INDUCED BY COLD TREATMENT. *Journal of Experimental Botany,* 41**,** 1405-1413.

DVORAK, J. & AKHUNOV, E. D. 2005. Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *aegilops-triticum* alliance. *Genetics,* 171**,** 323-332.

EAGLES, H. A., CANE, K. & TREVASKIS, B. 2011. Veery wheats carry an allele of *Vrn-A1* that has implications for freezing tolerance in winter wheats. *Plant Breeding,* 130**,** 413-418.

ENDO, T. R. & GILL, B. S. 1996. The deletion stocks of common wheat. *Journal of Heredity,* 87**,** 295-307.

ENSMINGER, I., BUSCH, F. & HUNER, N. P. A. 2006. Photostasis and cold acclimation: sensing low temperature through photosynthesis. *Physiologia Plantarum,* 126**,** 28-44.

EVANNO, G., REGNAUT, S. & GOUDET, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology,* 14**,** 2611-2620.

FAO 2016. Food and Agriculture Organization of the United Nations Statistics Divisions. 2016 ed. http://www.fao.org/faostat/en/#home: Food and Agriculture Organization of the United Nations

FELDMAN, M. & KISLEV, M. E. 2007. Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. *Israel Journal of Plant Sciences,* 55**,** 207-221.

FRANCIA, E., RIZZA, F., CATTIVELLI, L., STANCA, A. M., GALIBA, G., TOTH, B., HAYES, P. M., SKINNER, J. S. & PECCHIONI, N. 2004. Two loci on chromosome 5H determine low-temperature tolerance in a 'Nure' (winter) x 'Tremois' (spring) barley map. *Theoretical and Applied Genetics,* 108**,** 670-680.

FU, D. L., SZUCS, P., YAN, L. L., HELGUERA, M., SKINNER, J. S., VON ZITZEWITZ, J., HAYES, P. M. & DUBCOVSKY, J. 2005. Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Molecular Genetics and Genomics,* 273**,** 54-65.

GALIBA, G., QUARRIE, S. A., SUTKA, J., MORGOUNOV, A. & SNAPE, J. W. 1995. RFLP MAPPING OF THE *VERNALIZATION* (*VRN1*) AND F*ROST-RESISTANCE* (*FR1*) GENES ON CHROMOSOME 5A OF WHEAT. *Theoretical and Applied Genetics,* 90**,** 1174-1179.

GANA, J. A., SUTTON, F. & KENEFICK, D. G. 1997. cDNA structure and expression patterns of a low-temperature-specific wheat gene *tacr7*. *Plant Molecular Biology,* 34**,** 643-650.

GANESHAN, S., SHARMA, P., YOUNG, L., KUMAR, A., FOWLER, D. B. & CHIBBAR, R. N. 2011. Contrasting cDNA-AFLP profiles between crown and leaf tissues of cold-acclimated wheat plants indicate differing regulatory circuitries for low temperature tolerance. *Plant Molecular Biology,* 75**,** 379-398.

GILMOUR, S. J., ZARKA, D. G., STOCKINGER, E. J., SALAZAR, M. P., HOUGHTON, J. M. & THOMASHOW, M. F. 1998. Low temperature regulation of the *Arabidopsis CBF* family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. *Plant Journal,* 16**,** 433-442.

GOODMAN, M. M. & STUBER, C. W. 1983. RACES OF MAIZE .6. ISOZYME VARIATION AMONG RACES OF MAIZE IN BOLIVIA. *Maydica,* 28**,** 169-187.

GRIFFITH, M. & YAISH, M. W. F. 2004. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends in Plant Science,* 9**,** 399-405.

GULICK, P. J., DROUIN, S., YU, Z. H., DANYLUK, J., POISSON, G., MONROY, A. F. & SARHAN, F. 2005. Transcriptome comparison of winter and spring wheat responding to low temperature. *Genome,* 48**,** 913-923.

GUO, Z. A., SONG, Y. X., ZHOU, R. H., REN, Z. L. & JIA, J. Z. 2010. Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. *New Phytologist,* 185**,** 841-851.

GUSTA, L. V., WISNIEWSKI, M., NESBITT, N. T. & GUSTA, M. L. 2004. The Effect of Water, Sugars, and Proteins on the Pattern of Ice Nucleation and Propagation in Acclimated and Nonacclimated Canola Leaves. *Plant Physiology,* 135**,** 1642-1653.

HAN, Z. G., WANG, C. B., SONG, X. L., GUO, W. Z., GOU, J. Y., LI, C. H., CHEN, X. Y. & ZHANG, T. Z. 2006. Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theoretical and Applied Genetics,* 112**,** 430-439.

HIGO, K., UGAWA, Y., IWAMOTO, M. & KORENAGA, T. 1999. Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research,* 27**,** 297-300.

HUANG, S., SIRIKHACHORNKIT, A., SU, X. J., FARIS, J., GILL, B., HASELKORN, R. & GORNICKI, P. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilop*s complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 8133-8138.

ISHIKAWA, G., YONEMARU, J., SAITO, M. & NAKAMURA, T. 2007. PCR-based landmark unique gene (PLUG) markers effectively assign homoeologous wheat genes to A, B and D genomes. *Bmc Genomics,* 8.

IWGSC, T. I. W. G. S. C. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science,* 345.

JAGLO, K. R., KLEFF, S., AMUNDSEN, K. L., ZHANG, X., HAAKE, V., ZHANG, J. Z., DEITS, T. & THOMASHOW, M. F. 2001. Components of the *Arabidopsis C-repeat/dehydration-responsive element binding factor* cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiology,* 127**,** 910-917.

JIA, J. Z., ZHAO, S. C., KONG, X. Y., LI, Y. R., ZHAO, G. Y., HE, W. M., APPELS, R., PFEIFER, M., TAO, Y., ZHANG, X. Y., JING, R. L., ZHANG, C., MA, Y. Z., GAO, L. F., GAO, C., SPANNAGL, M., MAYER, K. F. X., LI, D., PAN, S. K., ZHENG, F. Y., HU, Q., XIA, X. C., LI, J. W., LIANG, Q. S., CHEN, J., WICKER, T., GOU, C. Y., KUANG, H. H., HE, G. Y., LUO, Y. D., KELLER, B., XIA, Q. J., LU, P., WANG, J. Y., ZOU, H. F., ZHANG, R. Z., XU, J. Y., GAO, J. L., MIDDLETON, C., QUAN, Z. W., LIU, G. M., WANG, J., YANG, H. M., LIU, X., HE, Z. H., 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.

JORDAN, K. W., WANG, S. & LUN, Y. 2015. A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biology,* 16.

KÄLLBERG, M., WANG, H., WANG, S., PENG, J., WANG, Z., LU, H. & XU, J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nature Protocols,* 7**,** 1511.

KAPLAN, B., DAVYDOV, O., KNIGHT, H., GALON, Y., KNIGHT, M. R., FLUHR, R. & FROMM, H. 2006. Rapid Transcriptome Changes Induced by Cytosolic Ca^{2+} Transients Reveal ABRE-Related Sequences as Ca^{2+}-Responsive cis Elements in *Arabidopsis*. *Plant Cell,* 18**,** 2733-2748.

KATO, K. & YAMAGATA, H. 1988. Method for evaluation of chilling requirement and narrow-sense earliness of wheat cultivars. *Japanese Journal of Breed,* 38.

KATOH, K., KUMA, K.-I., TOH, H. & MIYATA, T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research,* 33**,** 511-518.

KATOH, K., MISAWA, K., KUMA, K. & MIYATA, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research,* 30**,** 3059-3066.

KAUFMANN, K., MELZER, R. & THEISSEN, G. 2005. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene,* 347**,** 183-198.

KAUR, N., STREET, K., MACKAY, M., YAHIAOUI, N. & KELLER, B. 2008. Molecular approaches for characterization and use of natural disease resistance in wheat. *European Journal of Plant Pathology,* 121**,** 387-397.

KEDDIE, J. S., CARROLL, B. J., THOMAS, C. M., REYES, M. E. C., KLIMYUK, V., HOLTAN, H., GRUISSEM, W. & JONES, J. D. G. 1998. Transposon tagging of the *Defective embryo and meristems* gene of tomato. *Plant Cell,* 10**,** 877-887.

KEILWAGEN, J., KILIAN, B., OZKAN, H., BABBEN, S., PEROVIC, D., MAYER, K. F. X., WALTHER, A., POSKAR, C. H., ORDON, F., EVERSOLE, K., BOERNER, A., GANAL, M., KNUEPFFER, H., GRANER, A. & FRIEDEL, S. 2014. Separating the wheat from the chaff - a strategy to utilize plant genetic resources from ex situ genebanks. *Scientific Reports,* 4.

KILIAN, B. & GRANER, A. 2012. NGS technologies for analyzing germplasm diversity in genebanks*. *Briefings in Functional Genomics,* 11**,** 38-50.

KNOX, A. K., DHILLON, T., CHENG, H. M., TONDELLI, A., PECCHIONI, N. & STOCKINGER, E. J. 2010. *CBF* gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. *Theoretical and Applied Genetics,* 121**,** 21-35.

KNOX, A. K., LI, C. X., VAGUJFALVI, A., GALILBA, G., STOCKINGER, E. J. & DUBCOVSKY, J. 2008. Identification of candidate *CBF* genes for the frost tolerance locus Fr-A(m)2 in *Triticum monococcum*. *Plant Molecular Biology,* 67**,** 257-270.

KOČOVÁ, M., HOLÁ, D., WILHELMOVÁ, N. & ROTHOVÁ, O. 2009. The influence of low-temperature on the photochemical activity of chloroplasts and activity of antioxidant enzymes in maize leaves. *Biologia Plantarum,* 53**,** 475-483.

KOCSY, G., ATHMER, B., PEROVIC, D., HIMMELBACH, A., SZUCS, A., VASHEGYI, I., SCHWEIZER, P., GALIBA, G. & STEIN, N. 2010. Regulation of gene expression by chromosome 5A during cold hardening in wheat. *Molecular Genetics and Genomics,* 283**,** 351-363.

KOEMEL, J. E., GUENZI, A. C., ANDERSON, J. A. & SMITH, E. L. 2004. Cold hardiness of wheat near-isogenic lines differing in *vernalization* alleles. *Theoretical and Applied Genetics,* 109**,** 839-846.

KORESSAAR, T. & REMM, M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics,* 23**,** 1289-1291.

KOSOVA, K., PRASIL, I. T. & VITAMVAS, P. 2008. The relationship between *vernalization*- and *photoperiodically-regulated* genes and the development of frost tolerance in wheat and barley. *Biologia Plantarum,* 52**,** 601-615.

KUMAR, S., STECHER, G., PETERSON, D. & TAMURA, K. 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics,* 28**,** 2685-2686.

KUMAR, S. V. & WIGGE, P. A. 2010. H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in *Arabidopsis*. *Cell,* 140**,** 136-147.

LAUDENCIA-CHINGCUANCO, D., GANESHAN, S., YOU, F., FOWLER, B., CHIBBAR, R. & ANDERSON, O. 2011. Genome-wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (*Triticum aestivum* L.). *Bmc Genomics,* 12.

LAW, C. N., SUTKA, J. & WORLAND, A. J. 1978. GENETIC STUDY OF DAY-LENGTH RESPONSE IN WHEAT. *Heredity,* 41**,** 185-191.

LI, Y., BOECK, A., HASENEYER, G., KORZUN, V., WILDE, P., SCHOEN, C.-C., ANKERST, D. P. & BAUER, E. 2011a. Association analysis of frost tolerance in rye using candidate genes and phenotypic data from controlled, semi-controlled, and field phenotyping platforms. *Bmc Plant Biology,* 11.

LI, Y. L., HASENEYER, G., SCHON, C. C., ANKERST, D., KORZUN, V., WILDE, P. & BAUER, E. 2011b. High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *Bmc Plant Biology,* 11.

LIBRADO, P. & ROZAS, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25**,** 1451-1452.

LING, H. Q., ZHAO, S. & LIU, D. 2013. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature,* 496.

LIU, J., HUANG, S. M., SUN, M. Y., LIU, S. Y., LIU, Y. M., WANG, W. X., ZHANG, X. R., WANG, H. Z. & HUA, W. 2012. An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods,* 8.

LIU, Q., KASUGA, M., SAKUMA, Y., ABE, H., MIURA, S., YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell,* 10**,** 1391-1406.

MARÈ, C., MAZZUCOTELLI, E., CROSATTI, C., FRANCIA, E., STANCA, A. M. & CATTIVELLI, L. 2004. Hv-WRKY38: a new transcription factor involved in cold- and drought-response in barley. *Plant Molecular Biology,* 55**,** 399-416.

MATSUSHIKA, A., MAKINO, S., KOJIMA, M. & MIZUNO, T. 2000. Circadian waves of expression of the APRR1/TOC1 family of pseudo-response regulators in *Arabidopsis thaliana*: Insight into the plant circadian clock. *Plant and Cell Physiology,* 41**,** 1002-1012.

MCCORMACK, E. & BRAAM, J. 2003. Calmodulins and related potential calcium sensors of *Arabidopsis*. *New Phytologist,* 159**,** 585-598.

MCINTOSH, R. A., YMAZAKI, Y., DUBCOVSKY, J., ROGERS, J., MORRIS, C., APPELS, R. & XIA, X. C. 2013. Catalogue of Gene Sympols for Wheat. *In:* MCINTOSH, R. A. (ed.) *Morphological and Physiological Traits.*

MCKERSIE, B. D. & LESHEM, Y. Y. 1994. Stress and stress coping in cultivated plants, Dordrecht, Kluwer Academic Publishers.

MICHALEK, W., WESCHKE, W., PLEISSNER, K. P. & GRANER, A. 2002. EST analysis in barley defines a unigene set comprising 4,000 genes. *Theoretical and Applied Genetics,* 104**,** 97-103.

MILLER, A. K., GALIBA, G. & DUBCOVSKY, J. 2006. A cluster of 11 *CBF* transcription factors is located at the frost tolerance locus *Fr-A^{m}2* in *Triticum monococcum*. *Molecular Genetics and Genomics,* 275**,** 193-203.

MOHSENI, S., CHE, H., DJILLALI, Z., DUMONT, E., NANKEU, J. & DANYLUK, J. 2012. Wheat *CBF* gene family: identification of polymorphisms in the *CBF* coding sequence. *Genome,* 55**,** 865-881.

MONROY, A. F., DRYANOVA, A., MALETTE, B., OREN, D. H., FARAJALLA, M. R., LIU, W., DANYLUK, J., UBAYASENA, L. W. C., KANE, K., SCOLES, G. J., SARHAN, F. & GULICK, P. J. 2007. Regulatory gene candidates and gene expression analysis of cold acclimation in winter and spring wheat. *Plant Molecular Biology,* 64**,** 409-423.

MOSCOU, M. J. & BOGDANOVE, A. J. 2009. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science,* 326**,** 1501-1501.

MOTOMURA, Y., KOBAYASHI, F., IEHISA, J. C. M. & TAKUMI, S. 2013. A major quantitative trait locus for cold-responsive gene expression is linked to frost-resistance gene *Fr-A2* in common wheat. *Breeding Science,* 63**,** 58-67.

NEUMANN, K., KOBILJSKI, B., DENČIĆ, S., VARSHNEY, R. K. & BÖRNER, A. 2011. Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Molecular Breeding,* 27**,** 37-58.

NICOT, N., CHIQUET, V., GANDON, B., AMILHAT, L., LEGEAI, F., LEROY, P., BERNARD, M. & SOURDILLE, P. 2004. Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *Theoretical and Applied Genetics,* 109**,** 800-805.

OHNO, R., TAKUMI, S. & NAKAMURA, C. 2001. Expression of a cold-responsive *Lt-Cor* gene and development of freezing tolerance during cold acclimation in wheat (*Triticum aestivum* L.). *Journal of Experimental Botany,* 52**,** 2367-2374.

ORVAR, B. L., SANGWAN, V., OMANN, F. & DHINDSA, R. S. 2000. Early steps in cold sensing by plant cells. The role of actin cytoskeleton and membrane fluidity. *Plant Journal,* 23.

PEARSON, K. 1914. On certain errors with regard to multiple correlation occasionally made by those who have not adequately studied this subject. *Biometrika,* 10.

PENG, J. & XU, J. 2011. RaptorX: Exploiting structure information for protein alignment by statistical inference. *Proteins-Structure Function and Bioinformatics,* 79**,** 161-171.

PENG, J. H. H., SUN, D. F. & NEVO, E. 2011. Domestication evolution, genetics and genomics in wheat. *Molecular Breeding,* 28**,** 281-301.

PENG, Y.-L., WANG, Y.-S., CHENG, H., SUN, C.-C., WU, P., WANG, L.-Y. & FEI, J. 2013. Characterization and expression analysis of three CBF/DREB1 transcriptional factor genes from mangrove *Avicennia marina*. *Aquatic Toxicology,* 140**,** 68-76.

PEROVIC, J., SILVAR, C., KOENIG, J., STEIN, N., PEROVIC, D. & ORDON, F. 2013. A versatile fluorescence-based multiplexing assay for CAPS genotyping on capillary electrophoresis systems. *Molecular Breeding,* 32**,** 61-69.

POCZAI, P., CERNAK, I., GORJI, A. M., NAGY, S., TALLER, J. & POLGAR, Z. 2010. DEVELOPMENT OF INTRON TARGETING (IT) MARKERS FOR POTATO AND CROSS-SPECIES AMPLIFICATION IN *SOLANUM NIGRUM* (*SOLANACEAE*). *American Journal of Botany,* 97**,** E142-E145.

POND, S. L. K., FROST, S. D. W. & MUSE, S. V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics,* 21**,** 676-679.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-959.

PRUITT, K., TATUSOVA, T., BROWN, G. & MAGLOTT, D. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research***,** D130 - D135.

PURANIK, S., ACAJJAOUI, S., CONN, S., COSTA, L., CONN, V., VIAL, A., MARCELLIN, R., MELZER, R., BROWN, E., HART, D., THEISSEN, G., SILVA, C. S., PARCY, F., DUMAS, R., NANAO, M. & ZUBIETA, C. 2014. Structural Basis for the Oligomerization of the MADS Domain Transcription Factor SEPALLATA3 in *Arabidopsis*. *Plant Cell,* 26**,** 3603-3615.

RAATS, D., FRENKEL, Z., KRUGMAN, T., DODEK, I., SELA, H., SIMKOVA, H., MAGNI, F., CATTONARO, F., VAUTRIN, S., BERGES, H., WICKER, T., KELLER, B., LEROY, P., PHILIPPE, R., PAUX, E., DOLEZEL, J., FEUILLET, C., KOROL, A. & FAHIMA, T. 2013. The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. *Genome Biology,* 14.

RAMASAMY, R. K., RAMASAMY, S., BINDROO, B. B. & NAIK, V. G. 2014. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus,* 3**,** 431.

REDDY, L., ALLAN, R. E. & CAMPBELL, K. A. G. 2006. Evaluation of cold hardiness in two sets of near-isogenic lines of wheat (*Triticum aestivum*) with polymorphic vernalization alleles. *Plant Breeding,* 125**,** 448-456.

REIF, J. C., MELCHINGER, A. E. & FRISCH, M. 2005. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Science,* 45**,** 1-7.

RODER, M. S., PLASCHKE, J., KONIG, S. U., BORNER, A., SORRELLS, M. E., TANKSLEY, S. D. & GANAL, M. W. 1995. ABUNDANCE, VARIABILITY AND CHROMOSOMAL LOCATION OF MICROSATELLITES IN WHEAT. *Molecular & General Genetics,* 246**,** 327-333.

ROZAS, J., SANCHEZ-DELBARRIO, J. C., MESSEGUER, X. & ROZAS, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics,* 19**,** 2496-2497.

RUELLAND, E., VAULTIER, M.-N., ZACHOWSKI, A. & HURRY, V. 2009. Chapter 2 Cold Signalling and Cold Acclimation in Plants. *Advances in Botanical Research.* Academic Press.

SANDVE, S. R., RUDI, H., ASP, T. & ROGNLI, O. A. 2008. Tracking the evolution of a cold stress associated gene family in cold tolerant grasses. *BMC Evolutionary Biology,* 8**,** 245.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA SEQUENCING WITH CHAIN-TERMINATING INHIBITORS. *Proceedings of the National Academy of Sciences of the United States of America,* 74**,** 5463-5467.

SANGWAN, V., FOULDS, I., SINGH, J. & DHINDSA, R. S. 2001. Cold-activation of *Brassica napus BN115* promoter is mediated by structural changes in membranes and cytoskeleton, and requires Ca^{2+} influx. *Plant Journal,* 27.

SANTELLI, E. & RICHMOND, T. J. 2000. Crystal structure of MEF2A core bound to DNA at 1.5 angstrom resolution. *Journal of Molecular Biology,* 297**,** 437-449.

SANTRA, D. K., SANTRA, M., ALLAN, R. E., CAMPBELL, K. G. & KIDWELL, K. K. 2009. Genetic and Molecular Characterization of Vernalization Genes *Vrn-A1*, *Vrn-B1*, and *Vrn-D1* in Spring Wheat Germplasm from the Pacific Northwest Region of the USA. *Plant Breeding,* 128**,** 576-584.

SCARTH, R. & LAW, C. N. 1983. THE LOCATION OF THE PHOTOPERIOD GENE, *PPD2* AND AN ADDITIONAL GENETIC-FACTOR FOR EAR-EMERGENCE TIME ON CHROMOSOME-2B OF WHEAT. *Heredity,* 51**,** 607-619.

SEARS, E. R. 1966. Nullisomictetrasomic combinations in hexaploid wheat. *In:* RILEY, R. & LEWIS, K. R. (eds.) *Chromosome manipulations and plant genetics.*

SEKI, M., CHONO, M., MATSUNAKA, H., FUJITA, M., ODA, S., KUBO, K., KIRIBUCHI-OTOBE, C., KOJIMA, H., NISHIDA, H. & KATO, K. 2011. Distribution of photoperiod-insensitive alleles *Ppd-B1a* and *Ppd-D1a* and their effect on heading time in Japanese wheat cultivars. *Breeding Science,* 61**,** 405-412.

SHCHERBAN, A. B., EFREMOVA, T. T. & SALINA, E. A. 2012. Identification of a new *Vrn-B1* allele using two near-isogenic wheat lines with difference in heading time. *Molecular Breeding,* 29**,** 675-685.

SILVAR, C., PEROVIC, D., CASAS, A. M., IGARTUA, E. & ORDON, F. 2011. Development of a cost-effective pyrosequencing approach for SNP genotyping in barley. *Plant Breeding,* 130**,** 394-397.

SKINNER, J., SZUCS, P., VON ZITZEWITZ, J., MARQUEZ-CEDILLO, L., FILICHKIN, T., STOCKINGER, E. J., THOMASHOW, M. F., CHEN, T. H. H. & HAYES, P. M. 2006. Mapping of barley homologs to genes that regulate low temperature tolerance in *Arabidopsis*. *Theoretical and Applied Genetics,* 112**,** 832-842.

SKINNER, J. S., VON ZITZEWITZ, J., SZUCS, P., MARQUEZ-CEDILLO, L., FILICHKIN, T., AMUNDSEN, K., STOCKINGER, E. J., THOMASHOW, M. F., CHEN, T. H. H. & HAYES, P. M. 2005. Structural, functional, and phylogenetic characterization of a large *CBF* gene family in barley. *Plant Molecular Biology,* 59**,** 533-551.

SMITH, D. B. & FLAVELL, R. B. 1975. CHARACTERIZATION OF WHEAT GENOME BY RENATURATION KINETICS. *Chromosoma,* 50**,** 223-242.

SOLOVYEV, V., SHAHMURADOV, I. & SALAMOV, A. 2010. Identification of Promoter Regions and Regulatory Sites. *In:* LADUNGA, I. (ed.) *Computational Biology of Transcription Factor Binding.* Humana Press.

SOLTESZ, A., SMEDLEY, M., VASHEGYI, I., GALIBA, G., HARWOOD, W. & VAGUJFALVI, A. 2013. Transgenic barley lines prove the involvement of *TaCBF14* and *TaCBF15* in the cold acclimation process and in frost tolerance. *Journal of Experimental Botany,* 64**,** 1849-1862.

SONG, Q. J., SHI, J. R., SINGH, S., FICKUS, E. W., COSTA, J. M., LEWIS, J., GILL, B. S., WARD, R. & CREGAN, P. B. 2005. Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical and Applied Genetics,* 110**,** 550-560.

STEIN, N., HERREN, G. & KELLER, B. 2001. A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. *Plant Breeding,* 120**,** 354-356.

STOCKINGER, E. J., GILMOUR, S. J. & THOMASHOW, M. F. 1997. *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a *cis*-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 1035-1040.

STOCKINGER, E. J., SKINNER, J. S., GARDNER, K. G., FRANCIA, E. & PECCHIONI, N. 2007. Expression levels of barley *Cbf* genes at the *Frost resistance-H2* locus are dependent upon alleles at *Fr-H1* and *Fr-H2*. *Plant Journal,* 51**,** 308-321.

STRAYER, C., OYAMA, T., SCHULTZ, T. F., RAMAN, R., SOMERS, D. E., MAS, P., PANDA, S., KREPS, J. A. & KAY, S. A. 2000. Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science,* 289**,** 768-771.

SUTKA, J., GALIBA, G., VAGUJFALVI, A., GILL, B. S. & SNAPE, J. W. 1999. Physical mapping of the *Vrn-A1* and *Fr1* genes on chromosome 5A of wheat using deletion lines. *Theoretical and Applied Genetics,* 99**,** 199-202.

SUTTON, F., CHEN, D. G., GE, X. J. & KENEFICK, D. 2009. *Cbf* genes of the *Fr-A2* allele are differentially regulated between long-term cold acclimated crown tissue of freeze-resistant and - susceptible, winter wheat mutant lines. *Bmc Plant Biology,* 9.

SUYAMA, M., TORRENTS, D. & BORK, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research,* 34**,** W609-W612.

TALANOVA, V. V., TITOV, A. F., TOPCHIEVA, L. V., MALYSHEVA, I. E., VENZHIK, Y. V. & FROLOVA, S. A. 2009. Expression of WRKY transcription factor and stress protein genes in wheat plants during cold hardening and ABA treatment. *Russian Journal of Plant Physiology,* 56**,** 702-708.

TAN, G.-M., XU, L., BU, D.-B., FENG, S.-Z. & SUN, N.-H. 2006. Improvement of performance of MegaBlast algorithm for DNA sequence alignment. *Journal of Computer Science and Technology,* 21**,** 973-978.

TANKSLEY, S. D. & MCCOUCH, S. R. 1997. Seed banks and molecular maps. Unlocking genetic potential from the wild. *Science,* 277.

TAOKA, K., OHKI, I. & TSUJI, H. 2011. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature,* 476.

THEISSEN, G., KIM, J. T. & SAEDLER, H. 1996. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *Journal of Molecular Evolution,* 43**,** 484-516.

TREMBLAY, K., OUELLET, F., FOURNIER, J., DANYLUK, J. & SARHAN, F. 2005. Molecular Characterization and Origin of Novel Bipartite Cold-regulated Ice Recrystallization Inhibition Proteins from Cereals. *Plant and Cell Physiology,* 46**,** 884-891.

TREVASKIS, B., BAGNALL, D. J., ELLIS, M. H., PEACOCK, W. J. & DENNIS, E. S. 2003. MADS box genes control vernalization-induced flowering in cereals. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 13099-13104.

TURNER, A., BEALES, J., FAURE, S., DUNFORD, R. P. & LAURIE, D. A. 2005. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science,* 310**,** 1031-1034.

UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Research,* 40.

