Theor Appl Genet (2015) 128:131–143
DOI 10.1007/s00122-014-2417-5

ORIGINAL PAPER

A yield-associated gene *TaCWI*, in wheat: its function, selection and evolution in global breeding revealed by haplotype analysis

Yanmiao Jiang · Qiyan Jiang · Chenyang Hao · Jian Hou · Lanfen Wang · Hongna Zhang · Suna Zhang · Xinhong Chen · Xueyong Zhang

Received: 10 July 2014 / Accepted: 11 October 2014 / Published online: 4 November 2014

Abstract

Key message **Wheat anther-specific invertase genes were haplotyped in wheat. Strong allelic selection occurred during wheat polyploidization, domestication and breeding because of their association with yield traits.**

Abstract Plant invertase hydrolyzes sucrose into glucose and fructose. Cell wall invertase (CWI), one of the three types of invertase, is essential for plant development. Based on isolated *TaCWI* genes from chromosomes 4A, 5B and 5D, two SNPs were detected in the promoter region of *TaCWI-4A*, and four SNPs and two Indels were present in the *TaCWI-5D* gene. No polymorphism was detected in *TaCWI-5B* coding or promoter regions. CAPS markers *caps4A* and *caps5D* were developed to discriminate haplotypes of *TaCWI-4A* and *TaCWI-5D*. Marker/trait association analysis indicated that *Hap-5D-C* at *TaCWI-5D* was significantly associated with higher thousand kernel weight (TKW) in 348 Chinese modern cultivars grown in multiple environments. Geographic distributions and changes over time of favored haplotypes showed that *Hap-5D-C* was the most frequent haplotype in modern cultivars and was strongly positively selected in six major wheat production regions worldwide. However, selection for haplotypes at *TaCWI-4A* was not so evident, possibly due to balancing effects of the two haplotypes on TKW and grain number per spike (GN). In rainfed production regions, *Hap-4A-C* was favored because it brought more seeds, but in well irrigated conditions, *Hap-4A-T* was favored in modern breeding because of higher TKW. Evolutionary analysis among wheat and its relatives showed that genetic diversity of *TaCWI* genes on chromosomes 4A and 5D declined dramatically in progression from the diploid level to modern polyploid cultivars. There was strong allelic selection during polyploidization, domestication and breeding.

Communicated by Mark E. Sorrells.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2417-5) contains supplementary material, which is available to authorized users.

Y. Jiang · X. Chen (✉)
Shaanxi Key Laboratory of Genetic Engineering for Plant Breeding, College of Agronomy, Northwest A and F University, Yangling 712100, Shaanxi, China
e-mail: cxh2089@126.com

Y. Jiang
e-mail: jiangyanm12@163.com

Y. Jiang · Q. Jiang · C. Hao · J. Hou · L. Wang · H. Zhang · S. Zhang · X. Zhang (✉)
Key Laboratory of Crop Gene Resources and Germplasm Enhancement, Ministry of Agriculture/The National Key Facility for Crop Gene Resources and Genetic Improvement/Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China
e-mail: zhangxueyong@caas.cn

Abbreviations

ANOVA	One-way analysis of variance
ETN	Effective tiller number
GN	Grain number per spike
HD	Heading date
MD	Maturity date
PH	Plant height
TKW	Thousand kernel weight
SpL	Spike length

Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most important staple food crops worldwide. High yield is the

major breeding objective because of the increasing human population and decreasing cultivation areas. Generally, wheat yield is closely associated with grain number per unit area and grain weight (Kumbhar et al. 1983). In China, a 140–160 kg per hectare increase in yield can be obtained when TKW is increased by just one gram (Tian et al. 2006). However, the genes influencing grain weight and underlying molecular mechanisms controlling wheat grain weight are still largely unknown.

Recent studies showed that CWI (cell wall invertase) is closely related to crop yield. CWI is one of three types of invertase (β-fructofuranosidase, EC 3.2.1.26), a hydrolase that catalyzes the irreversible conversion of sucrose to glucose and fructose. In most plants, sucrose as one of the assimilated carbon forms is transported from photosynthetic to non-photosynthetic tissues (Sturm and Tang 1999) where it can be used directly or cleaved into hexoses by sucrose synthase or invertase (Copenald 1990). CWI is considered to be one of the key enzymes involved in establishing sink strength in various sink tissues (Sturm 1996; Sturm and Tang 1999). In higher plants, invertases were classified into cytoplasmic, vacuolar, and cell wall types based on solubility, localization and pH optima (Sturm 1996; Tymowska-Lalanne and Kreis 1998). There are two common features on their amino acid sequences: the β-fructofuranosidase motif (NDPNG/A) and a Cys residue (MWECP) (Sturm 1999).

In rice the *OsGIF* gene (*OsCIN2*) is required for carbon partitioning during early grain development. In *gif1* mutant grains, a reduced rate of grain filling rate started 3 days after pollination (DAP). The final grain weight of the *gif1* mutant was 24 % lower than that of wild type at 30 DAP (Wang et al. 2008). In maize (*Zea mays* L.) lack of expression of endosperm-specific cell wall-invertase gene *Incw2* in the pedicel and endosperm caused interruption photoassimilate transport into developing kernels, resulting in a seed weight that was only one-fifth of the normal weight (Miller and Chourey 1992; Cheng et al. 1996; Kang et al. 2009). In carrot (*Daucus carota* L.) antisense expression of cell-wall invertase caused the dry weight leaf-to-root ratio of cell wall invertase antisense plants to change from 1:3 to 17:1 (Tang et al. 1999). Besides influencing vegetative growth and grain filling, invertase also affects anther development. Down-regulation of anther-specific *OsINV4* expression in the tapetums of rice anthers by cold treatment of young microspores decreased anther vitality and resulted in increased numbers of empty grains and yield loss (Oliver et al. 2005). In rice, eight *CWIs* (Cho et al. 2005) were identified. Studies on function and mechanism of wheat CWI genes are very limited compared to rice and tobacco. Ma et al. (2012) reported that *TaCwi-A1* was associated with kernel weight. Using a cell wall invertase-specific motif sequence as a probe, Webster et al. (2012) screened a wheat genomics database to find the unidentified *IVR1* isoforms. Other reports focused on the relationship between invertase and anther development following cold, water and heat stresses (Dorion et al. 1996; Koonjul et al. 2005; Ji et al. 2010).

The main objectives of the present study were to characterize allelic variation at *TaCWI* loci on chromosomes 4A, 5B and 5D, develop functional markers to characterize variation in the genes, and evaluate the association of the variation and yield. Haplotype frequencies in varieties released in the major global wheat production areas during the last century were also examined. We also considered haplotype evolution during the polyplodization and domestication of wheat.

