

RESEARCH ARTICLE

Open Access



# 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

Steve Babben<sup>1,2</sup>, Edgar Schliephake<sup>1</sup>, Philipp Janitza<sup>2</sup>, Thomas Berner<sup>3</sup>, Jens Keilwagen<sup>3</sup>, Michael Koch<sup>4</sup>, Fernando Alberto Arana-Ceballos<sup>5</sup>, Sven Eduard Templer<sup>1,6</sup>, Yuriy Chesnokov<sup>7</sup>, Tatyana Pshenichnikova<sup>8</sup>, Jörg Schondelmaier<sup>9</sup>, Andreas Börner<sup>5</sup>, Klaus Pillen<sup>10</sup>, Frank Ordon<sup>1</sup> and Dragan Perovic<sup>1\*</sup> 

## Abstract

**Background:** Understanding the genetic basis of frost tolerance (FT) in wheat (*Triticum aestivum* L.) is essential for preventing yield losses caused by frost due to cellular damage, dehydration and reduced metabolism. FT is a complex trait regulated by a number of genes and several gene families. Availability of the wheat genomic sequence opens new opportunities for exploring candidate genes diversity for FT. Therefore, the objectives of this study were to identify SNPs and insertion-deletion (indels) in genes known to be involved in frost tolerance and to perform association genetics analysis of respective SNPs and indels on FT.

**Results:** Here we report on the sequence analysis of 19 candidate genes for FT in wheat assembled using the Chinese Spring IWGSC RefSeq v1.0. Out of these, the tandem duplicated C-repeat binding factors (*CBF*), i.e. *CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *CBF-A18*, the vernalisation response gene *VRN-A1*, *VRN-B3*, the photoperiod response genes *PPD-B1* and *PPD-D1* revealed association to FT in 235 wheat cultivars. Within six genes (*CBF-A3*, *CBF-A15*, *VRN-A1*, *VRN-B3*, *PPD-B1* and *PPD-D1*) amino acid (AA) substitutions in important protein domains were identified. The amino acid substitution effect in *VRN-A1* on FT was confirmed and new AA substitutions in *CBF-A3*, *CBF-A15*, *VRN-B3*, *PPD-B1* and *PPD-D1* located at highly conserved sites were detected. Since these results rely on phenotypic data obtained at five locations in 2 years, detection of significant associations of FT to AA changes in *CBF-A3*, *CBF-A15*, *VRN-A1*, *VRN-B3*, *PPD-B1* and *PPD-D1* may be exploited in marker assisted breeding for frost tolerance in winter wheat.

**Conclusions:** A set of 65 primer pairs for the genes mentioned above from a previous study was BLASTed against the IWGSC RefSeq resulting in the identification of 39 primer combinations covering the full length of 19 genes. This work demonstrates the usefulness of the IWGSC RefSeq in specific primer development for highly conserved gene families in hexaploid wheat and, that a candidate gene association genetics approach based on the sequence data is an efficient tool to identify new alleles of genes important for the response to abiotic stress in wheat.

**Keywords:** *Triticum aestivum* L., Frost tolerance (FT), Candidate genes, Association studies, SNP, Indel, Haplotypes, *CBF*, *VRN*, *PPD1*

\* Correspondence: [dragan.perovic@julius-kuehn.de](mailto:dragan.perovic@julius-kuehn.de)

<sup>1</sup>Julius Kühn-Institut (JKI), Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Erwin-Baur-Str. 27, 06484 Quedlinburg, Saxony-Anhalt, Germany

Full list of author information is available at the end of the article



## Background

Wheat (*Triticum aestivum* L.) is on the world wide level the crop grown on the largest acreage and is of prime importance for human nutrition and animal feed. Worldwide production of 729 million tones in 2014, ranks wheat globally the third most important crop, next to maize (*Zea mays* L.) and rice (*Oryza sativa* L.). Production of wheat in Europe was 249.13 mt with an average yield of 4.25 t/ha in 2014. At the same time in North America 84.68 mt of wheat were harvested with an average yield of 2.994 t/ha, while in Asia the average yield was 3.095 t/ha resulting in a production of 315.71 mt [1]. In North America, North and Eastern Europe and Russia wheat production is exposed to low temperature and frost damage occurs frequently. Depending of the time of sowing wheat is known as winter or spring wheat, respectively, which differ in the vernalisation requirement and frost hardiness. Winter wheat varieties require an extended exposure to cold temperatures, typically 4 to 8 weeks at less than 4 °C, to induce flowering (vernalization) while spring wheat varieties do not have such a requirement [2]. Vernalization requirements keep these plants safely dormant over winter resulting in a longer vegetation period and higher yields of winter wheat.

### Wheat genome

Bread wheat is an allohexaploid species ( $2n = 6 \times = 42$ ) with an AABBDD genome derived from two independent hybridisations [3–5]. This complex genome of about 17 Giga-base pairs (Gbp) has a repeat content of approximately 80% [6] with a gene density of on average one gene per 87 to 184 Kilo-base pairs (Kbp) [7]. The acquisition of this large genome sequence became feasible since the introduction of next Generation Sequencing (NGS) technologies.

Nowadays, the efforts of the International Wheat Genome Sequencing Consortium (IWGSC), in the development of a physical map and a reference sequence facilitates many downstream applications, i.e. development of high throughput genotyping platforms [8], efficient development of genome specific primers [9], exome sequencing in large genebank collections [10] etc. The wheat reference sequence, survey sequence and physical map are available at many public data bases. The CerealsDB web page, created by members of the Functional Genomics Group at the University of Bristol [11], includes online resources of genomic information, i.e. varietal SNPs, DArT markers, and EST sequences all linked to a draft genome sequence of the cultivar Chinese Spring [12]. The Unité de Recherche Génomique Info (URGI) web portal includes datasets such as chromosome survey sequences, reference sequences, physical maps, genetic maps, polymorphisms, genetic

resources, extensive phenotypic data and various genomic arrays [13]. The chromosomal sequence information is available from the IWGSC [14]. All mentioned databases are useful for the identification of homologous chromosome sequences in bread wheat. Summarising, a lot of sequence information of wheat sorted chromosome arms [15–17], *T. urartu* [18] and *Ae. tauschii* [19] have been published during the past few years and are integrated in the public databases mentioned above. Today, the current version of the Chinese Spring IWGSC RefSeq v1.0 [14] facilitates the exploitation of the wheat sequence in basic science as well as in applied wheat breeding.

### 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 [8], the Affymetrix 820 K [20] and the breeding 35 K axiom [21] 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 [22].

### Cold stress signalling and transcriptional regulation

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 the second pathway involves vernalisation. Plant cells can perceive cold stress through the membrane rigidification effect. This leads to  $Ca^{2+}$  influx into the cytosol and these characteristic  $Ca^{2+}$  signatures are detected by calcium binding proteins (CBPs) [23, 24]. The induction of cold response genes starts with the activation of so-called inducer of CBF expression (*ICE*) genes. These genes are MYC-type basic helix–loop–helix transcription factors which bind on MYC recognition sites of C-repeat binding factor (*CBF*) promoters and consequently activate the expression of these genes [25]. The *CBF* transcription factors are members of the APE-TALA2/Ethylene response element binding protein (*AP2/EREBP*) family of DNA-binding proteins [26, 27]. The *AP2/EREBP* DNA-binding protein domain comprises a structure with three  $\beta$ -strands and one  $\alpha$ -helix [28–30]. Furthermore, PKK/RPAGR<sub>x</sub>KFxETRHP and DSAWR motifs are present, which are typical features of *CBF* proteins [31]. 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 [32–34]. The *CRT/DRE*

element is a highly conserved CCGAC sequence in the promoter of cold- and dehydration-responsive genes [35].

The flowering pathway is involved in FT because it contains vernalisation (*VRN*) and photoperiod response (*PPD*) genes that contributed to low temperature acclimatisation [36, 37]. The *PPD1* is a member of the PRR (pseudo response regulator) protein family and interacts with CONSTANS (CO) [38]. This family possesses a pseudo-receiver domain and a CONSTANS motif [39, 40]. Downstream the *VRN* genes are localised, i.e. *VRN1*, *VRN2* and *VRN3*. *VRN1* encodes a MADS-box transcription factor, *VRN2* is similar to a putative zinc finger and a CCT domain and *VRN3* encodes a RAF kinase inhibitor like protein [41–45].

Both pathways, the flowering and the cold response pathway, are connected by the interaction of *VRN1* and *CBF* genes. For example, the *VRN1* gene is able to reduce the transcript levels of *CBFs* and *COR* genes under long day conditions [36].

#### 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 [46, 47]. Zhao et al. [48] described an additional FT Quantitative Trait Locus (QTL) on chromosome 5B in wheat germplasm from central Europe. Due to the importance of cold acclimatisation in winter and spring wheat, the locus FR1 was physically and genetically mapped [49, 50]. However, it is not clear whether FR1 is an independent gene or is based on a pleiotropic effect of *VRN1* [36, 51]. As a consequence of the presence of the A, B and D genome, there are three *VRN1* homologous genes (*VRN-A1*, *VRN-B1* and *VRN-D1*) not having the same impact on vernalisation. Wheat plants with a dominant *VRN-A1* allele are spring type and do not need vernalisation for flowering, while the dominant *VRN-B1* and *VRN-D1* alleles result also in spring habit, but they are weaker than *VRN-A1* [52]. The difference among the spring (dominant *VRN-A1* alleles) and winter varieties (recessive *vrn-A1* alleles) is based on a C/T single nucleotide polymorphism (SNP) in the fourth exon of the *VRN-A1* gene. The winter varieties carrying an ambiguous nucleotide Y (C/T) are more frost tolerant than genotypes carrying the *VRN-B1* or *VRN-D1* allele which both confer higher frost tolerance than spring varieties carrying a C at the respective SNP in *VRN-A1* [2, 53–56].

Plenty of studies identified the FR-A2 locus on chromosome 5A as the most important locus involved in FT in wheat [57–59]. The FR-A2 locus comprises at minimum 11 *CBF* genes and is located approximately 30 cM proximal to *VRN1* [46, 47, 60]. Two independent studies illustrate that *CBF-A3*, which is located in the FR-A2 locus, plays an important role in wheat FT

[47, 61]. Knox et al. [62] analysed the FR-A<sup>m2</sup> locus of diploid *Triticum monococcum* (A<sup>m</sup> 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. [59] identified *CBF14* and *CBF15* as FT associated in *Triticum monococcum* and Soltesz et al. [63] confirmed this for *Triticum aestivum*. Additionally, Kocsy et al. [64] identified three genes, i.e. *Tacr7* (*Triticum aestivum* cold-regulated 7), *Cab* (calcium-binding EF-hand family protein-like) and *Dem* (Defective embryo and meristems) being differentially expressed during cold hardening in wheat. In addition, on the transcriptome level FT signaling is much more complex. Hundreds to thousands of wheat genes were identified to be significantly up- or downregulated under low temperature [34, 65–69].

The aim of this study was therefore to (i) identify SNPs and indels (insertion-deletion) in genes known to be involved in frost tolerance in wheat and (ii) to conduct a candidate gene based association genetics approach to get information on the effect of respective SNPs and indels on FT.

## Methods

### Plant material and DNA extraction

A diverse set of 235 bread wheat genotypes was used for PCR amplification, amplicon sequencing and association genetics studies. One hundred and seventy-nine cultivars, 48 lines and 8 doubled haploid (DH) lines originating from 28 countries from five continents (Additional file 1: Table S1) were analysed. 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 [70] and parental lines of Western European hybrid breeding programs were included. For the physical mapping of PCR amplicons to chromosomes and chromosome segments, 21 nulli-tetrasomic lines (NT-lines) and 46 deletion-lines were used [9, 71, 72]. The DNA was extracted at the three leaf stage according to Stein et al. [73].

### 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 (2.5m<sup>2</sup>) trial with one replication. FT was evaluated as winter survival in percent (%), 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 with an average air temperature under -15 °C in the period from December 1st to April 30th of each year was calculated. The correlation coefficient  $r$  [74] 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 (under 0.15) and/or variance (under 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.