UPADHYAY, S. K., KUMAR, J., ALOK, A. & TULI, R. 2013. RNA-Guided Genome Editing for Target Gene Mutations in Wheat. *G3-Genes Genomes Genetics,* 3**,** 2233-2238.

VAGUJFALVI, A., APRILE, A., MILLER, A., DUBCOVSKY, J., DELUGU, G., GALIBA, G. & CATTIVELLI, L. 2005. The expression of several *Cbf* genes at the *Fr-A2* locus is linked to frost resistance in wheat. *Molecular Genetics and Genomics,* 274**,** 506-514.

VAGUJFALVI, A., GALIBA, G., CATTIVELLI, L. & DUBCOVSKY, J. 2003. The cold-regulated transcriptional activator *Cbf3* is linked to the frost-tolerance locus *Fr-A2* on wheat chromosome 5A. *Molecular Genetics and Genomics,* 269**,** 60-67.

WANG, L. H., LI, G. Y., PENA, R. J., XIA, X. C. & HE, Z. H. 2010. Development of STS markers and establishment of multiplex PCR for *Glu-A3* alleles in common wheat (T*riticum aestivum* L.). *Journal of Cereal Science,* 51**,** 305-312.

WANG, S., WONG, D., FORREST, K., ALLEN, A., CHAO, S., 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., INTERNATIONAL WHEAT GENOME SEQUENCING, C., 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. 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. Z., XU, H. B., ZHU, H. L., TAO, Y., ZHANG, G. X., ZHANG, L. X., ZHANG, C. Q., ZHANG, Z. Z. & MA, Z. Q. 2014b. Classification and expression diversification of wheat *dehydrin* genes. *Plant Science,* 214**,** 113-120.

WATERHOUSE, A. M., PROCTER, J. B., MARTIN, D. M. A., CLAMP, M. & BARTON, G. J. 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* 25**,** 1189-1191.

WELSH, J. R., KEIM, D. L., PIRASTEH, B. & RICHARDS, R. D. 1973. GENETIC CONTROL OF PHOTOPERIOD RESPONSE IN WHEAT. *In:* SEARS, E. R. & SEARS, L. M. S. (eds.) *PROCEEDINGS OF THE FOURTH INTERNATIONAL WHEAT GENETICS SYMPOSIUM.* UNIVERSITY OF MISSOURI, COLORADO, USA.

WILKINSON, P., WINFIELD, M., BARKER, G., ALLEN, A., BURRIDGE, A., COGHILL, J. & EDWARDS, K. 2012. CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC Bioinformatics,* 13**,** 219.

WINFIELD, M. O., ALLEN, A. M., BURRIDGE, A. J. & BARKER, G. L. A. 2016. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnology Journal,* 14.

WINFIELD, M. O., LU, C. G., WILSON, I. D., COGHILL, J. A. & EDWARDS, K. J. 2009. Cold- and light-induced changes in the transcriptome of wheat leading to phase transition from vegetative to reproductive growth. *Bmc Plant Biology,* 9.

WINFIELD, M. O., LU, C. G., WILSON, I. D., COGHILL, J. A. & EDWARDS, K. J. 2010. Plant responses to cold: transcriptome analysis of wheat. *Plant Biotechnology Journal,* 8**,** 749-771.

WRIGHT, S. 1978. Evolution and the genetics of populations : a treatise in four volumes. Vol. 4, Chicago; London, University of Chicago Press.

WU, J. H., HONG, P. Y. & LIU, W. T. 2009. Quantitative effects of position and type of single mismatch on single base primer extension. *Journal of Microbiological Methods,* 77**,** 267-275.

WU, Y., DEY, R., HAN, A., JAYATHILAKA, N., PHILIPS, M., YE, J. & CHEN, L. 2010. Structure of the MADS-box/MEF2 Domain of MEF2A Bound to DNA and Its Implication for Myocardin Recruitment. *Journal of Molecular Biology,* 397**,** 520-533.

WÜRSCHUM, T., LONGIN, C. F. H., HAHN, V., TUCKER, M. R. & LEISER, W. L. 2017. Copy number variations of *CBF* genes at the *Fr-A2* locus are essential components of winter hardiness in wheat. *Plant Journal,* 89**,** 764-773.

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., HELGUERA, M., KATO, K., FUKUYAMA, S., SHERMAN, J. & DUBCOVSKY, J. 2004a. Allelic variation at the *VRN*-1 promoter region in polyploid wheat. *Theoretical and Applied Genetics,* 109**,** 1677-1686.

YAN, L., LOUKOIANOV, A., BLECHL, A., TRANQUILLI, G., RAMAKRISHNA, W., SANMIGUEL, P., BENNETZEN, J. L., ECHENIQUE, V. & DUBCOVSKY, J. 2004b. The Wheat *VRN2* Gene Is a Flowering Repressor Down-Regulated by Vernalization. *Science,* 303**,** 1640-1644.

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.

YEH, S., MOFFATT, B. A., GRIFFITH, M., XIONG, F., YANG, D. S. C., WISEMAN, S. B., SARHAN, F., DANYLUK, J., XUE, Y. Q., HEW, C. L., DOHERTY-KIRBY, A. & LAJOIE, G. 2000. Chitinase Genes Responsive to Cold Encode Antifreeze Proteins in Winter Cereals. *Plant Physiology,* 124**,** 1251-1264.

ZHANG, H. N., SREENIVASULU, N., WESCHKE, W., STEIN, N., RUDD, S., RADCHUK, V., POTOKINA, E., SCHOLZ, U., SCHWEIZER, P., ZIEROLD, U., LANGRIDGE, P., VARSHNEY, R. K., WOBUS, U. & GRANER, A. 2004. Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant Journal,* 40**,** 276-290.

ZHAO, Y. S., GOWDA, M., WURSCHUM, T., LONGIN, C. F. H., KORZUN, V., KOLLERS, S., SCHACHSCHNEIDER, R., ZENG, J., FERNANDO, R., DUBCOVSKY, J. & REIF, J. C. 2013. Dissecting the genetic architecture of frost tolerance in Central European winter wheat. *Journal of Experimental Botany,* 64**,** 4453-4460.

ZHU, J., PEARCE, S., BURKE, A., SEE, D., SKINNER, D., DUBCOVSKY, J. & GARLAND-CAMPBELL, K. 2014. Copy number and haplotype variation at the *VRN-A1* and central *FR-A2* loci are associated with frost tolerance in hexaploid wheat. *Theoretical and Applied Genetics***,** 1-15.

Appendix

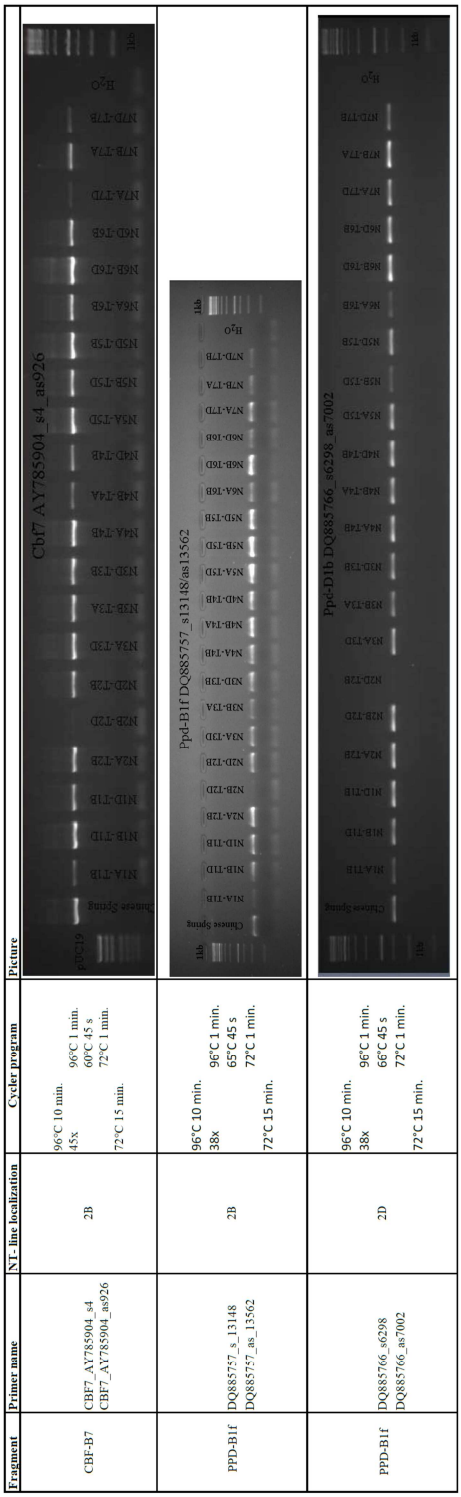
List of Figures

List of Tables

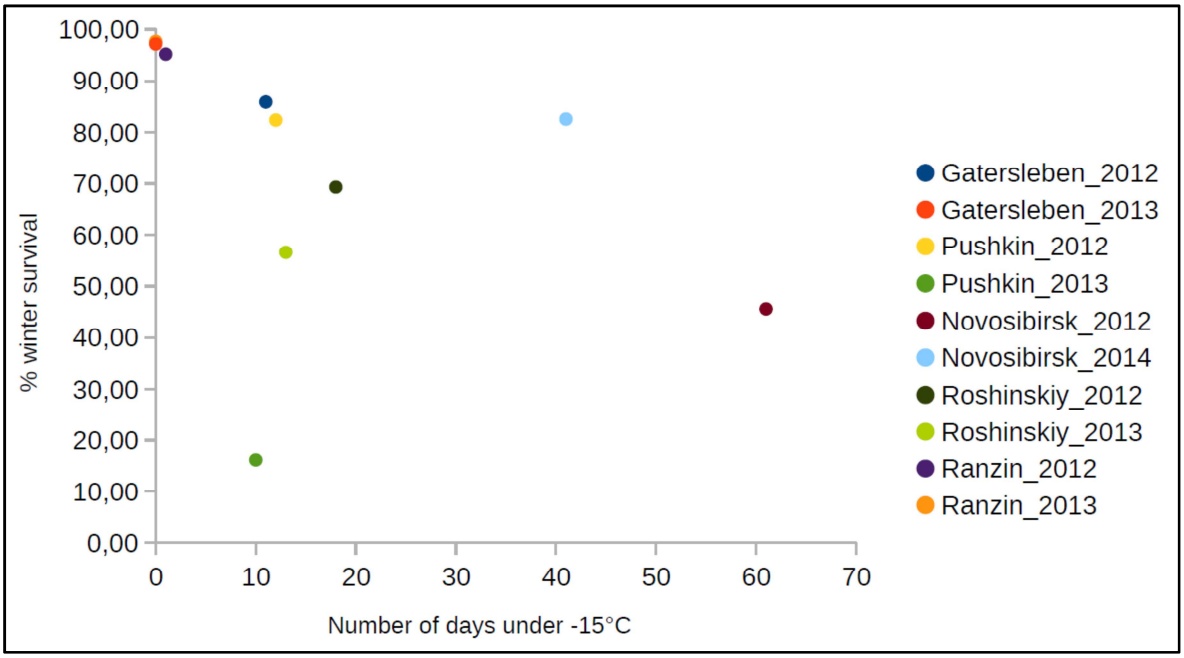
List of Formula

List of Data

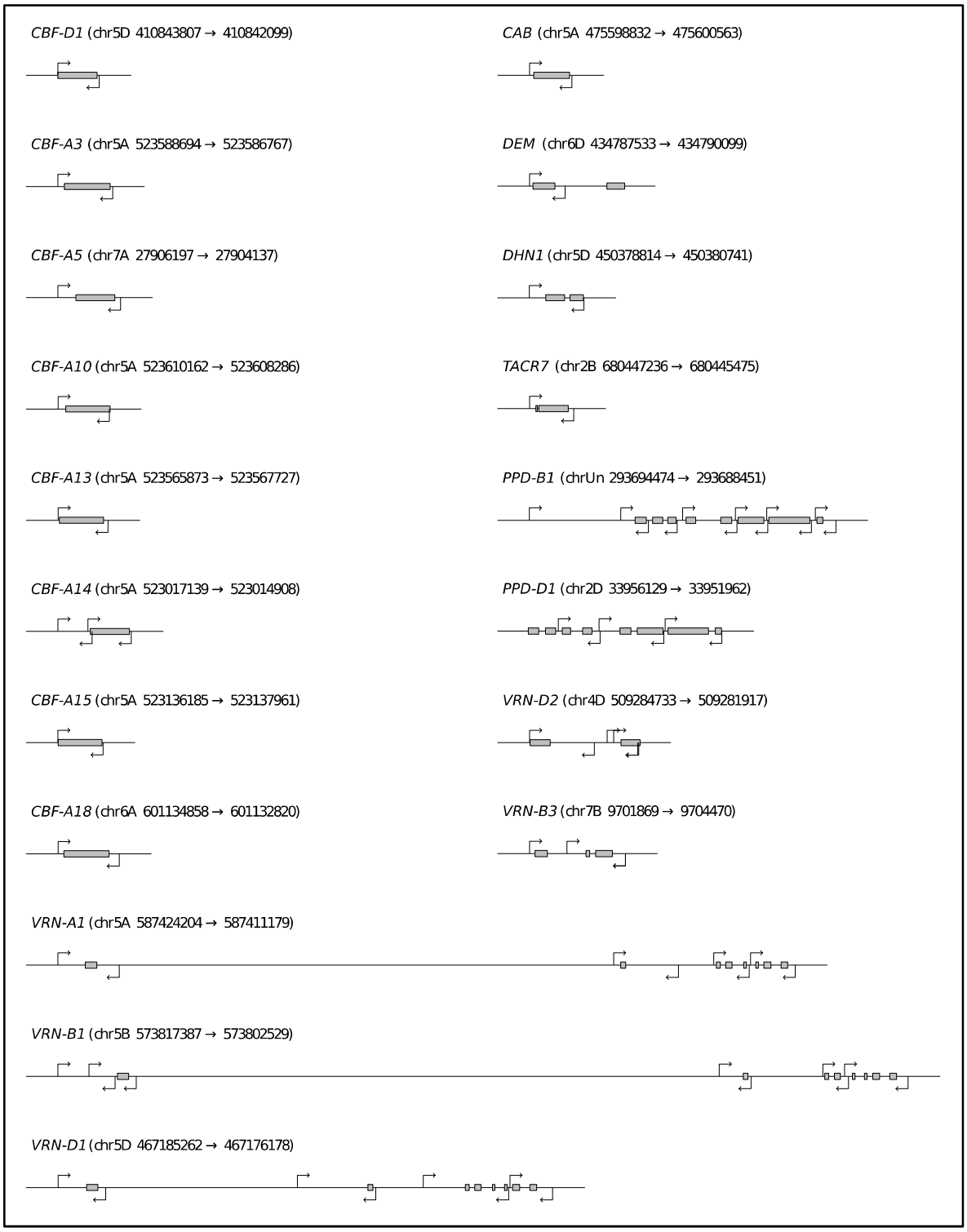
Supplementary Figures



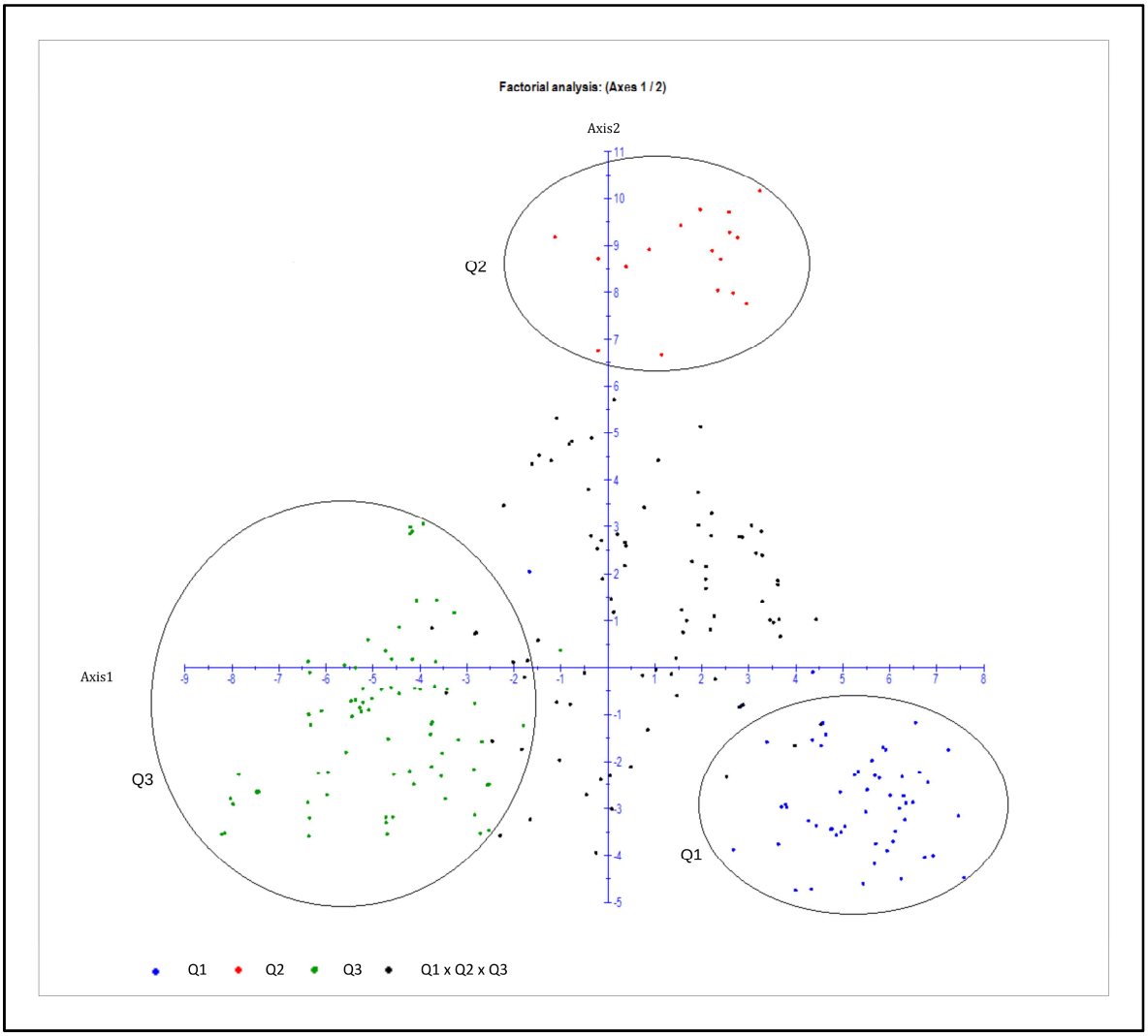
Supplemental Figure 1: Chromosome localisation and cycler programs of the PCR fragments CBF7, PPD-B1f and PPD-D1b via NT-lines (Babben et al., 2015).



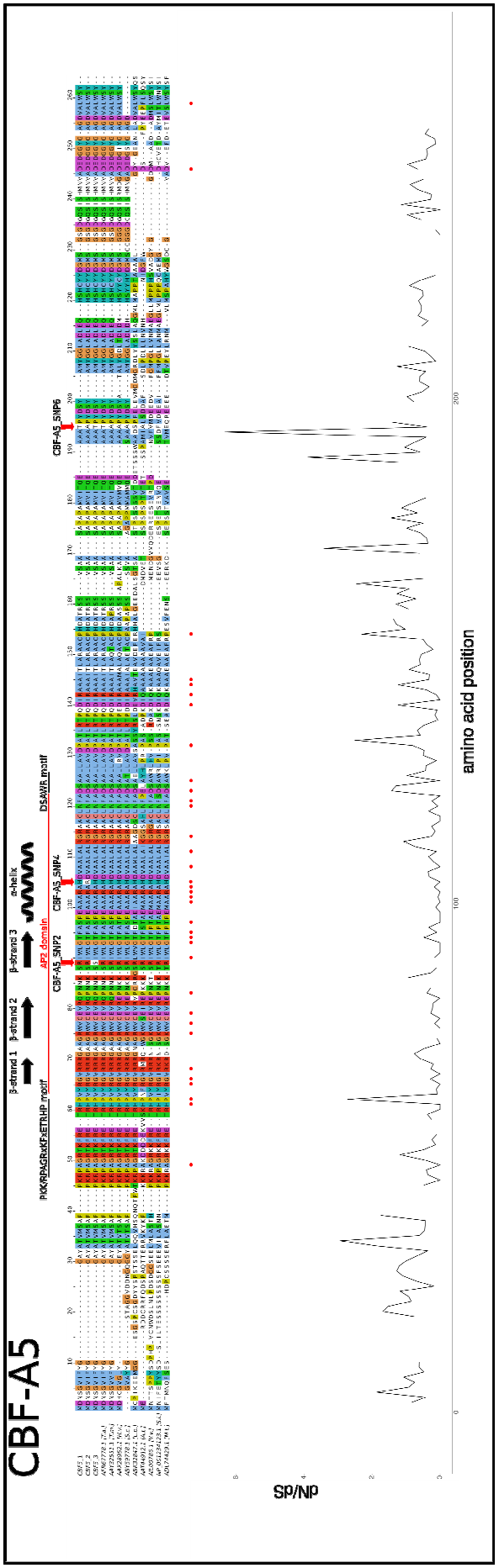
Supplemental Figure 2: Scatter plot of mean winter survival and number of days under – 15 °C from ten environments (Babben et al., 2018).



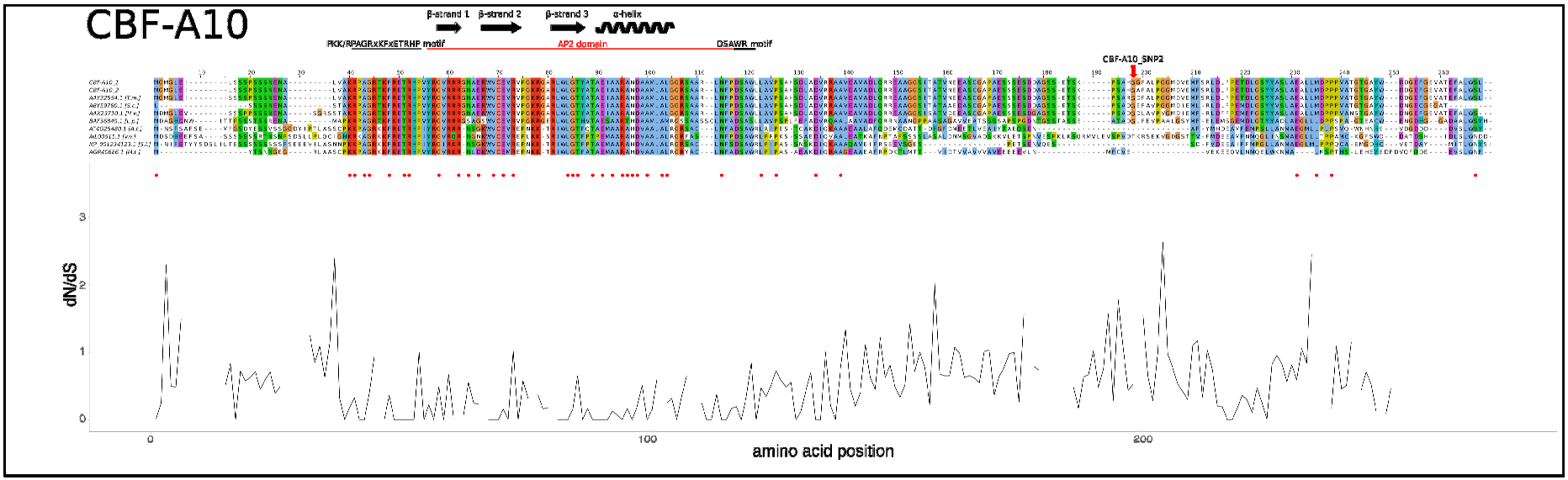
Supplemental Figure 3: Localisation of 19 candidate genes and corresponding primers in the Chinese Spring Reference assembly v1.0. Next to the gene names, the chromosome number and physical position of the Chinese Spring reference assembly v1.0 are shown in brackets. For each gene the structure is illustrated below. The black lines indicate the genomic sequences, the grey boxes the exons, the black arrows rightward the forward primers and the black arrows leftwards the reverse primers (Babben et al., 2018).



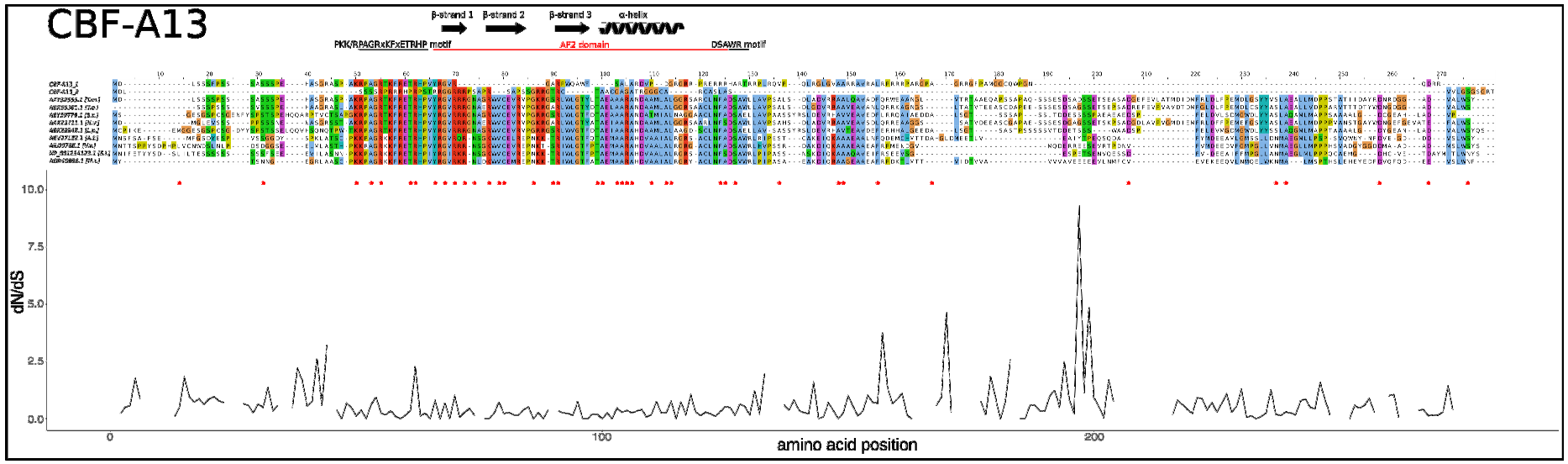
Supplemental Figure 4: Principal coordinate analysis of 235 wheat cultivars. Three sub-populations based on geographical origin were shown in three colours. Blue, red and green indicate the cultivars from Europe, North America and Asia, respectively. The black dots indicate the mixture germplasm from three sub-populations (Babben et al., 2018).