Materials and methods

Cloning and chromosome mapping of *TaCWI*s

The *OsGIF* (*OsCWI2*) cDNA sequence (GenBank accession No: GU797938) was used for blasting wheat ESTs in NCBI, and a full-length wheat cell wall invertase cDNA (GenBank no. AF030420) was found. Based on this sequence, primers were designed to amplify genomic sequences from *T. aestivum* cv. Chinese Spring (CS), *Triticum urartu*, *Aegilops speltoides* and *Aegilops tauschii*.

To obtain the sequence of the *TaCWI* promoter region, a Chinese Spring bacterial artificial chromosome (BAC) library was screened by a PCR-based method. The DNA of the target single BAC clone was isolated by a large construct kit (Qiagen) and a shot-gun library was constructed for sequencing. Based on sequence polymorphisms of *TaCWI-A*, *-B*, and *-D*, genome-specific primers (Table 1) were designed for mapping *TaCWI* in a set of CS nulli-tetrasomic lines.

TaCWI tissue expression analysis and promoter element analysis

Total RNA was extracted from roots, stems, leaves, leaf sheaths, anthers, spikelets, seeds (5 and 14 days after pollination), ovaries and glumes using TRIzol reagent. DNA was removed by digestion with *DNAaseI* (Fermentas) before reverse transcription. The mRNAs were used for tissue-specific *TaCWI* expression analysis. First-strand cDNA was synthesized using M-MLV transcriptase (Invitrogen) according to the manufacturer's instructions. 1 μL of product was used as a PCR template. This was performed with SYBR Premix Ex-Taq [TaKaRa Biotechnology (Dalian) Co. Ltd, Product Code: DRR041A] by primer cwi-rt and GAPDH as a control. RT-PCR were performed in total volumes of 20 μL, including 1 μL cDNA, 10 μL of 2× SYBR Premix Ex-Taq, and 0.4 μL of each primer (10 μM). Putative functional promoter elements were analyzed using the PLACE database (Higo et al. 1999).

Table 1 Primer sequences used in this study

Primer name	Sequence (5′-3′)	Tm (°C)	Purpose
Ags	For AGGGCGTCCGACCAAAGTG	60	A genome-specific
	Rev GCGACCTAGCGTGTATCAAGGAG		
Bgs	For AGTACGTATGTTACTCACTCGACAC	59	B genome-specific
	Rev TGACTCGCCATTGTTGAAGAC		
Dgs	For CATGTGCCTCTAAAATTAGGTTATG	58	D genome-specific
	Rev CGCCCAGATGATGTTCC		
cwi-rt	For TGTCAAGAGCGGCGAGTT	60	Expression analysis
	Rev CGAGCAGCCAGAGTCCAA		
GAPDH	For AGGGTGGTGCAAAGAAGGTCA	60	Expression control
	Rev GATCCCCACTCGTTGTCGTA		

Table 2 Genes used in the phylogenic analysis

Species	Gene	Accession number	Species	Gene	Accession number
Wheat	*CWINV1SM*	AB196522	Maize	*ZmIncw4*	AF043347
	TaCWINV2SM	AB196523		*Zm2G095725 P02*	AFW58192
T. urartu	*TuCWI1*	gi473890652		*Zm2G174249*	DAA55698
	TuCWI2	gi474416113		*Zm2G018716_P01*	NP_001145776
	TuCWI3	gi474404617	Rice	*OsCIN1–OsCIN8*	AY578158-65
	TuCWI4	gi474270132	Barley	*HvCWINV1*	CAD58960
	TuCWI5	gi474403046		*Hv-AK364085*	BAJ95288
Ae. tauschii	*AeCWI1*	gi475504165		*Hv-MLOC_57254*	BAK08333
	AeCWI2	gi475519402		*Hv-MLOC_64782*	BAJ88320
	AeCWI3	gi475544102		*Hv-MLOC_80162*	BAJ93390
	AeCWI4	gi475503462	*B. distachyon*	*BdCWI7-like*	XP_003580762
	AeCWI7	gi475612419		*BdCWI4-like*	XP_003565145
Maize	*ZmIncw1*	U17695		*BdCWI3-like*	XP_003581207
	ZmIncw2	AF050631		*BdCWI2-like*	XP_003579704
	ZmIncw3	AF043346		*BdCWI1-like*	XP_003575078

Table 3 Wheat cultivars used in analysis of *TaCWI* sequence polymorphism

Cultivar	*CAPS-4A*	*CAPS-5D*	Cultivar	*CAPS-4A*	*CAPS-5D*
Zhongyou9507	T	G	Nongda139	T	C
Jinmai8	T	C	Mingxian169	T	G
Beijing8	T	C	Jingyang60	C	C
Shijiazhuang54	T	C	Aifeng3	T	C
Yannong15	T	C	Lumai1	T	C
Laizhou953	T	C	Beijing15	C	G
Lankao906	C	C	Jinmai11	C	C
Xuzhou22	T	C	CS	T	C

Phylogenetic tree construction of orthologous *CWI* genes in monocots

The intact cell wall invertase gene sequences of *Oryza sativa*, *T. aestivum*, *T. urartu* and *Aegilops tauschii, Zea mays* (maize), *Hordeum vulgare* (barley), and *Brachypodium distachyon* were used for phylogenic analysis. The deduced amino acid sequences of the CWIs were aligned using the ClustalW program (Thompson et al. 1994). Phylogenetic and molecular evolutionary analyses were conducted with software MEGA 4 (Kumar et al. 2008). Accession numbers of sequences used to construct the phylogenetic tree are listed in Table 2.

TaCWI diversity in wheat cultivars

Sixteen cultivars (Table 3) with significant differences in TKW were selected for detecting polymorphism in *TaCWI*. Genomic DNA was extracted from young leaves using the CTAB method following Chen and Ronalds (1999). Genome-specific primer pairs were designed by the software Primer Premier 5.0 (Premier Biosoft International, Palo Alto, CA) and all primers were synthesized by Sangon Biotech Co. Ltd (China). PCR was performed in a total volume of 15 μL, including 10 μM of each primer, 120 μM of dNTP, 30 ng of template, 0.75 units LA Tag, and 7.5 μL 2× GC Buffer (Takara Bio). The PCR procedure was 95 °C for 3 min, followed by 35 cycles of 95 °C for 30 s, annealing (55–65) for 45 s, and extension at 72 °C (30 s to 2 min), and final extension at 72 °C for 10 min. The PCR products were separated by electrophoresis in agarose gels, and the target bands were recovered and cloned into the pEASY-T1 vector and transformed into DH5α competent *E. coli* cells by the heat shock method (Beijing TransGen Biotech Co., Ltd). Positive clones were selected and sequenced by an ABI 3730XI DNA Analyzer (Applied Biosystems). The sequences were assembled by the SeqMan program and compared by MegAlign (DNASTAR Pro, Madison, WI).