#### PCR amplification, fragment analysis and re-sequencing

Primer development and testing was conducted according to Babben et al. [9]. Furthermore, 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 nulli-tertasomic (NT) lines [9]. 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. PCR amplification was performed in 20 µl reaction volume (RV) by using two polymerases, i.e. FIREPol® DNA polymerase (Solis BioDyne, Tartu, Estonia) and MyTaq™ DNA polymerase (BIOLINE, Luckenwalde, Germany). The master mix for a single PCR reaction comprised 0.4 U FIREPol® DNA Polymerase, 1 x Buffer B, 2.5 mM MgCl<sub>2</sub> (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 MyTaq Reaction Buffer B (that comprised 1 mM dNTPs and 3 mM MgCl<sub>2</sub>) (BIOLINE, Luckenwalde, Germany) and 0.25 pmol primers. The PCR fragment amplification was conducted in a GeneAmp® PCR System 9700 (Applied Biosystems, Darmstadt, Germany) using various PCR profiles (Additional file 2: Table S2). PCR fragments were separated by agarose gel electrophoresis and analysed using the imaging system Gel Doc™ XR and the Quantity One® 1-D analysis software (4.6.2) (Bio-Rad, Hercules, USA) and subsequently sequenced by the company Microsynth AG (Balgach, Switzerland) using the Sanger sequencing method [75].

#### 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) [76]. MAFFT parameters were set as default. The polymorphisms between the 235 genotypes were detected automatically via a small multiple sequence alignment (MSA) script (Additional file 3: Data S1) 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 SNPs 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).

#### 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 [8, 77]. 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 [78]. 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  $\Delta k$  based on the rate of change in  $\ln P(D)$  between successive  $k$ -values [79]. Wheat lines with probabilities  $\geq 0.5$  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 [80] and the Principal coordinate analysis (PCoA) by using the software package DARwin [81]. The kinship (K) matrix was calculated on a modified Roger's distance [82–84] by using R version 3.2.1 free software [85]. The Roger's distance was calculated as follows:

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

where  $p_{ij}$  and  $q_{ij}$  are allele frequencies of the  $j$ th allele at the  $i$ th locus,  $n$  number of alleles at the  $i$ th locus and  $m$  number of loci.

### 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 [86, 87]. 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  $-\log_{10}$  of  $P$  ( $P$ -value of 0.05) according to Li et al. [22, 77] 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 [86, 87]. 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.

### Sequence analysis

The translation of associated gene sequences into amino acid (AA) sequences, homologue AA sequence search, 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 [88], NCBI (National Center for Biotechnology) protein BLAST [89] and RaptorX [90–92]. 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, uSeqId-value and  $P$ -value were taken into account. A uGDT > 50, uSeqId > 30 and a  $P$ -value less than  $1 \times 10^{-3}$  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 [93, 94]. Haplotype and reference sequences were used to generate sequence alignments by applying the L-INS-i option in MAFFT [95]. To obtain the best fitting substitution model, the model test application in MEGA-CC was used [96]. 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 [97]. The codon alignments and the phylogenetic tree were used to compute dN/dS for each variable site using the SLAC method in HyPhy [94].

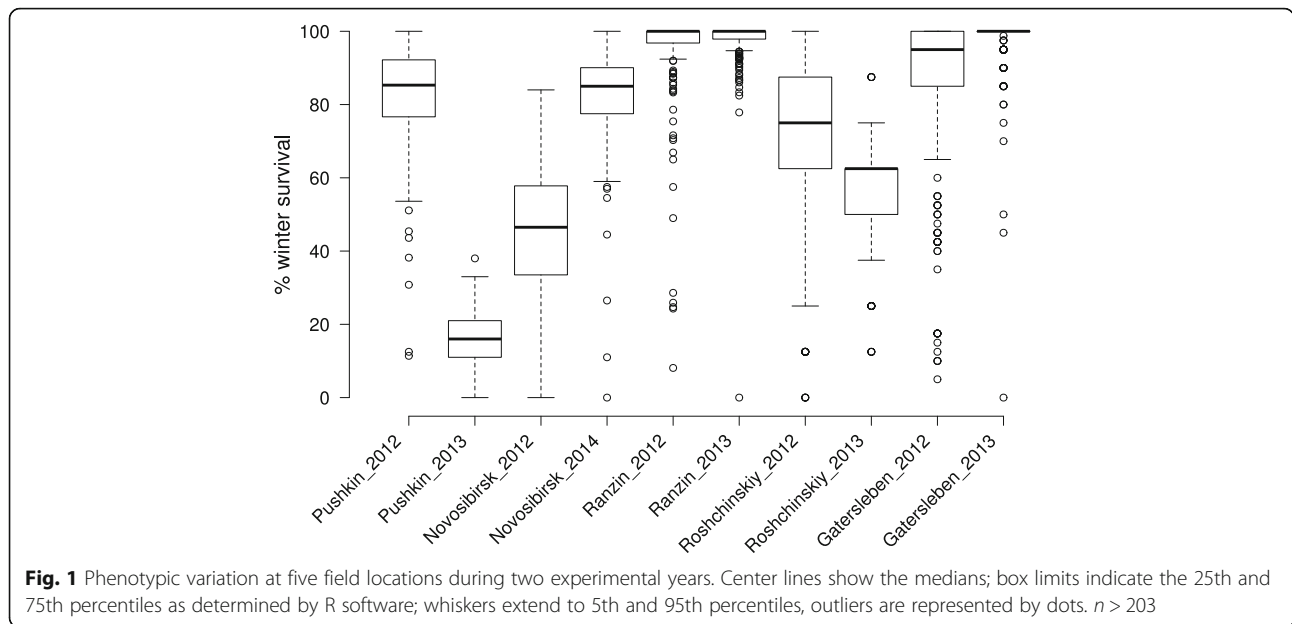
### In silico promoter analysis

Identification of promoter regions and regulatory sites was performed using the internet databases SOGO from the National Institute of Agrobiological Sciences [98, 99] and Softberry [100, 101].

## Results

### Phenotypic data analysis

Phenotypic data of five field locations were obtained during 2 years (Fig. 1 and Additional file 4: Figure S1). These phenotypic data sets and the established co-variable (number of days under  $-15^\circ\text{C}$ ) 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 (Additional file 4: Figure S1). A reason for this may be the very few days under  $-15^\circ\text{C}$  at Pushkin in 2013 (10 days) but 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 under  $-15^\circ\text{C}$  (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 (Additional file 5: Figure S2). This indicates that at these locations in the respective years phenotypic data are influenced by additional factors and are not entirely due to differences in FT. An additional characteristic of phenotypic data is the CV and variance (in %) calculation of each environment per year. High CVs (over 0.15) and/or variances (over 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 between 0.08 and 0.15 and variances between 54.3 and 149.4. The environment Pushkin\_2013 shows a low variance with 51 but a high CV value of 0.44. These values resulted from a very low winter survival with a low standard deviation (Table 1). In conclusion of the phenotypic data analysis, environments Ranzin\_2013, Gatersleben\_2013, Pushkin\_2013 and Novosibirsk\_2014 were excluded from

further analysis because results are not really related to FT. Out of the data sets Ranzin\_2012, Gatersleben\_2012, Pushkin\_2012, Novosibirsk\_2012, Roshchinskiy\_2012 and Roshchinskiy\_2013 the LS means were calculated (Additional file 1: Table S1).

**Candidate gene polymorphisms**

A set of 65 specific primer pairs from the previous study [9] was BLASTed against the IWGSC RefSeq allowing the identification of 39 primer combinations that cover the full length of 19 genes and their structural analyses. An optimised set of 39 primer pairs was used for PCR amplification, amplicon sequencing 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*,

**Table 1** Variance and coefficient of variation of frost tolerance data

Location/Year	Size	Missing data	Mean frost tolerance 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
Roshchinskiy_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

since BLASTN revealed seven and five 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 (Additional file 6: Figure S3). 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 0.30. 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 and four (*CBF-D1*, *Dhn1*, *VRN2* and *Dem*) (Table 2) were monomorphic. In total 254 polymorphic sites, i.e. 221 SNPs and 33 indels were identified. The SNP number per gene ranged from 0 to 97 and the indel number from 0 to 12. Over 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 (Fig. 2). The number of haplotypes ranged between two and six and the haplotype diversity (Hd) between 0.07 and 0.68 (Table 3).

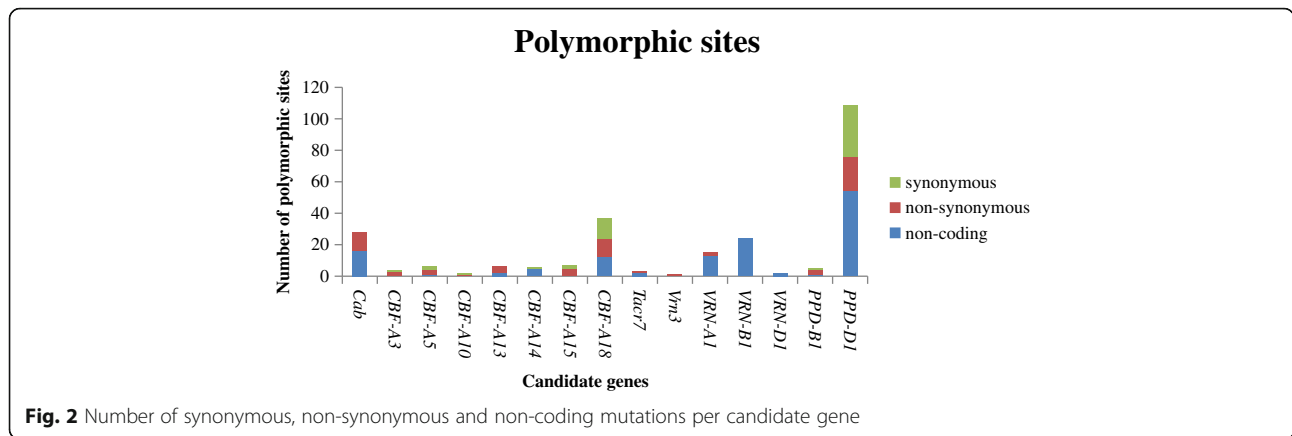
### Population structure and kinship

Genetic profiles were obtained by using the ILLUMINA Infinium iSelect 90 k wheat chip. These data were kindly provided by the Dr. A. Börner from IPK Gatersleben and will be used for GWAS analysis in an additional paper. 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 Fig. 3 and Additional file 7: Figure S4. 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

**Table 2** List of analysed FT candidate genes, sequence length, number of specific PCR fragments and detected mutations

Gene	Chr. position	Number 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
<i>CBF-D1</i>	5D	1	709	639	639	1.00	no	0
<i>CBF-A3</i>	5A	1	790	741	741	1.00	yes	4
<i>CBF-A5</i>	7A	1	1027	633	633	1.00	yes	6
<i>CBF-A10</i>	5A	1	867	720	720	1.00	yes	2
<i>CBF-A13</i>	5A	1	855	720	720	1.00	yes	6
<i>CBF-A14</i>	5A	2	610/748	639	639	1.00	yes	6
<i>CBF-A15</i>	5A	1	786	726	726	1.00	yes	7
<i>CBF-A18</i>	6A	1	1045	738	738	1.00	yes	37
<i>Dhn1</i>	5D	1	936	420	638	0.66	no	0
<i>VRN-A1</i>	5A	4	1039/1097/621/770	735	11,414	0.06	yes	15
<i>VRN-B1</i>	5B	5	600/817/662/467/1077	735	5498	0.13	yes	24
<i>VRN-D1</i>	5D	4	814/1359/1429/705	735	11,550	0.06	yes	2
<i>VRN-D2</i>	4D	2	976/534	639	1650	0.39	no	0
<i>VRN-B3</i>	7B	2	1602/994	534	1258	0.42	yes	1
<i>Cab</i>	5A	1	728	n.a.	n.a.	n.a.	yes	28
<i>Dem</i>	6B/6D	1	616	n.a.	n.a.	n.a.	no	0
<i>Tacr7</i>	2B	1	765	n.a.	n.a.	n.a.	yes	3
<i>PPD-B1</i>	2B	6	1600/954/927/571/773/378	1995	3053	0.65	yes	5
<i>PPD-D1</i>	2D	3	726/1094/966	1983	3141	0.63	yes	109
total		39	34,034	13,332	43,758	0.30		255

Chr. chromosome, bp base pairs, CDS coding DNA sequence, n.a. not available



**Fig. 2** Number of synonymous, non-synonymous and non-coding mutations per candidate gene

may be due to germplasm exchange and it use in breeding programs.