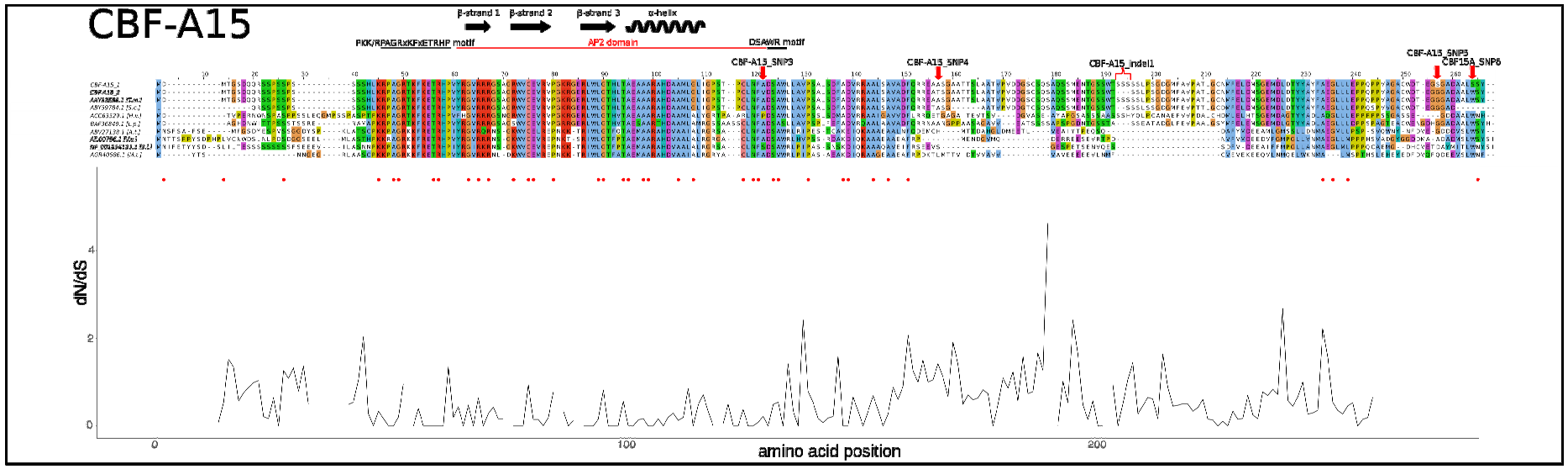
Supplemental Figure 5: Amino acid alignment and nucleotide divergence rates (dN/dS) of CBF-A5 and nine homologous amino acid sequences. Illustrated are alignments of two haplotype AA sequences of CBF-A5 and nine homologous plant AA sequences. The numbers above the alignment indicate the sites of AAs. The black line above the alignment illustrates the PKK/RPAGRxKFxETRHP and DSAWR motif, the red line the AP2 domain, the black arrows the β-strands and the black spiral the α-helix. The description of red arrows, black line und red dots is according to Fig. 16 (Babben et al., 2018).



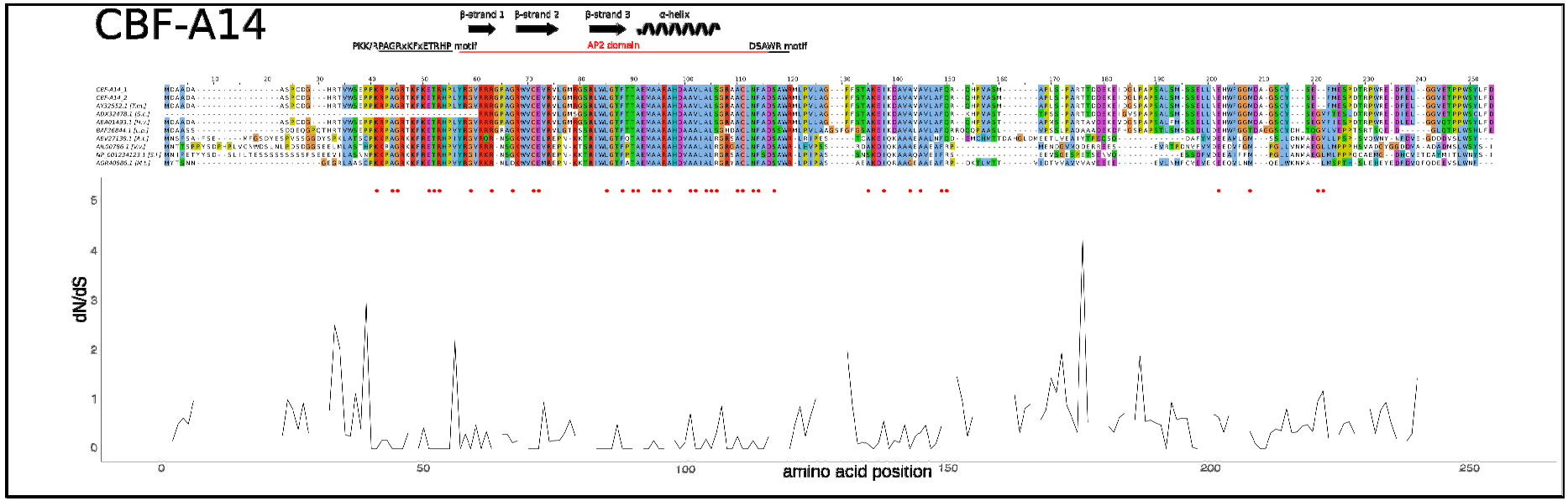
Supplemental Figure 6: Amino acid alignment and nucleotide divergence rates (dN/dS) of CBF-A10 and nine homologous amino acid sequences. Illustrated are alignments of two haplotype AA sequences of CBF-A10 and nine homologous plant AA sequences. The numbers above the alignment indicate the sites of AAs. The black line above the alignment illustrates the PKK/RPAGRxKFxETRHP and DSAWR motif, the red line the AP2 domain, the black arrows the β-strands and the black spiral the α-helix. The description of red arrow, black line and red dots is according to Fig. 16 (Babben et al., 2018).



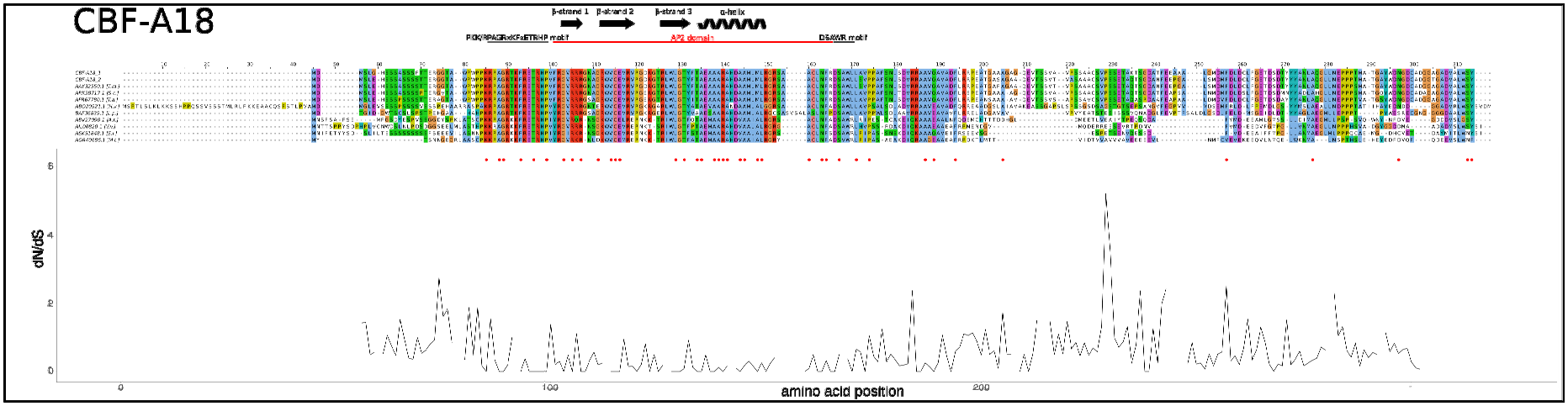
Supplemental Figure 7: Amino acid alignment and nucleotide divergence rates (dN/dS) of CBF-A13 and nine homologous amino acid sequences. Illustrated are alignments of two haplotype AA sequences of CBF-A13 and nine homologous plant AA sequences. The numbers above the alignment illustrate the sites of AAs. The black line above the alignment illustrates the PKK/RPAGRxKFxETRHP and DSAWR motif, the red line the AP2 domain, the black arrows the β-strands and the black spiral the α-helix. The description of black line and red dots is according to Fig. 16 (Babben et al., 2018).



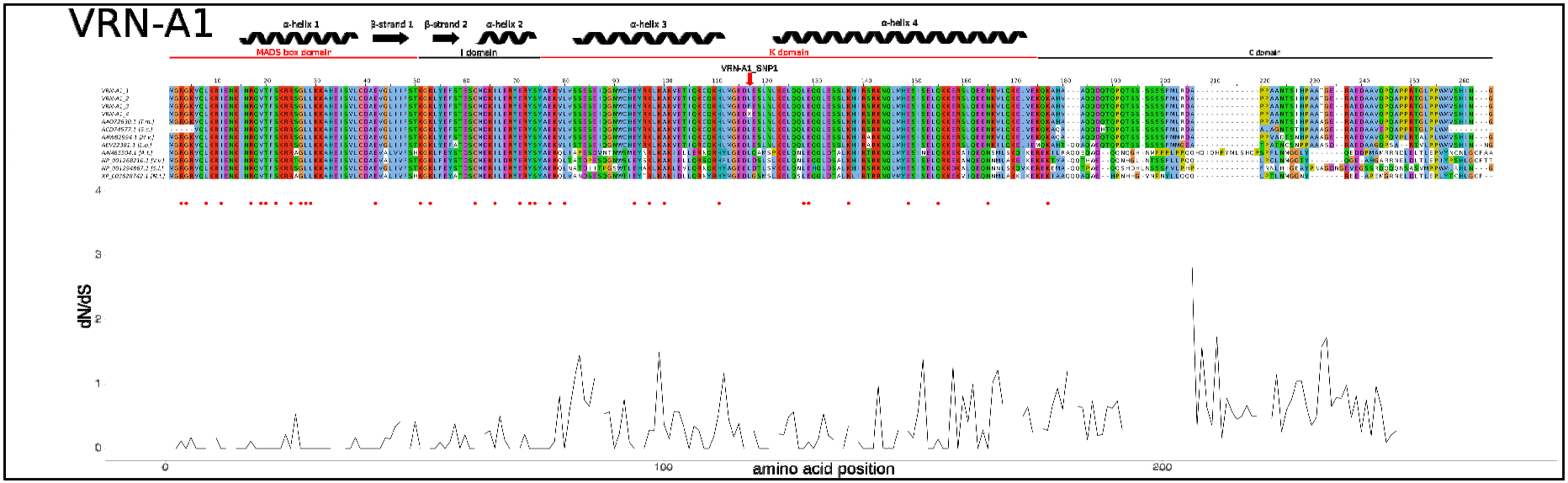
Supplemental Figure 8: Amino acid alignment and nucleotide divergence rates (dN/dS) of CBF-A15 and nine homologous amino acid sequences. Illustrated are alignments of two haplotype AA sequences of CBF-A15 and nine homologous plant AA sequences. The numbers above the alignment illustrate the sites of AAs. The black line above the alignment illustrates the PKK/RPAGRxKFxETRHP and DSAWR motif, the red line the AP2 domain, the black arrows the β-strands and the black spiral the α-helix. The description of red arrows, black line and red dots is according to Fig. 16 (Babben et al., 2018).



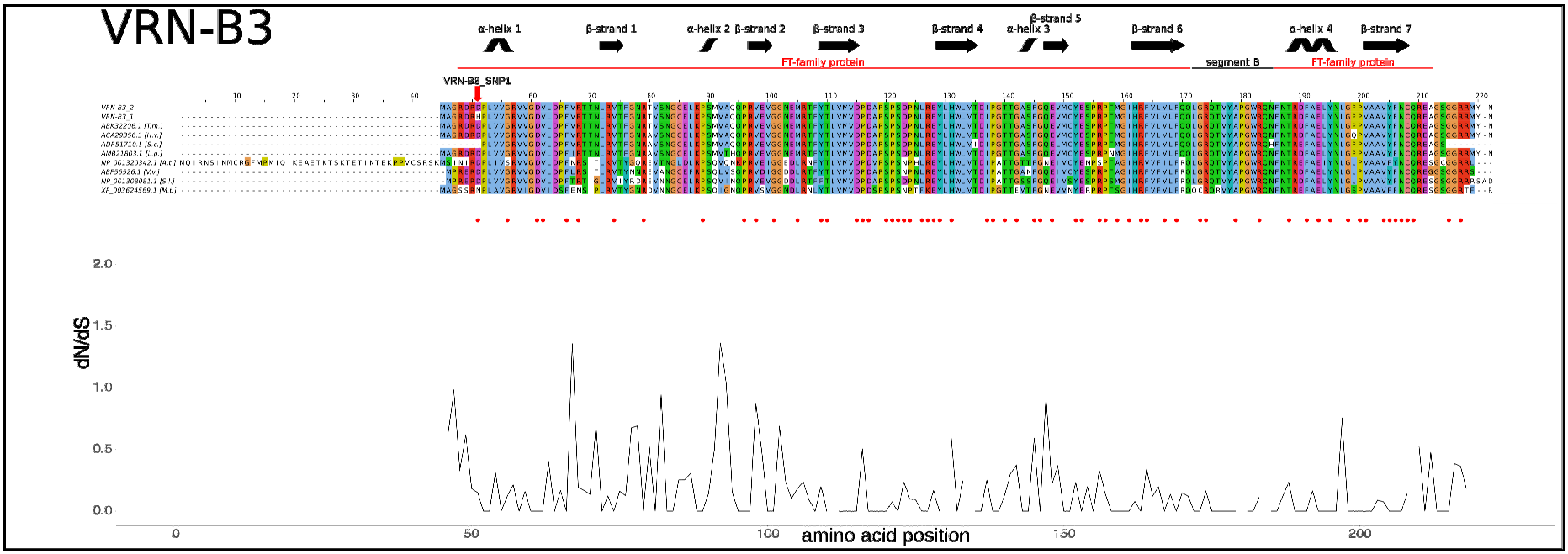
Supplemental Figure 9: Amino acid alignment and nucleotide divergence rates (dN/dS) of CBF-A14 and nine homologous amino acid sequences. Illustrated are alignments of two haplotype AA sequences of CBF-A14 and nine homologous plant AA sequences. The numbers above the alignment illustrate the sites of AAs. The black line above the alignment illustrates the PKK/RPAGRxKFxETRHP and DSAWR motif, the red line the AP2 domain, the black arrows the β-strands and the black spiral the α-helix. The description of black line and red dots is according to Fig. 16 (Babben et al., 2018).



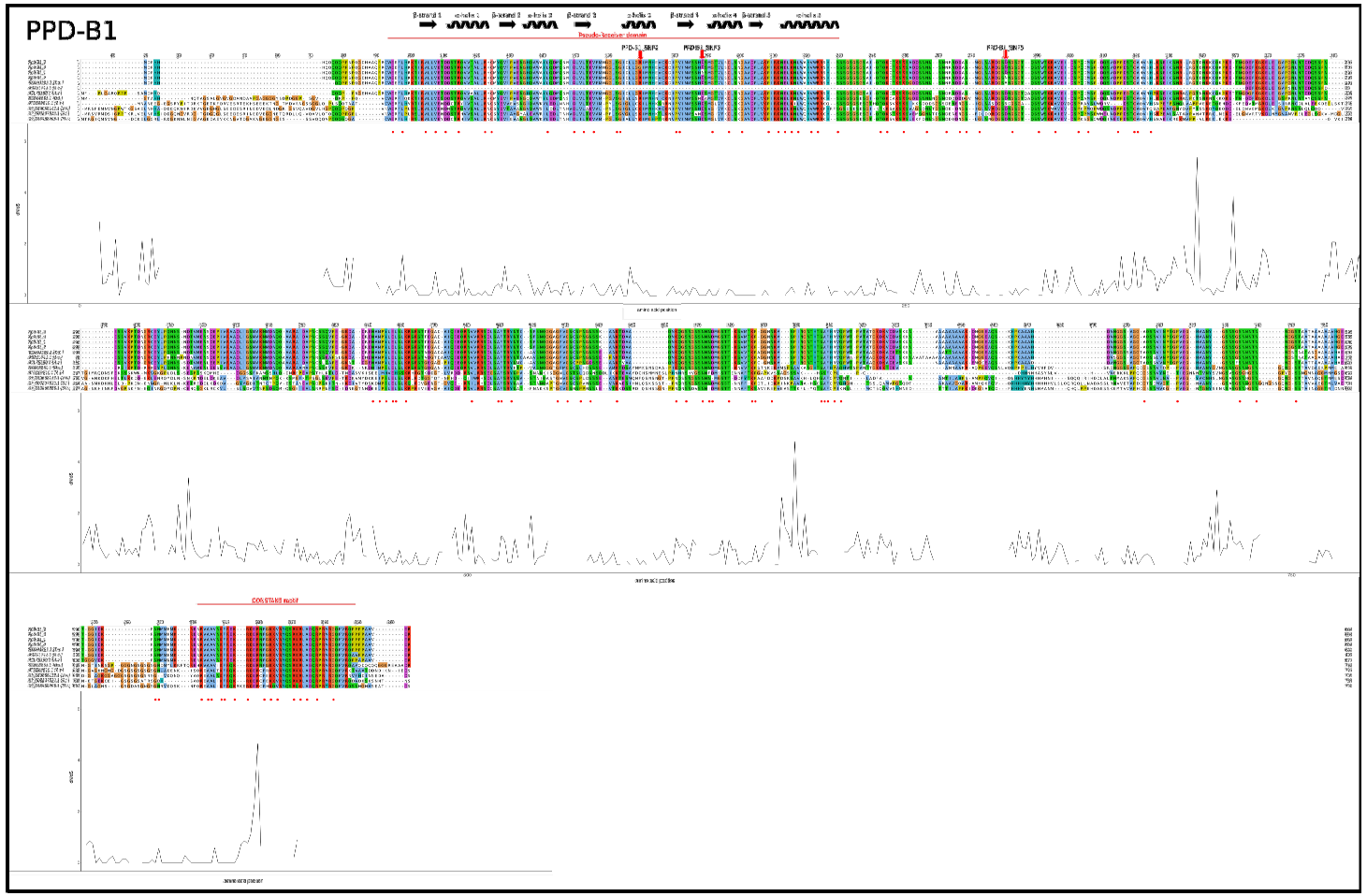
Supplemental Figure 10: Amino acid alignment and nucleotide divergence rates (dN/dS) of CBF-A18 and nine homologous amino acid sequences. Illustrated are alignments of two haplotype AA sequences of CBF-A18 and nine homologous plant AA sequences. The numbers above the alignment illustrate the sites of AAs. The black line above the alignment illustrates the PKK/RPAGRxKFxETRHP and DSAWR motif, the red line the AP2 domain, the black arrows the β-strands and the black spiral the α-helix. The description of black line und red dots is according to Fig. 16 (Babben et al., 2018).



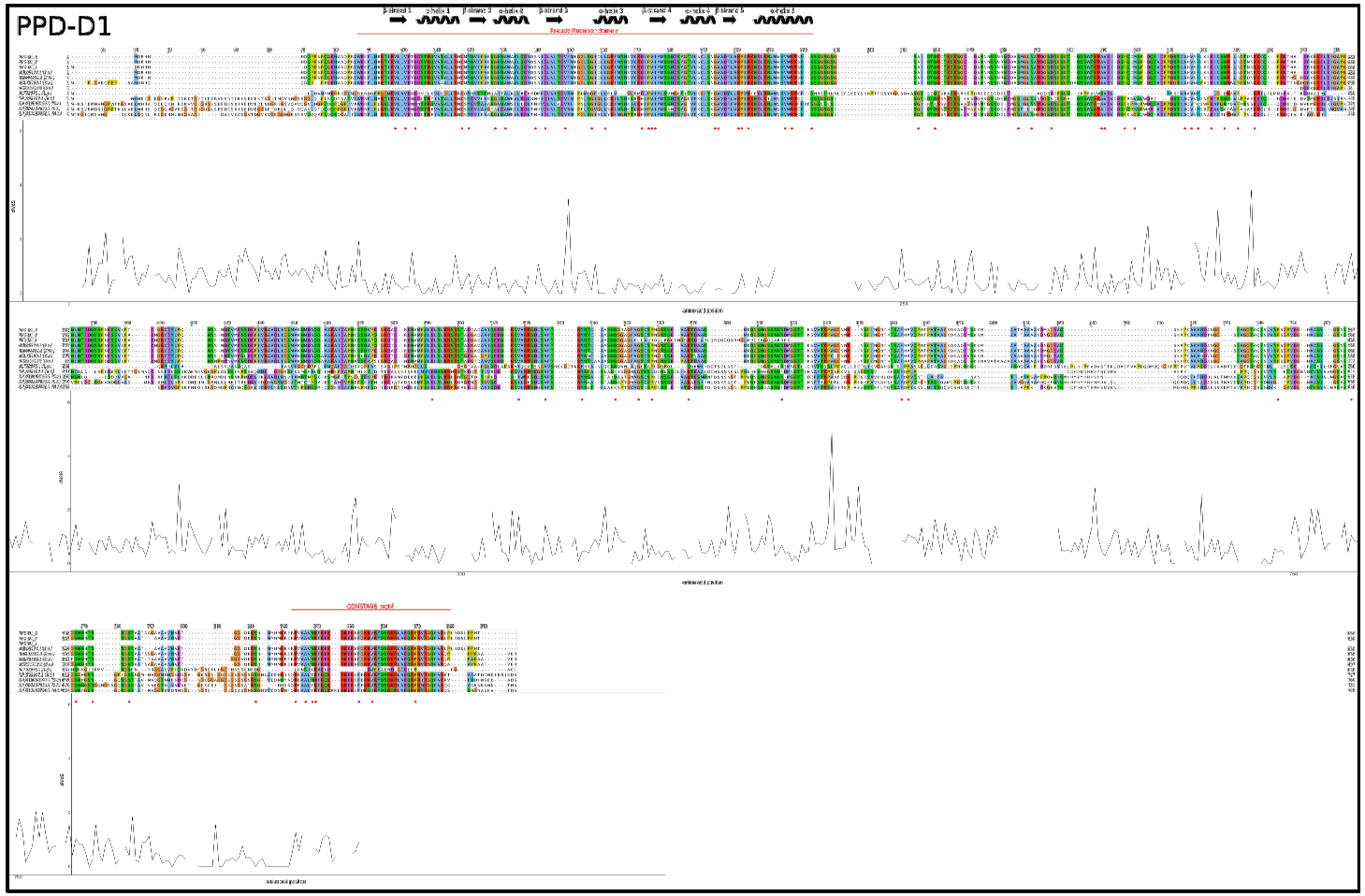
Supplemental Figure 11: Amino acid alignment and nucleotide divergence rates (dN/dS) of VRN-A1 and nine homologous amino acid sequences. Illustrated are alignments of three haplotype AA sequences of VRN-A1 nine homologous plant AA sequences. The numbers above the alignment illustrate the sites of AAs. The black line above the alignment illustrates the I domain and C domain, the red line the MADS box domain and K domain, the black arrows the β-strands and the spirals the α-helices. The description of red arrow, black line and red dots is according to Fig. 16 (Babben et al., 2018).



Supplemental Figure 12: Amino acid alignment and nucleotide divergence rates (dN/dS) of VRN-B3 and nine homologous amino acid sequences. Illustrated are alignments of three haplotype AA sequences of VRN-B3 nine homologous plant AA sequences. The numbers above the alignment illustrate the sites of AAs. The black line above the alignment illustrates the segment B, the red line the flowering time (FT)-family protein, the black arrows the β-strands and the spirals the α-helices. The description of red arrow, black line and red dots is according to Fig. 16 (Babben et al., 2018).



Supplemental Figure 13: Amino acid alignment and nucleotide divergence rates (dN/dS) of PPD-B1 and nine homologous amino acid sequences. Illustrated are alignments of four haplotype AA sequences of PPD-B1 and nine homologous plant AA sequences. The numbers above the alignment illustrate the sites of AAs. The red line above the alignment illustrates the Pseudo Receiver domain and COSTANS motif. The description of red arrows, black line and red dots is according to Fig. 16 (Babben et al., 2018).



Supplemental Figure 14: Amino acid alignment and nucleotide divergence rates (dN/dS) of PPD-D1 and nine homologous amino acid sequences. Illustrated are alignments of four haplotype AA sequences of PPD-D1 and nine homologous plant AA sequences. The numbers above the alignment illustrate the sites of AAs. The red line above the alignment illustrates the Pseudo Receiver domain and COSTANS motif. The description of black line and red dots is according to Fig. 16 (Babben et al., 2018).

Supplementary Tables

Supplemental Table 1: Plant material of NT- and Deletion-lines.

Name	Modification	Chromosome
	NT-lines	
TA 3257	(N1A-T1B)	1A
TA 3260	(N1B-T1D)	1B
TA 3262	(N1D-T1B)	1D
TA 3263	(N2A-T2B)	2A
TA 3266	(N2B-T2D)	2B
TA 3268	(N2D-T2B)	2D
TA 3270	(N3A-T3D)	3A
TA 3271	(N3B-T3A)	3B
TA 3274	(N3D-T3B)	3D
TA 3277	(N4A-T4B)	4A
TA 3275	(N4B-T4A)	4B
TA 3279	(N4D-T4B)	4D
TA 3063	(N5A-T5D)	5A
TA 3065	(N5B-T5D)	5B
TA 3067	(N5D-T5B)	5D
TA 3152	(N6A-T6B)	6A
TA 3155	(N6B-T6D)	6B
TA 3157	(N6D-T6B)	6D
TA 3281	(N7A-T7D)	7A
TA 3283	(N7B-T7A)	7B
TA 3286	(N7D-T7B)	7D
	Deletion-lines	
TA 4519 Line 6	CSdel 2BL-6	2BL
TA 4518 Line 1	CSdel 2BS-1	2BS
TA 4518 Line 3	CSdel 2BS-3	2BS
TA 4518 Line 4	CSdel 2BS-4 del 2BL-4 del 3DL-2	2BS
TA 4521 Line 9	CSdel 2DL-9	2DL
TA 4520 Line 5	CSdel 2DS-5	2DS
TA 4533 Line 13	CSdel 4DL-13	4DL
TA 4533 Line 9	CSdel 4DL-9 del 1BS-18	4DL
TA 4532 Line 1	CSdel 4DS-1	4DS
TA 4532 Line 3	CSdel 4DS-3	4DS
TA 4541 Line 8	CSdel 6AL-8 del 1BL-2 del 4DS-2	4DS/6AL
TA 4535 Line 12	CSdel 5AL-12	5AL
TA 4535 Line 17	CSdel 5AL-17	5AL
TA 4535 Line 23	CSdel 5AL-23	5AL
TA 4534 Line 3	CSdel 5AS-3	5AS
TA 4534 Line 10	CSdel 5AS-10 del 1BL-3 del 6DL-11	5AS/6DL
TA 4537 Line 9	CSdel 5BL-9	5BL
TA 4537 Line 16	CSdel 5BL-16	5BL
TA 4536 Line 5	CSdel 5BS-5	5BS
TA 4536 Line 6	CSdel 5BS-6	5BS

Supplemental Table 1: (Continued)

Name	Modification	Chromosome
	Deletion-lines	
TA 4536 Line 8	CSdel 5BS-8	5BS
TA 4536 Line 4	CSdel 5BS-4 del 7AS-1	5BS/7AS
TA 4539 Line 1	CSdel 5DL-1	5DL
TA 4539 Line 5	CSdel 5DL-5	5DL
TA 4539 Line 7	CSdel 5DL-7	5DL
TA 4538 Line 5	CSdel 5DS-5	5DS
TA 4531 Line 5	CSdel 4BL-5 del 6AL-4	6AL
TA 4540 Line 1	CSdel 6AS-1	6AS
TA 4543 Line 3	CSdel 6BL-3	6BL
TA 4543 Line 5	CSdel 6BL-5	6BL
TA 4543 Line 6	CSdel 6BL-6	6BL
TA 4542 Line 2	CSdel 6BS-2	6BS
TA 4542 Line 3	CSdel 6BS-3 del 6DL-12	6BS/6DL
TA 4528 Line 4	CSdel 4AS-4 del 6DL-10	6DL
TA 4545 Line 6	CSdel 6DL-6	6DL
TA 4544 Line 4	CSdel 6DS-4 del 1BS-19	6DS
TA 4544 Line 6	CSdel 6DS-6	6DS
TA 4526 Line 6	CSdel 3DS-6 del 4AL-11 del 7AL-18	7AL
TA 4529 Line 13	CSdel 4AL-13 del 7AL-21	7AL
TA 4547 Line 1	CSdel 7AL-1	7AL
TA 4511 Line 5	CSdel 1AL-5 del 7AS-5	7AS
TA 4546 Line 8	CSdel 7AS-8 del 7AL-17	7AS/7AL
TA 4524 Line 8	CSdel 3BS-8 del 4AS-3 del 7BL-10	7BL
TA 4549 Line 7	CSdel 7BL-7 del 1DS-3	7BL
TA 4529 Line 5	CSdel 4AL-5 del 7BL-6	7BL
TA 4548 Line 1	CSdel 7BS-1	7BS

Supplemental Table 2: Candidate gene specific primers, primer specificity, PCR fragments, used polymerases, cycler programs and primers for fragment re-sequencing.