SNP marker development and linkage mapping

SNP sites were transformed to cleaved amplified polymorphic sequence (CAPS) markers (King et al. 1993). The digestion sites of restriction endonucleases were detected by NEB-cutter V2.0 (Vincze et al. 2003). Genomic-specific primers (Table 1) were used to amplify specific genome sequences. Restriction endonucleases used in the study were produced by Fermentas life sciences.

Recombinant inbred line (RIL) populations with 184 lines derived from Xiaoyan54 × Jing411 and 135 lines derived from Nanda 2419 × Wangshuibai, was used for genetic linkage mapping of *TaCWI-4A* and *TaCWI-5D,* respectively.

Marker/trait association analyses

The Chinese wheat mini-core collection (MCC, 262 wheat accessions) and 348 modern cultivars released since the 1940s were employed for functional validation of the *TaCWI* markers (Zhang et al. 2002; Dong et al. 2003; Hao et al. 2008). These accessions were planted in several growing seasons. Each accession was planted in 2 × 1 m rows spaced 25 cm apart, with 40 plants in each row. Field management followed local agricultural practices. Various traits, including effective tiller number (ETN), TKW, GN, HD, MD, PH, spike length (SpL), spikelet number (SN), and first internode length (FIN) were measured in three cropping seasons to examine the association of each trait with *TaCWI* haplotype in 89 MCC modern cultivars, 156 MCC landraces and 348 Chinese modern cultivars (Hou et al. 2014; Qin et al. 2014).

TKW, GN and yield data for nested sets of selected breeding lines (21 *Hap-4A-C*, 17 *Hap-4A-T* and 6 C/T heterozygotes) and were used to estimate the genetic effects of each haplotype of *TaCWI-4A* on GN and TKW. These lines had Han 6172 and Zhoumai 18 as their predominant genetic backgrounds. They were planted and evaluated in 9.2-m^2 plots under irrigation at the CAAS Experiment Station at Xinxiang, Henan Province, in 2012 and 2013. Each line was sown in 6.1 × 1.5 m, with 25 cm between lines. Their pedigrees, haplotypes, fertile tiller numbers, GN, TKW, and yield are given in Table S1. Six near isogenic lines (3 *Hap-4A-C*, 3 *Hap-4A-T* and 3 C/T heterozygotes) selected from inbred lines derived from a Han 6172 was used to further validate the effect of haplotypes at *TaCWI-4A*. They were planted in yield plots in the 2013–14 crop seasons.

Estimation of haplotype effects

One-way analyses of variance (ANOVA) were performed using SPSS for Windows version 16.0 (SPSS Inc., Chicago, IL, USA) to separately determine differences in agronomic traits between haplotypes of landraces and modern cultivars. Previous study had revealed that the Chinese landraces and modern cultivars clustered into two subgroups (Hao et al. 2008).

Geographical distribution of *TaCWI* haplotypes in global wheat cultivars

The Chinese wheat production area is divided into ten ecological zones based on growing season, cultivar type and environment (Zhang et al. 2002; Zhuang 2003). Zones I–VIII produce more than 85 % of the national wheat production. The 158 landraces in the MCC and 348 modern cultivars were used to analyze the distribution of *TaCWI* haplotypes in China. Frequencies of *TaCWI-4A* and *TaCWI-5D* haplotypes were also estimated in cultivar groups released in USA, Canada, Mexico, Australia, Europe, and the former USSR in the last century.

European cultivars were obtained from the Clermont-Ferrand Genetic Resources Center (http://www.clermont.inra.fr/umr-asp) as part of an agreement between the French National Institute for Agronomical Research (INRA) and the Chinese Academy of Agricultural Science (CAAS). In addition, 436 wheat cultivars from USA, 54 from Canada, 53 from Mexico, 83 from Russia, 51 from Australia were used to study the global distribution of *TaCWI* haplotypes.

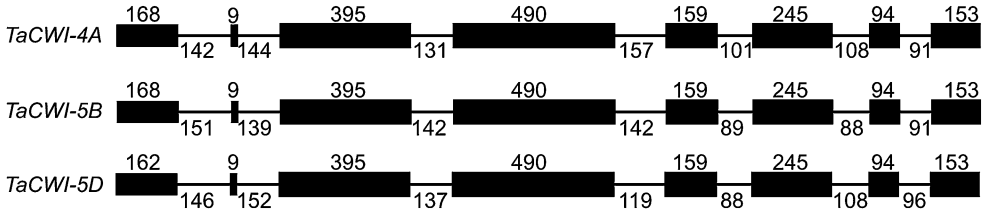

Fig. 1 Exon–intron structures of the three *TaCWI* homoeologous genes in Chinese Spring. *Numbered solid blocks* denote exons, and the *lines* between exons represent introns. *Numbers* under exons and introns indicate size (bp)

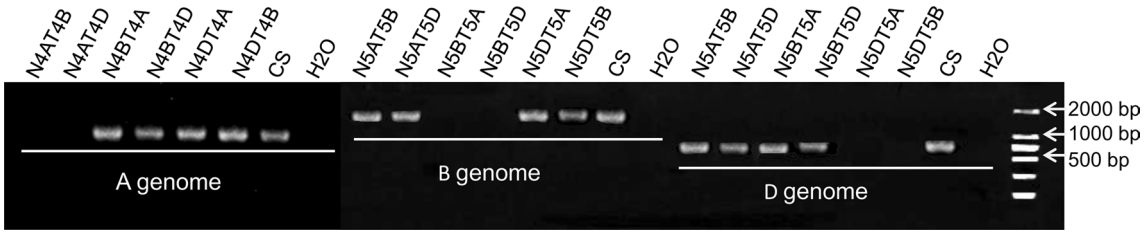

Fig. 2 PCR-based chromosome mapping of the three *TaCWI* homoeologous genes in CS nulli-tetrasomic lines

Nucleotide diversity in *TaCWIs* in bread wheat and its diploid and tetraploid relatives

The promoters (about 2,000 bp upstream of ATG) and gene sequences of *TaCWIs* in bread wheat and its diploid and tetraploid relatives (Table S2) were compared in nucleotide sequence. Sequence alignment was carried out by Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Tests of nucleotide diversity (π), *Tajima's D* and population differentiation were carried out by DnaSP5.10 (http://www.ub.es/dnasp). Haplotype networks were constructed based on the *TaCWI* DNA sequences using the program TCS1.21 (Clement et al. 2000).