**SNP/indel association analysis**

The association analysis was performed using 254 polymorphic sites identified in 15 candidate genes. After 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. In the SNP/indel association, 27 statistically significant ( $P < 0.05$ ) polymorphic sites (21 SNPs and six indels) in six candidate genes were identified (Table 4). Four SNPs and two indels were located in the promoter

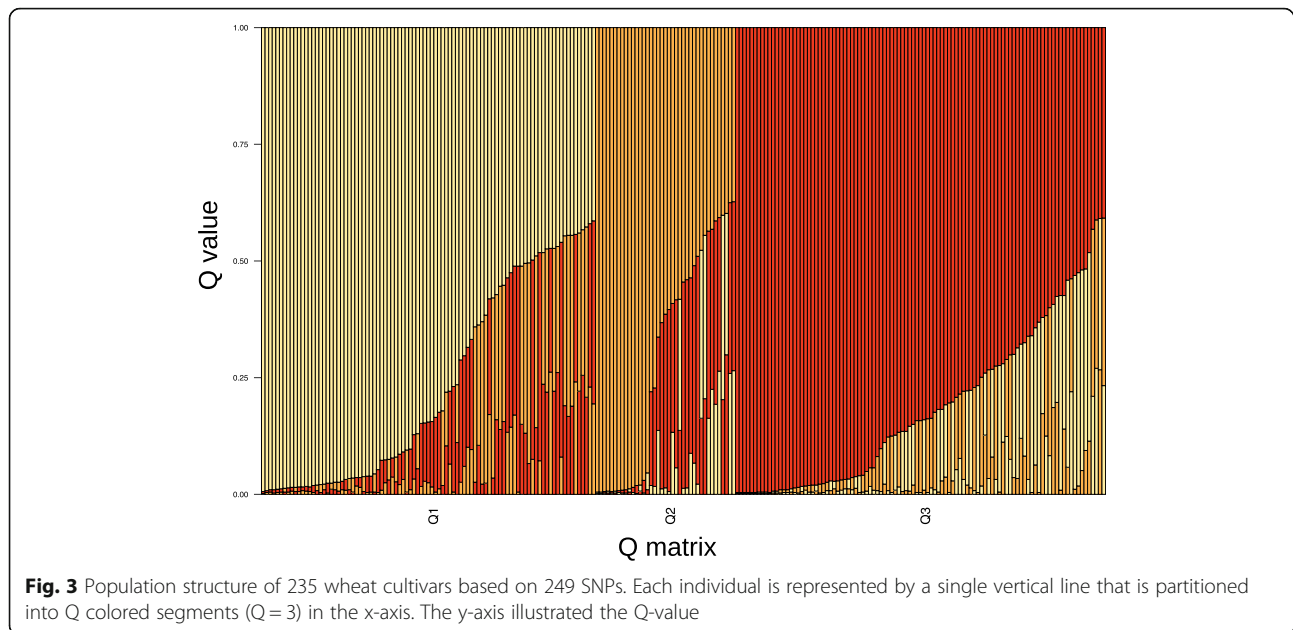
region of *CBF-A5*, *CBF-A13* and *CBF-A14*. The remaining 17 SNPs and four indels were located in exon regions of five *CBFs*, *PPD-D1*, *VRN-A1* and *VRN-B3*. 11 of these SNPs are non-synonymous. Six associated genes were located on wheat chromosome 5A and one each on chromosome 2D, 7A and 7B. Allelic effects of significantly associated polymorphic sites on FT ranged from 0.11% to 15.01 (Fig. 4, Table 4, Additional file 8: Table S3). With an LD of  $r^2 = 0.92$  to 1, the statistical significances of associated SNPs and indels from the five *CBF* genes on chromosome 5A (*CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15*) is very high. The LD plot of all used SNPs/indels for association calculation is shown in Fig. 5.

**Table 3** Polymorphic candidate genes, location of polymorphisms, amino acid change and haplotype diversity

Gene	Chr.	Acces-sions	Poly-morphic sites	SNPs	in-dels	Promotor	Intron	Exon	3' UTR	Non-coding	Non-syno-nymous	Syno-nymous	Haplo-types	Hd
<i>CBF-A3</i>	5A	235	4	4	0	0	0	4	0	0	3	1	2	0.30
<i>CBF-A5</i>	7A	235	6	5	1	1	0	5	0	1	3	2	4	0.54
<i>CBF-A10</i>	5A	235	2	2	0	0	0	2	0	0	1	1	2	0.30
<i>CBF-A13</i>	5A	235	6	4	2	1	0	5	0	2	4	0	3	0.30
<i>CBF-A14</i>	5A	235	6	5	1	5	0	1	0	5	0	1	3	0.30
<i>CBF-A15</i>	5A	235	7	6	1	0	0	7	0	0	5	2	2	0.30
<i>CBF-A18</i>	6A	235	37	34	3	1	0	25	11	12	12	13	3	0.20
<i>VRN-A1</i>	5A	235	15	10	5	7	6	2	0	13	2	0	6	0.19
<i>VRN-B1</i>	5B	235	24	21	3	20	4	0	0	24	0	0	5	0.07
<i>VRN-D1</i>	5D	235	2	2	0	0	1	0	1	2	0	0	3	0.14
<i>VRN-B3</i>	7B	235	1	1	0	0	0	1	0	0	1	0	3	0.50
<i>Cab</i>	5A	235	28	24	4	5	0	12	11	16	12	0	6	0.64
<i>Tacr7</i>	2B	235	3	2	1	0	0	1	2	2	1	0	3	0.40
<i>PPD-B1</i>	2B	235	5	5	0	1	0	4	0	1	3	1	6	0.37
<i>PPD-D1</i>	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





#### 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.

#### 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-B1* 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 5). *CBF-A18* is the only exon haplotype which is 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 0.68% to 33.95 (Fig. 6, Table 5, Additional file 8: Table S3).

#### 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 (Additional file 9: Table S4). 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  $\beta$ -strands and one  $\alpha$ -helix (2gcc). The parameters of protein structure modelling are shown in Table 6. The *CBF-A13\_1* haplotype protein shows a low  $P$ -value of  $4.79 \times 10^{-4}$ , and a high uGDT of 51, and a uSeqId of 15 extent. 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. The AA sequences between two *CBF-A3* haplotypes revealed three AA substitutions. The *CBF-A3\_SNP2* [C/G] on CDS site 263 leads to an Ala/Gly change, *CBF-A3\_SNP3* [A/G] on CDS site 367 to a Ser/Gly change and *CBF-A3\_SNP4* [C/T] on CDS site 407 to a Ser/Phe change. The AA change of *CBF-A3\_SNP2* is located in the  $\alpha$  helix of the AP2 domain. The *CBF-A3\_SNP3* and *CBF-A3\_SNP4* is located upstream of the identified domains/motifs. The

**Table 4** Statistics of significantly associated SNPs and indels

Gene/Polymorphism name	CDS and promotor site	P-value	-Log of P	Polymorphism		Effect in %		Observations <sup>a</sup>			
CBF-A3_SNP1	222	6.28E-10	9.20	A	C	15.01	0.00	193	42		
CBF-A3_SNP2	263	6.28E-10	9.20	C	G	15.01	0.00	193	42		
CBF-A3_SNP3	367	6.28E-10	9.20	A	G	15.01	0.00	193	42		
CBF-A3_SNP4	407	6.28E-10	9.20	C	T	15.01	0.00	193	42		
CBF-A5_indel1	-83	6.00E-3	2.22	C	-	-4.52	0.00	141	84		
CBF-A10_SNP1	471	6.28E-10	9.20	C	T	15.01	0.00	193	42		
CBF-A10_SNP2	518	6.28E-10	9.20	G	C	15.01	0.00	193	42		
CBF-A13_SNP1	-11	6.28E-10	9.20	T	C	15.01	0.00	193	42		
CBF-A13_indel3	19	6.28E-10	9.20	T	-	15.01	0.00	193	42		
CBF-A13_SNP2	92	6.28E-10	9.20	G	A	15.01	0.00	193	42		
CBF-A13_indel2	133	6.28E-10	9.20	-	C	15.01	0.00	193	42		
	134	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	135	6.28E-10	9.20	-	T	15.01	0.00	193	42		
	136	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	137	6.28E-10	9.20	-	C	15.01	0.00	193	42		
	138	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	139	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	140	6.28E-10	9.20	-	C	15.22	0.00	6.75	193	41	1
	141	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	142	6.28E-10	9.20	-	C	15.01	0.00	193	42		
	143	6.28E-10	9.20	-	A	15.01	0.00	193	42		
	144	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	145	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	146	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	147	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	148	6.28E-10	9.20	-	C	15.01	0.00	193	42		
	149	6.28E-10	9.20	-	A	15.01	0.00	193	42		
	150	6.28E-10	9.20	-	A	15.01	0.00	193	42		
	151	6.28E-10	9.20	-	C	15.01	0.00	193	42		
	152	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	153	6.28E-10	9.20	-	C	15.01	0.00	193	42		
	154	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	155	4.00E-10	9.40	-	G	15.01	0.00	193	42		
	156	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	157	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	158	6.28E-10	9.20	-	C	15.01	0.00	193	42		
	159	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	160	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	161	6.28E-10	9.20	-	T	15.01	0.00	193	42		
	162	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	163	6.28E-10	9.20	-	G	15.01	0.00	193	42		
	164	6.28E-10	9.20	-	G	15.01	0.00	193	42		
CBF-A13_SNP3	294	6.28E-10	9.20	C	A	15.01	0.00	193	42		

**Table 4** Statistics of significantly associated SNPs and indels (Continued)

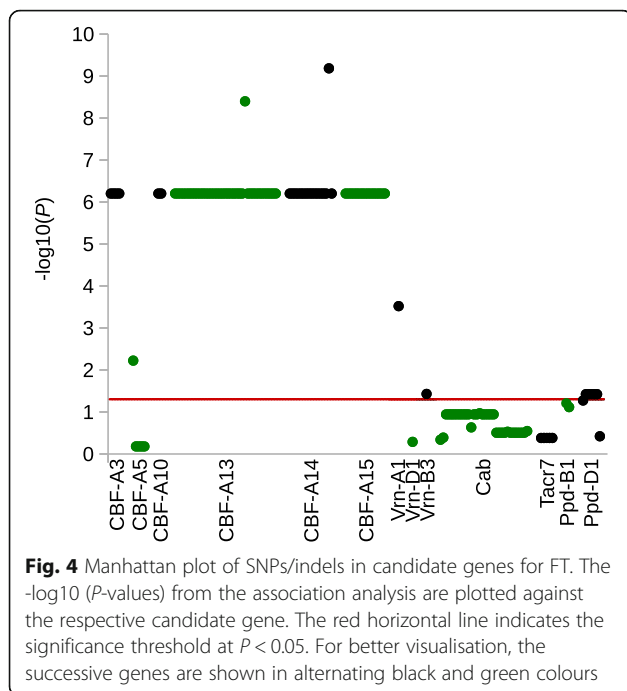
Gene/Polymorphism name	CDS and promotor site	P-value	-Log of P	Polymorphism			Effect in %			Observations <sup>a</sup>		
CBF-A14_indel1	-506	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	-505	6.28E-10	9.20	T	-		15.01	0.00		193	42	
	-504	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	-503	6.28E-10	9.20	A	-		15.01	0.00		193	42	
	-502	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	-501	6.28E-10	9.20	T	-		15.01	0.00		193	42	
	-500	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	-499	6.28E-10	9.20	T	-		15.01	0.00		193	42	
	-498	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	-497	6.28E-10	9.20	A	-		15.01	0.00		193	42	
	-496	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	-495	6.28E-10	9.20	T	-		15.01	0.00		193	42	
	CBF-A14_SNP1	-449	6.28E-10	9.20	T	C		15.01	0.00		193	42
CBF-A14_SNP2	-158	6.28E-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	6.28E-10	9.20	G	A		15.01	0.00		193	42	
CBF-A15_SNP1	84	6.28E-10	9.20	C	T		15.01	0.00		193	42	
CBF-A15_SNP2	243	6.28E-10	9.20	T	C		15.01	0.00		193	42	
CBF-A15_SNP3	293	6.28E-10	9.20	C	T		15.01	0.00		193	42	
CBF-A15_SNP4	397	6.28E-10	9.20	G	A		15.01	0.00		193	42	
CBF-A15_indel1	503	6.28E-10	9.20	T	-		15.01	0.00		193	42	
	504	6.28E-10	9.20	C	-		15.01	0.00		193	42	
	505	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	506	6.28E-10	9.20	T	-		15.01	0.00		193	42	
	507	6.28E-10	9.20	C	-		15.01	0.00		193	42	
	508	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	509	6.28E-10	9.20	T	-		15.01	0.00		193	42	
	510	6.28E-10	9.20	C	-		15.01	0.00		193	42	
	511	6.28E-10	9.20	G	-		15.01	0.00		193	42	
	CBF-A15_SNP5	694	6.28E-10	9.20	A	G		15.01	0.00		193	42
CBF-A15_SNP6	716	6.28E-10	9.20	C	G		15.01	0.00		193	42	
PPD-D1_indel1	1266	3.77E-2	1.42	-	C		3.96	0.00		68	166	
	1267	3.77E-2	1.42	-	G		3.96	0.00		68	166	
	1268	3.77E-2	1.42	-	T		3.96	0.00		68	166	
	1269	3.77E-2	1.42	-	C		3.96	0.00		68	166	
	1270	3.77E-2	1.42	-	G		3.96	0.00		68	166	
VRN-A1_SNP1	349	3.03E-4	3.52	C	T	Y	0.00	1.35	0.11	24	2	209
VRN-B3_SNP1	19	3.76E-2	1.43	C	G		3.05	0.00		105	128	

CDS coding DNA sequence, P probability, Log logarithm

<sup>a</sup>number of genotypes per allele

AA site of CBF-A3\_SNP2 is significant for negative selection (Fig. 7). 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 change which are located in the β strand 3 or the α helix of the AP2 domain, respectively. Additionally, CBF-A5\_SNP4 site underlies negative selection (Additional file 10: Figure S5). The AA haplotypes of CBF-A10 show an AA change 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 (Additional file 11: Figure S6). The haplotypes of CBF-A13 show completely different AA sequences. The single bp deletion on CDS site 19 of CBF-A13\_indel results in a frame shift that changes every AA from the seventh AA onwards. Furthermore, a stop codon at position 74 AAs 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 homologue 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 homologue AA sequences, no assumptions about positive or negative selection can be made (Additional file 12: Figure S7).