Fragment	Primer name	Primer sequence (5` - 3`)	Polymerase	Cycler program no.	Primer specificity	Sequencing primer
Cbf1	AF376136_s1	TTTTTGACGCTGCAACTGAT	FIRE Pol.	4	unspecific	Yes
	AF376136_as709	TTTACCGAGGGAGTAGTTTCCA			5D specific	Yes
Cbf5	TmCBF5_F [*2]	CGATGCAAAGTGTGCAATTC	Bioline	4	n.a	Yes
	AY951947_as1691	ACTAGCTCATGCGAATATGGTGT			n.a	Yes
Cbf7	AY785904_s4	TTCTAGTCCACCTAGCTACAGGC	Bioline	5	unspecific	Not good for sequencing
	AY785904_as926	CACTAGCAAAGCAATTCATGAGC			2A/2B specific	Not good for sequencing
Cbf10	AY951950_s1522	ACATCTCACACACTCCACAGATG	Bioline	10	5A specific	Yes
	Cbf4B_R [*3]	GCAGAATCGGCTACAAGCTCCAG			5A/5B specific	Yes
Cbf13	Cbf5_F [*3]	CAGAGCAGAATCAGATGGGGAATC	FIRE Pol.	10	5A/5B specific	Yes
	AY951951_as1964	GCTAAGCTCACACTCCTCGATAA			5A specific	Yes
Cbf14	AY951948_s_565	TAAACTCGCTGCTTAATTACCCC	Bioline	4	5A specific	Yes
	AY951948_as_1312	ATATTTGGTGGAACAGAAGCAGA			5A specific	Yes
	AY951948_s_528	CAGCATCCATCTCTCTCAAATCT	Bioline	4	5A/5D specific	No
	AY951948_as_1299	CAGAAGCAGAGAAACCGTCTAAA			5A specific	No
	AY951948_s_565	TAAACTCGCTGCTTAATTACCCC	Bioline	4	5A specific	No
	AY951948_as_1299	CAGAAGCAGAGAAACCGTCTAAA			5A specific	No
	AY951948_s_528	CAGCATCCATCTCTCTCAAATCT	Bioline	4	5A/5D specific	No
	AY951948_as_1312	ATATTTGGTGGAACAGAAGCAGA			5A specific	No
Cbf15	EF028765_s_90	ACCGACCACCTGCAGTACC	Bioline	4	5A specific	Yes
	EF028765_as_875	TTGTTCCATGCATAGAGTCAAAG			unspecific	Yes
Cbf18	AY951946_s400	CGTATAAATACGCACACGCACTA	Bioline	8	6A specific	Yes
	AY951946_as1445	ACATGGTGGAGGGATCTTTTTAT			6A specific	Yes
Dhn1	ScDhn1_F [*1]	CCACGTAGCACGCACGCTGT	Bioline	1	n.a	Yes
	AF043087_as1808	TCGGAACATAGAGAAGACACACA			n.a	Yes
Vrn-A1b	VRN1-A_F [*4]	GAAAGGAAAAATTCTGCTCG	FIRE Pol.	1	5A specific	Yes
	AY747600_as1083	GATTACCGTCTTAACCCTTCCAC			5A specific	Yes

Supplemental table 2: (Continued)

Fragment	Primer name	Primer sequence (5` - 3`)	Polymerase	Cycler program no.	Primer specificity	Sequencing primer
Vrn-A1c	AY747600_s9072	CATGAAACAACGCATTACAGAAA	FIRE Pol.	1	5A specific	Yes
	AY747600_as10169	CAGATAGAACTGGTTGGATCCCT			5A specific	Yes
Vrn-A1d	AY747600_s_10698	TTTCTGTCATTGTTCCTTCCTGT	FIRE Pol.	1	5A specific	Yes
	AY747600_as_11318	CAAGCTAAGGCTTCATGACAAGT			5A specific	Yes
	AY747600_s_10718	TGTCCCACCCAAAGTTAGTAATG	Bioline	12	5A/5D specific	No
	AY747600_as_11390	AACGATGTAATGAGGTTACGTGC			5A specific	No
	AY747600_s_10698	TTTCTGTCATTGTTCCTTCCTGT	Bioline	12	5A specific	No
	AY747600_as_11390	AACGATGTAATGAGGTTACGTGC			5A specific	No
	AY747600_s_10718	TGTCCCACCCAAAGTTAGTAATG	FIRE Pol.	1	5A/5D specific	No
	AY747600_as_11318	CAAGCTAAGGCTTCATGACAAGT			5A specific	No
Vrn-A1e	AY747600_s_11297	CTTGTCATGAAGCCTTAGCTTGT	FIRE Pol.	8	5A specific	Yes
	AY747600_as_12066	GCTGCAGCTTGCTACTTTACTCT			unspecific	Yes
	AY747600_s_11297	CTTGTCATGAAGCCTTAGCTTGT	FIRE Pol.	8	5A specific	No
	AY747600_as_12099	AAACTGAGGTGGACAAAGTGAAA			unspecific	No
Vrn-B1b	AY747603_s18	AGGCCTAGGGTACAGTAGAATAGTAG	FIRE Pol.	1	5B specific	Yes
	AY747606_as820	CAAACGGAATCAACCAAACAG			5A/5D specific	Yes
Vrn-B1c	AY747603_s3097	TCTGAGCAGAATTATACTTACCTTGC	FIRE Pol.	1	5B specific	Yes
	AY747606_as9488	AGATCATCTGATATCGGCAAAAA			5D specific	Yes
Vrn-B1d	AY747603_s_4783	CCTTCCTGTTCCACTCAAAGTTA	FIRE Pol.	10	5B specific	Yes
	AY747603_as_5249	TTTTTAACTGTGAAGAGCATATGACTAA			unspecific	Yes
Vrn-B1e	AY747603_s5134	AAACAAGAAAAACACTTGCAGAGA	Bioline	1	5A/5B specific	Yes
	AY747603_as6211	ATTACATGGTAAATTGAGCCCAG			5B/5D specific	Yes
Vrn-D1b	AY747606_s6	TTCCCTTCTACTAGGCATAGGGT	FIRE Pol.	1	5D specific	Yes
	AY747606_as820	CAAACGGAATCAACCAAACAG			5A/5D specific	Yes
Vrn-D1c	AY747606_s8129	GTGTTGGTAGAAGGCTAGAAGCA	Bioline	1	5D specific	Yes
	AY747606_as9488	AGATCATCTGATATCGGCAAAAA			5D specific	Yes
Vrn-D1d	AY747606_s10179	GACCTCACGCCAATTTTTGT	Bioline	1	5D specific	Yes

Supplemental table 2: (Continued)

Fragment	Primer name	Primer sequence (5` - 3`)	Polymerase	Cycler program no.	Primer specificity	Sequencing primer
Vrn-D1d	AY747606_as11608	TACGAAACAATTTAGACCGGTTG			5A/5D specific	Yes
Vrn-D1e	AY747606_s11586	CAACCGGTCTAAATTGTTTCGTA	FIRE Pol.	1	5D specific	Yes
	AY747606_as12291	TTAATTCACATAAACAACATCCCACTA			5D specific	Yes
Vrn2a	AY485977_s306	AAACAAGCAAACGTTGGAGTTAG	FIRE Pol.	1	4D specific	Yes
	AY485977_as1282	AATAAGCAATTTCCTGATGCAAA			4D specific	Yes
Vrn2b	AY485977_s_1542	CAACACTGAATGAAAATGGATCA	FIRE Pol.	8	4D specific	No
	AY485977_as_1985	GAACCATCCGAGGTGAAGTTTA			4D specific	No
	AY485977_s_1542	CAACACTGAATGAAAATGGATCA	FIRE Pol.	8	4D specific	No
	AY485977_as_1972	TGAAGTTTACTAGGATCATGGGG			4D specific	No
	AY485977_s_1439	CCATAGAGCAATTGAGTTTGGAC	FIRE Pol.	21	unspecific	No
	AY485977_as_1972	TGAAGTTTACTAGGATCATGGGG			4D specific	No
Vrn2a/b	AY485977_s306	AAACAAGCAAACGTTGGAGTTAG	Bioline	9	4D specific	Yes
	AY485977_as_1972	TGAAGTTTACTAGGATCATGGGG			4D specific	Yes
Vrn3a	DQ890162_s_1430	AAGGAGTACTAGAGCGGCGAG	FIRE Pol.	12	7B/7D specific	No
	DQ890162_as_1915	TGTGGTGAGCACTTTCAGAGATA			7B specific	No
	DQ890162_s_1552	TTCCTCAATTCACAGCTTACTCC	FIRE Pol.	12	7B/7D specific	Yes
	DQ890162_as_1915	TGTGGTGAGCACTTTCAGAGATA			7B specific	Yes
Vrn3b	DQ890162_s2159	TCTTAAATACTCTCTCCGTCCGA	Bioline	18	7B specific	Yes
	DQ890162_as3153	AAGCCATTGATCTAGGGTTCAC			7B specific	Yes
	DQ890162_s2396	GAAGTACACTTATTCGTGGACGG	FIRE Pol.	19	7B specific	No
	DQ890162_as3153	AAGCCATTGATCTAGGGTTCAC			7B specific	No
Vrn3a/b	DQ890162_s_1552	TTCCTCAATTCACAGCTTACTCC	FIRE Pol.	11	7B/7D specific	No
	DQ890162_as3153	AAGCCATTGATCTAGGGTTCAC			7B specific	No
Cab b	contig22616_s209	TTTTGCGAAAGCACACTTATACA	FIRE Pol.	1	n.a	Yes
	contig22616_as938	GAAGCATCGCCAGCTATAAATAC			n.a	Yes
Cab d	contig22616_s209	TTTTGCGAAAGCACACTTATACA	FIRE Pol.	17	n.a	No
	contig22616_as828	CAGTTGCAGCAGAGAGATTCTT			n.a	No

Supplemental table 2: (Continued)

Fragment	Primer name	Primer sequence (5` - 3`)	Polymerase	Cycler program no.	Primer specificity	Sequencing primer
Dem	CD937801_s29	ATACCATCGGCAACTCCTCTG	Bioline	1	n.a	Yes
	contig1013618_as520	CCATTATGGATAGCGAAATTTGA			n.a	Yes
Tacr7 b	contig4120743_s26	CAACCAAAACTCGCCTATAAAAG	FIRE Pol.	1	unspecific	Yes
	contig2688312_as455	AATCGGAGAGGAAGCTCTCTTTA			2B specific	Yes
Tacr7 c	contig4120743_s271	CGAGGAGAAGGTTTGGGGTT	FIRE Pol.	1	2B/2D specific	No
	contig2688312_as455	AATCGGAGAGGAAGCTCTCTTTA			2B specific	No
Tacr7 d	BJ246882_s196	GTCGGCGAGGAGAAGGTTTT	FIRE Pol.	1	2B specific	No
	contig2688312_as455	AATCGGAGAGGAAGCTCTCTTTA			2B specific	No
Ppd-B1c	DQ885757_s11028	TCCTTCCAGCTTACTAGTGCATC	FIRE Pol.	1	unspecific	Yes
	DQ885757_as11954	ATCACCTGGAAAACATATTGGAA			2B specific	Yes
Ppd-B1d	DQ885757_s11883	AACTGAACCAAAAGCCTGCTACT	FIRE Pol.	1	2B specific	Yes
	DQ885757_as12453	GTACCTTGCAAAGAATGAAAACG			unspecific	Yes
Ppd-B1e	DQ885757_s_12390	CCTTTGTGAATCCTTAAATCATCC	FIRE Pol.	10	unspecific	Yes
	DQ885757_as_13162	AACAGAGAACAAACGAAATCGG			2B specific	Yes
Ppd-B1f	DQ885757_s_13184	GGGCTTATTCATGATAGCTGATG	FIRE Pol.	13	2B specific	Yes
	DQ885757_as_13562	ATCGACTCCGCACTTCTACTATG			2A/2B specific	Yes
	DQ885757_s_13148	CGTTTGTTCTCTGTTCTTCGTTT	FIRE Pol.	12	2A/2B specific	No
	DQ885757_as_13625	ACCGTTACACAGGTTCAGACATT			2B specific	No
	DQ885757_s_13184	GGGCTTATTCATGATAGCTGATG	Bioline	13	2B specific	No
	DQ885757_as_13625	ACCGTTACACAGGTTCAGACATT			2B specific	No
	DQ885757_s_13148	CGTTTGTTCTCTGTTCTTCGTTT	Bioline	13	2A/2B specific	No
	DQ885757_as_13562	ATCGACTCCGCACTTCTACTATG			2A/2B specific	No
Ppd-D1 Prom	DQ885766_s3601 [†6]	CTTGTCCAACTCCCAATCTAGTG	FIRE Pol.	1	2D specific	Not sequenced
	DQ885766_as4689 [†6]	TCCTCCCCTGTTTCTTTTTACTC			2D specific	Not sequenced
	DQ885766_s4578	TCGTCCATCCAAAGATACTGATT	Bioline	4	2D specific	No
	DQ885766_as5712 [†6]	AGTACGCTGCCGTGAGTAATAAT			2B/2D specific	No
	DQ885766_s4450 [†6]	CATACTCCCTCCGTTTCTTCTTT	FIRE Pol.	6	2D specific	Not sequenced

Supplemental table 2: (Continued)

Fragment	Primer name	Primer sequence (5` - 3`)	Polymerase	Cycler program no.	Primer specificity	Sequencing primer
	DQ885766_as5712 [†6]	AGTACGCTGCCGTGAGTAATAAT			2B/2D specific	Not sequenced
	DQ885766_s4578	TCGTCCATCCAAAGATACTGATT	FIRE Pol.	6	2D specific	No
	DQ885766_as5700	TGAGTAATAATCGAACCTCGGTC			2B/2D specific	No
Ppd-D1a	DQ885766_s5689	ATTATTACTCACGGCAGCGTACT	FIRE Pol.	7	2B/2D specific	No
	DQ885766_as6299 [†6]	TACTGAAACATTTTAGGGCCAAG			2D specific	No
	DQ885766_s5677 [†6]	GACCGAGGTTCGATTATTACTCA	FIRE Pol.	8	2B/2D specific	Not sequenced
	DQ885766_as6299 [†6]	TACTGAAACATTTTAGGGCCAAG			2D specific	Not sequenced
Ppd-D1a2	DQ885766_s5766 [†6]	CAACATGTTTCCTCTTGGAGC	FIRE Pol.	1	2A/2D specific	Not sequenced
	DQ885766_as6535 [†6]	GAACAGAGTCAAACACCATCAGA			2D specific	Not sequenced
Ppd-D1b	DQ885766_s6298	TATCAGGTTCATTTGCTTCAGTG	FIRE Pol.	14	2A/2D specific	No
	DQ885766_as7002 [†6]	ATGGACAAATTGACCTCTAGTGC			unspecifc	No
	DQ885766_s6277 [†6]	CTTGGCCCTAAAATGTTTCAGTA	Bioline	6	2D specific	Not sequenced
	DQ885766_as7002 [†6]	ATGGACAAATTGACCTCTAGTGC			unspecifc	Not sequenced
	DQ885766_s6277 [†6]	CTTGGCCCTAAAATGTTTCAGTA	Bioline	20	2D specific	No
	DQ885766_as6963	GCCATTCAGTTTTATCTAGCTTCC			unspecific	No
Ppd-D1c	DQ885766_s7244	TGACAAGTATCTGCATCTGAACC	FIRE Pol.	1	2D specific	No
	DQ885766_as8033 [†6]	GATTCGCAAAGGACACTGATATT			unspecific	No
Ppd-D1d	DQ885766_s6939 [†6]	GGAAGCTAGATAAAACTGAATGGC	FIRE Pol.	10	2D specific	Yes
	DQ885766_as8033 [†6]	GATTCGCAAAGGACACTGATATT			unspecific	Yes
Ppd-D1e	DQ885766_s8011 [†6]	AATATCAGTGTCCTTTGCGAATC	FIRE Pol.	1	2A/2D specific	Yes
	Ppd-D1exon8_R1 [*5]	gtctaaatagtaggtactagg			2D specific	Yes
Ppd-D1 3`UTR	DQ885766_s8771	CTGCTCTCTGTTCTTGGTTTCAT	FIRE Pol.	1	2D specific	Yes
	DQ885766_as9720	ACCTCCCTGACGAAAAGCTC			2D specific	Yes

Supplemental table 2: (Continued)

No.	Cycler program						
1	TD 62-56	94°C 5 min.			final	94°C 5 min.	
		12x	94°C 30 s			30x	94°C 30 s
			62°C - 56°C 30 s	-0.5°C/cycle			56°C 30 s
			72°C 30 s				72°C 30 s
		72°C 10 min.				72°C 10 min.	
2	TD 66-62	94°C 5 min.			final	94°C 5 min.	
		8x	94°C 30 s			30x	94°C 30 s
			66°C - 62°C 30 s	-0.5°C/cycle			62°C 30 s
			72°C 30 s				72°C 30 s
		72°C 10 min.				72°C 10 min.	
3		96°C 10 min.					
		38x	96°C 1 min.				
			58°C 45 s				
			72°C 1 min.				
		72°C 15 min.					
4		96°C 10 min.					
		38x	96°C 1 min.				
			60°C 45 s				
			72°C 1 min.				
		72°C 15 min.					
5		96°C 10 min.					
		45x	96°C 1 min.				
			60°C 45 s				
			72°C 1 min.				
		72°C 15 min.					

Supplemental table 2: (Continued)

No.	Cycler program	
6	94°C 5 min.	
	38x	94°C 30 s
		61°C 30 s
		72°C 1 min.
	72°C 10 min.	
7	96°C 10 min.	
	38x	96°C 1 min.
		61°C 45 s
		72°C 1 min.
	72°C 15 min.	
8	96°C 10 min.	
	40x	96°C 1 min.
		62°C 45 s
		72°C 1 min.
	72°C 15 min.	
9	94°C 5 min.	
	40x	94°C 30 s
		62°C 45 s
		72°C 1 min 45 s
	72°C 10 min.	
10	96°C 10 min.	
	38x	96°C 1 min.
		63°C 45 s
		72°C 1 min.
	72°C 15 min.	
11	96°C 10 min.	
	38x	96°C 1 min.

Supplemental table 2: (Continued)

No.	Cycler program		
			63°C 45 s
			72°C 1 min 30 s
		72°C 15 min.	
12		96°C 10 min.	
		38x	96°C 1 min.
			64°C 45 s
			72°C 1 min.
		72°C 15 min.	
13		96°C 10 min.	
		38x	96°C 1 min.
			65°C 45 s
			72°C 1 min.
		72°C 15 min.	
14		96°C 10 min.	
		38x	96°C 1 min.
			66°C 45 s
			72°C 1 min.
		72°C 15 min.	
15		96°C 10 min.	
		38x	96°C 1 min.
			68°C 45 s
			72°C 1 min.
		72°C 15 min.	

Supplemental table 2: (Continued)

No.	Cycler program		
16		96°C 10 min.	
		38x	96°C 1 min.
			55°C 45 s
			72°C 1 min.
		72°C 15 min.	
17		94°C 10 min.	
		38x	94°C 30 s
			57°C 30 s
			72°C 30 s
		72°C 15 min.	
18		94°C 5 min.	
		45x	94°C 30 s
			60°C 30 s
			72°C 1 min.
		72°C 10 min.	
19		94°C 10 min.	
		38x	94°C 30 s
			63°C 30 s
			72°C 30 s
		72°C 10 min.	
20		96°C 10 min.	
		38x	96°C 1 min.
			65.5°C 45 s
			72°C 1 min.

Supplemental table 2: (Continued)

No.	Cycler program		
		72°C 15 min.	
21		94°C 5 min.	
		40x	94°C 1min.
			65.5°C 1min.
			72°C 1 min.
		72°C 10 min.	

Primer names with † are developed in course of this work but published from Keilwagen et al. (2014). Primer names with * as already published were used in combination with primers with † and without labels. [1] (Li et al., 2011b); [2][(Miller et al., 2006); [3] (Vagujfalvi et al., 2005); [4] (Yan et al., 2004a); [5] (Beales et al., 2007); [6] (Keilwagen et al., 2014)

Supplemental Table 3: Plant material, origin, LS means of winter survival [%] and winter survival [%] of each tested environment.

					Winter survival in % for each tested location and year									
Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
7017	IC&G SB RAS	RUS	line	77.48	80.40	7.00	39.90	87.00	98.80	98.60	100.00	37.50	100.00	100.00
7109	IC&G SB RAS	RUS	line	82.34	97.50	24.00	39.60	74.50	100.00	100.00	100.00	50.00	100.00	100.00
7114	IC&G SB RAS	RUS	line	83.15	89.20	22.00	59.20	91.50	99.00	98.80	87.50	50.00	100.00	100.00
88-85	HLWWC	RUS	line	83.81	82.80	18.00	64.70	98.50	95.70	100.00	100.00	62.50	95.00	100.00
88-86	HLWWC	RUS	line	80.06	81.30	16.00	49.30	89.50	88.90	100.00	100.00	62.50	100.00	100.00
Acciaio	CoreCollection NoviSad via IPK	ITA	cultivar	29.48	74.50	8.00	21.10	77.50	24.70	100.00	12.50	12.50	15.00	100.00
Ai-bian	CoreCollection NoviSad via IPK	JPN	cultivar	31.41	73.60	6.00	30.90	67.00	24.30	99.10	12.50	25.00	10.00	85.00
AKTEUR	GSA	GER	cultivar	80.53	92.60	9.00	41.70	97.50	100.00	100.00	75.00	62.50	100.00	100.00
Al'bina 45	HLWWC	RUS	cultivar	84.85	92.20	7.00	51.30	88.50	100.00	100.00	100.00	62.50	100.00	100.00
Almari	GSA	PL	cultivar	82.06	92.10	10.00	59.80	77.00	100.00	100.00	75.00	62.50	90.00	100.00
Ana	CoreCollection NoviSad via IPK	HRV	cultivar	63.09	81.80	24.00	27.50	93.00	71.60	100.00	62.50	37.50	85.00	90.00
Antonovka	GSA	BUL	cultivar	60.77	71.60	11.00	47.80	97.00	88.60	92.50	75.00	37.50	40.00	90.00
APACHE	GSA	FRA	cultivar	64.00	68.10	7.00	38.80	91.00	87.30	98.50	25.00	37.50	95.00	90.00
ARKTIS	GSA	GER	cultivar	75.98	84.20	11.50	41.10	91.50	100.00	91.05	87.50	62.50	80.00	100.00
ATHLON	GSA	GER	cultivar	77.00	77.50	33.00	43.50	94.50	95.70	86.05	75.00	62.50	100.00	45.00
AURA	NGB	FIN	cultivar	82.28	85.70	21.00	44.80	84.00	100.00	100.00	87.50	75.00	100.00	100.00
AURELE	GSA	FRA	cultivar	69.68	77.20	17.00	29.90	71.00	95.50	100.00	75.00	37.50	90.00	100.00
Avalon	CoreCollection NoviSad via IPK	GBR	cultivar	59.31	62.40	13.00	20.40	83.50	85.40	98.70	37.50	50.00	85.00	100.00
Bagrationovskaya	IC&G SB RAS	RUS	cultivar	25.35	12.50	0.00	0.00	0.00	100.00	0.00	12.50	n.a.	5.00	0.00
Bankut 1205	CoreCollection NoviSad via IPK	HUN	cultivar	76.67	65.90	4.00	52.80	63.00	100.00	100.00	75.00	75.00	90.00	100.00
Bashkirskaya 10	VIR	RUS	cultivar	79.38	73.20	21.00	61.80	85.50	100.00	100.00	n.a.	50.00	100.00	100.00
Batuta	GSA	PL	cultivar	74.64	71.40	21.00	54.70	79.00	100.00	100.00	62.50	62.50	85.00	100.00
BCD 1302/83	CoreCollection NoviSad via IPK	MDA	line	84.73	100.00	10.00	47.80	94.00	100.00	98.90	100.00	62.50	95.00	85.00
Benni multifloret	CoreCollection NoviSad via IPK	USA	cultivar	77.00	88.90	16.00	37.00	82.00	97.10	94.45	75.00	50.00	100.00	100.00
Bezenchukskaja 380	GSA	RUS	cultivar	80.21	92.70	17.75	57.80	83.25	100.00	98.89	59.38	62.50	90.00	100.00
Bezenchukskaya 380	VIR	RUS	cultivar	77.32	78.90	18.50	64.50	83.00	98.50	100.00	n.a.	37.50	90.00	100.00
Bezenchukskaya 616	VIR	RUS	cultivar	73.19	95.00	11.00	57.80	89.50	97.00	100.00	n.a.	n.a.	55.00	100.00

Supplemental Table 3: (Continued)

Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Winter survival in % for each tested location and year									
					Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
Bezenchukskaya 790	VIR	RUS	cultivar	79.71	84.40	23.00	62.00	84.00	100.00	92.90	n.a.	50.00	90.00	100.00
Bezostaja 1	NGB CoreCollection	RUS	cultivar	85.05	92.60	9.00	58.10	88.00	100.00	100.00	87.50	62.50	100.00	100.00
Bezostaja 1	NoviSad via IPK	RUS	cultivar	84.53	98.70	5.00	49.40	85.00	100.00	100.00	87.50	62.50	100.00	100.00
BEZOSTAYA (1?)	HLWWC	RUS	cultivar	75.39	82.70	14.88	49.20	82.38	98.60	98.74	68.75	50.00	87.50	97.50
Bezostaya 1	VIR	RUS	cultivar	74.06	72.50	23.00	41.30	95.00	96.00	100.00	n.a.	62.50	95.00	100.00
Biruza	GSA	RUS	cultivar	65.26	81.70	12.75	65.00	81.88	100.00	98.66	62.50	62.50	17.50	98.75
Biryuza	VIR	RUS	cultivar	77.98	98.90	10.00	37.60	89.50	97.10	100.00	n.a.	50.00	95.00	100.00
Bogatka	GSA	PL	cultivar	72.99	69.70	14.00	41.90	91.50	100.00	96.15	75.00	62.50	85.00	100.00
BORG	NGB CoreCollection	SWE	cultivar	73.50	81.30	25.00	27.90	88.00	100.00	100.00	87.50	50.00	90.00	100.00
Brigant	NoviSad via IPK	GBR	cultivar	69.53	76.00	16.00	23.10	80.00	97.80	100.00	50.00	62.50	95.00	95.00
BRILLIANT	GSA	GER	cultivar	71.80	57.40	19.00	50.80	87.50	100.00	100.00	75.00	37.50	95.00	100.00
BUENNO	GSA	FRA	cultivar	54.83	89.60	1.00	23.20	100.00	89.30	100.00	25.00	25.00	47.50	85.00
BULAVA	HLWWC	RUS	cultivar	78.72	94.00	15.00	32.50	90.00	98.80	100.00	75.00	62.50	100.00	100.00
BULK02R2B	HLWWC CoreCollection	USA	line	85.56	96.20	11.00	69.70	80.50	98.50	100.00	75.00	62.50	95.00	100.00
Cajeme 71	NoviSad via IPK CoreCollection	MEX	cultivar	74.84	89.50	18.00	38.10	86.50	97.10	92.30	37.50	62.50	100.00	85.00
Capelle Desprez	NoviSad via IPK	FRA	cultivar	36.64	70.50	7.00	0.60	59.00	57.50	94.45	12.50	12.50	42.50	85.00
CDC BUTEO	HLWWC	CAN	cultivar	89.57	90.50	24.00	76.50	92.50	100.00	100.00	87.50	75.00	100.00	100.00
CDC FALCON	HLWWC	CAN	cultivar	83.77	86.20	20.00	69.40	93.00	100.00	99.00	75.00	62.50	95.00	100.00
Cheyenne	NGB CoreCollection	USA	cultivar	77.59	67.80	15.33	56.00	92.50	100.00	99.60	66.67	62.50	100.00	100.00
Cook	NoviSad via IPK	AUS	cultivar	48.73	54.50	9.00	24.90	76.00	88.40	92.95	25.00	37.50	45.00	100.00
Crina	GSA	ROM	cultivar	64.59	100.00	28.00	62.10	92.00	100.00	100.00	37.50	50.00	17.50	100.00
DH01-25-135*R	HLWWC	CAN	DH line	86.31	91.00	22.00	66.00	82.00	100.00	100.00	87.50	62.50	100.00	100.00
DH01-25-199*R	HLWWC	CAN	DH line	83.65	88.40	7.00	63.80	86.00	97.70	98.40	75.00	62.50	100.00	100.00
DH01-29-33*R	HLWWC	CAN	DH line	81.84	77.40	30.00	62.30	86.00	100.00	98.80	75.00	75.00	95.00	90.00
DH01-29-167	HLWWC	CAN	DH line	89.07	95.80	10.00	75.00	87.00	100.00	100.00	87.50	62.50	100.00	100.00
DH01-32-13	HLWWC	CAN	DH line	81.03	74.40	6.00	56.20	85.50	100.00	100.00	87.50	62.50	100.00	100.00
DH02-15-54	HLWWC	CAN	DH line	84.18	83.20	21.00	66.90	98.50	100.00	98.45	100.00	62.50	90.00	95.00
DH99-39-55-5*	HLWWC	CAN	DH line	70.00	54.60	23.00	34.80	88.00	98.50	100.00	75.00	50.00	100.00	100.00

Supplemental Table 3: (Continued)

Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Winter survival in % for each tested location and year									
					Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
DH99-55-342-4	HLWWC	CAN	DH line	79.44	81.20	19.00	29.00	90.00	100.00	93.40	100.00	75.00	100.00	95.00
DISCUS	GSA	GER	cultivar	60.12	96.10	24.00	50.20	73.50	87.30	98.75	62.50	37.50	17.50	85.00
Doneko	GSA	RUS	cultivar	81.19	81.60	25.00	72.60	89.50	95.80	100.00	75.00	62.50	87.50	100.00
Donska polupat.	CoreCollection NoviSad via IPK	RUS	cultivar	81.69	96.10	25.00	49.10	88.50	100.00	100.00	87.50	50.00	95.00	100.00
Durin	CoreCollection NoviSad via IPK	FRA	cultivar	45.52	38.20	12.00	22.30	75.00	65.00	97.30	62.50	37.50	52.50	95.00
Dzhangal	VIR	RUS	cultivar	80.31	81.50	10.00	57.60	78.50	100.00	97.15	n.a.	50.00	100.00	100.00
Ermak	GSA	RUS	cultivar	61.69	82.90	7.00	54.80	69.00	96.20	100.00	62.50	62.50	12.50	95.00
Ernie	GSA	USA	cultivar	74.38	83.20	25.00	41.00	80.50	95.60	100.00	62.50	62.50	90.00	100.00
Ershovskaya 11	VIR	RUS	cultivar	81.18	92.50	22.00	48.60	88.50	95.70	100.00	n.a.	62.50	100.00	100.00
FAMULUS	GSA	GER	cultivar	79.36	86.70	15.00	43.40	93.00	98.40	92.55	75.00	62.50	100.00	100.00
Favoritka	GSA	RUS	cultivar	77.39	94.10	22.00	37.50	75.50	98.20	100.00	62.50	62.50	95.00	100.00
FI-400	NGB	KG	cultivar	80.62	89.00	8.00	43.50	91.00	96.10	100.00	87.50	62.50	100.00	100.00
Filatovka	IC&G SB RAS	RUS	cultivar	87.96	91.40	19.00	72.60	81.50	100.00	100.00	87.50	75.00	95.00	100.00
Finezja	GSA	PL	cultivar	75.52	74.90	18.00	55.90	79.50	83.60	92.70	62.50	75.00	95.00	100.00
Flair	GSA	GER	cultivar	62.60	74.70	6.00	26.20	97.50	100.00	100.00	62.50	62.50	50.00	100.00
Florida	CoreCollection NoviSad via IPK	USA	cultivar	83.14	87.90	12.00	53.20	74.00	100.00	98.60	100.00	50.00	100.00	95.00
Goldfield	GSA	USA	cultivar	77.40	88.90	11.00	38.60	86.50	96.20	100.00	75.00	62.50	95.00	100.00
GSA1	GSA	FRA	line	68.53	76.30	17.00	26.10	84.00	91.90	97.05	75.00	62.50	80.00	100.00
GSA3	GSA	FRA	line	71.91	87.40	16.00	32.60	81.50	100.00	82.45	75.00	50.00	77.50	100.00
GSA5	GSA	FRA	line	59.92	69.20	15.00	16.30	67.00	92.00	97.75	75.00	50.00	60.00	100.00
GSA7	GSA	FRA	line	69.53	72.70	24.00	35.60	91.50	93.50	100.00	62.50	50.00	90.00	90.00
GSA10	GSA	FRA	line	58.74	79.10	11.00	39.00	83.50	94.80	100.00	12.50	50.00	50.00	100.00
GSA12	GSA Kursk 2009/10	GER	line	69.08	81.30	18.00	28.80	90.00	97.00	90.45	37.50	50.00	95.00	100.00
Guberniya	VIR	RUS	cultivar	70.71	43.60	7.00	55.40	91.00	97.00	98.00	n.a.	50.00	100.00	100.00
Helios	CoreCollection NoviSad via IPK	USA	cultivar	68.21	84.80	22.00	19.30	76.50	97.40	100.00	62.50	37.50	90.00	100.00
Holly E	CoreCollection NoviSad via IPK	USA	cultivar	41.01	59.60	23.00	13.60	26.50	78.60	100.00	0.00	37.50	35.00	95.00
Irkutskaya ozimaya	IC&G SB RAS	RUS	cultivar	83.52	71.50	27.00	64.00	79.00	100.00	100.00	100.00	75.00	95.00	100.00

Supplemental Table 3: (Continued)

Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Winter survival in % for each tested location and year									
					Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
ISENGRAIN	GSA CoreCollection	FRA	cultivar	71.09	73.60	12.00	51.10	80.00	98.70	100.00	37.50	37.50	95.00	95.00
Ivanka	NoviSad via IPK	SRB	cultivar	53.74	61.40	17.00	26.80	92.50	98.40	100.00	25.00	50.00	45.00	100.00
JULIUS	GSA	GER	cultivar	71.71	73.50	17.25	28.00	90.13	100.00	100.00	75.00	50.00	95.00	97.50
JYVÄ	NGB	FIN	cultivar	83.75	95.80	21.00	50.60	84.00	97.70	98.70	87.50	62.50	100.00	100.00
Kalach 60	VIR	RUS	cultivar	85.10	91.30	18.00	69.40	57.00	98.80	100.00	n.a.	62.50	95.00	100.00
Karabalikskaya Osimaya	GSA	KAZ	cultivar	87.07	94.70	16.00	58.60	78.00	100.00	94.45	100.00	75.00	95.00	100.00
KARABALYKSKAYA 101	HLWWC	KAZ	cultivar	87.37	91.90	22.00	70.30	76.00	100.00	97.85	87.50	62.50	100.00	100.00
KARABALYKSKAYA OSTISTAYA	HLWWC	KAZ	cultivar	86.82	87.50	25.00	68.40	86.00	97.40	97.85	100.00	62.50	100.00	100.00
Kazanskaya 285	VIR	RUS	cultivar	77.01	86.50	20.00	32.90	94.50	98.60	100.00	n.a.	62.50	100.00	100.00
Kazanskaya 560	VIR	RUS	cultivar	81.15	90.00	6.00	46.70	90.00	100.00	100.00	n.a.	62.50	100.00	100.00
Kirgiskaja 16	GSA	RUS	cultivar	80.11	95.60	19.00	57.90	86.50	100.00	97.35	62.50	62.50	85.00	100.00
Kirija	GSA	UKR	cultivar	79.09	86.20	19.00	58.40	86.50	100.00	91.05	62.50	75.00	82.50	90.00
Kobra plus	GSA	PL	cultivar	79.15	88.30	15.00	49.10	78.00	100.00	100.00	75.00	62.50	90.00	100.00
Kohelia	GSA	PL	cultivar	80.57	90.20	8.00	55.50	84.00	100.00	100.00	62.50	62.50	95.00	100.00
KOMSOMOLSKAYA 103	HLWWC	KAZ	cultivar	57.28	60.90	6.00	12.40	44.50	86.00	100.00	50.00	25.00	90.00	100.00
Komsomolskaya 75	GSA	KAZ	cultivar	78.38	79.70	9.50	41.40	70.00	100.00	100.00	75.00	87.50	90.00	100.00
Korweta	GSA	PL	cultivar	71.17	73.70	30.00	51.20	84.50	96.30	100.00	37.50	62.50	85.00	95.00
KP134-3	HLWWC	RUS	line	84.47	100.00	18.00	50.60	86.50	97.10	100.00	87.50	62.50	100.00	100.00
Kristall	GSA	BUL	cultivar	63.17	85.70	17.00	41.80	88.50	100.00	95.00	50.00	37.50	45.00	100.00
KRUZHINKA/MV IRMA	HLWWC	UKR	cultivar	80.65	100.00	12.00	41.20	85.50	100.00	96.75	62.50	62.50	100.00	100.00
Kuibyshevka	VIR	RUS	cultivar	84.30	94.80	22.00	62.20	84.50	98.90	100.00	n.a.	62.50	95.00	100.00
Kulundinka	IC&G SB RAS CoreCollection	RUS	cultivar	79.48	91.50	23.00	76.30	95.00	96.30	100.00	87.50	75.00	52.50	100.00
L - 1	NoviSad via IPK CoreCollection	HUN	line	76.44	92.70	3.00	48.40	90.50	97.80	100.00	62.50	50.00	87.50	100.00
L 1/91	NoviSad via IPK CoreCollection	SRB	line	51.03	67.10	14.00	10.90	63.50	96.30	100.00	12.50	50.00	50.00	100.00
Lambriego Inia	NoviSad via IPK	CHL	cultivar	72.47	75.30	24.00	42.60	90.00	92.40	100.00	62.50	62.50	90.00	100.00
Legenda	GSA	PL	cultivar	77.86	79.40	14.00	57.80	92.50	100.00	100.00	62.50	62.50	90.00	100.00

Supplemental Table 3: (Continued)

Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Winter survival in % for each tested location and year									
					Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
LEIFFER	GSA	GER	cultivar	75.72	92.70	4.00	45.00	93.00	100.00	100.00	75.00	37.50	85.00	95.00
LINNA	NGB	FIN	cultivar	82.65	92.00	12.00	47.90	87.50	98.80	100.00	75.00	75.00	100.00	100.00
Liona	GSA	UKR	cultivar	72.70	76.40	19.00	18.90	91.50	98.70	100.00	75.00	75.00	95.00	100.00
LUTESCENS410H48	HLWWC	KAZ	cultivar	88.66	79.80	14.00	84.00	96.50	98.70	100.00	100.00	62.50	100.00	100.00
LUTESCENS410H53	HLWWC	KAZ	cultivar	89.68	89.90	26.00	77.70	74.00	100.00	100.00	87.50	75.00	100.00	100.00
M-31*/Jo3088	IC&G SB RAS	RUS	line	76.25	95.00	5.00	33.50	81.50	97.00	93.25	87.50	50.00	87.50	100.00
M-31/Cloud	IC&G SB RAS	RUS	line	84.32	92.90	24.00	54.10	80.00	100.00	100.00	87.50	62.50	100.00	100.00
M-31/Flex	IC&G SB RAS	RUS	line	83.80	95.00	25.00	53.60	72.00	100.00	96.65	n.a.	62.50	100.00	100.00
M-31/Fox	IC&G SB RAS	RUS	line	83.54	91.30	13.00	47.90	84.50	96.40	84.70	100.00	n.a.	100.00	100.00
M-31/Holley	IC&G SB RAS	RUS	line	78.43	97.30	26.00	29.60	71.50	96.90	100.00	100.00	37.50	100.00	100.00
Malahit	GSA	RUS	cultivar	84.57	92.00	6.00	62.60	90.50	100.00	96.65	75.00	62.50	100.00	100.00
MATRIX	GSA	GER	cultivar	76.40	82.20	15.00	43.90	70.00	100.00	100.00	75.00	62.50	87.50	95.00
Mikon	GSA	GER	cultivar	84.22	78.50	17.00	69.90	97.50	96.90	100.00	87.50	75.00	95.00	100.00
Min. Dwarf	CoreCollection NoviSad via IPK	AUS	cultivar	54.41	51.10	19.00	10.50	78.50	83.30	100.00	62.50	37.50	77.50	100.00
Mina	CoreCollection NoviSad via IPK	SRB	cultivar	70.39	77.70	15.00	30.40	97.00	98.10	96.35	62.50	50.00	90.00	100.00
Mironovska 808	CoreCollection NoviSad via IPK	UKR	cultivar	80.95	95.40	10.00	44.80	76.50	100.00	100.00	87.50	62.50	90.00	100.00
Mironovskaya 808	HLWWC	UKR	cultivar	89.33	90.10	17.00	69.50	92.50	100.00	98.80	100.00	75.00	100.00	100.00
Moskovskaja 56	GSA	RUS	cultivar	83.57	82.60	13.00	60.70	91.50	100.00	96.60	75.00	75.00	100.00	100.00
Moskovskaya 39	GSA	RUS	cultivar	86.28	93.60	14.00	56.90	80.00	100.00	99.50	87.50	75.00	100.00	100.00
Moskovskaya 40	VIR	RUS	cultivar	85.55	95.30	30.00	67.40	100.00	100.00	100.00	n.a.	50.00	100.00	100.00
Moskovskaya 56	VIR	RUS	cultivar	83.96	98.50	19.00	53.20	98.00	97.60	100.00	n.a.	62.50	100.00	100.00
Mulan	GSA	GER	cultivar	78.42	88.30	19.00	41.70	73.00	100.00	100.00	62.50	62.50	100.00	100.00
Muszelka	GSA	PL	cultivar	67.53	64.30	25.00	36.70	69.50	98.30	100.00	50.00	62.50	82.50	100.00
Muza	GSA	PL	cultivar	84.51	92.50	31.00	61.70	87.00	100.00	100.00	62.50	75.00	100.00	100.00
Naridana	GSA	PL	cultivar	72.58	78.30	16.50	32.50	82.50	100.00	100.00	62.50	62.50	90.00	95.00
Narobna	GSA	PL	cultivar	75.32	86.30	17.00	22.00	71.00	100.00	100.00	75.00	62.50	100.00	100.00
Neda	GSA	BUL	cultivar	62.67	79.20	20.00	53.00	75.00	92.80	92.85	62.50	25.00	45.00	85.00
Nemchinovskaya 24	VIR	RUS	cultivar	79.53	100.00	10.00	36.80	83.50	98.80	95.70	n.a.	50.00	100.00	100.00
Nemchinovskaya 57	VIR	RUS	cultivar	81.12	91.80	38.00	44.70	90.50	100.00	100.00	n.a.	62.50	100.00	100.00

Supplemental Table 3: (Continued)

					Winter survival in % for each tested location and year									
Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
NI98414	HLWWC CoreCollection	USA	line	84.99	81.90	10.00	62.20	88.50	100.00	92.30	87.50	75.00	100.00	100.00
Nizija	NoviSad via IPK CoreCollection	SRB	cultivar	43.88	90.50	18.00	37.10	78.50	25.90	98.55	62.50	50.00	10.00	100.00
Norin 10	NoviSad via IPK	JPN	cultivar	67.31	81.10	19.00	47.70	75.50	66.90	98.00	62.50	50.00	85.00	95.00
Norstar	GSA CoreCollection	USA	cultivar	85.32	82.70	16.00	63.10	93.00	100.00	100.00	87.50	75.00	100.00	100.00
Nov.Crvena	NoviSad via IPK CoreCollection	SRB	cultivar	68.98	85.70	14.00	33.50	74.00	98.60	100.00	62.50	62.50	65.00	100.00
Nova banatka	NoviSad via IPK	SRB	cultivar	78.09	81.80	16.00	46.50	71.50	100.00	98.80	62.50	62.50	100.00	100.00
Novosibirskaya 32	IC&G SB RAS CoreCollection	RUS	cultivar	86.41	92.20	28.00	52.80	94.50	100.00	100.00	100.00	75.00	100.00	100.00
NS 22/92	NoviSad via IPK CoreCollection	SRB	line	67.91	97.70	26.00	33.30	90.50	97.70	100.00	62.50	50.00	55.00	100.00
NS 46/90	NoviSad via IPK CoreCollection	SRB	line	68.16	81.30	11.00	38.60	85.50	96.30	100.00	62.50	37.50	75.00	100.00
NS 55-25	NoviSad via IPK CoreCollection	SRB	line	71.89	97.50	22.00	48.30	91.50	96.60	100.00	87.50	62.50	42.50	100.00
NS 602	NoviSad via IPK CoreCollection	SRB	line	36.69	63.60	14.00	14.20	85.50	28.60	100.00	37.50	37.50	40.00	100.00
NS 63-24	NoviSad via IPK CoreCollection	SRB	line	54.28	78.20	14.00	21.60	76.50	70.80	100.00	62.50	50.00	45.00	95.00
NS 66/92	NoviSad via IPK CoreCollection	SRB	line	76.27	87.90	18.00	47.50	80.00	96.30	100.00	75.00	50.00	87.50	100.00
NS 79/90	NoviSad via IPK	SRB	line	56.29	72.70	20.00	18.80	76.00	75.40	100.00	62.50	37.50	65.00	100.00
Odesska 267	GSA	UKR	cultivar	85.39	80.30	13.00	69.60	86.50	100.00	98.60	87.50	87.50	90.00	100.00
PÄRL II	NGB	SWE	cultivar	84.54	94.60	21.50	58.50	81.50	100.00	99.10	87.50	62.50	95.00	100.00
PEREGRINE	HLWWC CoreCollection	CAN	line	79.99	81.60	10.00	51.10	89.50	98.90	100.00	75.00	62.50	100.00	100.00
Phoenix	NoviSad via IPK	USA	cultivar	75.04	89.40	5.00	42.50	73.50	98.80	100.00	62.50	37.50	95.00	100.00
PICARD	GSA CoreCollection	FRA	cultivar	61.64	78.20	0.00	18.10	81.00	94.80	100.00	50.00	25.00	80.00	95.00
PKB Krupna	NoviSad via IPK CoreCollection	SRB	cultivar	60.90	45.40	6.00	24.70	89.00	97.30	100.00	62.50	62.50	75.00	100.00
Pobeda	NoviSad via IPK	SRB	cultivar	59.19	93.20	4.00	27.80	91.50	70.30	100.00	62.50	62.50	42.50	80.00
Podjaka	GSA	UKR	cultivar	82.23	83.80	21.00	57.80	90.00	100.00	95.45	75.00	75.00	95.00	95.00

Supplemental Table 3: (Continued)

					Winter survival in % for each tested location and year									
Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
Podoljanka	GSA	UKR	cultivar	82.14	94.30	22.00	54.30	78.00	100.00	86.75	62.50	62.50	100.00	100.00
Polisska 90	GSA	UKR	cultivar	79.67	75.00	31.00	55.00	91.00	100.00	97.90	75.00	62.50	100.00	100.00
Povolzhskaya 86	VIR	RUS	cultivar	81.22	98.50	13.00	38.50	70.50	100.00	98.55	n.a.	62.50	100.00	100.00
Powaga	GSA	UKR	cultivar	79.61	92.70	12.00	33.30	87.50	100.00	92.70	87.50	87.50	85.00	100.00
Purd.39120	CoreCollection NoviSad via IPK	USA	line	73.94	78.90	13.00	27.40	84.00	100.00	98.30	87.50	50.00	95.00	100.00
Rannyaya 12	VIR	RUS	cultivar	77.50	84.10	29.00	49.30	88.00	98.10	100.00	n.a.	50.00	95.00	95.00
Renesansa	CoreCollection NoviSad via IPK	SRB	cultivar	68.93	86.90	14.00	31.30	87.50	94.40	97.30	37.50	37.50	95.00	100.00
RENODLAT SAMMETSVETE	NGB	SWE	cultivar	82.03	92.30	19.00	42.10	92.00	100.00	100.00	87.50	75.00	95.00	100.00
Resurs	VIR	RUS	cultivar	74.59	91.30	18.00	25.80	79.50	100.00	100.00	n.a.	62.50	90.00	95.00
RIDA	NGB	NOR	cultivar	79.92	91.50	21.00	33.50	91.50	100.00	100.00	87.50	62.50	100.00	95.00
Ridit	NGB	USA	cultivar	81.98	72.50	15.00	69.10	80.00	100.00	100.00	75.00	62.50	100.00	100.00
Roma	GSA	PL	cultivar	76.04	70.00	10.00	51.90	79.00	100.00	98.40	75.00	62.50	90.00	100.00
Roughrider	GSA	USA	cultivar	84.86	97.40	27.00	52.30	87.50	100.00	100.00	75.00	75.00	100.00	100.00
Rusalka	CoreCollection NoviSad via IPK	BUL	cultivar	81.62	86.80	7.00	62.90	73.00	100.00	100.00	75.00	62.50	90.00	85.00
Rywalka	GSA	PL	cultivar	75.82	89.80	18.00	33.50	77.50	100.00	93.95	62.50	50.00	100.00	100.00
S01-31-12	HLWWC	CAN	line	85.40	94.20	24.00	64.50	85.50	100.00	100.00	87.50	50.00	100.00	100.00
S01-249-8*R	HLWWC	CAN	line	78.22	77.00	14.00	45.80	71.00	100.00	91.30	87.50	50.00	100.00	100.00
S01-249-14*R	HLWWC	CAN	line	70.35	55.70	21.00	43.90	84.00	96.30	100.00	75.00	37.50	100.00	95.00
S01-285-7*R	HLWWC	CAN	line	77.10	57.80	15.00	53.10	86.50	100.00	100.00	87.50	62.50	100.00	100.00
S01-285-20*R	HLWWC	CAN	line	83.01	77.90	19.00	68.90	90.50	100.00	97.80	75.00	62.50	100.00	100.00
S01-360-1	HLWWC	CAN	line	77.78	76.10	8.00	44.40	57.50	100.00	100.00	75.00	62.50	100.00	100.00
Sailor	GSA	GER	cultivar	81.06	98.70	10.00	43.30	89.50	100.00	100.00	75.00	62.50	95.00	90.00
Santa	VIR	RUS	cultivar	74.05	76.90	16.00	42.80	87.00	100.00	100.00	n.a.	62.50	85.00	100.00

Supplemental Table 3: (Continued)

Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Winter survival in % for each tested location and year									
					Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
Saratovskaya 17	VIR CoreCollection	RUS	cultivar	77.65	81.80	16.00	44.10	86.50	100.00	100.00	n.a.	62.50	95.00	100.00
Sava	NoviSad via IPK	SRB	cultivar	62.84	61.00	12.00	28.50	77.50	97.60	100.00	25.00	50.00	90.00	100.00
SD98444/SD97060	HLWWC	USA	line	79.96	86.50	21.00	45.00	76.50	100.00	100.00	87.50	50.00	100.00	100.00
SERI82	HLWWC	MEX	cultivar	10.04	11.40	5.00	2.40	11.00	8.10	77.85	25.00	12.50	10.00	75.00
Severodonetskaya Yubileinaya	VIR	RUS	cultivar	78.40	88.00	24.00	44.40	74.00	97.00	100.00	n.a.	62.50	95.00	95.00
Simila	GSA	CZ	cultivar	70.24	62.40	15.00	38.60	85.50	93.10	100.00	62.50	62.50	95.00	100.00
SKAGEN	GSA	GER	cultivar	79.90	88.70	12.25	40.40	89.25	100.00	96.19	79.17	62.50	100.00	97.50
Skipetr	GSA	RUS	cultivar	86.35	100.00	16.00	69.60	88.00	100.00	100.00	75.00	50.00	100.00	100.00
Skorospekla 35	VIR	RUS	cultivar	78.65	78.10	33.00	47.40	81.50	100.00	100.00	n.a.	62.50	100.00	100.00
Slawa	GSA	PL	cultivar	81.42	96.00	17.00	41.50	76.50	100.00	100.00	87.50	62.50	95.00	100.00
Smuga	GSA	PL	cultivar	81.57	90.00	10.00	50.10	66.00	98.20	100.00	75.00	75.00	95.00	100.00
Smuglyanka	VIR CoreCollection	RUS	cultivar	85.79	92.30	12.00	72.90	85.00	98.50	96.60	n.a.	50.00	100.00	95.00
Sofija	NoviSad via IPK	SRB	cultivar	79.35	88.60	26.00	51.10	81.50	100.00	100.00	75.00	50.00	95.00	100.00
ST.ERYHTR 1282-08	HLWWC	UKR	line	84.93	84.90	26.00	65.20	94.50	100.00	93.90	87.50	62.50	100.00	100.00
ST.ERYHTR 1334-07	HLWWC	UKR	line	83.60	92.40	30.00	65.40	80.50	96.90	96.55	87.50	50.00	95.00	100.00
Sukces	GSA CoreCollection	PL	cultivar	87.19	93.00	26.00	68.40	87.00	100.00	100.00	75.00	75.00	100.00	100.00
Suwon 92	NoviSad via IPK	IND	cultivar	62.05	72.30	8.00	17.20	71.50	87.40	94.70	37.50	50.00	90.00	100.00
Suzor'e	NGB	BLR	cultivar	81.61	85.90	8.00	52.60	95.00	100.00	98.50	87.50	62.50	95.00	100.00
Svetoch	VIR	RUS	cultivar	81.13	96.90	24.00	39.70	81.00	100.00	100.00	n.a.	62.50	100.00	100.00
Svilena	GSA	BUL	cultivar	73.05	81.00	21.00	50.90	95.00	100.00	90.00	75.00	37.50	77.50	90.00
SW MAXI	GSA	GER	cultivar	77.43	98.80	19.00	34.30	95.50	97.10	98.50	62.50	62.50	95.00	90.00
SW TATAROS	GSA CoreCollection	GER	cultivar	68.21	78.30	11.00	24.40	84.00	100.00	100.00	50.00	62.50	82.50	95.00
Szegedi 768	NoviSad via IPK	HUN	cultivar	60.65	74.40	7.00	24.00	82.50	94.10	98.70	25.00	37.50	80.00	100.00
TARSO	GSA	GER	cultivar	83.03	92.00	19.00	59.80	82.00	100.00	83.35	87.50	50.00	95.00	50.00
TIGER	GSA	GER	cultivar	78.14	88.30	15.00	53.10	90.50	98.60	100.00	62.50	50.00	95.00	100.00