Results

TaCWIs with high homology were mapped on wheat chromosomes 4A, 5B and 5D

The genome sequences of *TaCWIs* were obtained based on the full-length wheat cell wall invertase cDNA (GenBank No. AF030420). Both A and B genome cDNA are 1,755 bp and predicted to encode 585 amino acids. The D genomic cDNA is 1,707 bp, and encodes 569 amino acids. The predicted protein sequence included an NDPNG motif and MWECP sequences, which are hallmarks of cell wall invertases (Fig. S1). Comparisons of *TaCWI* cDNAs with PCR amplified genomic DNA sequences by MegAlign program, showed that the three *TaCWI* orthologues share a homologous structure with 8 exons and 7 introns. The second exon is 9 bp, consistent with reported CWI gene characteristics (Sturm 1999). The lengths of the introns differed among the three genes (Fig. 1), which were located on chromosomes 4A, 5B, and 5D (Fig. 2) using genome-specific primer pairs (Table 1). Part of chromosome 4A in modern common wheat carries a translocated segment from 5A (Liu et al. 1992, discussed later).

TaCWI-5 has anther-specific expression

Spatial expression of the *TaCWI* was examined by real time-PCR using primer pair cwi-rt and GAPDH as control (Table 1). Transcripts were detected in various wheat tissues including roots, stems, leaves, leaf sheaths, anthers, spikelets, seeds at DAP5 and DAP14, ovaries and glumes. The result suggested that *TaCWI* was expressed only in anthers (Fig. 3a).

The upstream sequences of the three *TaCWI* orthologous genes were analyzed using PLACE database. Motifs POLLEN1LELAT52 (AGAAA) (Bate and Twell 1998) and GTGANTG10 (GTGA) (Rogers et al. 2001), cis-acting regulatory elements known to be necessary for pollen/anther-specific expression, were detected (Fig. 3b). All three promoters contain multiple copies of both motifs, suggesting that genes driven by them should be pollen- or anther-specific.

CWI genes with the same function clustered into the same subgroup

To compare the phylogenetic relationship of the *TaCWIs* with CWIs from other species, we constructed a phylogenetic tree using the neighbor-joining method based on the CWI protein sequences (Fig. S2). The phylogenetic tree displayed

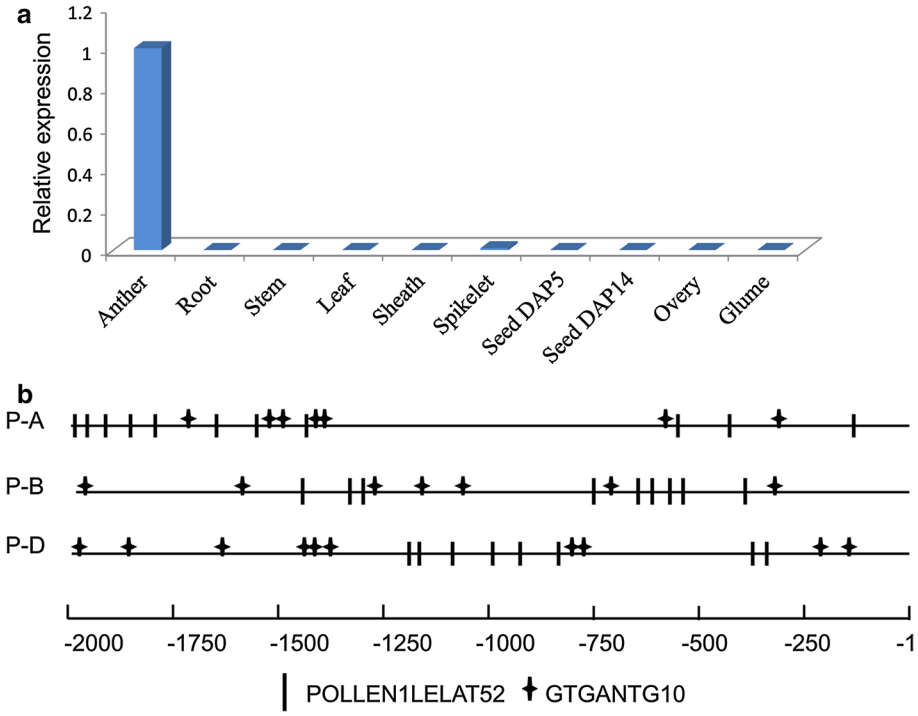

Fig. 3 Expression profiles of *TaCWI* in different tissues (**a**), and anther-specific cis-regulated elements of the homoeologous *TaCWI* promoters (**b**)

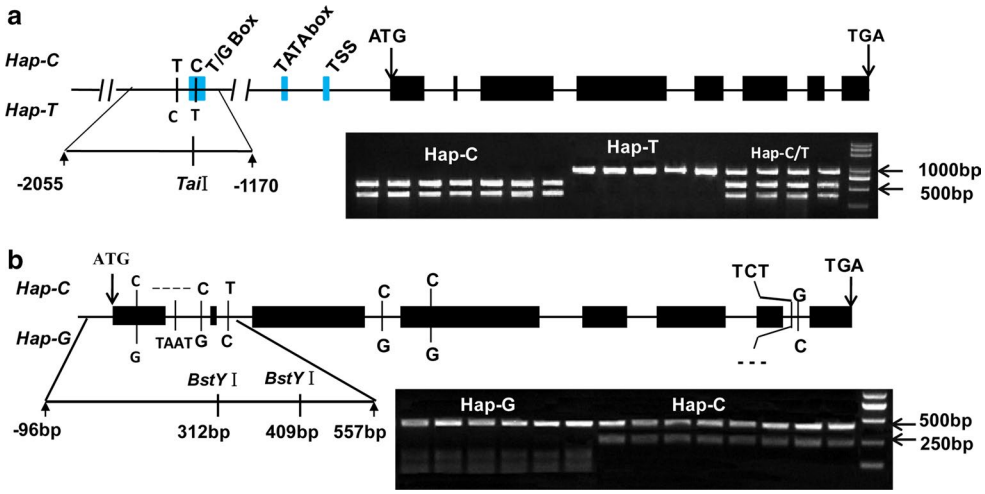

Fig. 4 Marker development of *TaCWI-4A* and *TaCWI-5D*. **a** SNPs at −1,523 and −1,527 bp are labeled. PCR product profiles restrictively digested by *Tai*I are shown. **b** Six SNPs and two indels are labeled. Profiles are for PCR products restrictively digested by *Bst*YI

five major groups: CWI families 1–4 and others. *TaCWIs* were clustered into CWI family 3 with *OsCWI3*, *ZmCWI3* and *BdCWI3*. *OsCWI3* has anther-specific expression (Cho et al. 2005). *ZmCWI3* expresses in 4- and 25-day-old seedling roots, 4-day-old seedling shoots, pollen, anther and cupules of male and female reproductive tissues (Kim et al. 2000). In this study, *TaCWI* expression was detected only in anthers.