The AA analysis of two CBF-A15 haplotypes revealed four AA changes and one AA indel. A nine-nucleotide insertion within the coding region (sites 503 to 511) resulted in three additional Ser residues in the haplotype with CBF\_indel. The AA change 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 change. The last AA changes 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\_indel1 are located upstream of the identified domains/motifs.

All five AA substitutions regions show no positive or negative selection (Fig. 8a, Additional file 13: Figure S8). The CBF-A14 AA haplotypes do not result in AA changes (Additional file 14: Figure S9). The CBF-A18 haplotypes carry 14 AA substitutions. However, none of these is associated with FT or underlies negative or positive selection (Additional file 15: Figure S10).

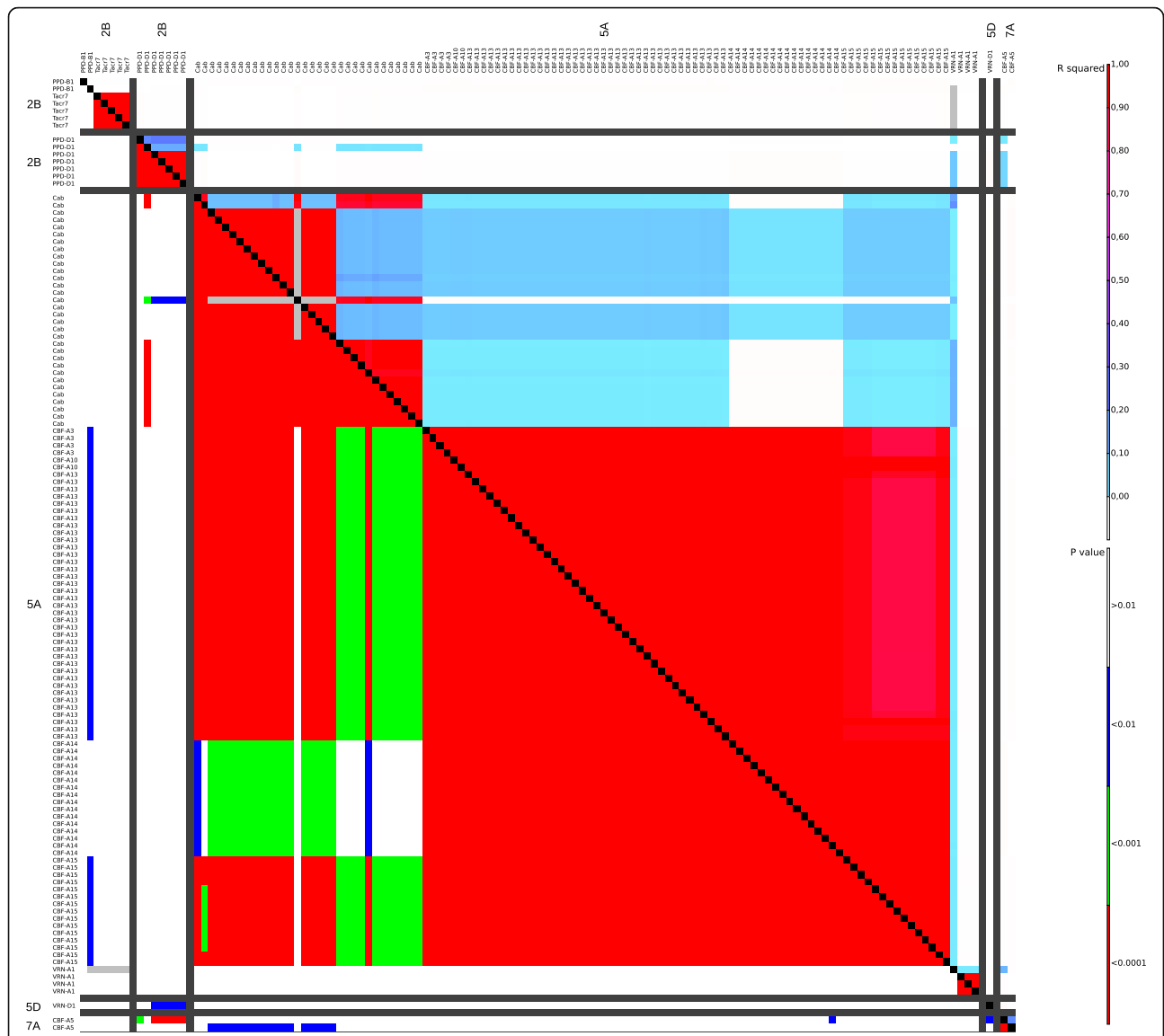
#### Vernalisation genes (*VRN-A1* and *VRN-B3*)

Nine homologue AA sequences of VRN-A1 and VRN-B3 were identified (Additional file 9: Table S4). 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. [102] and Kaufmann et al. [103] 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 6. The VRN-A1\_SNP1 [C/T/Y] on CDS site 349 generated an AA change of Leu/Phe in the K domain between  $\alpha$  helix three and  $\alpha$  helix four. Furthermore, in the region of this AA substitution, no positive or negative selection was identified (Fig. 8b, Additional file 16: Figure S11).

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 6). This flowering time family protein contains four  $\alpha$  helices, seven  $\beta$  strands and one segment B [104]. The VRN-B3\_SNP1 [C/G] on CDS site 19 generates an AA change of His/Asp directly before the start of the  $\alpha$  helix 1. Furthermore, this site is subject to significant negative selection (Fig. 8c, Additional file 17: Figure S12).

#### Photoperiod response genes (*PPD-B1* and *PPD-D1*)

The association study revealed that the gene PPD-B1 is significantly associated to FT only in the haplotype analysis and PPD-D1 in both analyses. The NCBI protein BLASTX of the PPD-B1 and PPD-D1 AA haplotype sequences showed nine homologues (Additional file 9: Table S4). 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 [39, 40]. In the RaptorX analysis the protein models of a putative response regulator domain (3t6k) and response regulator receiver domain (3jte) were identified (Table 6). These domains contain five  $\alpha$  helices and five  $\beta$  strands. The PPD-B1 haplotype AA



**Fig. 5** 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 coloured square in the below triangle represents the intensity of LD expressed by *P*-values for each pairwise comparison between polymorphic sites. Each coloured square in the top triangle represents the intensity of LD expressed by *r*<sup>2</sup> for each pairwise comparison between polymorphic sites. On the right side the legend for *r*<sup>2</sup> values and *P*-values are shown

exhibited three AA changes. The PPD-B1\_SNP2 [A/G] on CDS site 304 generated an AA change of Arg/Gly and PPD-B1\_SNP3 [A/G] on CDS site 368 generated an AA change of Asn/Asp. Both are located in the Pseudo-Receiver domain. The PPD-B1\_SNP2 is located in the  $\alpha$  helix 3 and PPD-B1\_SNP3 between  $\beta$  strand 4 and  $\alpha$  helix 4. The third AA change 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 (Fig. 8d, Additional file 18: Figure S13). 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 motif. All

other AA changes between haplotype two and three did not show significant association in the SNP/indel association study (Additional file 19: Figure S14).

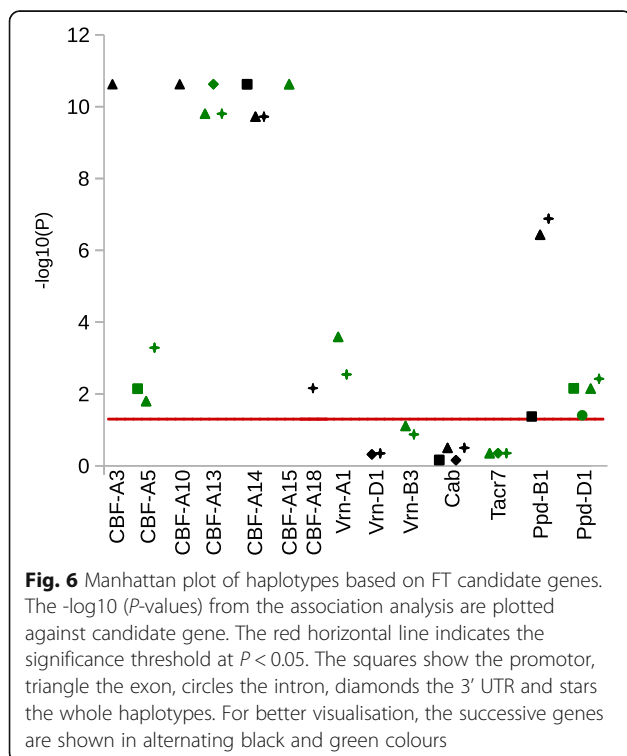
**In silico promoter analysis**

The significantly associated CBF-A5\_indel1 [C/-], which is located 83 bp upstream of the start codon, entails no changes on the promoter regulatory sites and has no influence on the transcription of the *CBF-A5* gene. Also, the associated polymorphic sites CBF-A13\_SNP1 of the *CBF-A13* promoter and the sites CBF-A14\_indel1, CBF-A14\_SNP1, CBF-A14\_SNP2 and CBF-A14\_SNP3 of the *CBF-A14* promoter show no regulatory site

**Table 5** Statistics of haplotypes significantly associated to FT

Haplotype name	Chr.	P-value	-Log of P	Haplotype	FT effect of haplotypes in %					Observations <sup>a</sup>						
CBF-A3 ex.	5A	2.37E-11	10.63	1 2	15.01	0.00				193	42					
CBF-A5 pro.	7A	7.17E-3	2.14	1 2	4.44	0.00				82	143					
CBF-A5 ex.	7A	1.58E-2	1.80	1 2 3	19.51	18.09	0.00			204	28	3				
CBF-A5	7A	5.16E-4	3.29	1 2 3 4	24.18	19.29	18.88	0.00		55	27	140	3			
CBF-A10 ex.	5A	2.37E-11	10.63	1 2	15.01	0.00				193	42					
CBF-A13 ex.	5A	1.57E-10	9.80	1 2 3	15.22	0.00	8.47			193	41	1				
CBF-A13 pro.	5A	2.37E-11	10.63	1 2	15.01	0.00				193	42					
CBF-A13	5A	1.57E-10	9.80	1 2 3	15.22	0.00	8.47			193	41	1				
CBF-A14 pro.	5A	2.37E-11	10.63	1 2	15.01	0.00				193	42					
CBF-A14 ex.	5A	1.90E-10	9.72	1 2 3	17.01	2.11	0.00			193	40	2				
CBF-A14	5A	1.90E-10	9.72	1 2 3	17.01	2.11	0.00			193	40	2				
CBF-A15 ex.	5A	2.37E-11	10.63	1 2	15.01	0.00				193	42					
CBF-A18	6A	6.88E-3	2.16	1 2 3	9.32	6.66	0.00			209	9	17				
PPD-B1 pro.	2B	4.23E-2	1.37	1 2	5.50	0.00				213	22					
PPD-B1 ex.	2B	3.69E-7	6.43	1 2 3 4	18.97	0.00	28.28	33.39		3	5	207	20			
PPD-B1	2B	1.31E-7	6.88	1 2 3 4 5	19.03	0.00	33.95	29.07	22.63	3	5	20	185	22		
PPD-D1 pro.	2D	6.95E-3	2.16	1 2 3	14.38	10.75	0.00			141	80	4				
PPD-D1 in.	2D	3.93E-2	1.41	1 2 3	18.66	19.56	0.00			162	69	3				
PPD-D1 ex.	2D	7.05E-3	2.15	1 2 3	21.80	18.18	0.00			68	163	3				
PPD-D1	2D	3.78E-3	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 ex.	5A	2.57E-4	3.59	1 2 3 4	0.00	0.68	13.12	11.89		9	15	2	209			
VRN-A1	5A	2.87E-3	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, ex. exon, pro. Promotor, in. intron, P probability, Log logarithm, FT frost tolerance  
<sup>a</sup>number of genotypes per allele



**Fig. 6** Manhattan plot of haplotypes based on FT candidate genes. The  $-\log_{10}(P)$ -values from the association analysis are plotted against candidate gene. The red horizontal line indicates the significance threshold at  $P < 0.05$ . The squares show the promotor, triangle the exon, circles the intron, diamonds the 3' UTR and stars the whole haplotypes. For better visualisation, the successive genes are shown in alternating black and green colours

changes which are indicative for a modification of the gene transcription.