Supplemental Table 3: (Continued)

Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Winter survival in % for each tested location and year									
					Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
TJB 990-15	CoreCollection NoviSad via IPK	GBR	line	65.58	72.10	12.00	28.50	82.00	92.70	88.00	62.50	37.50	85.00	100.00
Tonacja	GSA	PL	cultivar	68.96	72.90	6.00	54.40	72.50	96.70	100.00	75.00	62.50	52.50	100.00
TRANSIT	GSA	GER	cultivar	74.33	94.00	12.00	33.10	85.00	100.00	100.00	50.00	50.00	95.00	100.00
Triple Dirk B (GK 12)	CoreCollection NoviSad via IPK	AUS	line	44.70	56.90	8.00	21.40	66.00	84.30	98.40	12.50	25.00	42.50	95.00
Triple Dirk B (GK 775)	CoreCollection NoviSad via IPK	AUS	line	87.18	91.60	18.00	69.70	67.50	100.00	100.00	87.50	62.50	100.00	100.00
Tulsa	GSA	GER	cultivar	77.08	81.50	22.00	41.70	92.00	98.90	100.00	87.50	50.00	95.00	100.00
TÜRKIS	GSA	GER	cultivar	74.71	80.20	16.00	38.00	89.00	98.20	100.00	62.50	62.50	95.00	100.00
Turnia	GSA	PL	cultivar	79.33	83.70	20.00	49.60	72.50	100.00	96.30	75.00	62.50	95.00	100.00
Ul'yanovka	IC&G SB RAS	RUS	cultivar	82.19	82.50	16.00	58.90	90.50	100.00	100.00	n.a.	62.50	100.00	100.00
UMKA	HLWWC	RUS	cultivar	89.31	91.80	14.00	73.90	88.00	100.00	100.00	100.00	62.50	100.00	100.00
Vakka	NGB	FIN	cultivar	85.69	91.70	16.00	49.70	85.00	100.00	87.25	100.00	75.00	100.00	95.00
Vel	CoreCollection NoviSad via IPK	USA	line	57.34	64.30	9.00	27.80	87.00	96.30	96.45	62.50	50.00	42.50	100.00
Viktoriya 95	VIR	RUS	cultivar	80.41	93.20	27.00	51.80	92.50	98.40	100.00	n.a.	62.50	90.00	100.00
Volzhskaya K	VIR	RUS	cultivar	87.23	96.50	5.00	80.00	83.50	100.00	88.35	n.a.	37.50	100.00	100.00
Wagrein	n/a	AUT	cultivar	82.69	91.40	15.50	56.40	96.00	99.00	100.00	75.00	n.a.	97.50	100.00
Wesley/SD97049	HLWWC	USA	line	73.22	93.40	31.00	43.10	100.00	100.00	100.00	87.50	62.50	55.00	100.00
Wydma	GSA	PL	cultivar	73.90	69.60	7.00	30.30	84.50	100.00	100.00	87.50	62.50	95.00	100.00
ZAURALSKAYA OZIMAYA	HLWWC	RUS	cultivar	83.05	84.00	16.00	62.70	79.00	99.00	98.20	87.50	62.50	95.00	100.00
Zemka	VIR	RUS	cultivar	80.97	90.10	17.00	45.80	68.00	100.00	100.00	n.a.	62.50	100.00	100.00
Zentos	GSA	GER	cultivar	69.63	30.80	17.00	42.70	96.00	100.00	100.00	75.00	75.00	100.00	95.00
ZG K 3/82	CoreCollection NoviSad via IPK	HRV	line	47.90	65.40	8.00	33.10	68.50	49.00	97.85	0.00	25.00	80.00	95.00
ZG K 238/82	CoreCollection NoviSad via IPK	HRV	line	60.33	78.30	11.00	25.10	73.00	95.00	86.65	0.00	37.50	85.00	80.00
ZG K T 159/82	CoreCollection NoviSad via IPK	HRV	line	44.64	53.60	2.00	8.60	54.50	83.90	100.00	12.50	37.50	52.50	90.00
Zhemchuzhina Povolzhiya	VIR	RUS	cultivar	82.19	98.70	18.00	44.10	77.00	98.50	100.00	n.a.	62.50	100.00	100.00

Supplemental Table 3: (Continued)

Genotype	Source	Country code	Status genotype	LSmean in % winter survival	Winter survival in % for each tested location and year									
					Pushkin 2012	Pushkin 2013	Novosibirsk 2012	Novosibirsk 2014	Ranzin 2012	Ranzin 2013	Roshchinskiy 2012	Roshchinskiy 2013	Gatersleben 2012	Gatersleben 2013
Zlatica	GSA	BUL	cultivar	70.15	87.40	19.00	28.70	86.00	100.00	100.00	62.50	37.50	85.00	100.00
ZOBEL	GSA	GER	cultivar	57.35	85.30	13.00	40.60	67.00	100.00	89.45	50.00	37.50	17.50	70.00
ZOLOTAVA/DAR ZERNOGRADA	HLWWC	UKR	cultivar	79.16	79.80	9.00	49.20	68.00	98.50	93.75	87.50	50.00	100.00	100.00
Zvonniza	GSA	RUS	cultivar	83.76	94.20	27.00	68.80	80.00	97.50	100.00	62.50	75.00	90.00	100.00

n.a.: not available

Supplemental Table 4: Primer specificity and mismatches to compared the three sub-genomes of functional and correct localised PCR fragments via *in silico* alignments.

Gene	NT- and deletion line localisation	Primer name	Specificity	A genome			B genome			D genome		
				Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction
CBF1	5DL	AF376136_s1	unspecific	no contig available			0			0		
		AF376136_as709	5D specific	no contig available			9	1, 2, 4, 5, 6, 9, 10 of 22	13 and 21 of 22	0		
CBF5	7AS	TmCBF5_F [*2]	n.a	no contig available			no contig available			no contig available		
		AY951947_as1691	n.a	0			no contig available			no contig available		
CBF7	2BL	AY785904_s4	unspecific	1		3 of 23	1		3 of 23	0		
		AY785904_as926	2A/2B specific	0			0			4	2 of 23	4, 7, 21 of 23
CBF10	5AL	AY951950_s1522	5A specific	no contig available			3		3, 19, 22 of 23	3		3, 17, 23 of 23
		Cbf4B_R [*3]	5A/5B specific	no contig available			1		15 of 23	4		8, 9, 14, 15 of 23
CBF13	5AL	Cbf5_F [*3]	5A/5B specific	no contig available			0			10	3 and 14 of 24	1, 5, 6, 9, 10, 15, 16, 24 of 24
		AY951951_as1964	5A specific	no contig available			3		16, 19, 20 of 23	3	4 of 23	1 and 20 of 23
CBF14	5AL	AY951948_s_565	5A specific	0			3	2 of 23	4 and 21 of 23	1		21 of 23
		AY951948_as_1312	5A specific	0			2	21 of 23	20 of 23	2	20 of 23	1 of 23
		AY951948_s_528	5A/5D specific	0			13	6, 7, 8, 9, 13, 18 of 23	1, 4, 5, 14, 16, 19, 21 of 23	0		
		AY951948_as_1299	5A specific	0			11	8 bp insertion	7, 11, 13, 16, 17, 18, 20, 21, 22, 23 of 23	3	7 of 23	17 and 22 of 23
CBF15	5AL	EF028765_s_90	5A specific	no contig available			1		18 of 19	no contig available		
		EF028765_as_875	unspecific	0			0			1		1 of 23

Supplemental Table 4: (Continued)

Gene	NT- and deletion line localisation	Primer name	Specificity	A genome			B genome			D genome		
				Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction
CBF18	6AL	AY951946_s400	6A specific	0			no contig available			3		8, 11, 22 of 23
		AY951946_as1445	6A specific	0			no contig available			inside gap		
Dhn1	5DL	ScDhn1_F [*1]	n.a	no contig available			no contig available			no contig available		
		AF043087_as1808	n.a	no contig available			no contig available			no contig available		
VRN-A1	5AL	VRN1-A_F [*4]	5A specific	0			5	11 of 20	1, 8, 14, 20 of 20	5	11 and 20 of 20	1, 8, 14 of 20
		AY747600_as1083	5A specific	0			3		7, 13, 17 of 23	2		11 and 17 of 23
		AY747600_s9072	5A specific	0			3		9, 11, 13 of 23	2		13 and 20 of 23
		AY747600_as10169	5A specific	0			1		21 of 23	1		21 of 23
		AY747600_s_10698	5A specific	0			2		5 and 11 of 23	2		5 and 11 of 23
		AY747600_as_11318	5A specific	0			3		1, 11, 15 of 23	4		1, 10, 11, 15 of 23
		AY747600_s_10718	5A/5D specific	0			2		4 and 9 of 23	0		
		AY747600_as_11390	5A specific	0			1		18 of 23	1		18 of 23
		AY747600_s_11297	5A specific	0			4		8, 12, 22, 23 of 23	5		8, 12, 13, 22, 23 of 23
		AY747600_as_12066	unspecific	0			1		1 of 23	0		
		AY747600_as_12099	unspecific	0			0			1		7 of 23
VRN-B1	5BL	AY747603_s18	5B specific	2		23 and 26 of 26	0			2		5 and 26 of 26
		AY747603_s3097	5B specific	3	5 and 10 of 27	20 of 27	0			10	5, 12, 15, 16, 23 of 27	2, 9, 10, 17, 21 of 27
		AY747603_s_4783	5B specific	2		10 and 15 of 23	0			2		4 and 9 of 23
		AY747603_as_5249	unspecific	1		10 of 28	0			1		10 of 28
		AY747603_s5134	5A/5B specific	1		4 of 24	0			3		18, 21, 23 of 24

Supplemental Table 4: (Continued)

Gene	NT- and deletion line localisation	Primer name	Specificity	A genome			B genome			D genome		
				Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction
VRN-B1	5BL	AY747603_as6211	5B/5D specific	5	2 and 20 of 23	4, 8, 22 of 23	0			0		
VRN-D1	5DL	AY747606_s6	5D specific	2		10 and 17 of 23	1		17 of 23	0		
		AY747606_as820	5A/5D specific	1		12 of 21	2		2 and 6 of 21	0		
		AY747606_s8129	5D specific	8	15 of 23	1, 2, 4, 13, 16, 22, 23 of 23	5	13 and 16 of 23	20, 21, 23 of 23	0		
		AY747606_as9488	5D specific	2		4 and 21 of 23	1		21 of 23	0		
		AY747606_s10179	5D specific	4	19 of 20	7, 9, 10 of 20	3	19 of 20	9 and 10 of 20	0		
		AY747606_as11608	5A/5D specific	1		5 of 23	2		14 and 21 of 23	0		
		AY747606_s11586	5D specific	1		19 of 23	2		3 and 10 of 23	0		
		AY747606_as12291	5D specific	2	18 of 27	25 of 27	3	20 of 27	24 and 27 of 27	0		
VRN-D2	4D	AY485977_s306	4D specific	no contig available			4		3, 14, 16, 20 of 23	0		
		AY485977_as1282	4D specific	no contig available			3		2, 5, 20 of 23	0		
		AY485977_s_1542	4D specific	no contig available			4		3, 5, 17, 20 of 23	0		
		AY485977_as_1985	4D specific	no contig available			2		8 and 15 of 22	0		
		AY485977_s_1439	unspecific	no contig available			1		1 of 23	1		1 of 23
		AY485977_as_1972	4D specific	no contig available			2		2 and 20 of 23	0		
VRN-B3	7BS	DQ890162_s_1430	7B/7D specific	2		8 and 18 of 21	0			0		
		DQ890162_as_1915	7B specific	13	10, 11, 12, 13, 14,15 18, 19, 20 of 23	6, 21, 22, 23 of 23	0			2		6 and 21 of 23
		DQ890162_s_1552	7B/7D specific	2		21 and 23 of 23	0			1		10 of 23

Supplemental Table 4: (Continued)

Gene	NT- and deletion line localisation	Primer name	Specificity	A genome			B genome			D genome		
				Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction
VRN-B3	7BS	DQ890162_s2159	7B specific	10	9 and 23 of 23	10, 12, 13, 14, 16, 17, 19, 20 of 23	0			8	11, 16, 17, 19 of 23	12, 14, 21, 22 of 23
		DQ890162_as3153	7B specific	inside gap			0			2	22 of 22	19 of 22
		DQ890162_s2396	7B specific	inside gap			0			inside gap		
Cab	5AL	contig22616_s209	n.a	0			no contig available			no contig available		
		contig22616_as938	n.a	no contig available			no contig available			no contig available		
		contig22616_as828	n.a	no contig available			no contig available			no contig available		
Dem	6BS/6DL	CD937801_s29	n.a	no contig available			no contig available			no contig available		
		contig1013618_as520	n.a	no contig available			no contig available			no contig available		
Tacr7	2B	contig4120743_s26	unspecific	0			0			0		
		contig4120743_s271	2B/2D specific	2	12 of 22	11 of 21	1		12 of 21	1	9 of 21	
		contig2688312_as455	2B specific	no contig available			0			2		9 and 17 of 23
		BJ246882_s196	2B specific	1		16 of 20	0			3	1, 2, 4 of 20	
PPD-B1	2BS	DQ885757_s11028	unspecific	1		6 of 23	0			0		
		DQ885757_as11954	2B specific	2		18 and 20 of 23	0			8	5 and 16 of 23	13, 18, 19, 20, 21, 23 of 23
		DQ885757_s11883	2B specific	4	10 of 23	15, 21, 22 of 23	0			3		16, 17, 21 of 23
		DQ885757_as12453	unspecific	0			0			0		
		DQ885757_s_12390	unspecific	1		7 of 24	0			1		7 of 24
		DQ885757_as_13162	2B specific	1		16 of 22	0			2		9 and 12 of 22
		DQ885757_as_13562	2A/2B specific	1		2 of 23	0			1		16 of 23
		DQ885757_s_13148	2A/2B specific	1		4 of 24	0			7	15 of 23	6, 11, 13, 20, 22, 23 of 23
		DQ885757_as_13625	2B specific	2		1, 8 of 23	0			2		1 and 8 of 23

Supplemental Table 4: (Continued)

Gene	NT- and deletion line localisation	Primer name	Specificity	A genome			B genome			D genome		
				Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction
PPD-B1	2BS	DQ885757_s_13184	2B specific	5		3, 4, 7, 11, 16 of 23	0			2		4 and 7 of 23
PPD-D1	2DS	DQ885766_s3601 [†6]	2D specific	2		1 and 14 of 23	6	20 of 23	8, 10, 11, 15, 17 of 23	0		
		DQ885766_as4689 [†6]	2D specific	5		1, 8, 17, 20, 23 of 23	7	18 and 23 od 23	1, 6, 9, 12, 13, of 23	0		
		DQ885766_s4578	2D specific	7	3 and 20 of 23	1, 10, 12, 14, 17 of 23	7	11, 12, 19 of 23	7, 15, 22, 23 of 23	0		
		DQ885766_s4450 [†6]	2D specific	3		2, 20, 22 of 23	4	17 of 23	1, 8, 20 of 23	0		
		DQ885766_as5712 [†6]	2B/2D specific	2		16 and 20 of 23	0			0		
		DQ885766_as5700	2B/2D specific	2		4 and 8 of 23	0			0		
		DQ885766_s5689	2B/2D specific	2		4 and 8 of 23	0			0		
		DQ885766_s5677 [†6]	2B/2D specific	2		16 and 20 of 23	0			0		
		DQ885766_as6299 [†6]	2D specific	2		17 and 21 of 23	4		13, 17, 19, 21 of 23	0		
		DQ885766_s5766 [†6]	2A/2D specific	1		5 of 21	1		21 of 21	0		
		DQ885766_as6535 [†6]	2D specific	3	9 of 23	8 and 12 of 23	6	4 and 15 of 23	3, 7, 9, 11 of 23	0		
		DQ885766_s6298	2A/2D specific	0			5	20 of 23	9, 15, 16, 18 of 23	0		
		DQ885766_s6277 [†6]	2D specific	2		3 and 7 of 23	4		3, 5, 7, 11 of 23	0		
		DQ885766_as7002 [†6]	unspecifc	1		13 of 23	0			0		
		DQ885766_as6963	unspecific	1		1 of 24	1		3 of 24	0		
		DQ885766_s7244	2D specific	2		4 and 23 of 23	2		4 and 23 of 23	0		
		DQ885766_s6939 [†6]	2D specific	1		24 of 24	1		22 of 24	0		
		DQ885766_as8033 [†6]	unspecific	0			1		6 of 23	0		
		DQ885766_s8011 [†6]	2A/2D specific	0			1		18 of 23	0		

Supplemental Table 4: (Continued)

Gene	NT- and deletion line localisation	Primer name	Specificity	A genome			B genome			D genome		
				Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction	Differences	Position of indels in 5` to 3' direction	Positions of SNPs in 5` to 3' direction
		Ppd-D1exon8_R1 [*5]	2D specific	4	10, 11, 12 of 21	13 of 21	4	18 of 21	13, 20, 21 of 21	0		
		DQ885766_s8771	2D specific	3		1, 4, 16 of 23	3		1, 4, 16 of 23	0		
		DQ885766_as9720	2D specific	7		3, 6, 8, 9, 10, 13, 20 of 20	4		8, 11, 17, 20 of 20	0		

Primers assigned † are developed in course of this study and published in Keilwagen et al. (2014). Already published primers with * assigned were used in combination with new developed primers. The column differences describe the numbers of SNPs/indels between primers at sequence level of A, B and D genomes. The columns position of indels and SNPs in 5' to 3` direction describes the position of the differences between primers at sub-genomes (indels and SNP) from primer 5´ to 3´ end direction. [1] (Li et al., 2011b); [2][(Miller et al., 2006); [3] (Vagujfalvi et al., 2005); [4] (Yan et al., 2004a); [5] (Beales et al., 2007); [6] (Keilwagen et al., 2014)

Supplemental Table 5: Used oligos of PCR and sequencing primers.

Fragment	Primer name	Primer sequence (5` - 3`)	Polymerase	Cycler program no.	Sequencing primer	Sequencing primer sequence (5` - 3`)
CBF-A3	CBF3_F[7]CBF3_F[7]	ATCCCACACTCTCGCTCAAG	Bioline	6	AY951949_short_s491	CAATCACCCAGCACTCAGC
	CBF3_R[7]CBF3_R[7]	GCTGGGAATTATCGGCACTA			AY951949_short_as1280	GGAATTATCGACTGTACTAGC
CBF-A5	TmCBF5_F[2]	CGATGCAAAGTGTGCAATTC	Bioline	4	AY951947_s905	GTGTCTGCAGATAAGCAAAC
	AY951947_as1691[8]	ACTAGCTCATGCGAATATGGTGT			AY951947_short_as1691	ACTAGCTCATGCGAATATGGTG
CBF-A10	AY951950_s1522[8]	ACATCTCACACACTCCACAGATG	Bioline	7	AY951950_s1628	GACTGCTCAGAAAGGAAGCG
	Cbf4B_R[3]Cbf4B_R[3]	GCAGAATCGGCTACAAGCTCCAG			MS new_60-R	GCTCCAGAGCGCGAACTC
CBF-A13	Cbf5_F[3]Cbf5_F[3]	CAGAGCAGAATCAGATGGGGAATC	FIRE Pol.	7	Cbf5_short_F	CAGAGCAGAATCAGATG
	AY951951_as1964[8]	GCTAAGCTCACACTCCTCGATAA			AY951951_as1878	GTTCGCCATGGATCTTTCC
CBF-A14a	CBF14a_F[7]CBF14a_F[7]	GCCGATATAAGCGAGCTGTC	FIRE Pol.	2	same	
	ScCbf14 _R(F)[10]	CGTCCTGTGCCCATCAC			same	
CBF-A14b	AY951948_s_565[8]	TAAACTCGCTGCTTAATTACCCC	Bioline	3	same	
	AY951948_as_1312[8]	ATATTTGGTGGAACAGAAGCAGA			same	
CBF-A15	EF028765_s_90[8]	ACCGACCACCTGCAGTACC	Bioline	3	same	
	EF028765_as_875[8]	TTGTTCCATGCATAGAGTCAAAG			same	
CBF-A18	AY951946_s400[8]	CGTATAAATACGCACACGCACTA	Bioline	5	AY951946_s474	CTAGATACCACCAGCGAGCC
	AY951946_as1445[8]	ACATGGTGGAGGGATCTTTTTAT			MS alt_65-R.1	CTCCGAAACATATAAGTC
VRN-A1b	VRN1-A_F[4]	GAAAGGAAAAATTCTGCTCG	FIRE Pol.	1	same	
	AY747600_as1083[8]	GATTACCGTCTTAACCCTTCCAC			AY747600_as834	GGAATCTACCAAACAGCAGCG
VRN-A1c	AY747600_s9072[8]	CATGAAACAACGCATTACAGAAA	FIRE Pol.	1	AY747600_short_s9072	CATGAAACAACGCATTACAG
	AY747600_as10169[8]	CAGATAGAACTGGTTGGATCCCT			AY747600_as9776	TTGTGATTTTCACTGATGCACAC
VRN-A1d	AY747600_s_10698[8]	TTTCTGTCATTGTTCCTTCCTGT	FIRE Pol.	1	same	
	AY747600_as_11318[8]	CAAGCTAAGGCTTCATGACAAGT			same	
VRN-A1e	AY747600_s_11297[8]	CTTGTCATGAAGCCTTAGCTTGT	FIRE Pol.	5	same	
	AY747600_as_12066[8]	GCTGCAGCTTGCTACTTTACTCT			same	
VRN-B1a	Vrn-B1_Pr1_F[1]	TACCCCTGCTACCAGTGCCT	FIRE Pol.	1	Pr1_short_F	TACCCCTGCTACCAGTG
	Vrn-B1_Pr2_R[1]	GGCCAACCCTACACCCCAAG			Pr2_short_R	GGCCAACCCTACACC

Supplemental Table 5: (Continued)

Fragment	Primer name	Primer sequence (5` - 3`)	Polymerase	Cycler program no.	Sequencing primer	Sequencing primer sequence (5` - 3`)
VRN-B1b	AY747603_s18[8]	AGGCCTAGGGTACAGTAGAATAGTAG	FIRE Pol.	1	AY747603_s22	CTAGGGTACAGTAGAATAGTAG
	AY747606_as820[8]	CAAACGGAATCAACCAAACAG			same	
VRN-B1c	AY747603_s3097[8]	TCTGAGCAGAATTATACTTACCTTGC	FIRE Pol.	1	same	
	AY747606_as9488[8]	AGATCATCTGATATCGGCAAAAA			AY747606_short_as9488	AGATCATCTGATATCGGC
VRN-B1d	AY747603_s_4783[8]	CCTTCCTGTTCCACTCAAAGTTA	FIRE Pol.	6	same	
	AY747603_as_5249[8]	TTTTTAACTGTGAAGAGCATATGACTAA			same	
VRN-B1e	AY747603_s5134[8]	AAACAAGAAAAACACTTGCAGAGA	Bioline	1	AY747603_s5224	AGTCATATGCTCTTCACAG
	AY747603_as6211[8]	ATTACATGGTAAATTGAGCCCAG			AY747603_as6030	TAGTACGCTTATATGGGCTG
VRN-D1b	AY747606_s6[8]	TTCCCTTCTACTAGGCATAGGGT	FIRE Pol.	1	AY747606_short_s6	TTCCCTTCTACTAGGCATAG
	AY747606_as820[8]	CAAACGGAATCAACCAAACAG			same	
VRN-D1c	AY747606_s8129[8]	GTGTTGGTAGAAGGCTAGAAGCA	Bioline	1	AY747606_s8676	TTTCTGTGTGTTGTGGGTACTTG
	AY747606_as9488[8]	AGATCATCTGATATCGGCAAAAA			AY747606_short_as9488	AGATCATCTGATATCGGC
VRN-D1d	AY747606_s10179[8]	GACCTCACGCCAATTTTTGT	Bioline	1	AY747606_10830	CCTTCCTGTTCCACTCAAAG
	AY747606_as11608[8]	TACGAAACAATTTAGACCGGTTG			same	
VRN-D1e	AY747606_s11586[8]	CAACCGGTCTAAATTGTTTCGTA	FIRE Pol.	1	AY747606_short_s11586	CAACCGGTCTAAATTGTTTCG
	AY747606_as12291[8]	TTAATTCACATAAACAACATCCCACTA			MS alt_43-R	ACTAGAGACGGGTATCATGG
VRN-B3a	DQ890162_s_1552[8]	TTCCTCAATTCACAGCTTACTCC	FIRE Pol.	7	same	
	DQ890162_as3153[8]	AAGCCATTGATCTAGGGTTCAC			DQ890162_as2436	CATAAGCATATACTCCCTCCGTC
VRN-B3b	DQ890162_s2159[8]	TCTTAAATACTCTCTCCGTCCGA	Bioline	9	DQ890162_s2396	GAAGTACACTTATTCGTGGACGG
	DQ890162_as3153[8]	AAGCCATTGATCTAGGGTTCAC			MS alt_68_R	TTAATTAGCCGTTGATTG
Cab	contig22616_s209[8]	TTTTGCGAAAGCACACTTATACA	FIRE Pol.	1	contig22616_short_s209	TTTGCGAAAGCACACTTATAC
	contig22616_as938[8]	GAAGCATCGCCAGCTATAAATAC			same	
Tacr7	contig4120743_s26[8]	CAACCAAAACTCGCCTATAAAAG	FIRE Pol.	1	same	
	contig2688312_as455[8]	AATCGGAGAGGAAGCTCTCTTTA			contig2688312_short_as455	AATCGGAGAGGAAGCTCTC
PPD-B1a	TaPpd-B1proF[9]	ACACTAGGGCTGGTCGAAGA	FIRE Pol.	1	TaPpd-short_B1proF	ACACTAGGGCTGGTCGAAG
	TaPpd-B1int1R1[9]	CCGAGCCAGTGCAAATTAAC			DQ885757_as9994	GGGGGCAGTAGCAGACAAG

Supplemental Table 5: (Continued)

Fragment	Primer name	Primer sequence (5` - 3`)	Polymerase	Cycler program no.	Sequencing primer	Sequencing primer sequence (5` - 3`)
PPD-B1b	Ppd-B1exon3SNP_F1[5]	agacgattcattccgctcc	FIRE Pol.	1	DQ885757_s10233	AACATGTTTCCTCTTGGAGTCAG
	Ppd-B1exon3SNP_R1[5]	tctgaatgatgatacaccatg			same	
PPD-B1c	DQ885757_s11028[8]	TCCTTCCAGCTTACTAGTGCATC	FIRE Pol.	1	MS new_15-F	CTAGTGCATCTTTTACTG
	DQ885757_as11954[8]	ATCACCTGGAAAACATATTGGAA			DQ885757_as11867	TTGAGCACTTGAGCATACCATAC
PPD-B1d	DQ885757_s11883[8]	AACTGAACCAAAAGCCTGCTACT	FIRE Pol.	1	DQ885757_short_s11883	AACTGAACCAAAAGCCTG
	DQ885757_as12453[8]	GTACCTTGCAAAGAATGAAAACG			MS new_51-R	GTATGCAACATATGGATG
PPD-B1e	DQ885757_s_12390[8]	CCTTTGTGAATCCTTAAATCATCC	FIRE Pol.	6	same	
	DQ885757_as_13162[8]	AACAGAGAACAAACGAAATCGG			same	
PPD-B1f	DQ885757_s_13184[8]	GGGCTTATTCATGATAGCTGATG	FIRE Pol.	8	same	
	DQ885757_as_13562[8]	ATCGACTCCGCACTTCTACTATG			same	
PPD-D1b	DQ885766_s6277[6]	CTTGGCCCTAAAATGTTTCAGTA	Bioline	4	same	
	DQ885766_as7002[6]	ATGGACAAATTGACCTCTAGTGC			same	
PPD-D1c	DQ885766_s6939[6]	GGAAGCTAGATAAAACTGAATGGC	FIRE Pol.	6	same	
	DQ885766_as8033[6]	GATTCGCAAAGGACACTGATATT			same	
PPD-D1d	DQ885766_s8011[6]	AATATCAGTGTCCTTTGCGAATC	FIRE Pol.	1	same	
	Ppd-D1exon8_R1[5]	gtctaaatagtaggtactagg			DQ885766_as8805	CCAATCAGAAGAATGAAACCAAG

No.	Cycler program					
1	TD 62-56	94°C 5 min.			94°C 5 min.	
		12x	94°C 30 s		30x	94°C 30 s
			62°C - 56°C 30 s	-0.5°C/cycle		56°C 30 s
			72°C 30 s			72°C 30 s
		72°C 10 min.			72°C 10 min.	