Diversity *in TaCWIs* on chromosomes 4A and 5D mainly occurred in the promoter regions and introns in common wheat

Nucleotide diversities including SNPs and Indels were detected at *TaCWI-4A* and *TaCWI-5D* in wheat cultivars (Fig. 4), but *TaCWI-5B* was fully conserved. Two SNPs were found in the *TaCWI-4A* promoter; one SNP at −1,523(C/T), and the other one at −1,527 (T/C) (Fig. 4a). The two SNPs formed two haplotypes (TAAAC/CAAAT), which were named *Hap-4A-C* and *Hap-4A-T* according to the -1,527 SNP, respectively, in the latter description (Fig. 4a). There is a T/G-box (AACGTG), which is related to jasmonate-regulated plant defense responses (Boter et al. 2004) at the mutated site of *Hap-4A-C*; this is lacking in *Hap-4A-T*. Six SNPs and two Indels were found in *TaCWI-5D* forming two haplotypes (C/TAAT/G/C/G/C/—/C and G/—/C/T/C/G/TCT/G), which were, respectively, named *Hap-5D-G* and *Hap-5D-C* based on the second SNP (G/C) (Fig. 4b).

Table 4 Phenotypic traits associated with haplotypes at *TaCWI-4A* in Chinese wheat cultivars (mean ± SE)

Trait	*Hap-4A-C*	*Hap-4A-T*	*F*	*P*	Sub-population	Year
TKW	32.611 ± 0.401	37.516 ± 1.926	14.231	0.000**	Landraces	2006
HD	186.125 ± 0.614	190.444 ± 2.007	5.535	0.020*	Landraces	2006
SN	22.539 ± 0.29	20.714 ± 0.266	7.048	0.009**	Modern cultivars	2002

F ratio and probability based on one-way ANOVA

TKW 1,000 grain weight, *HD* heading date, *SN* spikelet number

* P < 0.05; ** P < 0.01

Table 5 Yield performance of selected nested breeding lines of Handan 6172 and Zhoumai 18 with different haplotypes at *TaCWI-4A*, their mean TKW and GN in yield plots

Year	Trait	Hap-T	Hap-C	*F*	*P*
2012	GN	35.88 ± 1.34	39.49 ± 0.80	6.09	0.03*
	TKW	44.90 ± 1.13	42.17 ± 0.70	4.71	0.05*
	Yield	9.62 ± 0.23	9.48 ± 0.24	0.16	0.69
2013	GN	36.02 ± 0.92	36.68 ± 0.95	0.24	0.63
	TKW	44.19 ± 0.89	42.87 ± 0.84	1.15	0.29
	Yield	8.10 ± 0.18	7.86 ± 0.23	0.65	0.43

* P < 0.05; ** P < 0.01

Table 6 Difference of TKW and GN betweenhaplotypes of *TaCWI-4A* inner-isogenic lines selected from Handan 6172 in 2013–14 crop seasons

Trait	*Hap-T*	*Hap-C*	*F*	*P*
TKW	52.28 ± 0.21	49.48 ± 1.24	4.97	0.04*
GN	37.43 ± 1.04	41.28 ± 0.41	11.77	0.03*
SL	8.81 ± 0.10	8.58 ± 0.08	3.14	0.15
ETN	6.55 ± 0.12	6.17 ± 0.16	3.81	0.12
SN	19.30 ± 0.16	18.9 ± 0.18	2.74	0.17
Plot yield	9.33 ± 0.25	9.98 ± 0.07	6.21	0.07
FSN/m^2	883.06 ± 23.00	847.57 ± 11.46	1.07	0.41

* P < 0.05; ** P < 0.01

Haplotyping and linkage mapping

CAPS marker *caps4A* was developed to discriminate the two haplotypes of *TaCWI-4A*. The genomic-specific primer pair Ags (Table 1) was used to amplify an 885-bp fragment at *TaCWI-4A*. The restriction enzyme *Tai*I cut the *Hap-4A-C* fragment into 354 and 531 bp parts, but did not cut the *Hap-4A-T* product. The two types were easily distinguished after electrophoresis in 1.5 % agarose gels (Fig. 4a). *TaCWI-4A* was mapped on chromosome 4AL 2.7 cM distal to *Barc170* in the RILs derived from Xiaoyan 54 × Jing 411 (Fig. S3).

Marker *caps5D* was similarly developed. A 649-bp fragment was amplified by primer pair Dgs (Table 1) from *TaCWI-5D*. The restriction enzyme *Bst*YI cut *Hap-5D-G* amplicons into three fragments of 409, 143 and 97 bp, whereas *Hap-5D-C* was cut into two fragments, 405 and 244 bp. The two haplotypes can be easily distinguished after electrophoresis in a 2 % agarose gel (Fig. 4b). *TaCWI-5D* was distally located in chromosome 5DL at 1.7 cM distal to *Xwmc443* and 4.8 cM proximal to *Xgwm272* (Fig. S4) in a RIL population derived from Nanda 2419 × Wangshuibai.

Haplotypes at *TaCWI-5D* and *TaCWI-4A* were associated with multiple agronomic traits

The 262 MCC accessions and 348 Chinese modern cultivars were genotyped using markers *caps4A* and *caps5D*. Associations of *TaCWI-4A* and *TaCWI-5D* with eight agronomic traits in the Chinese wheat mini core collection were determined. Because entries in the MCC were clustered into landrace and modern cultivar subsets (Hao et al. 2008), separate association analyses were carried out separately for each subgroup.

No significant difference between *Hap-4A-C* and *Hap-4A-T* at *TaCWI-4A* was found in modern cultivars except a significant effect on SN in 2002 (Table 4). In landraces, there were significant differences in TKW, HD in 2006 (Table 4). Analysis of the two haplotypes in a set of selected nested breeding lines indicated that *Hap-4A-T* was associated with higher TKW and *Hap-4A-C* with higher GN in 2 years (Table 5, Table S1). This was further proved in near-isogenic lines inbreed-selected from a Han 6172 (Table 6). All these results suggested that *TaCWI-4A* regulates the balance of TKW and GN.

For *TaCWI-5D* there were no significant differences in eight traits between the two haplotypes in landraces. However, in the modern cultivar group, differences in HD, MD and PH were significant in 2006. Differences between the two haplotypes for SpL in 2005 and MD in 2003 were also significant. Furthermore, differences in HD, MD, PH, ETN, TKW were significant in all three environments for the 348 modern cultivars (Table 7). The *Hap-5D-C* cultivars have higher TKW, lower PH, earlier HD and shorter MD. Therefore, *Hap-5D-C* is a favored haplotype in Chinese wheat breeding.