**Discussion**

After hybridisation, domestication and 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 [105]. Therefore, sequencing of candidate genes in large collections is a tool to rediscover hidden genetic diversity and use this in breeding.

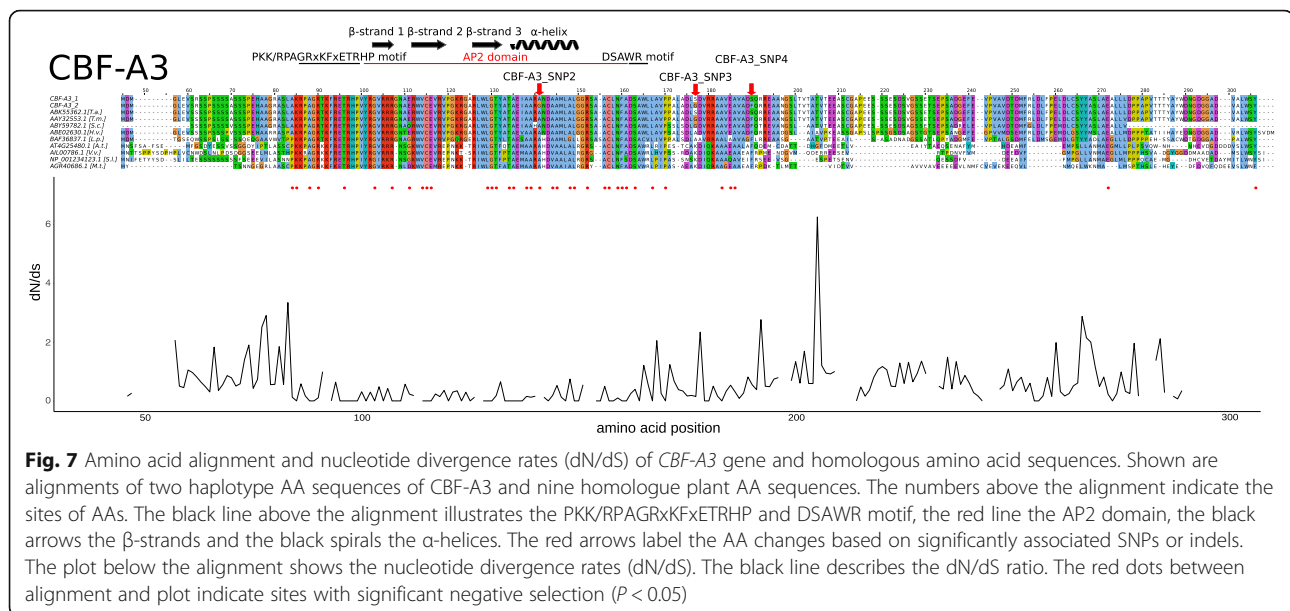
**LD and diversity**

The genotype group from Asia shows a very small diversity with 25 polymorphic sites out of 81 genotypes. That implies that these genotypes are very close to each other based on the analyzed genes. The polymorphic sites of the different 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 38 genotypes. That implies that these

**Table 6** Protein structure modelling of associated genes

Protein haplotypes	P-value	uGDT (GDT)	uSeqId (SeqId)	Score	Domain/motif	Literature
CBF-A3_1	3.03E <sup>-4</sup>	71 (62)	30 (26)	51	2gcc (AP2)	[28]
CBF-A3_2	1.52E <sup>-4</sup>	71 (61)	29 (25)	51	2gcc (AP2)	[28]
CBF-A5_1	3.22E <sup>-4</sup>	69 (67)	30 (29)	51	2gcc (AP2)	[28]
CBF-A5_2	5.75E <sup>-4</sup>	56 (54)	30 (29)	54	2gcc (AP2)	[28]
CBF-A5_3	9.71E <sup>-4</sup>	57 (56)	29 (28)	56	2gcc (AP2)	[28]
CBF-A10_1	2.78E <sup>-4</sup>	70 (66)	31 (29)	51	2gcc (AP2)	[28]
CBF-A10_2	2.88E <sup>-4</sup>	71 (67)	31 (29)	50	2gcc (AP2)	[28]
CBF-A13_1	4.79E <sup>-4</sup>	51 (38)	15 (11)	36	2gcc (AP2)	[28]
CBF-A13_2	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CBF-A14_1	1.97E <sup>-4</sup>	71 (69)	29 (28)	52	2gcc (AP2)	[28]
CBF-A14_2	1.97E <sup>-4</sup>	71 (69)	29 (28)	52	2gcc (AP2)	[28]
CBF-A15_1	3.80E <sup>-4</sup>	70 (66)	29 (27)	50	2gcc (AP2)	[28]
CBF-A15_2	2.95E <sup>-4</sup>	70 (70)	29 (29)	49	2gcc (AP2)	[28]
CBF-A18_1	1.33E <sup>-3</sup>	55 (49)	29 (26)	52	2gcc (AP2)	[28]
CBF-A18_2	8.72E <sup>-4</sup>	55 (49)	29 (26)	54	2gcc (AP2)	[28]
PPD-B1	1.24E <sup>-4</sup>	105(62)	32(19)	130	3t6k (putative response regulator domain)	n.a.
	8.74E <sup>-5</sup>	101(60)	27(16)	134	3jte (response regulator receiver domain)	[111]
PPD-B2	1.21E <sup>-4</sup>	104(61)	32(19)	129	3t6k (putative response regulator domain)	n.a.
	8.12E <sup>-5</sup>	101(59)	27(16)	134	3jte (response regulator receiver domain)	[111]
PPD-B3	1.06E <sup>-4</sup>	103(60)	31(18)	129	3t6k (putative response regulator domain)	n.a.
	7.21E <sup>-5</sup>	102(59)	27(16)	134	3jte (response regulator receiver domain)	[111]
PPD-B4	1.28E <sup>-4</sup>	104(61)	32(19)	130	3t6k (putative response regulator domain)	n.a.
	8.98E <sup>-5</sup>	102(59)	27(16)	134	3jte (response regulator receiver domain)	[111]
PPD-D1_1	1.05E <sup>-4</sup>	104 (47)	32 (15)	131	3t6k (putative response regulator domain)	n.a.
	7.11E <sup>-5</sup>	102 (46)	27 (12)	136	3jte (response regulator receiver domain)	[111]
PPD-D1_2	1.14E <sup>-4</sup>	105 (63)	32 (19)	130	3t6k (putative response regulator domain)	n.a.
PPD-D1_3	1.16E <sup>-4</sup>	104(62)	32 (19)	129	3t6k (putative response regulator domain)	n.a.
	8.07E <sup>-5</sup>	101 (60)	27 (16)	134	3jte (response regulator receiver domain)	[111]
VRN-A1_1	8.02E <sup>-3</sup>	84 (83)	39 (39)	84	4ox0 (K domain)	[110]
	1.02E <sup>-4</sup>	78 (103)	38 (50)	53	3kov (MADS domain/I domain)	[112]
	1.30E <sup>-4</sup>	75 (98)	38 (50)	52	1egw (MADS domain/I domain)	[113]
VRN-A1_2	8.47E <sup>-3</sup>	83 (82)	39 (39)	84	4ox0 (K domain)	[110]
	7.90E <sup>-5</sup>	73 (103)	38 (50)	54	3kov (MADS domain/I domain)	[112]
	9.11E <sup>-5</sup>	74 (98)	38 (50)	53	1egw (MADS domain/I domain)	[113]
VRN-A1_3	7.95E <sup>-3</sup>	84 (83)	38 (38)	84	4ox0 (K domain)	[110]
	7.06E <sup>-5</sup>	78 (103)	38 (50)	54	3kov (MADS domain/I domain)	[112]
	1.08E <sup>-4</sup>	75 (99)	38 (50)	53	1egw (MADS domain/I domain)	[113]
VRN-B3_1	7.92E <sup>-13</sup>	154 (87)	147 (83)	166	3axy (Hd3A protein)	[104]
VRN-B3_2	1.66E <sup>-12</sup>	154 (87)	148 (84)	168	3axy (Hd3A protein)	[104]

P: probability, uGDT: unnormalized 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 normalized by the protein (or domain) sequence length and multiplied by 100, n.a.: not available



genotypes are very distant to each other based on the analyzed genes. The gene cluster of *CBFs* (*CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15*) on chromosome 5A shows a very high LD of  $r^2 = 1$ . Consequently, this gene cluster has not been divided by meiotic events but 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. Consequently the gene cluster (*CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15*), *Cab* and *VRN-A1* are inherited independently.

#### Association study

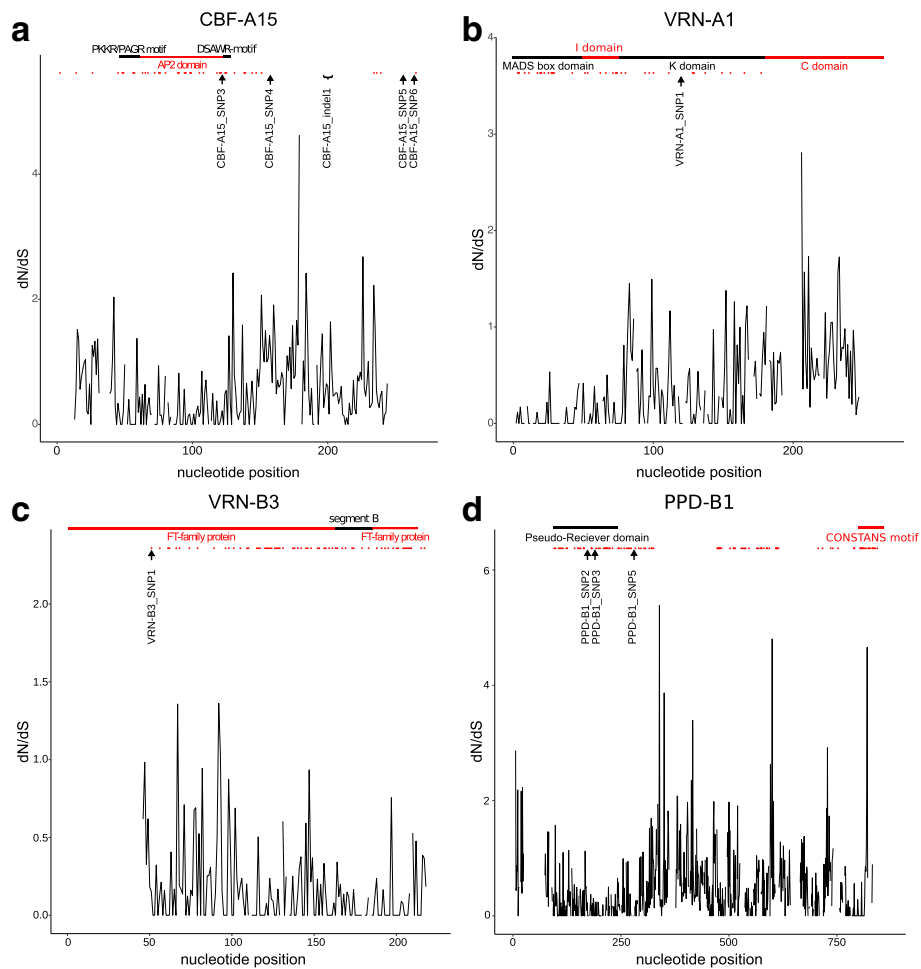
FT is a highly complex and important trait of winter wheat that is usually studied using QTL and expression profiling approaches [47, 106, 107]. In this study we are presenting, to our best knowledge, the first large scale candidate gene based association analysis of FT in wheat.

Accordingly, we identified significantly associated polymorphisms (SNPs/indels) in eleven studied 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 were investigated in this study, six revealed FT association in the SNP/indel and haplotype method. Associated polymorphisms at the five candidate genes *CBF-A3*, *CBF-A10*, *CBF-A13*, *CBF-A14* and *CBF-A15* are located at the FR-A2 locus [46, 47] 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 are in a very high LD ( $r^2 = 0.92$  to 1), indicating localisation at the same chromosomal region without or very rare recombination. This result confirms previous knowledge of strong 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 [50, 108]. 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 shows 15% better winter survival in comparison to group two (about 42 genotypes) (Table 4). Therefore, breeding efforts in combining two FT haplotypes are highly desirable towards creating elite cultivars exhibiting high FT.