Supplemental Table 5: (Continued)

No.	Cycler program					
2	TD 66-62	94°C 5 min.			94°C 5 min.	
		8x	94°C 30 s		30x	94°C 30 s
			66°C - 62°C 30 s	-0.5°C/cycle		62°C 30 s
			72°C 1 min.			72°C 1 min.
		72°C 10 min.			72°C 10 min.	
3		96°C 10 min.				
		38x	96°C 1 min.			
			60°C 45 s			
			72°C 1 min.			
		72°C 15 min.				
4		94°C 5 min.				
		38x	94°C 30s			
			61°C 30 s			
			72°C 1 min.			
		72°C 10 min.				
5		96°C 10 min.				
		40x	96°C 1 min.			
			62°C 45 s			
			72°C 1 min.			
		72°C 15 min.				
6		96°C 10 min.				
		38x	96°C 1 min.			
			63°C 45 s			
			72°C 1 min.			
		72°C 15 min.				

Supplemental Table 5: (Continued)

No.	Cycler program		
7		96°C 10 min.	
		38x	96°C 1 min.
			63°C 45 s
			72°C 1 min 30 s
		72°C 15 min.	
8		96°C 10 min.	
		38x	96°C 1 min.
			65°C 45 s
			72°C 1 min.
		72°C 15 min.	
9		94°C 5 min.	
		45x	94°C 30 s
			60°C 30 s
			72°C 1 min.
		72°C 10 min.	

no.: not available; [1] (Shcherban et al., 2012); [2] (Miller et al., 2006); [3] (Vagujfalvi et al., 2005); [4] (Yan et al., 2004a); [5] (Beales et al., 2007); [6] (Keilwagen et al., 2014); [7] (Knox et al., 2008); [8] (Babben et al., 2015); [9] (Seki et al., 2011); [10] (Li et al., 2011b)

Supplemental Table 6: All involved polymorphic sites and haplotypes of association study and whose *P*-values und FT effects.

Gene/Polymorphic site	*P*-value	-Log of *P*
*Ppd-B1*_polymorphic_site_1	6.19E^{-02}	1.208309
*Ppd-B1*_polymorphic_site_2	7.71E^{-02}	1.113171
*TACR7*_polymorphic_site_1	4.16E^{-01}	0.380562
*TACR7*_polymorphic_site_2	4.16E^{-01}	0.380562
*TACR7*_polymorphic_site_3	4.16E^{-01}	0.380562
*TACR7*_polymorphic_site_4	4.16E^{-01}	0.380562
*TACR7*_polymorphic_site_5	4.16E^{-01}	0.380562
*Ppd-D1*_polymorphic_site_InDel	5.34E^{-02}	1.272459
*Ppd-D1*_polymorphic_site_TE	3.81E^{-01}	0.418972
*Ppd-D1*_polymorphic_site_83	3.77E^{-02}	1.423543
*Ppd-D1*_polymorphic_site_84	3.77E^{-02}	1.423543
*Ppd-D1*_polymorphic_site_85	3.77E^{-02}	1.423543
*Ppd-D1*_polymorphic_site_86	3.77E^{-02}	1.423543
*Ppd-D1*_polymorphic_site_87	3.77E^{-02}	1.423543
*CAB*_polymorphic_site_1	4.59E^{-01}	0.338358
*CAB*_polymorphic_site_2	4.08E^{-01}	0.388914
*CAB*_polymorphic_site_8	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_9	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_10	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_11	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_12	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_13	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_14	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_15	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_16	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_17	2.33E^{-01}	0.632551
*CAB*_polymorphic_site_18	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_19	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_20	1.07E^{-01}	0.969279
*CAB*_polymorphic_site_21	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_22	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_23	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_24	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_25	1.14E^{-01}	0.942448
*CAB*_polymorphic_site_27	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_37	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_38	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_39	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_40	2.96E^{-01}	0.528987
*CAB*_polymorphic_site_41	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_42	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_43	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_44	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_45	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_46	3.12E^{-01}	0.506277
*CAB*_polymorphic_site_48	2.85E^{-01}	0.545338
*CBF-A3*_polymorphic_site_1	6.28E^{-10}	9.201881
*CBF-A3*_polymorphic_site_2	6.28E^{-10}	9.201881
*CBF-A3*_polymorphic_site_3	6.28E^{-10}	9.201881
*CBF-A3*_polymorphic_site_4	6.28E^{-10}	9.201881
*CBF-A10*_polymorphic_site_1	6.28E^{-10}	9.201881
*CBF-A10*_polymorphic_site_2	6.28E^{-10}	9.201881
*CBF-A13*_polymorphic_site_1	6.28E^{-10}	9.201881

Supplemental Table 6: (Continued)

Gene/Polymorphic site	*P*-value	-Log of *P*
*CBF-A13*_polymorphic_site_3	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_4	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_5	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_6	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_7	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_8	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_9	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_10	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_11	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_12	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_13	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_14	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_15	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_16	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_17	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_18	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_19	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_20	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_21	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_22	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_23	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_24	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_25	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_26	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_27	$4.00E^{-09}$	8.397766
*CBF-A13*_polymorphic_site_28	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_29	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_30	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_31	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_32	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_33	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_34	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_35	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_36	$6.28E^{-10}$	9.201881
*CBF-A13*_polymorphic_site_37	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_1	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_2	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_3	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_4	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_5	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_6	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_7	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_8	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_9	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_10	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_11	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_12	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_13	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_14	$6.28E^{-10}$	9.201881
*CBF-A14*_polymorphic_site_16	$6.57E^{-10}$	9.182382
*CBF-A14*_polymorphic_site_19	$6.28E^{-10}$	9.201881
*CBF-A15*_polymorphic_site_1	$6.28E^{-10}$	9.201881
*CBF-A15*_polymorphic_site_2	$6.28E^{-10}$	9.201881
*CBF-A15*_polymorphic_site_3	$6.28E^{-10}$	9.201881

Supplemental Table 6: (Continued)

Gene/Polymorphic site	*P*-value	-Log of *P*
*CBF-A15*_polymorphic_site_4	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_5	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_6	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_7	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_8	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_9	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_10	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_11	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_12	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_13	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_14	6.28E^{-10}	9.201881
*CBF-A15*_polymorphic_site_15	6.28E^{-10}	9.201881
*VRN-A1*_polymorphic_site_30	3.03E^{-04}	3.519102
*VRN-D1*_polymorphic_site_3	5.16E^{-01}	0.287165
*CBF-A5*_polymorphic_site_1	6.00E^{-03}	2.221849
*CBF-A5*_polymorphic_site_3	6.61E^{-01}	0.179601
*CBF-A5*_polymorphic_site_4	6.61E^{-01}	0.179601
*CBF-A5*_polymorphic_site_5	6.61E^{-01}	0.179601
*CBF-A5*_polymorphic_site_6	6.61E^{-01}	0.179601
*VRN-B3*_polymorphic_site_1	3.76E^{-02}	1.425043
Haplotype name		
CAB promotor haplotype	6.87E^{-01}	0.163037
CAB exon haplotype	3.16E^{-01}	0.500849
CAB 3` haplotype	6.96E^{-01}	0.157247
CAB haplotype	3.16E^{-01}	0.500849
CBF-A3 exon haplotype	2.38E^{-11}	10.62392
CBF-A5 promotor haplotype	7.19E^{-03}	2.143271
CBF-A5 exon haplotype	1.58E^{-02}	1.801343
CBF-A5 haplotype	5.17E^{-04}	3.286434
CBF-A10 exon haplotype	2.38E^{-11}	10.62392
CBF-A13 exon haplotype	1.58E^{-10}	9.801838
CBF-A13 3` haplotype	2.38E^{-11}	10.62392
CBF-A13 haplotype	1.58E^{-10}	9.801838
CBF-A14 promotor haplotype	2.38E^{-11}	10.62392
CBF-A14 exon haplotype	1.91E^{-10}	9.719786
CBF-A14 haplotype	1.91E^{-10}	9.719786
CBF-A15 exon haplotype	2.38E^{-11}	10.62392
CBF-A18 haplotype	6.88E^{-03}	2.162412
PPD-B1 promotor haplotype	4.23E^{-02}	1.373454
PPD-B1 exon haplotype	3.69E^{-07}	6.432703
PPD-B1 haplotype	1.31E^{-07}	6.881438
PPD-D1 promotor haplotype	6.96E^{-03}	2.157391
PPD-D1 intron haplotype	3.93E^{-02}	1.40616
PPD-D1 exon haplotype	7.05E^{-03}	2.151811
PPD-D1 haplotype	3.78E^{-03}	2.422508
TACR7 exon haplotype	4.45E^{-01}	0.352118
TACR7 3` haplotype	4.45E^{-01}	0.352118
TACR7 haplotype	4.45E^{-01}	0.352118
VRN-B3 exon haplotype	7.80E^{-02}	1.107905
VRN-B3 haplotype	1.35E^{-01}	0.870665
VRN-A1 exon haplotype	2.57E^{-04}	3.589543
VRN-A1 haplotype	2.87E^{-03}	2.542118
VRN-D1 3` haplotype	4.80E^{-01}	0.318804
VRN-D1 haplotype	4.51E^{-01}	0.345583

Supplemental Table 7: Identities and e-values of FT associated AA sequences and homologue AA sequences.

Gene/AA comparison	Identity AA in %	e-value
CBF-A3		
haplotype1 – haplotype2	99	$4.00E^{-179}$
haplotype1 – ABK55362.1 [*Triticum aestivum*]	99	$2.00E^{-180}$
haplotype1 – AAY32553.1 [*Triticum monococcum*]	98	$2.00E^{-178}$
haplotype1 – ABY59782.1 [*Secale cereale*]	90	$2.00E^{-132}$
haplotype1 – ABE02630.1 [*Hordeum vulgare*]	81	$6.00E^{-130}$
haplotype1 – BAF36837.1 [*Lolium perenne*]	69	$7.00E^{-88}$
haplotype1 – AT4G25480.1 [*Arabidopsis thaliana*]	41	$1.00E^{-41}$
haplotype1 – AGR40686.1 [*Medicago truncatula*]	64	$5.00E^{-41}$
haplotype1 – NP_001234123.1 [*Solanum lycopersicum*]	54	$5.00E^{-42}$
haplotype1 – AIL00786.1 [*Vitis vinifera*]	42	$4.00E^{-40}$
haplotype2 – ABK55362.1 [*Triticum aestivum*]	98	$3.00E^{-178}$
haplotype2 – AAY32553.1 [*Triticum monococcum*]	99	$4.00E^{-180}$
haplotype2 – ABY59782.1 [*Secale cereale*]	90	$7.00E^{-134}$
haplotype2 – ABE02630.1 [*Hordeum vulgare*]	81	$2.00E^{-130}$
haplotype2 – BAF36837.1 [*Lolium perenne*]	69	$3.00E^{-88}$
haplotype2 – AT4G25480.1 [*Arabidopsis thaliana*]	43	$2.00E^{-41}$
haplotype2 – AGR40686.1 [*Medicago truncatula*]	61	$7.00E^{-41}$
haplotype2 – NP_001234123.1 [*Solanum lycopersicum*]	55	$2.00E^{-42}$
haplotype2 – AIL00786.1 [*Vitis vinifera*]	42	$1.00E^{-40}$
CBF-A5		
haplotype1 – haplotype2	99	$8.00E^{-157}$
haplotype1 – haplotype3	99	$6.00E^{-158}$
haplotype2 – haplotype3	99	$6.00E^{-156}$
haplotype1 – AFR67778.1 [*Triticum aestivum*]	100	$1.00E^{-158}$
haplotype1 – AAY32551.1 [*Triticum monococcum*]	98	$2.00E^{-155}$
haplotype1 – ABY59778.1 [*Secale cereale*]	76	$2.00E^{-88}$
haplotype1 – AAX28952.1 [*Hordeum vulgare*]	81	$7.00E^{-104}$
haplotype1 – ABK32847.1 [*Lolium perenne*]	68	$1.00E^{-42}$
haplotype1 – AAT44912.1 [*Arabidopsis thaliana*]	55	$9.00E^{-25}$
haplotype1 – ADL74429.1 [*Medicago truncatula*]	48	$4.00E^{-46}$
haplotype1 – NP_001234123.1 [*Solanum lycopersicum*]	68	$2.00E^{-45}$
haplotype1 – AIL00786.1 [*Vitis vinifera*]	45	$2.00E^{-44}$
haplotype2 – AFR67778.1 [*Triticum aestivum*]	99	$1.00E^{-156}$
haplotype2 – AAY32551.1 [*Triticum monococcum*]	98	$9.00E^{-155}$
haplotype2 – ABY59778.1 [*Secale cereale*]	76	$6.00E^{-89}$
haplotype2 – AAX28952.1 [*Hordeum vulgare*]	81	$2.00E^{-104}$
haplotype2 – ABK32847.1 [*Lolium perenne*]	67	$3.00E^{-41}$
haplotype2 – AAT44912.1 [*Arabidopsis thaliana*]	54	$2.00E^{-23}$
haplotype2 – ADL74429.1 [*Medicago truncatula*]	47	$1.00E^{-44}$
haplotype2 – NP_001234123.1 [*Solanum lycopersicum*]	67	$4.00E^{-44}$
haplotype2 – AIL00786.1 [*Vitis vinifera*]	44	$5.00E^{-43}$
haplotype3 – AFR67778.1 [*Triticum aestivum*]	99	$8.00E^{-157}$
haplotype3 – AAY32551.1 [*Triticum monococcum*]	98	$8.00E^{-155}$
haplotype3 – ABY59778.1 [*Secale cereale*]	76	$5.00E^{-89}$
haplotype3 – AAX28952.1 [*Hordeum vulgare*]	81	$2.00E^{-104}$
haplotype3 – ABK32847.1 [*Lolium perenne*]	67	$2.00E^{-41}$
haplotype3 – AAT44912.1 [*Arabidopsis thaliana*]	54	$1.00E^{-23}$
haplotype3 – ADL74429.1 [*Medicago truncatula*]	47	$9.00E^{-45}$
haplotype3 – NP_001234123.1 [*Solanum lycopersicum*]	67	$3.00E^{-44}$
haplotype3 – AIL00786.1 [*Vitis vinifera*]	44	$4.00E^{-43}$

Supplemental Table 7: (Continued)

Gene/AA comparison	Identity AA in %	e-value
CBF-A10		
haplotype1 – haplotype2	99	1.00E^{-177}
haplotype1 – AFR67783.1 [*Triticum aestivum*]	100	1.00E^{-177}
haplotype1 – AAY32554.1 [*Triticum monococcum*]	99	3.00E^{-134}
haplotype1 – ABY59780.1 [*Secale cereale*]	93	9.00E^{-148}
haplotype1 – AAX23710.1 [*Hordeum vulgare*]	89	1.00E^{-150}
haplotype1 – BAF36840.1 [*Lolium perenne*]	57	9.00E^{-75}
haplotype1 – AT4G25480.1 [*Arabidopsis thaliana*]	68	2.00E^{-37}
haplotype1 – AGR40686.1 [*Medicago truncatula*]	61	8.00E^{-38}
haplotype1 – NP_001234123.1 [*Solanum lycopersicum*]	56	1.00E^{-41}
haplotype1 – AIL00515.1 [*Vitis vinifera*]	39	1.00E^{-36}
haplotype2 – AFR67783.1 [*Triticum aestivum*]	99	6.00E^{-177}
haplotype2 – AAY32554.1 [*Triticum monococcum*]	99	4.00E^{-174}
haplotype2 – ABY59780.1 [*Secale cereale*]	93	4.00E^{-148}
haplotype2 – AAX23710.1 [*Hordeum vulgare*]	89	1.00E^{-150}
haplotype2 – BAF36840.1 [*Lolium perenne*]	57	4.00E^{-75}
haplotype2 – AT4G25480.1 [*Arabidopsis thaliana*]	68	1.00E^{-37}
haplotype2 – AGR40686.1 [*Medicago truncatula*]	61	7.00E^{-38}
haplotype2 – NP_001234123.1 [*Solanum lycopersicum*]	56	9.00E^{-42}
haplotype2 – AIL00515.1 [*Vitis vinifera*]	38	1.00E^{-36}
CBF-A13		
haplotype1 – haplotype2	98	2.00E^{-18}
haplotype1 – ABK55361.1 [*Triticum aestivum*]	94	5.00E^{-16}
haplotype1 – AAY32555.1 [*Triticum monococcum*]	98	1.00E^{-18}
haplotype1 – ABY59779.1 [*Secale cereale*]	69	1.00E^{-13}
haplotype1 – AAX23711.1 [*Hordeum vulgare*]	74	8.00E^{-16}
haplotype1 – ABK32848.1 [*Lolium perenne*]	63	3.00E^{-12}
haplotype1 – ABV27132.1 [*Arabidopsis thaliana*]	62	1.00E^{-10}
haplotype1 – AGR40686.1 [*Medicago truncatula*]	80	5.00E^{-10}
haplotype1 – NP_001234123.1 [*Solanum lycoper*sicum]	81	1.00E^{-09}
haplotype1 – AIL00786.1 [*Vitis vinifera*]	57	1.00E^{-08}
haplotype2 – ABK55361.1 [*Triticum aestivum*]	37	0.37
haplotype2 – AAY32555.1 [*Triticum monococcum*]	100	5.50E^{+01}
haplotype2 – ABY59779.1 [*Secale cereale*]	39	1.30
haplotype2 – AAX23711.1 [*Hordeum vul*gare]	34	0.16
haplotype2 – ABK32848.1 [*Lolium perenne*]	39	0.50
haplotype2 – ABV27132.1 [*Arabidopsis thaliana*]	36	1.50E^{+01}
haplotype2 – AGR40686.1 [*Medicago truncatula*]	27	n.a.
haplotype2 – NP_001234123.1 [*Solanum lycopersicum*]	33	1.20E^{+01}
haplotype2 – AIL00786.1 [*Vitis vinifera*]	29	7.10E^{+01}
CBF-A14		
haplotype1 – haplotype2	100	2.00E^{-160}
haplotype1 – ABK55380.1 [*Triticum aestivum*]	100	2.00E^{-160}
haplotype1 – AY32552.1 [*Triticum monococcum*]	99	2.00E^{-159}
haplotype1 – ADX32478.1 [*Secale cereale*]	92	4.00E^{-114}
haplotype1 – ABA01493.1 [*Hordeum vulgare*]	94	7.00E^{-151}
haplotype1 – BAF36844.1 [*Lolium perenne*]	68	3.00E^{-82}
haplotype1 – ABV27138.1 [*Arabidopsis thaliana*]	66	5.00E^{-46}

Supplemental Table 7: (Continued)

Gene/AA comparison	Identity AA in %	e-value
CBF-A14		
haplotype1 – AGR40686.1 [*Medicago truncatula*]	61	$2.00E^{-39}$
haplotype1 – NP_001234123.1 [*Solanum lycopersicum*]	54	$1.00E^{-43}$
haplotype1 – AIL00786.1 [*Vitis vinifera*]	72	$2.00E^{-38}$
haplotype2 – ABK55380.1 [*Triticum aestivum*]	100	$2.00E^{-160}$
haplotype2 – AY32552.1 [*Triticum monococcum*]	99	$2.00E^{-159}$
haplotype2 – ADX32478.1 [*Secale cereale*]	92	$4.00E^{-114}$
haplotype2 – ABA01493.1 [*Hordeum vulgare*]	94	$7.00E^{-151}$
haplotype2 – BAF36844.1 [*Lolium perenne*]	68	$2.00E^{-82}$
haplotype2 – ABV27138.1 [*Arabidopsis thaliana*]	64	$2.00E^{-43}$
haplotype2 – AGR40686.1 [*Medicago truncatula*]	61	$2.00E^{-39}$
haplotype2 – NP_001234123.1 [*Solanum lycopersicum*]	54	$1.00E^{-43}$
haplotype2 – AIL00786.1 [*Vitis vinifera*]	72	$2.00E^{-38}$
CBF-A15		
haplotype1 - haplotype2	97	$2.00E^{-174}$
haplotype1 – ABK55368.1 [*Triticum aestivum*]	100	0.00
haplotype1 – AAY32556.1 [*Triticum monococcum*]	99	$5.00E^{-179}$
haplotype1 – ABY59784.1 [*Secale cereale*]	92	$3.00E^{-131}$
haplotype1 – ACC63529.1 [*Hordeum vulgare*]	87	$6.00E^{-128}$
haplotype1 – BAF36840.1 [*Lolium perenne*]	68	$8.00E^{-96}$
haplotype1 - ABV27138.1 [*Arabidopsis thaliana*]	55	$2.00E^{-39}$
haplotype1 – AGR40686.1 [*Medicago truncatula*]	57	$7.00E^{-36}$
haplotype1 - NP_001234123.1 [*Solanum lycopersicum*]	41	$1.00E^{-41}$
haplotype1 - AIL00786.1 [*Vitis vinifera*]	41	$2.00E^{-38}$
haplotype2 – ABK55368.1 [*Triticum aestivum*]	97	$2.00E^{-174}$
haplotype2 – AAY32556.1 [*Triticum monococcum*]	98	$1.00E^{-176}$
haplotype2 – ABY59784.1 [*Secale cereale*]	91	$5.00E^{-130}$
haplotype2 – ACC63529.1 [*Hordeum vulgare*]	87	$9.00E^{-129}$
haplotype2 – BAF36840.1 [*Lolium perenne*]	69	$7.00E^{-99}$
haplotype2 - ABV27138.1 [*Arabidopsis thaliana*]	38	$2.00E^{-40}$
haplotype2 – AGR40686.1 [*Medicago truncatula*]	56	$3.00E^{-35}$
haplotype2 - NP_001234123.1 [*Solanum lycopersicum*]	40	$5.00E^{-42}$
haplotype2 - AIL00786.1 [*Vitis vinifera*]	41	$5.00E^{-40}$
CBF-A18		
haplotype1 - haplotype2	94	$9.00E^{-161}$
haplotype1 – AFR67780.1 [*Triticum aestivum*]	89	$1.00E^{-143}$
haplotype1 – AAY32550.1 [*Triticum monococcum*]	95	$2.00E^{-161}$
haplotype1 – ARK38717.1 [*Secale cereale*]	91	$4.00E^{-144}$
haplotype1 – ABO20921.1 [*Hordeum vulgare*]	63	$1.00E^{-81}$
haplotype1 – BAF36839.1 [*Lolium perenne*]	57	$8.00E^{-77}$
haplotype1 – ABV27090.1 [*Arabidopsis thaliana*]	59	$3.00E^{-41}$
haplotype1 – AGR40686.1 [*Medicago truncatula*]	58	$4.00E^{-39}$
haplotype1 - ASK51848.1 [*Solanum lycopersicum*]	44	$3.00E^{-49}$
haplotype1 - AIL00820.1 [*Vitis vinifera*]	43	$3.00E^{-43}$
haplotype2 – AFR67780.1 [*Triticum aestivum*]	89	$9.00E^{-146}$
haplotype2 – AAY32550.1 [*Triticum monococcum*]	98	0.00
haplotype2 –ARK38717.1 [*Secale cereale*]	90	$1.00E^{-142}$
haplotype2 – ABO20921.1 [*Hordeum vulgare*]	61	$2.00E^{-82}$

Supplemental Table 7: (Continued)

Gene/AA comparison	Identity AA in %	e-value
CBF-A18		
haplotype2 – BAF36839.1 [*Lolium perenne*]	57	$4.00E^{-75}$
haplotype2 – ABV27090.1 [*Arabidopsis thaliana*]	59	$2.00E^{-41}$
haplotype2 – AGR40686.1 [*Medicago truncatula*]	58	$2.00E^{-39}$
haplotype2 - ASK51848.1 [*Solanum lycopersicum*]	44	$8.00E^{-49}$
haplotype2 - AIL00820.1 [*Vitis vinifera*]	44	$2.00E^{-42}$
VRN-A1		
haplotype1 - haplotype2	99	0.00
haplotype1 - haplotype3	99	0.00
haplotype2 - haplotype3	99	0.00
haplotype1 – AAW73219.1 [*Triticum turgidum*]	100	0.00
haplotype1 – AAO72630.1 [*Triticum monococcum*]	99	0.00
haplotype1 – ACD74577.1 [*Secale cereale*]	96	$1.00E^{-153}$
haplotype1 – AAW82994.1 [*Hordeum vulgare*]	95	$4.00E^{-148}$
haplotype1 – AEV22381.1 [*Lolium perenne*]	87	$9.00E^{-145}$
haplotype1 – AAM65504.1 [*Arabidopsis thaliana*]	68	$3.00E^{-94}$
haplotype1 – XP_003628742.1 [*Medicago truncatula*]	57	$1.00E^{-91}$
haplotype1 - NP_001294867.1 [*Solanum lycopersicum*]	60	$1.00E^{-101}$
haplotype1 - NP_001268210.1 [*Vitis vinifera*]	62	$4.00E^{-104}$
haplotype2 – AAW73219.1 [*Triticum turgidum*]	99	0.00
haplotype2 – AAO72630.1 [*Triticum monococcum*]	99	0.00
haplotype2 – ACD74577.1 [*Secale cereale*]	96	$7.00E^{-173}$
haplotype2 – AAW82994.1 [*Hordeum vulgare*]	95	$5.00E^{-148}$
haplotype2 – AEV22381.1 [*Lolium perenne*]	87	$7.00E^{-145}$
haplotype2 – AAM65504.1 [*Arabidopsis thaliana*]	68	$4.00E^{-94}$
haplotype2 – XP_003628742.1 [*Medicago truncatula*]	57	$1.00E^{-91}$
haplotype2 - NP_001294867.1 [*Solanum lycopersicum*]	61	$1.00E^{-101}$
haplotype2 - NP_001268210.1 [*Vitis vinifera*]	63	$8.00E^{-104}$
haplotype3 – AAW73219.1 [*Triticum turgidum*]	99	0.00
haplotype3 – AAO72630.1 [*Triticum monococcum*]	99	0.00
haplotype3 – ACD74577.1 [*Secale cereale*]	95	$2.00E^{-150}$
haplotype3 – AAW82994.1 [*Hordeum vulgare*]	94	$1.00E^{-145}$
haplotype3 – AEV22381.1 [*Lolium perenne*]	87	$2.00E^{-142}$
haplotype3 – AAM65504.1 [*Arabidopsis thaliana*]	67	$1.00E^{-91}$
haplotype3 – XP_003628742.1 [*Medicago truncatula*]	56	$3.00E^{-89}$
haplotype3 - NP_001294867.1 [*Solanum lycopersicum*]	60	$3.00E^{-99}$
haplotype3 - NP_001268210.1 [*Vitis vinifera*]	63	$2.00E^{-101}$
VRN-B3		
haplotype1 - haplotype2	99	$1.00E^{-134}$
haplotype1 – ABK32205.1 [*Triticum aestivum*]	99	$2.00E^{-134}$
haplotype1 – ABK32206.1 [*Triticum monococcum*]	99	$2.00E^{-134}$
haplotype1 – ADR51710.1 [*Secale cereale*]	97	$3.00E^{-121}$
haplotype1 – ACA29356.1 [*Hordeum vulgare*]	98	$7.00E^{-133}$
haplotype1 – AMB21803.1 [*Lolium perenne*]	97	$1.00E^{-131}$
haplotype1 – NP_001320342.1 [*Arabidopsis thaliana*]	70	$5.00E^{-99}$
haplotype1 – XP_003624569.1 [*Medicago truncatula*]	73	$7.00E^{-99}$
haplotype1 - NP_001308081.1 [*Solanum lycopersicum*]	83	$3.00E^{-110}$
haplotype1 - ABF56526.1 [*Vitis vinifera*]	82	$1.00E^{-110}$