Haplotypes frequencies in global breeding in the past century

Three hundred and eighty-four European cultivars, 436 American cultivars, 54 Canadian cultivars, 53 Mexican cultivars,

Table 7 Phenotypic traits associated with *TaCWI*-5D haplotypes in 348 Chinese modern cultivars (mean ± SE)

Trait	2002				2005				2010			
	Hap-G	*Hap-C*	*F*	*P*	Hap-G	Hap-C	*F*	*P*	Hap-G	Hap-C	*F*	*P*
HD	180.591 ± 1.375	176.333 ± 0.327	13.201	0.000**	201.893 ± 0.849	199.188 ± 0.215	6.795	0.001**	219.214 ± 0.677	217.423 ± 0.191	3.674	0.026*
MD	230.682 ± 1.599	226.058 ± 0.472	7.948	0.005**	238.92 ± 0.925	236.99 ± 0.206	3.251	0.040*	254.679 ± 0.733	253.138 ± 0.171	3.721	0.025*
PH	103.727 ± 4.692	91.532 ± 1.202	8.14	0.005**	105.071 ± 2.991	92.157 ± 1.043	7.692	0.001**	101.881 ± 2.208	89.688 ± 0.968	7.333	0.001**
ETN	9.409 ± 0.68	6.988 ± 0.153	19.239	0.000**	11.356 ± 0.945	8.866 ± 0.196	6.399	0.002**	13.533 ± 0.751	11.599 ± 0.214	4.299	0.014*
TKW	37.762 ± 1.061	42.909 ± 0.417	12.778	0.000**	34.628 ± 1.318	40.016 ± 0.377	9.021	0.000**	36.264 ± 1.174	40.255 ± 0.36	5.433	0.005**

F ratio and probability based on one-way ANOVA

ETN effective tiller number, *TKW* 1,000 grain weight, *HD* heading date, *MD* maturity date, *PH* plant height

* $P < 0.05$; ** $P < 0.01$

83 Russian cultivars and 51 Australian cultivars were genotyped using the two markers. We also used their release times to examine haplotype frequency trends over time. The *Hap-4A-T* frequency was relatively lower and variable, but *Hap-5D-C* showed an increasing trend from 78 % in the 1940s to 94 % in 1990s in Chinese varieties (Fig. S6). It approached fixation in European and North American cultivars, with only two and three accessions, respectively, carrying *Hap-5D-G*.

Distribution of favored haplotypes in global wheat cultivars

Hap-4A-C predominated in both Chinese landraces and modern cultivars, and in European cultivars (Fig. 5a–c). Globally, *Hap-4A-C* was favored in breeding; it was almost fixed in Australian cultivars (Fig. 7a).

The frequency of *Hap-5D-C* in modern cultivars (Fig. 6b) was much higher than in landraces (Fig. 6a), especially in the Chinese main wheat areas I and II. This suggested that *Hap-5D-C* was strongly and positively selected in breeding in China. There are fewer *Hap-5D-G* cultivars in other countries, especially in Mexican and Australian cultivars (Fig. 7b). Therefore, it was negatively selected in global breeding.

Selection of haplotypes at *TaCWI-4A* in rainfed areas

In the predominantly rainfed production areas of the China, Northern Spring Wheat Region and Australia, *TaCWI-4A-C* was the dominant haplotype with frequencies reaching 87.2 and 97.8 %, respectively (Fig. 7a, S5), much higher than frequencies in other areas. It seems that in arid and semi-arid areas genotypes with more grains per spike may be favored to achieve higher yield.

Nucleotide diversity at *TaCWI* loci in bread wheat and its diploid and tetraploid relatives

Diversity of *TaCWI* in A genome diploid relatives ($\pi = 0.00125$) was higher than in tetraploid relatives ($\pi = 0.0005$) and hexaploids ($\pi = 0.00021$) (Fig. S7a). *TaCWI-B* diversity was very low in all three B genome subsets. Diversity was very narrow in the B genomes of tetraploid relatives ($\pi = 0.00017$) (Fig. S7b). The diversity of *TaCWI-D* in wheat ($\pi = 0.00076$) was lower than in *Ae. tauschii* ($\pi = 0.00094$) (Fig. S7c). Diversity apparently declined during wheat polyploidization.

The reason for decreasing diversity may be explained by the Tajima's test (Fig. S8). The deviation of *Tajima's D* from zero at several sites indicates that bottleneck or purifying selection occurred. Both phenomena lead to decreased diversity.

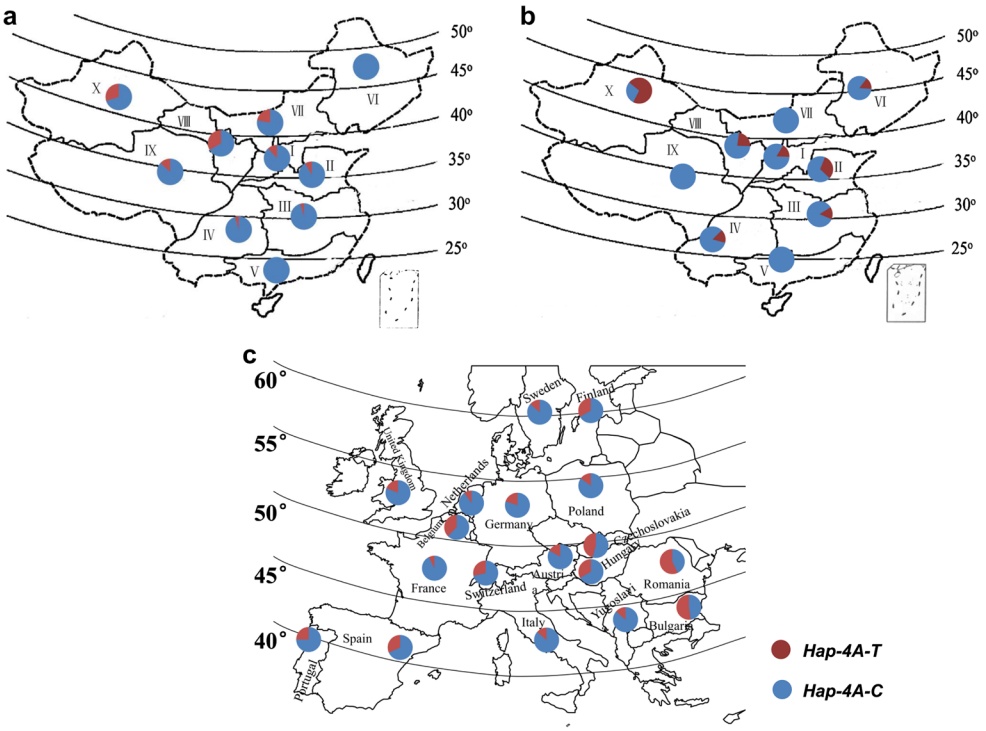

Fig. 5 Distribution of haplotypes at *TaCWI*-4A in China and Europe. **a** Chinese landraces; **b** Chinese modern cultivars; **c** European cultivars; *I* Northern winter wheat region; *II* Yellow and Huai River valley winter wheat region; *III* Low and middle Yangtze River valley winter wheat region; *IV* Southwestern winter wheat region; *V* Southern winter wheat region; *VI* Northeastern spring wheat region; *VII* Northern spring wheat region; *VIII* Northwestern spring wheat region; *IX* Qinghai–Tibet spring–winter wheat region; *X* Xinjiang winter–spring wheat region