The *CBF-A3\_SNP2* [C/G] which is significantly associated with FT ( $P = 6.28 \times 10^{-10}$ ) results in Ala/Gly AA change. This AA is localised in the  $\alpha$  helix of the AP2 domain and shows a high conservation of Ala in the AA alignment with homologues. Only the AA haplotype two had a Gly which corresponds to a significant reduction in winter survival of 15.01% in the SNP/indel association study. Also haplotype one shows a 15.01% better winter survival compared to haplotype two with Gly in the haplotype association study (Tables 4 and 5). Allen et al. [28] describe that Ala contributes to the stabilisation of the protein structure by its hydrophobic side chain. Through the AA change from Ala to Gly it is possible that the Gly of AA haplotype two loses or impairs the functionality as a transcription factor. Additionally, the dN/dS ratio analysis shows that the *CBF-A3\_SNP2* site





**Fig. 8** 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/RPAGR 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 indicates the 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-Receiver 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 ( $P < 0.05$ )

underlies negative selection (Fig. 7), reflecting the association effects. In detail, haplotype two with an Gly on CBF-A3\_SNP2 site shows lower winter survival of 15.01% compared to haplotype one.

The SNP/indel association study of *CBF-A5* exon SNPs revealed no significant association. In contrast, the exon haplotype is significantly associated with a reduced winter survival of 19.51% respectively 18.01% of exon haplotype three (Ser on site CBF-A5\_SNP2) compared

to haplotype one respectively two. Due to the fact that the exon haplotype three consists only of three genotypes it was not identified in the SNP/indel association analysis (Table 5, Additional file 10: Figure S5).

The significantly associated SNPs and indels of *CBF-A13* may disorganise the protein structure very strongly. The one bp CBF-A13\_indel1 of haplotype two changed 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 and the AA haplotype one shows a very low uSeqId of 15 but an appropriate  $P$ -value of  $4.79 \times 10^{-4}$  compared to the AP2 domain. Therefore, it is most likely that both proteins are non-functional. However, haplotype one shows about 6.20% higher winter survival in the SNP/indel association study and in the haplotype association study (9.80%) 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. 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 change 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  $6.28 \times 10^{-10}$ . Ala is conserved at this position at the nine homologue AA sequences. The haplotype two which comprises Val in the AA sequence shows 15.01% less winter survival in the SNP/indel association study and 15.01% 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 change is unlikely. On the other hand, no statistically significant positive or negative selection at this site was detected.

Only the whole 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 5).

All other SNPs/indels of the seven *CBFs*, which resulted in AA changes, may be involved in FT but they are not located in the highly conserved AP2 domain. But, the 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 explain modifications in gene transcription.

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 change 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* [47, 59, 61–63]. All this leads to the

conclusion that SNP CBF-A3\_SNP2 is the most interesting *CBF* allele for FT improvement and at the same time CBF-A15\_SNP3 is the second most important one. Further details remain to be revealed by investigations in the future such as complementation of promising alleles in spring or winter wheat varieties or protein functionality analysis of these alleles.

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 increase FT by 1.35% respectively 0.11% in comparison to the genotypes with the base C. The exon haplotype association study shows stronger effects. In detail, haplotype three (T) and haplotype four (Y) increase FT by 12.44% respectively 11.21%, in comparison to haplotype two (C). Chen et al. [53] and Eagles et al. [55] also identified the C and Y allele and Diaz et al. [109] and Zhu et al. [107] described that the C allele is associated with a lower *VRN-A1* copy number and the Y allele with a higher copy number. In addition, Zhu et al. [107] described that an increase of the *VRN-A1* copy number is 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 change on a Leu conserved side in the K-domain of a MIKC-type transcription factor between  $\alpha$  helix three and  $\alpha$  helix four to a Phe. Puranik et al. [110] showed that this Leu stabilises the kink region between  $\alpha$  helix three and  $\alpha$  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 increases the angle between both  $\alpha$  helices due to its bulkiness and the attachment to the target sequence of the transcription factor is improved. 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  $3.76 \times 10^{-2}$ . This SNP generates a His/Asp AA change. The genotypes with His show a higher winter survival of 3.05% compared to those containing Asp (Table 4). Additionally this site shows a high conservation of Asp regarding all nine homologue AA sequences and underlies significant negative selection. All this data indicates that *VRN-B3* and respective homologous play a role in FT.

The polymorphisms of the *PPD-B1* gene merely show significance in the haplotype association study. The exon haplotype three shows 5.11% better winter survival than haplotype four. The difference between both haplotypes is the PPD-B1\_SNP2 [A/G] which generates an Arg/Gly change in the  $\alpha$  helix 3 of the Pseudo-Receiver domain.

The Arg is highly conserved for grasses [39, 40]. Therefore, it is possible that the Gly from haplotype three has a positive effect on FT. Another AA change 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 homologue AA of *Triticum aestivum* show Asn resulting in a 9.25% decreased winter survival in comparison to haplotype four. The haplotype two, which is associated with PPD-B1\_SNP5 [G/A], and an Asp/Asn AA change shows 28.28% 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 AAs downstream of the Pseudo-Receiver domain and is slightly conserved. Association 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. 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) (Additional file 18: Figure S13).

The associated indel of *PPD-D1* has an effect of 3.96% in the SNP/indel association study (Table 4). Haplotype one (with a deletion) shows 3.62% better winter survival compared to haplotype two in the haplotype association study (Table 5). As a result of this deletion, a stop codon on AA position 470 occurs and the CONSTANS motif is missing. The consequence is that the PPD-D1 protein cannot interact with the CO protein and the flowering control pathway is interrupted [38]. Furthermore, the interaction between the flowering time and FT pathway is disturbed (Additional file 19: Figure S14).

This study demonstrated polymorphisms significantly associated with FT and the importance of the AA changes of seven *CBF* gene family members and the *VRN-A1*, *VRN-B3*, *PPD-B1* and *PPD-D1* genes (Fig. 9). These results may be used to design highly frost tolerant wheat cultivars via gene engineering or classical

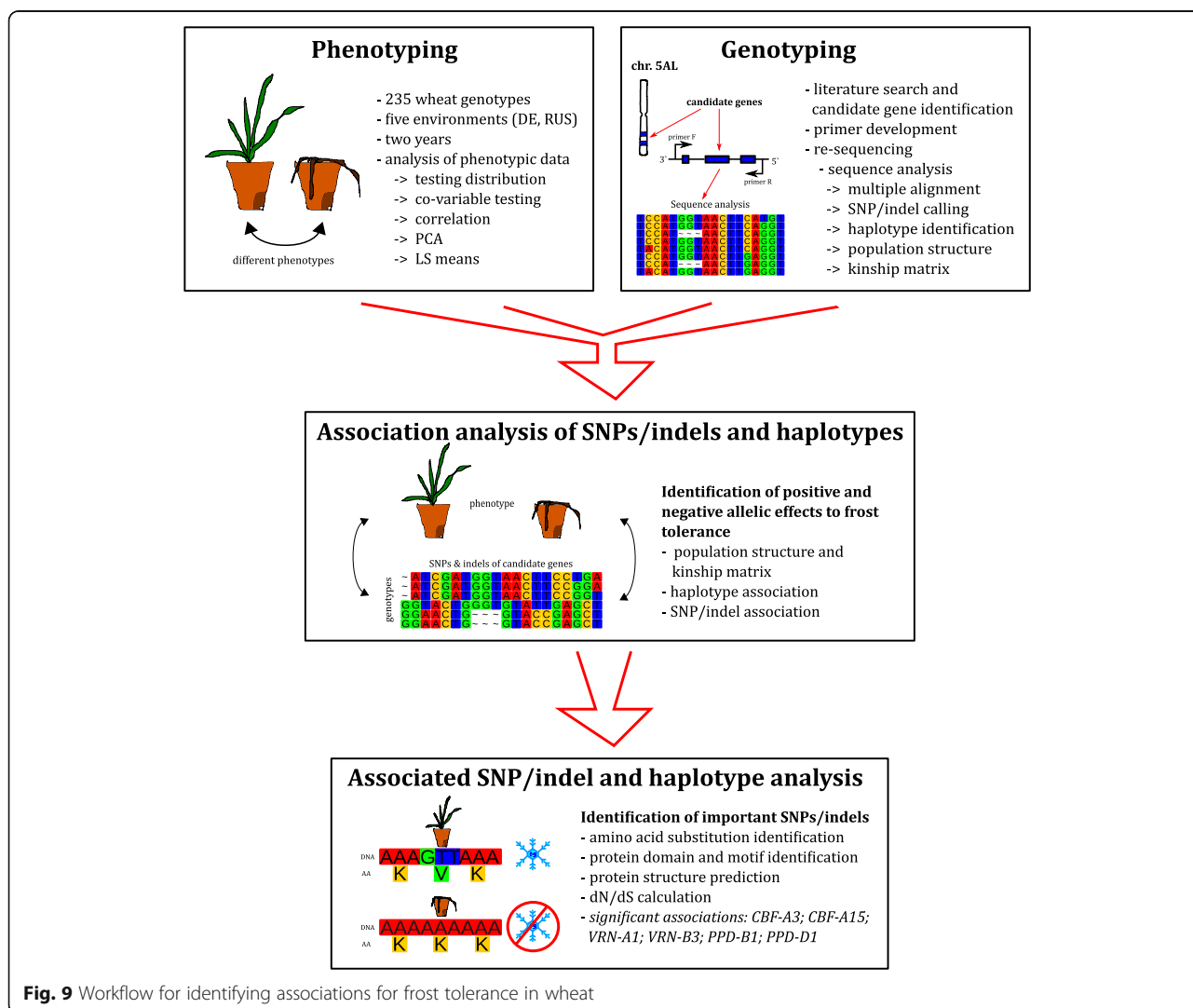


Fig. 9 Workflow for identifying associations for frost tolerance in wheat

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. We suggest 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* 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.

## Conclusion

In FT research and especially with respect to genome based breeding, it is important to identify genes and alleles involved in cold response. Based on the phenotypic data from German and Russian field trials, this study illustrates the importance of the *CBF* genes of the FR-A2 locus, *VRN-A1*, *VRN-B3*, *PPD-B1* and *PPD-D1* for FT of wheat. Polymorphisms in *CBF-A3*, *CBF-A5*, *CBF-A10*, *CBF-A13*, *CBF-A14*, *CBF-A15*, *CBF-A18*, *VRN-A1*, *VRN-B3*, *PPD-B1* and *PPD-D1* were identified which are significantly associated with FT. Besides this, we detected significantly associated polymorphisms and consequential AA substitutions in important cold response protein domains. The results of this study demonstrated that the candidate based association genetics approach is a very useful method to identify alleles positively contributing to frost tolerance in hexaploid wheat. The FT associated SNPs/indels and haplotypes identified may be used for developing diagnostic markers for marker assisted selection (MAS) for frost tolerance in wheat.