Supplemental Table 7: (Continued)

Gene/AA comparison	Identity AA in %	e-value
VRN-B3		
haplotype2 – ABK32205.1 [*Triticum aestivum*]	100	$2.00E^{-135}$
haplotype2 – ABK32206.1 [*Triticum monococcum*]	100	$2.00E^{-135}$
haplotype2 – ADR51710.1 [*Secale cereale*]	97	$3.00E^{-121}$
haplotype2 – ACA29356.1 [*Hordeum vulgare*]	99	$7.00E^{-134}$
haplotype2 – AMB21803.1 [*Lolium perenne*]	97	$1.00E^{-132}$
haplotype2 – NP_001320342.1 [*Arabidopsis thaliana*]	71	$4.00E^{-100}$
haplotype2 – XP_003624569.1 [*Medicago truncatula*]	73	$7.00E^{-99}$
haplotype2 - NP_001308081.1 [*Solanum lycopersicum*]	84	$4.00E^{-111}$
haplotype2 - ABF56526.1 [*Vitis vinifera*]	82	$2.00E^{-111}$
PPD-B1		
haplotype1 - haplotype2	99	0.00
haplotype1 - haplotype3	99	0.00
haplotype1 - haplotype4	99	0.00
haplotype2 - haplotype3	99	$3.00E^{-121}$
haplotype2 - haplotype4	99	$7.00E^{-133}$
haplotype3 - haplotype4	99	$1.00E^{-131}$
haplotype1 – ABL09468.1 [*Triticum aestivum*]	100	0.00
haplotype1 – BAM93386.1 [*Triticum monococcum*]	94	0.00
haplotype1 – ADR51713.1 [*Secale cereale*]	86	0.00
haplotype1 – ACL78560.1 [*Hordeum vulgare*]	89	0.00
haplotype1 – BAD38855.1 [*Oryza sativa*]	65	0.00
haplotype1 – AT5G02810.1 [*Arabidopsis thaliana*]	40	$5.00E^{-109}$
haplotype1 – XP_013468266.1 [*Medicago truncatula*]	43	$5.00E^{-142}$
haplotype1 – XP_004237533.1 [*Solanum lycopersicum*]	43	$5.00E^{-145}$
haplotype1 – XP_010658157.1 [*Vitis vinifera*]	45	$2.00E^{-143}$
haplotype2 – ABL09468.1 [*Triticum aestivum*]	99	0.00
haplotype2 – BAM93386.1 [*Triticum monococcum*]	94	0.00
haplotype2 – ADR51713.1 [*Secale cereale*]	86	0.00
haplotype2 – ACL78560.1 [*Hordeum vulgare*]	89	0.00
haplotype2 – BAD38855.1 [*Oryza sativa*]	65	0.00
haplotype2 – AT5G02810.1 [*Arabidopsis thaliana*]	40	$7.00E^{-110}$
haplotype2 – XP_013468266.1 [*Medicago truncatula*]	43	$5.00E^{-142}$
haplotype2 – XP_004237533.1 [*Solanum lycopersicum*]	43	$5.00E^{-145}$
haplotype2 – XP_010658157.1 [*Vitis vinifera*]	45	$2.00E^{-143}$
haplotype3 – ABL09468.1 [*Triticum aestivum*]	99	0.00
haplotype3 – BAM93386.1 [*Triticum monococcum*]	94	0.00
haplotype3 – ADR51713.1 [*Secale cereale*]	86	0.00
haplotype3 – ACL78560.1 [*Hordeum vulgare*]	89	0.00
haplotype3 – BAD38855.1 [*Oryza sativa*]	65	0.00
haplotype3 – AT5G02810.1 [*Arabidopsis thaliana*]	40	$3.00E^{-109}$
haplotype3 – XP_013468266.1 [*Medicago truncatula*]	43	$2.00E^{-142}$
haplotype3 – XP_004237533.1 [*Solanum lycopersicum*]	43	$3.00E^{-145}$
haplotype3 – XP_010658157.1 [*Vitis vinifera*]	45	$1.00E^{-143}$
haplotype4 – ABL09468.1 [*Triticum aestivum*]	99	0.00
haplotype4 – BAM93386.1 [*Triticum monococcum*]	94	0.00
haplotype4 – ADR51713.1 [*Secale cereale*]	86	0.00
haplotype4 – ACL78560.1 [*Hordeum vulgare*]	89	0.00

Supplemental Table 7: (Continued)

Gene/AA comparison	Identity AA in %	e-value
PPD-B1		
haplotype4 – BAD38855.1 [*Oryza sativa*]	65	0.00
haplotype4 – AT5G02810.1 [*Arabidopsis thaliana*]	40	$8.00E^{-110}$
haplotype4 – XP_013468266.1 [*Medicago truncatula*]	43	$6.00E^{-143}$
haplotype4 – XP_004237533.1 [*Solanum lycopersicum*]	43	$9.00E^{-146}$
haplotype4 – XP_010658157.1 [*Vitis vinifera*]	45	$3.00E^{-144}$
PPD-D1		
haplotype1 - haplotype2	99	0.00
haplotype1 - haplotype3	97	0.00
haplotype2 - haplotype3	97	0.00
haplotype1 – ABL09477.1 [*Triticum aestivum*]	99	0.00
haplotype1 – BAM93381.1 [*Triticum monococ*cum]	97	0.00
haplotype1 – ADR51713.1 [*Secale cereale*]	87	$4.00E^{-87}$
haplotype1 – ACL78560.1 [*Hordeum vulgare*]	91	0.00
haplotype1 – ALT32070.1 [*Lolium perenne*]	38	$2.00E^{-20}$
haplotype1 – NP_568107.1 [*Arabidopsis thaliana*]	43	$2.00E^{-83}$
haplotype1 – XP_013468266.1 [*Medicago truncatula*]	51	$3.00E^{-105}$
haplotype1 – XP_004237533.1 [*Solanum lycopersicum*]	46	$4.00E^{-97}$
haplotype1 – XP_010658157.1 [*Vitis vinifera*]	48	$2.00E^{-104}$
haplotype2 – ABL09477.1 [*Triticum aestivum*]	99	0.00
haplotype2 – BAM93381.1 [*Triticum monococcum*]	97	0.00
haplotype2 – ADR51713.1 [*Secale cereale*]	86	0.00
haplotype2 – ACL78560.1 [*Hordeum vulgare*]	89	0.00
haplotype2 – ALT32070.1 [*Lolium perenne*]	38	$4.00E^{-21}$
haplotype2 – NP_568107.1 [*Arabidopsis thaliana*]	42	$1.00E^{-90}$
haplotype2 – XP_013468266.1 [*Medicago trunca*tula]	44	$2.00E^{-140}$
haplotype2 – XP_004237533.1 [*Solanum lycopersicum*]	44	$4.00E^{-144}$
haplotype2 – XP_010658157.1 [*Vitis vinifera*]	45	$1.00E^{-147}$
haplotype3 – ABL09477.1 [*Triticum aestivum*]	97	0.00
haplotype3 – BAM93381.1 [*Triticum monococcum*]	98	0.00
haplotype3 – ADR51713.1 [*Secale cereale*]	90	0.00
haplotype3 – ACL78560.1 [*Hordeum vulgare*]	89	0.00
haplotype3 – ALT32070.1 [*Lolium perenne*]	38	$4.00E^{-21}$
haplotype3 – NP_568107.1 [*Arabidopsis thaliana*]	39	$6.00E^{-109}$
haplotype3 – XP_013468266.1 [*Medicago truncatula*]	44	$5.00E^{-142}$
haplotype3 – XP_004237533.1 [*Solanum lycopersicum*]	44	$1.00E^{-144}$
haplotype3 – XP_010658157.1 [*Vitis vinifera*]	45	$9.00E^{-144}$

AA: amino acid

Supplementary Data

Supplemental Data 1: Java script used to extract differences from a multiple sequence alignment (MSA).

```
import java.io.BufferedReader;
import java.io.BufferedWriter;
import java.io.FileReader;
import java.io.FileWriter;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Collections;
/**
 * This small Java class is used to extract differences from a multiple sequence alignment (MSA).
 */
public class MSAChecker2 {

	public static void main(String[] args) throws Exception {
		if( !(args.length == 7 || args.length == 5) ){
			System.out.println("Parameter: <input> <p> <ignoreN> <ignore> <tolerance> <polymorphism-output> [<minDifference> <fasta-output>]");
		}
		BufferedReader r = new BufferedReader( new FileReader( args[0]) );
		String line = r.readLine();
		StringBuffer seq = new StringBuffer();
		ArrayList<String> annot = new ArrayList<String>();
		ArrayList<char[]> sym = new ArrayList<char[]>();
		int l = -1, e, s;
		ArrayList<Integer> start = new ArrayList<Integer>();
		ArrayList<Integer> end = new ArrayList<Integer>();
		do {
			annot.add( line.substring(1) );
			seq.delete(0, seq.length());
			while( ((line=r.readLine()) != null) && line.charAt(0) != '>' ) {
				seq.append(line);
			}
			if( l == -1 ) {
				l = seq.length();
			} else {
				if( l != seq.length() ) {
					throw new Exception("Sequences have different length. Probably not aligned.");
				}
			}
			char[] c = new char[l];
			seq.getChars(0, l, c, 0);

			//represent upstream and downstream gap by blanks
			s = 0;
			while( s < c.length && c[s] == '-' ) {
				c[s]=' ';
				s++;
			}
			start.add(s);
			e = c.length-1;
			while( e>= 0 && c[e] == '-' ) {
```

```
                                c[e]=' ';
                                e--;
                        }
                        end.add(e);
                        sym.add( c );
                } while( line!=null );
                r.close();
                //TODO
                Collections.sort(start);
                Collections.sort(end);
                System.out.println( start );
                System.out.println( end );
                double p = Double.parseDouble(args[1]);
                int globalStart = start.get( (int)Math.ceil((start.size()-1)*p) );
                int globalEnd=end.get( (int)Math.floor((end.size()-1)*(1-p)) );
                System.out.println(l + "\t" + globalStart + "\t" + globalEnd );
                boolean ignoreN = Boolean.parseBoolean( args[2] );
                int ignore = Integer.parseInt( args[3] );
                if( globalEnd < globalStart ) {
                        throw new IllegalArgumentException( "There is no common central part of the
alignment, i.e. there is no region which includes sequence information of all sequences. [" + globalStart + ","
+ globalEnd + "]" );
                }
                char d;
                boolean[] conserved = new boolean[globalEnd-globalStart+1];
                char[] consensus = new char[conserved.length];
                boolean set;
                for( int i = 0; i < conserved.length; i++) {
                        conserved[i] = true;
                        set = false;
                        consensus[i] = '!';
                        for( int j = 0; j < annot.size(); j++) {
                                if( j != ignore ) {
                                        d = sym.get(j)[globalStart+i];
                                        if( d==' ' || (ignoreN && (d=='n' || d=='N')) ) {
                                                continue;
                                        }
                                        if( !set ) {
                                                consensus[i] = d;
                                                set = true;
                                        } else {
                                                if( d != consensus[i] ) {
                                                        /*
                                                        System.out.println(i);
                                                        for( int k = 0; k < sym.size(); k++ ) {
                                                                System.out.print(
sym.get(k)[globalStart+i] + "\t" );
                                                        }
                                                        System.out.println();/**/
                                                        conserved[i]=false;
                                                        consensus[i]='N';
                                                        break;
                                                }
                                        }
                                }
                        }
                }
```

```
int tolerance = Integer.parseInt(args[4]);
BufferedWriter w = new BufferedWriter( new FileWriter(args[5]));
w.append("position");
for( int k = 0; k < annot.size(); k++ ) {
	w.append("\t"+annot.get(k));
}
w.newLine();
ArrayList<int[][]> pos = new ArrayList<int[][]>();
int[] gaps = new int[sym.size()];
for( int i = 0; i < conserved.length; i++) {
	e = i;
	boolean next;
	do {
		while( e < conserved.length && !conserved[e] ) {
			e++;
		}
		next=false;
		if( e != i ) {
			for( int k = 1; k <= tolerance && e+k < conserved.length; k++ ) {
				if( !conserved[e+k] ) {
					next=true;
					e+=k;
					break;
				}
			}
		}
	} while( next );
	if( i != e ) {
		w.write( ""+(globalStart+i) );
		if( i+1 != e ) {
			w.append( "-"+(globalStart+e-1) );
		}
		int gap=0;
		for( int k = 0; k < sym.size(); k++ ) {
			gaps[k] = 0;
			w.append("\t");
			char[] x = sym.get(k);
			for( int j=i; j < e; j++ ) {
				char c = x[globalStart+j];
				if( conserved[j] ) {
					w.append(Character.toUpperCase( c ));
				} else {
					w.append(Character.toLowerCase( c ));
				}
				if( c=='-') {
					gaps[k]++;
				}
			}
			gap+=gaps[k];
		}
		w.newLine();
		//if( gap > 0 )//TODO
		{
			pos.add(new int[][]{{i,e},gaps.clone()});
		}
	}
	i = e;
```

```
                }
                w.close();
                if( args.length > 6 ) {
                        int minDist = Integer.parseInt(args[6]);
                        w = new BufferedWriter( new FileWriter(args[7]));
                        s=0;
                        for( int k = 0; k < pos.size(); k++ ) {
                                if( k+1 < pos.size() ) {
                                        e = pos.get(k+1)[0][0];
                                } else {
                                        e = consensus.length;
                                }
                                int[][] y = pos.get(k);
                                int max = 0, min=Integer.MAX_VALUE;
                                for( int j = 0; j < y[1].length; j++ ) {
                                        if( y[1][j] < min ) {
                                                min = y[1][j];
                                        }
                                        if( y[1][j] > max ) {
                                                max = y[1][j];
                                        }
                                }
                                if( max - min >= minDist ) {
                                        w.append(">alignment_"+(globalStart+s)+"-"+(globalStart+e-1)
+ "_" + (globalStart+y[0][0])+"-"+(globalStart+y[0][1]-1) + "_" + Arrays.toString(y[1]).replaceAll(" ","") );
                                        w.newLine();
                                        w.append(
                                                        new String(consensus, s, y[0][0]-s )
                                                        +  "["+new  String(consensus,  y[0][0],  y[0][1]-
y[0][0] ) + "]"
                                                        + new String(consensus, y[0][1], e-y[0][1] )
                                        );
                                        w.newLine();
                                }
                                s=y[0][1];
                        }
                        w.close();
                }
        }
}
```

List of abbreviations

AA	amino acid
ANCOVA	analysis of covariance
AP2/EREBP	APETALA2/ETHYLENE RESPONSE ELEMENT BINDING PROTEIN
as	antisense
BLAST	basic local alignment search tool
BLT	BARLEY LOW TEMPERATURE
BP	base pairs
CAB	CALCIUM-BINDING
CAPS	cleaved amplified polymorphic sequence
CBP	CALCIUM BINDING PROTEIN
CBF	C-REPEAT BINDING FACTOR
CDS	coding DNA sequence
CML	CAM-like
CNV	copy number variation
CO	CONSTANS
COR	cold-Responsive
COR/LEA	COLD RESPONSIVE/LATE EMBRYOGENESIS-ABUNDANT
CRT/DRE	C-repeat/dehydration-responsive element
CRISPR	clustered regularly interspaced short palindromic repeat
CV	coefficient of variation
DArT	diversity arrays technology
DEM	DEFECTIVE-EMBRYO AND MERISTEMS
DH	doubled haploid
DHN	DEHYDRIN
DNA	deoxyribonucleic acid
dN/dS	nucleotide divergence rate
EST	expressed sequence tag
FR	frost resistance
FT	frost tolerance
Gbp	Giga-base pairs
GDT	Global Distance Test
GLM	general linear model
GWAS	genome-wide association study
Hd	haplotype diversity
ICE	INDUCER OF CBF EXPRESSION
IFVCNS	Institute of Field and Vegetable Crops
INDEL	insertion-deletion
IT	intron targeting
IWGSC	international wheat genome sequencing consortium

KASP	competitive allele-specific PCR
kbp	kilo-base pairs
LD	linkage disequilibrium
LS	least-Squares
LT	LOW TEMPERATURE-RESPONSIVE
M	million
MAF	minor allele frequency
MAFFT	Multiple Alignment tool Fast Fourier Transform
MAS	marker assisted selection
MCMC	Markov Chain Monte Carlo
MEF	myocyte enhancer factor
MLM	mixed linear model
mRNA	messenger RNA
MSA	multiple sequence alignment
NCBI	National Center for Biotechnology Information
NT	nulli-tetrasomic
PCA	principal component analysis
PCOA	principal coordinate analysis
PCR	polymerase chain reaction
PPD	PHOTOPERIOD
PRR	PSEUDO RESPONSE REGULATOR
QTL	quantitative trait locus
RAB	RESPONSIVE TO ABSCISIC ACID
RNA	ribonucleic acid
RV	reaction volume
s	sense
SeqID	uSeqID normalized by the protein (or domain) sequence length and multiplied by 100
SNP	single nucleotide polymorphism
STS	sequence tagged sites
T	tonne
TACR7	TRITICUM AESTIVUM COLD COLD-REGULATED 7
TALEN	transcription activator-like effector nuclease
URGI	Unité de Recherche Génomique Info
UTR	untranslated region
K	number of populations
U	units
uGDT	unnormalised Global Distance Test
uSeqID	number of identical residues in the alignment
VRN	VERNALISATION
WLT	WHEAT LOW TEMPERATURE

German abstract

Das Verständnis der genetischen Struktur von hexaploidem Weizen (*Triticum aestivum* L.) ist elementar für die Entwicklung nützlicher und effektiver Methoden für genetische Analysen. In unterschiedlichen Regionen verursacht Frost Zellschäden, Austrocknung und reduzierten Metabolismus. Daher ist die Untersuchung der Frosttoleranz (FT) von Weizen wesentlich, um Ertragsverluste zu verhindern. Für solch ein komplexes Merkmal (reguliert durch eine Vielzahl von Genen und mehreren Genfamilien) und komplexe Spezies (kontrolliert durch drei Genomen) ist die Verfügbarkeit der genomischen Weizensequenzen und deren Verwendung zur Erforschung von Kandidatengenen von besonderem Interesse. Daher waren die Ziele dieser Studie die Entwicklung genomspezifischer Primer von FT-Kandidatengenen und deren Sequenzierung, die Identifikation von Polymorphismen und eine kandidatengenbasierte Assoziationsanalyse. Mit einer Erfolgsrate von >50% wurden 39 spezifische Primerpaare, welche 19 Kandidatengene entsprechen, entwickelt. *C-REPEAT-BINDING FACTOR* (*CBF*) Gene (*CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15* und *CBF-A18*), *VERNALISATION* Gene (*VRN-A1* und *VRN-B3*) und *PHOTOPERIOD* Gene (*PPD-B1* und *PPD-D1*) zeigten Assoziationen im Bezug auf FT anhand von 235 Weizensorten. Sechs dieser Gene zeigen Aminosäureaustausche in essentiellen Proteindomänen. Der Effekt des Aminosäureaustausches in *VRN-A1* wurde in der Literatur bereits beschrieben. Unbekannt waren jedoch die Aminosäureaustausche in *CBF-A3*, *CBF-A5*, *VRN-B3*, *PPD-B1* und *PPD-D1*, welche an hochkonservierten Positionen lokalisiert sind. Die Ergebnisse zeigen eine effektive Methode zum Ableiten von genom- und ortsspezifischen Primern in hexaploidem Weizen. Weiterhin wurden Aminosäureaustauschen in sechs FT-Genen identifiziert, die in der markergestützten Züchtung zur Entwicklung toleranterer Sorten genutzt werden können.

Acknowledgements

I am very grateful to my advisors Frank Ordon and Klaus Pillen. They guided me through my Bachelor and Master thesis and finally through my PhD thesis. Additionaly a heartfelt thanks to my co-supervisor Dragan Perovic, who guided me also through my Master and PhD thesis. All three gave me the freedom, suggestions and support to perform this presented PhD thesis. I am grateful for the expedient and instructive discussions. Finally, I can say, that this support made me a better scientist and cleared my vision for further tasks.

Furthermore I would like to thank Marcel Quint for giving me time, support and encouragement to finish my PhD thesis in my subsequent scientific work in the team of the department of Crop Physiology at the Martin Luther University Halle-Wittenberg.

Concerning the part of publications, which orginated from my PhD work, I would like to thank Michael Koch (Babben et al., 2015 and Babben et al., 2018), Edgar Schliephake, Philipp Janitza, Thomas Berner, Jens Keilwagen, Fernando Alberto Arana-Ceballos, Sven Eduard Templer, Yuriy Chesnokov, Tatyana Pshenichnikova, Jörg Schondelmaier and Andreas Börner (Babben et al., 2018) for the great collaboration.

All my research was realised at the Institut for Resistance Research and Stress Tolerance (RS) of the Julius Kühn-Institut (JKI) Federal Research Centre for Cultivated Plants. I would like to thank all members of the institute RS for support and specially Katy Niedung for excellent technical assistance. I would like to thank Deutsche Saatveredelung AG (DSV), Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Agrophysical Research Institute (AFI), Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences and German Seed Aliance (GSA) for the execution of the field experiments. I thank also the German Federal Ministry of Education and Research (BMBF) for funding the project FROWHEAT (0315953).

Finally, I would like to pronounce the deepest felt of graduation to my family and specially to my parents for the greatest support of my life. Without their support I would never have gotten this far. And at the end I want to thank my friends very much, with whom I can discuss any scientific problems, especially Konstantin Kahlo.

Curriculum vitae

Personal Data

Birthday	05 May 1984 in Finsterwalde, Germany
Address	Reilstraße 51, 06114 Halle (Saale), Germany
Nationality	German
Mobile	+49 175 5841777
E-mail	monstera@gmx.net

Education

Oct 2009 – Jan 2012	**Martin Luther University Halle-Wittenberg**
	Master of Science, Crop Sciences
	Halle (Saale), Germany
Master thesis	Creation of a barley NAM population in the BC_1S_0 generation and their genetic characterisation by using SSR marker
Oct 2006 – Sep 2009	**Martin Luther University Halle-Wittenberg**
	Bachelor of Science, Agricultural Sciences
	Halle (Saale), Germany
Bachelor thesis	Candidate gene mapping of *Soilborne Cereal Mosaic Virus* (SBCMV) resistance in wheat
Oct 2004 – Sep 2006	**Martin Luther University Halle-Wittenberg**
	Diploma, Bioengineering
	Halle (Saale), Germany
Aug 1996 – Jul 2004	**Städtische Gesamtschule Finsterwalde mit gymnasialer Oberstufe**
	Abitur
	Finsterwalde, Germany

Research Experience

Oct 2015 – present **Research Associate**

Martin Luther University Halle-Wittenberg, Institute of Agricultural and Nutritional Sciences, Halle (Saale), Germany

Research Topics: Thermomorphogenic root elongation response in Arabidopsis thaliana: a mutation approach

Lab – Marcel Quint

Jan 2012 – Mar 2015 **Research Associate**

Julius Kühn-Institut, Institute for Resistance Research and Stress Tolerance, Quedlinburg, Germany

Research topics: Evaluation of wheat prebreeding germplasm for frost tolerance via a genome wide and candidate gene based approach: FROWHEAT.

Lab – Frank Ordon

Teaching Experience

Oct 2016 – present **Master Lectures and Exercises**, Martin Luther University Halle-Wittenberg, Germany

Module: Pflanzenbiotechnologie

Apr 2016 – present **Bachelor Exercises and Lecture**, Martin Luther University Halle-Wittenberg, Germany

Module: Biologie der Nutzpflanzen

Oct 2015 – present **Master Exercises**, Martin Luther University Halle-Wittenberg, Germany

Module: Ertragsphysiologie

Scientific Societies

German Society for Plant Breeding (GPZ)

Date _____ Signature of the applicant _____

List of publications

BABBEN, **S**., PEROVIC, D., KOCH, M. & ORDON, F. 2015. An Efficient Approach for the Development of Locus Specific Primers in Bread Wheat (*Triticum aestivum* L.) and Its Application to Re-Sequencing of Genes Involved in Frost Tolerance. *Plos One,* 10**,** e0142746.

BABBEN, **S**., SCHLIEPHAKE, E., JANITZA, P., BERNER, T., KEILWAGEN, J., KOCH, M., ARANA-CEBALLOS, F. A., TEMPLER, S. E., CHESNOKOV, Y., PSHENICHNIKOVA, T., SCHONDELMAIER, J., BÖRNER, A., PILLEN, K., ORDON, F. & PEROVIC, D. 2018. Association genetics studies on frost tolerance in wheat (*Triticum aestivum* L.) reveal new highly conserved amino acid substitutions in CBF-A3, CBF-A15, VRN3 and PPD1 genes. *BMC Genomics,* 19**,** 409.

KEILWAGEN, J., KILIAN, B., OZKAN, H., **BABBEN**, **S**., PEROVIC, D., MAYER, K. F. X., WALTHER, A., POSKAR, C. H., ORDON, F., EVERSOLE, K., BOERNER, A., GANAL, M., KNUEPFFER, H., GRANER, A. & FRIEDEL, S. 2014. Separating the wheat from the chaff - a strategy to utilize plant genetic resources from ex situ genebanks. *Scientific Reports,* 4.

Eidesstatliche Erklärung / *Declaration under Oath*

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

Datum / *Date* Unterschrift des Antragstellers / *Signature of the applicant*

Teile dieser Arbeit sind veröffentlicht in:

BABBEN, S., PEROVIC, D., KOCH, M. & ORDON, F. 2015. An Efficient Approach for the Development of Locus Specific Primers in Bread Wheat (*Triticum aestivum* L.) and Its Application to Re-Sequencing of Genes Involved in Frost Tolerance. *Plos One,* 10**,** e0142746.

BABBEN, S., SCHLIEPHAKE, E., JANITZA, P., BERNER, T., KEILWAGEN, J., KOCH, M., ARANA-CEBALLOS, F. A., TEMPLER, S. E., CHESNOKOV, Y., PSHENICHNIKOVA, T., SCHONDELMAIER, J., BÖRNER, A., PILLEN, K., ORDON, F. & PEROVIC, D. 2018. Association genetics studies on frost tolerance in wheat (*Triticum aestivum* L.) reveal new highly conserved amino acid substitutions in CBF-A3, CBF-A15, VRN3 and PPD1 genes. *BMC Genomics,* 19**,** 409.