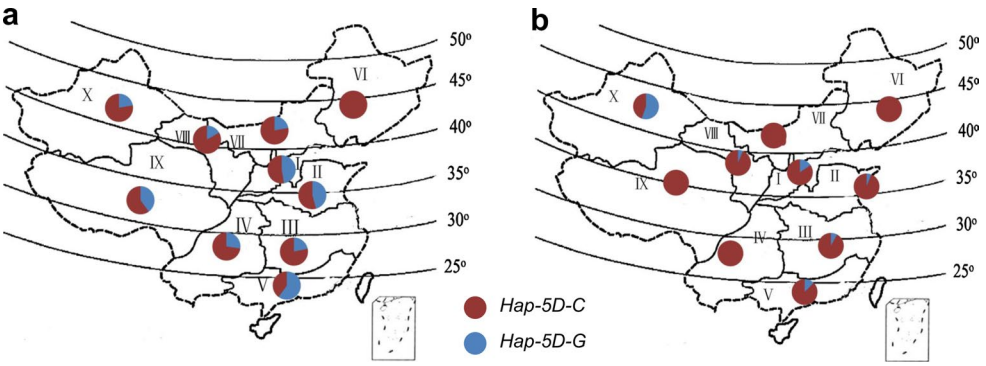

Fig. 6 Haplotype distributions of *TaCWI*-5D in ten Chinese wheat production zones. **a** Landraces; **b** modern cultivars

Genetic differences in *TaCWIs* in modern cultivars and landraces and diploid and tetraploid cultivars were compared. For the A genome (Fig. 8a), differences in F_{ST} values between diploid and tetraploid, and diploid and hexaploid genotypes were higher than 0.7. The differences in F_{ST} value between tetraploids and hexaploids were much lower (0.472 vs. 0.426). Between landraces and modern cultivars, the F_{ST} was only −0.086 suggesting that there was little diversity between the two subgroups. This indicated that strong selection for *TaCWI-5*A occurred during polyploidization. A very strong bottleneck occurred at *TaCWI-5B* in the progression from diploid to the tetraploid level; F_{ST} value differences reached 0.997 between diploid and tetraploid, 0.999 between diploid and hexaploid cultivars, but only 0.013 between tetraploid and hexaploid landraces, and 0.056 between tetraploid and modern cultivars. No obvious differentiation was detected between the landraces and modern cultivars (Fig. 8b). At *TaCWI-5*D, obvious differentiation occurred between *Ae. tauschii* and *T. aestivum*, but selection was not as strong as at *TaCWI-5*A and *TaCWI-5*B, with F_{ST} values reaching 0.308 and 0.368 between *Ae. tauschii* and landraces and modern cultivars of common wheat, respectively. For the D genome in common wheat, F_{ST} between modern cultivars and landraces was 0.05, suggesting that genetic differentiation occurred at this locus during wheat breeding (Fig. 8c).

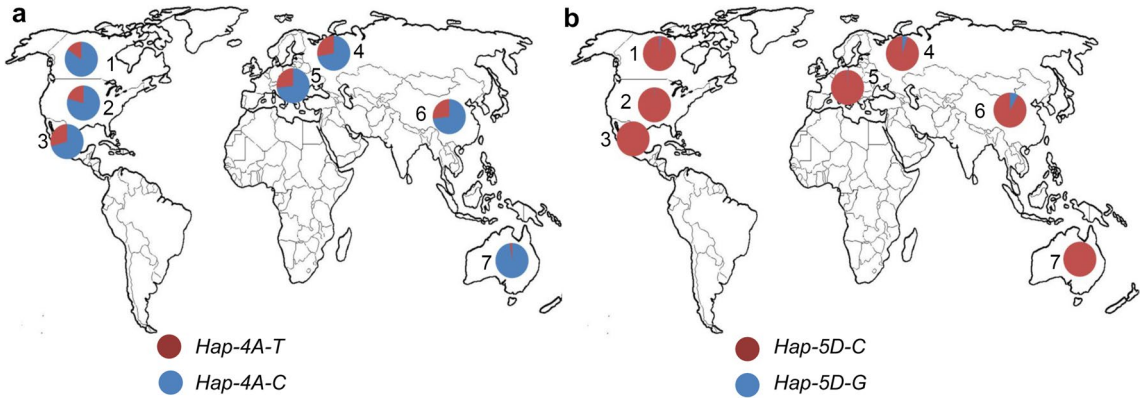

Fig. 7 Geographic distribution of haplotypes at *TaCWI-4A* (**a**) and *TaCWI-5D* (**b**) in seven major wheat production regions worldwide. *1* Canada, *2* USA, *3* Mexico, *4* Former USSR, *5* Europe, *6* China, *7* Australia

Haplotype networks of wheat and its relatives at *TaCWI loci*

The haplotype relationship between wheat relatives was analyzed by the computer program TCS. *TaCWIs* in the A genome were clustered into two unconnected sub-networks. Common and tetraploid wheats clustered into the same group (Fig. 9a) with *T. urartu* being a distinct set. Nine haplotypes at *TaCWI-4A* in the tetraploid and common wheat group, and three in *T. urartu* were detected in this study. The B genome relatives were also clustered into two sub-networks. Common wheat and tetraploid clustered into the same sub-networks (Fig. 9b) and diploids including *Ae. speltoides* and *Ae. longissima* were in a separate sub-network. For common wheat and tetraploid relatives, most species were the same haplotype at *TaCWI-B*. Fourteen haplotypes were found in *Ae. tauschii* (Fig. 9c). Distinct wide divergences existed among *TaCWI-D* haplotypes. There were scores of mutational steps between the most diverse pairs of lines. The results of haplotype networking showed that haplotypes underwent dramatic reductions in number during polyplodization and domestication of wheat.