## Additional files

**Additional file 1: Table S1.** Plant material, origin, LS means of winter survival [%] and winter survival [%] of respective environments. (XLSX 34 kb)

**Additional file 2: Table S2.** PCR and sequencing primers. <sup>1</sup>Shcherban et al., 2012; <sup>2</sup>Miller et al., 2006; <sup>3</sup>Vagujfalvi et al., 2005; <sup>4</sup>Yan et al., 2004; <sup>5</sup>Beales et al., 2007; <sup>6</sup>Keilwagen et al., 2014; <sup>7</sup>Knox et al., 2008; <sup>8</sup>Babben et al., 2015; <sup>9</sup>Seki et al., 2011; <sup>10</sup>Li et al., 2011b (XLSX 19 kb)

**Additional file 3: Data S1.** Java script used to extract differences from a multiple sequence alignment (MSA). (TXT 5 kb)

**Additional file 4: Figure S1.** Scatter plot of mean winter survival and number of days under  $-15^{\circ}\text{C}$  from ten environments. (PDF 24 kb)

**Additional file 5: Figure S2.** PCA plot of mean winter survival and number of days under  $-15^{\circ}\text{C}$  from ten environments. (PDF 52 kb)

**Additional file 6: Figure S3.** Localization 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 gray boxes the exons, the black arrows rightward the forward primers and the black arrows leftwards the reverse primers. (PDF 23 kb)

**Additional file 7: Figure S4.** Principal coordinate analysis of 235 wheat cultivars. Three sub-populations based on geographical origin were shown in three colors. Blue, red and green indicate the cultivars from Europe, North America and Asia, respectively. The black dots indicate the mixture gemplasm from three sub-populations. (PDF 42 kb)

**Additional file 8: Table S3.** Polymorphic sites and haplotypes of association study and their PH-values and FT effects. (XLSX 14 kb)

**Additional file 9: Table S4.** Identities and e values of FT associated AA sequences and homologous AA sequences. (XLSX 21 kb)

**Additional file 10: Figure S5.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *CBF-A5* and nine homologous amino acid sequences. Shown 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  $\beta$ -strands and the black spiral the  $\alpha$ -helix. The description of red arrows, black line and red dots is according to Fig. 6. (PDF 54 kb)

**Additional file 11: Figure S6.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *CBF-A10* gene and nine homologous amino acid sequences. 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  $\beta$ -strands and the black spiral the  $\alpha$ -helix. The description of red arrow, black line and red dots is according to Fig. 6. (PDF 53 kb)

**Additional file 12: Figure S7.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *CBF-A13* gene 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  $\beta$ -strands and the black spiral the  $\alpha$ -helix. The description of black line and red dots is according to Fig. 6. (PDF 53 kb)

**Additional file 13: Figure S8.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *CBF-A15* gene 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  $\beta$ -strands and the black spiral the  $\alpha$ -helix. The description of red arrows, black line and red dots is according to Fig. 6. (PDF 52 kb)

**Additional file 14: Figure S9.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *CBF-A14* gene 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  $\beta$ -strands and the black spiral the  $\alpha$ -helix. The description of black line and red dots is according to Fig. 6. (PDF 51 kb)

**Additional file 15: Figure S10.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *CBF-A18* gene 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  $\beta$ -strands and the black spiral the  $\alpha$ -helix. The description of black line and red dots is according to Fig. 6. (PDF 54 kb)

**Additional file 16: Figure S11.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *VRN-A1* gene and nine homologous amino acid sequences. Illustrated are alignments of three haplotype AA sequences of *VRN-A1* and 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 MAD5 box domain and K domain, the black arrows the  $\beta$ -strands and the spirals the  $\alpha$ -helices. The description of red arrow, black line and red dots is according to Fig. 6. (PDF 79 kb)

**Additional file 17: Figure S12.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *VRN-B3* gene 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  $\beta$ -strands and the spirals the  $\alpha$ -helices. The description of red arrow, black line and red dots is according to Fig. 6. (PDF 51 kb)

**Additional file 18: Figure S13.** Amino acid alignment and nucleotide divergence rates (dN/dS) of *PPD-B1* gene 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. 6. (PDF 8583 kb)

**Additional file 19: Figure S14.** AA alignment and nucleotide divergence rates (dN/dS) of *PPD-D1* gene 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. 6. (PDF 7539 kb)

## Abbreviations

AA: Amino acid; ANCOVA: Analysis of covariance; AP2/EREBP: APETALA2/Ethylene response element binding protein; BLAST: Basic local alignment search tool; CAB: Calcium-binding EF-hand family protein-like; CBF: C-repeat binding factor; CDS: Coding DNA sequence; CO: CONSTANS; COR/LEA: Cold-responsive/late embryogenesis-abundant; CRT/DRE: C-repeat/dehydration-responsive element; CV: Coefficient of variation; DEM: Defective embryo and meristems; DH: Doubled haploid; DN/DS: Nucleotide divergence rate; FR: Frost resistance; FT: Frost tolerance; GBP: Giga-base pairs; 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; IWGSC: International wheat genome sequencing consortium; KBP: Kilo-base pairs; LD: Linkage disequilibrium; LS: Least-Squares; M: Million; MAF: Minor allele frequency; MAFFT: Multiple Alignment tool Fast Fourier Transform; MCMC: Markov Chain Monte Carlo; MEF: Myocyte enhancer factor; MLM: Mixed linear model; MSA: Multiple sequence alignment; NGS: Next generation sequencing; NT: Nulli-tetrasomic; PCA: Principal component analysis; PCOA: Principal coordinate analysis; PPD: Photoperiod; PRR: Pseudo-response regulator; QTL: Quantitative trait locus; RV: Reaction volume; SNP: Single nucleotide polymorphism; T: Tonne; TACR7: *Triticum aestivum* cold-regulated 7; URGI: Unité de Recherche Génomique Info; UTR: Untranslated region K; number of populations; VRN: Vernalisation

## Acknowledgements

The authors thank the International Wheat Genome Sequencing Consortium for pre-publication access to IWGSC RefSeq v1.0, and Katy Niedung for excellent technical assistance. The authors thank to Majda Valjavec-Gratian from NCBI for help during preparation of sequences for submission to GenBank.

## Funding

The authors thank the German Federal Ministry of Education and Research (BMBF) for funding the project FROWHEAT (0315953). The field experiments in Novosibirsk were partially supported by ICG budget project #0324–2016-0001.

## Availability of data and materials

All data presented in this study, as well as GenBank accessions and cultivars will be publicly available. Sequences of individual haplotypes from representative

cultivars were deposited to GenBank under accession numbers MH264428-MH264501 to be available in the NCBI Nucleotide database.

## Authors' contributions

Conceived and designed the experiments: DP, FO and MK. Provided the experimental material: AB, MK, TP and YC. Performed the experiments: SB, FA, YC, TP, AB and MK. Analysed the data: SB, JK, TB, ES, PJ, SET and DP. Wrote the paper: SB, DP and FO. Study design, subject recruitment and sample preparation: DP, MK and FO. Data interpretation: SB, JK, TB, ES, PJ, SET, PK and DP. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

The material used in this study consisted mostly of registered varieties a/o lines publicly accessible. For all the accessions analysed no permissions or licenses were needed.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Julius Kühn-Institut (JKI), Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Erwin-Baur-Str. 27, 06484 Quedlinburg, Saxony-Anhalt, Germany. <sup>2</sup>Martin Luther University Halle-Wittenberg (MLU), Institute of Agricultural and Nutritional Sciences, Betty-Heimann-Str. 5, 06120 Halle (Saale), Saxony-Anhalt, Germany. <sup>3</sup>Julius Kühn-Institut (JKI), Federal Research Centre for Cultivated Plants, Institute for Biosafety in Plant Biotechnology, Erwin-Baur-Str. 27, 06484 Quedlinburg, Saxony-Anhalt, Germany. <sup>4</sup>Deutsche Saatveredelung AG (DSV), Weißburger Str. 5, 59557 Lippstadt, Nordrhein-Westfalen, Germany. <sup>5</sup>Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Resources Genetics and Reproduction, Correnstraße 3, 06466 Seeland OT Gatersleben, Saxony-Anhalt, Germany. <sup>6</sup>Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9B, 50931 Cologne, Nordrhein-Westfalen, Germany. <sup>7</sup>Agrophysical Research Institute (AFI), Grazhdanskii prosp. 14, 195220 St. Petersburg, Russia. <sup>8</sup>Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Prospekt Lavrentyeva 10, 630090 Novosibirsk, Russia. <sup>9</sup>Saaten-Union Biotech GmbH, Hovedisser Str. 94, 33818 Leopoldshoehe, Nordrhein-Westfalen, Germany. <sup>10</sup>Martin Luther University Halle-Wittenberg (MLU), Institute of Agricultural and Nutritional Sciences, Betty-Heimann-Str. 3, 06120 Halle (Saale), Saxony-Anhalt, Germany.

Received: 15 December 2017 Accepted: 14 May 2018

Published online: 29 May 2018

## References

- Food and Agriculture Organization of the United Nations. <http://www.fao.org/home/en/> (2014). Accessed 03 Feb 2015.
- Koemel JE, Guenzi AC, Anderson JA, Smith EL. Cold hardiness of wheat near-isogenic lines differing in vernalization alleles. *Theor Appl Genet.* 2004; 109:839–46.
- Dvorak J, Akhunov ED. Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the aegilops-triticum alliance. *Genetics.* 2005;171:323–32.
- Huang S, Sirikhachornkit A, Su XJ, Faris J, Gill B, Haselkorn R, Gornicki P. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci U S A.* 2002;99:8133–8.
- Peng JHH, Sun DF, Nevo E. Domestication evolution, genetics and genomics in wheat. *Mol Breed.* 2011;28:281–301.
- Smith DB, Flavell RB. Characterization of wheat genome by renaturation kinetics. *Chromosoma.* 1975;50:223–42.
- Choulet F, Wicker T, Rustenholz C, et al. Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and Transposable element spaces. *Plant Cell.* 2010;22:1686–701.
- Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, et al. Characterization of polyploid wheat genomic

- diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol J*. 2014;12(6):787–96.
9. Babben S, Perovic D, Koch M, Odon F. 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*. 2015;10:e0142746.
  10. Jordan KW, Wang S, Lun Y, et al. A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol*. 2015;16:48.
  11. School of Biological Sciences. University of Bristol. 2012. <http://www.cerealsdb.uk.net>. Accessed 01 Mar 2012.
  12. Wilkinson P, Winfield M, Barker G, Allen A, Burrige A, Coghill J, Edwards K. CerealsDB 2.0. An integrated resource for plant breeders and scientists. *BMC Bioinformatics*. 2012;13:219.
  13. Unité de Recherche Génomique Info. Centre de Versailles (2013). <http://wheat-urgi.versailles.inra.fr>. Accessed 20 Feb 2013.
  14. Unité de Recherche Génomique Info. Centre de Versailles (2013). <https://wheat-urgi.versailles.inra.fr/Seq-Repository/>. Accessed 20 Feb 2013.
  15. Brenchley R, Spannagl M, Pfeifer M, et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*. 2012;491:705–10.
  16. The International Wheat Genome Sequencing Consortium. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*. 2014;345:1251788.
  17. Raats D, Frenkel Z, Krugman T, et al. The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. *Genome Biol*. 2013;14:R138.
  18. Ling HQ, Zhao S, Liu D, et al. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*. 2013;496:87–90.
  19. Jia JZ, Zhao SC, Kong XY, et al. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*. 2013;496:91–5.
  20. Winfield MO, Allen AM, Burrige AJ, Barker GLA, et al. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol J*. 2016;14(5):1195–206.
  21. Allen AM, Winfield MO, Burrige AJ, et al. Characterization of a wheat breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnol J*. 2016;15(3):390–401.
  22. Li Y, Boeck A, Haseneyer G, Korzun V, Wilde P, Schoen CC, Ankerst DP, Bauer E. Association analysis of frost tolerance in rye using candidate genes and phenotypic data from controlled, semi-controlled, and field phenotyping platforms. *BMC Plant Biol*. 2011;11:146.
  23. Orvar BL, Sangwan V, Omann F, Dhindsa RS. Early steps in cold sensing by plant cells. The role of actin cytoskeleton and membrane fluidity. *Plant J*. 2000;23:785–94.
  24. Sangwan V, Foulds I, Singh J, Dhindsa RS. Cold-activation of *Brassica napus* BN115 promoter is mediated by structural changes in membranes and cytoskeleton, and requires Ca<sup>2+</sup> influx. *Plant J*. 2001;27:1–12.
  25. Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK. ICE1. A regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes Dev*. 2003;17:1043–54.
  26. Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. 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*. 1998;10:1391–406.
  27. Stockinger EJ, Gilmour SJ, Thomashow MF. *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. *Proc Natl Acad Sci U S A*. 1997;94:1035–40.
  28. Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M. 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 J*. 1998;17:5484–96.
  29. Dietz KJ, Vogel MO, Viehhauser A. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma*. 2010;245:3–14.
  30. Peng YL, Wang YS, Cheng H, Sun CC, Wu P, Wang LY, Fei J. Characterization and expression analysis of three CBF/DREB1 transcriptional factor genes from mangrove *Avicennia marina*. *Aquat Toxicol*. 2013;140:68–76.
  31. Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF. Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiol*. 2001;127:910–7.
  32. Allagulova CR, Gimalov FR, Shakirova FM, Vakhitov VA. The plant dehydrins. Structure and putative functions. *Biochemistry-Moscow*. 2003;68:945–51.
  33. Close TJ. Dehydrins. A commonality in the response of plants to dehydration and low temperature. *Physiol Plant*. 1997;100:291–6.
  34. Winfield MO, Lu CG, Wilson ID, Coghill JA, Edwards KJ. Plant responses to cold. Transcriptome analysis of wheat. *Plant Biotechnol J*. 2010;8:749–71.
  35. Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF. Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *Plant J*. 1998;16:433–42.
  36. Dhillon T, Pearce SP, Stockinger EJ, Distelfeld A, Li CX, Knox AK, Vashegyi I, Vagujfalvi A, Galiba G, Dubcovsky J. Regulation of freezing tolerance and flowering in temperate cereals. The VRN-1 connection. *Plant Physiol*. 2010;153:1846–58.
  37. Kato K, Yamagata H. Method for evaluation of chilling requirement and narrow-sense earliness of wheat cultivars. *Jpn J Breed*. 1988;38:172–86.
  38. Turner A, Beales J, Faure S, Dunford RP, Laurie DA. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science*. 2005;310:1031–4.
  39. Matsushika A, Makino S, Kojima M, Mizuno T. Circadian waves of expression of the APRR1/TOC1 family of pseudo-response regulators in *Arabidopsis thaliana*. Insight into the plant circadian clock. *Plant Cell Physiol*. 2000;41:1002–12.
  40. Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Mas P, Panda S, Kreps JA, Kay SA. Cloning of the *Arabidopsis* clock gene TOC1, an autoregulatory response regulator homolog. *Science*. 2000;289:768–71.
  41. Danyluk J, Kane NA, Breton G, Limin AE, Fowler DB, Sarhan F. TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. *Plant Physiol*. 2003;132:1849–60.
  42. Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES. MADS box genes control vernalization-induced flowering in cereals. *Proc Natl Acad Sci U S A*. 2003;100:13099–104.
  43. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene VRN1. *Proc Natl Acad Sci U S A*. 2003;100:6263–8.
  44. Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. *Science*. 2004;303:1640–4.
  45. Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J. The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proc Natl Acad Sci U S A*. 2006;103:19581–6.
  46. Francia E, Rizza F, Cattivelli L, Stanca AM, Galiba G, Toth B, Hayes PM, Skinner JS, Pecchioni N. Two loci on chromosome 5H determine low-temperature tolerance in a 'Nure' (winter) x 'Tremois' (spring) barley map. *Theor Appl Genet*. 2004;108:670–80.
  47. Vagujfalvi A, Galiba G, Cattivelli L, Dubcovsky J. The cold-regulated transcriptional activator Cbf3 is linked to the frost-tolerance locus Fr-A2 on wheat chromosome 5A. *Mol Gen Genomics*. 2003;269:60–7.
  48. Zhao YS, Gowda M, Wurschum T, et al. Dissecting the genetic architecture of frost tolerance in central European winter wheat. *J Exp Bot*. 2013;64:4453–60.
  49. Galiba G, Quarrie SA, Sutka J, Morgounov A, Snape JW. RFLP mapping of the vernalization (*vrn1*) and frost-resistance (*fr1*) genes on chromosome 5A of wheat. *Theor Appl Genet*. 1995;90:1174–9.
  50. Sutka J, Galiba G, Vagujfalvi A, Gill BS, Snape JW. Physical mapping of the *Vrn-A1* and *Fr1* genes on chromosome 5A of wheat using deletion lines. *Theor Appl Genet*. 1999;99:199–202.
  51. Stockinger EJ, Skinner JS, Gardner KG, Francia E, Pecchioni N. Expression levels of barley Cbf genes at the frost resistance-H2 locus are dependent upon alleles at *Fr-H1* and *Fr-H2*. *Plant J*. 2007;51:308–21.
  52. Santra DK, Santra M, Allan RE, Campbell KG, Kidwell KK. 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 Breed*. 2009;128:576–84.
  53. Chen YH, Carver BF, Wang SW, Zhang FQ, Yan LL. Genetic loci associated with stem elongation and winter dormancy release in wheat. *Theor Appl Genet*. 2009;118:881–9.

54. Chen Y, Carver BF, Wang S, Cao S, Yan L. Genetic regulation of developmental phases in winter wheat. *Mol Breed*. 2010;26:573–82.
55. Eagles HA, Cane K, Trevaskis B. Veery wheats carry an allele of Vrn-A1 that has implications for freezing tolerance in winter wheats. *Plant Breed*. 2011;130:413–8.
56. Reddy L, Allan RE, Campbell KAG. Evaluation of cold hardiness in two sets of near-isogenic lines of wheat (*Triticum aestivum*) with polymorphic vernalization alleles. *Plant Breed*. 2006;125:448–56.
57. Baga M, Chodaparambil SV, Limin AE, Pecar M, Fowler DB, Chibbar RN. Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. *Funct Integr Genomics*. 2007;7:53–68.
58. Motomura Y, Kobayashi F, Iehisa JCM, Takumi S. A major quantitative trait locus for cold-responsive gene expression is linked to frost-resistance gene Fr-A2 in common wheat. *Breed Sci*. 2013;63:58–67.
59. Vagujfalvi A, Aprile A, Miller A, Dubcovsky J, Delugu G, Galiba G, Cattivelli L. The expression of several Cbf genes at the Fr-A2 locus is linked to frost resistance in wheat. *Mol Gen Genomics*. 2005;274:506–14.
60. Miller AK, Galiba G, Dubcovsky J. A cluster of 11 CBF transcription factors is located at the frost tolerance locus Fr-a(m)2 in *Triticum monococcum*. *Mol Gen Genomics*. 2006;275:193–203.
61. Sutton F, Chen DG, Ge XJ, Kenefick D. 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 Biol*. 2009;9:34.
62. Knox AK, Li CX, Vagujfalvi A, Galiba G, Stockinger EJ, Dubcovsky J. Identification of candidate CBF genes for the frost tolerance locus Fr-a(m)2 in *Triticum monococcum*. *Plant Mol Biol*. 2008;67:257–70.
63. Soltesz A, Smedley M, Vashegyi I, Galiba G, Harwood W, Vagujfalvi A. Transgenic barley lines prove the involvement of TaCBF14 and TaCBF15 in the cold acclimation process and in frost tolerance. *J Exp Bot*. 2013; 64:1849–62.
64. Kocsy G, Athmer B, Perovic D, Himmelbach A, Szucs A, Vashegyi I, Schweizer P, Galiba G, Stein N. Regulation of gene expression by chromosome 5A during cold hardening in wheat. *Mol Gen Genomics*. 2010;283:351–63.
65. Gulick P, Drouin S, Yu Z, Poisson G, Monroy AF, Sarhan F. Transcriptome comparison of winter and spring wheat responding to low temperature. *Genome*. 2005;48:913–23.
66. Monroy AF, Dryanova A, Malette B, Oren DH, Farajalla MR, Liu W, Danyluk J, Ubayasena LWC, Kane K, Scoles GJ, Sarhan F, Gulick PJ. Regulatory gene candidates and gene expression analysis of cold acclimation in winter and spring wheat. *Plant Mol Biol*. 2007;64:409–23.
67. Laudencia-Chinguanco D, Ganeshan S, You F, Fowler B, Chibbar R, Anderson O. Genome-wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (*Triticum aestivum* L.). *BMC Genomics*. 2011;12:299.
68. Ganeshan S, Sharma P, Young L, Kumar A, Fowler DB, Chibbar RN. Contrasting cDNA-AFLP profiles between crown and leaf tissues of cold-acclimated wheat plants indicate differing regulatory circuitries for low temperature tolerance. *Plant Mol Biol*. 2011;75:379–98.
69. Winfield MO, Lu C, Wilson ID, Coghill JA, Edwards KJ. Cold- and light-induced changes in the transcriptome of wheat leading to phase transition from vegetative to reproductive growth. *BMC Plant Biol*. 2009;9:55.
70. Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A. Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol Breed*. 2011;27(1):37–58.
71. Endo TR, Gill BS. The deletion stocks of common wheat. *J Hered*. 1996; 87:295–307.
72. Sears ER. Nullisomic-tetrasomic combinations in hexaploid wheat. In: Riley R, Lewis KR, editors. *Chromosome manipulations and plant genetics*. New York: Springer Science+Business Media; 1966. p. 29–45.
73. Stein N, Herren G, Keller B. A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. *Plant Breed*. 2001;120:354–6.
74. Pearson K. On certain errors with regard to multiple correlation occasionally made by those who have not adequately studied this subject. *Biometrika*. 1914;10:181.
75. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74:5463–7.
76. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT. A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059–66.
77. Li YL, Haseneyer G, Schon CC, Ankerst D, Korzun V, Wilde P, Bauer E. High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biol*. 2011;11:6.
78. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
79. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE. A simulation study. *Mol Ecol*. 2005;14:2611–20.
80. Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*. 2014;3:431.
81. Perrier X, Jacquemoud-Collet JP. DARwin software. 2006. <http://darwin.cirad.fr/>.
82. Goodman MM, Stuber CW. Races of maize. 6. Isozyme variation among races of maize in Bolivia. *Maydica*. 1987;28:169–87.
83. Reif JC, Melchinger AE, Frisch M. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Sci*. 2005;45:1–7.
84. Wright S. Evolution and the genetics of populations. A treatise in four volumes. Vol. 4, variability within and among natural population. Chicago, London: University of Chicago Press; 1978. p. 91.
85. The R Project for Statistical Computing. R software. <https://www.r-project.org>.
86. Buckler Lab for Maize Genetics and Diversity. TASSEL software. <http://www.maizegenetics.net/tassel>.
87. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL. Software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;23:2633–5.
88. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25:1189–91.
89. National Center for Biotechnology Information. U.S. National Library of Medicine, Rockville pike. 2017. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Accessed 12 May 2017.
90. RaptorX: a Web Portal for Protein Structure and Function Prediction. Computational Institute at the University of Chicago, Chicago. 2017. <http://raptorx.uchicago.edu/StructurePrediction/predict/>. Accessed 15 May 2017.
91. Kaellberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*. 2012; 7:1511–22.
92. Peng J, Xu J. RaptorX. Exploiting structure information for protein alignment by statistical inference. *Proteins Struct Funct Bioinf*. 2011;79:161–71.
93. Delpot W, Poon AF, Frost SDW, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*. 2010;26:2455–7.
94. Kosakovsky Pond SL, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2005;21:676–9.
95. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005;33:511–8.
96. Kumar S, Stecher G, Peterson D, Tamura K. MEGA-CC: computing of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*. 2012;28:2685–6.
97. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006;34:W609–12.
98. SOGO Genome Information Database System for Innovation of Crop and Livestock Production. Institute of Crop Science NARO, Tsukuba. 2017. <https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?lang=en>. Accessed 09 Jul 2017.
99. Higo K, Ugawa Y, Iwamoto M, Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res*. 1999;27:297–300.
100. Softberry. Softberry Inc., Mount Kisco. 2017. [http://www.softberry.com/berry.phtml?topic=ann2\\_ann3&no\\_menu=on](http://www.softberry.com/berry.phtml?topic=ann2_ann3&no_menu=on). Accessed 09 Jul 2017.
101. Solovyyev V, Shahmuradov I, Salamov A. Identification of promoter regions and regulatory sites. In: Ladunga I, editor. *Computational biology of transcription factor binding*. Totowa: Springer Science+Business Media; 2010. p. 57–83.
102. Theissen G, Kim JT, Saedler H. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J Mol Evol*. 1996;43:484–516.
103. Kaufmann K, Melzer R, Theissen G. MIKC-type MADS-domain proteins. Structural modularity, protein interactions and network evolution in land plants. *Gene*. 2005;347:183–98.

104. Taoka K, Ohki I, Tsuji H, et al. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature*. 2011;476:332–5.
105. Tanksley SD, McCouch SR. Seed banks and molecular maps. Unlocking genetic potential from the wild. *Science*. 1997;277:1063–6.
106. Knox AK, Dhillon T, Cheng HM, Tondelli A, Pecchioni N, Stockinger EJ. CBF gene copy number variation at frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. *Theor Appl Genet*. 2010;121:21–35.
107. Zhu J, Pearce S, Burke A, See D, Skinner D, Dubcovsky J, Garland-Campbell K. Copy number and haplotype variation at the VRN-A1 and central FR-A2 loci are associated with frost tolerance in hexaploid wheat. *Theor Appl Genet*. 2014;127:1183–97.
108. Campoli C, Matus-Cadiz MA, Pozniak CJ, Cattivelli L, Fowler DB. Comparative expression of Cbf genes in the Triticeae under different acclimation induction temperatures. *Mol Gen Genomics*. 2009;282:141–52.
109. Diaz A, Zikhali M, Turner AS, Isaac P, Laurie DA. Copy number variation affecting the photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS One*. 2012;7:e33234.
110. Puranik S, Acaijaoui S, Conn S, et al. Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in *Arabidopsis*. *Plant Cell*. 2014;26:3603–15.
111. Bachhawat P, Stock AM. Crystal structures of the receiver domain of the response regulator PhoP from *Escherichia coli* in the absence and presence of the phosphoryl analog Berylofluoride. *J Bacteriol*. 2007;189:5987–95.
112. Wu Y, Dey R, Han A, Jayathilaka N, Philips M, Ye J, Chen L. Structure of the MADS-box/MEF2 domain of MEF2A bound to DNA and its implication for Myocardin recruitment. *J Mol Biol*. 2010;397:520–33.
113. Santelli E, Richmond TJ. Crystal structure of MEF2A core bound to DNA at 1.5 angstrom resolution. *J Mol Biol*. 2000;297:437–49.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