Discussion

TaCWI-4A is located in the 4A-5A-7B translocation

TaCWI-B and TaCWI-D were located on chromosomes 5B and 5D, respectively, and *TaCWI-A* was located on 4A by genomic-specific primers (Fig. 2). *TaCWI-4A* was mapped between *Xbarc170* and *Xbarc1136.2* on chromosome 4AL in the Xiaoyan 54 × Jing 411 RIL population (Fig. S3). Using the *CAPS-5D* marker, *TaCWI-5D* was mapped between *Xwmc443* and *Xgwm272*, on the long arm of chromosome 5D in Nanda 2419 × Wangshuibai population (Fig. S4). Scaffold 60450 was obtained by blasting the *TaCWI-5D* genomic sequence in the *Aegilops tauschii* draft genome sequence (Jia et al. 2013). Scaffold 60450 was mapped to bin 288, at the terminal region of 5DL. Webster et al. (2012) isolated the same invertase genes (*IVR1_4AL*, *IVR1_5B* and *IVR1_5DL*) based on the same cDNA. Based on the mapping results, we estimated that *TaCWI-4A* was present in the 5AL fragment that was translocated to the current 4AL chromosome in wheat, where it is flanked by a 7BS-derived fragment and 4AS (Liu et al. 1992).

Global wheat breeding has targeted the *TaCWI-5D* locus during improvement of yield and other traits

In the MCC landrace group *Hap-4A-T* was associated with higher TKW and a later HD. There was no significant difference between the two haplotypes in 348 modern cultivars. Hou et al. (2014) showed that genetic differences influenced haplotype selection and intensity in cultivar populations. The frequency of *Hap-4A-T* fluctuated between breeding eras with no evidence of selection worldwide. In the MCC modern cultivar group, *Hap-4A-T* was associated with fewer SN. Ji et al. (2010) reported that anther-specific cell wall invertase was related to the reduced grain number under drought stress. In their report, drought stress down regulated *CWI* expression and led to anther abortion and spikelet reduction. This supported the association found in the present study.

Significant selection occurred at the *TaCWI-5D* site in global breeding. *Hap-5D-C* is almost fixed in modern global cultivars. Evidence of selection was supported by changes in haplotype frequencies in the last 60 years. In landraces the frequency of *Hap-5D-C* was 66.5 %, but in modern cultivars released after the 1950s it was 91.3 % (Fig. S6). *Hap-5D-C* is associated with superior

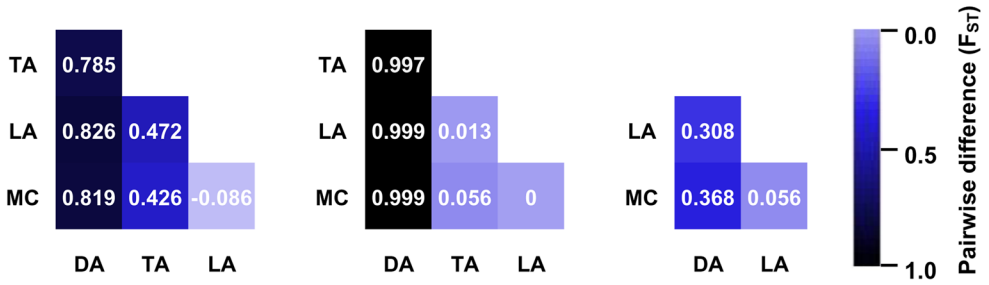

Fig. 8 Genetic distances (F_{ST}) at *TaCWI-4A* (**a**), *TaCWI-5B* (**b**) and *TaCWI-5D* (**c**) between pairs of populations. The *color gradient* represents changes in F_{ST} value from dark (1.0) to light blue (0.0). *DA* diploid accessions, *TA* tetraploid accessions, *MC* modern cultivars, *LA* landraces

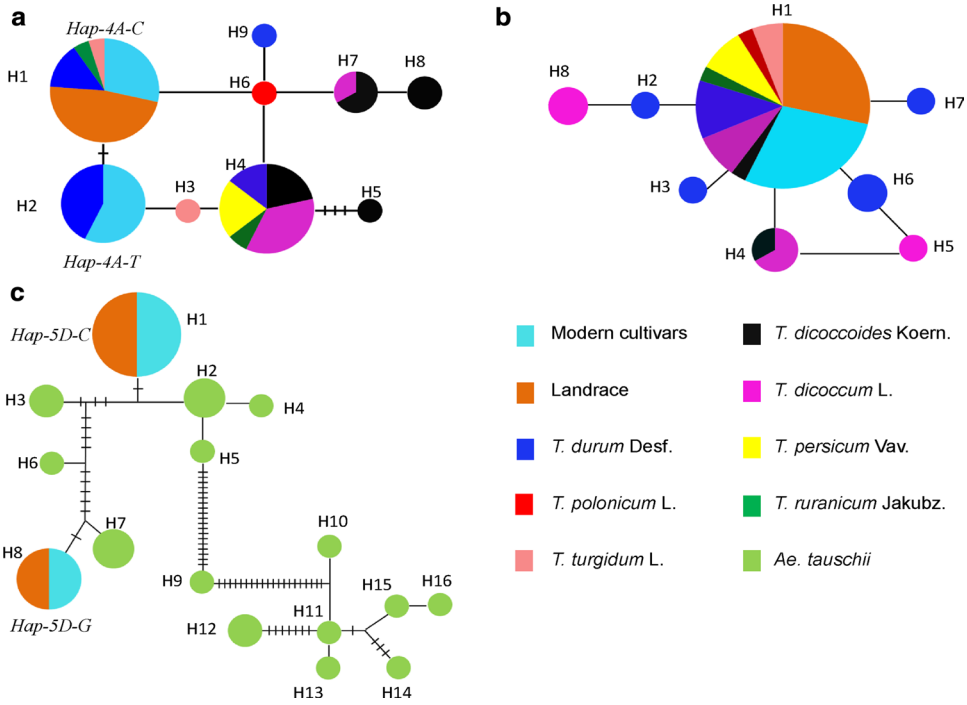

Fig. 9 Haplotype networks of *TaCWI-4A* (**a**), *TaCWI-5B* (**b**) and *TaCWI-5D* (**c**). Each *circle* represents a haplotype and the *circle* size is proportional to the number of accessions for a given haplotype. All haplotypes are separated from the nearest haplotype by a one-nucleotide difference. *Lines* between haplotypes represent likely mutational steps separating alleles. *Colors* represent different species (color figure online)

agronomic traits, assessed by increased TKW, HD, MD, PH and fewer ETN (Table 7). Moreover, QTLs for yield (Bordes et al. 2013) and heading date (Hanocq et al. 2007) were detected in the *Xgwm227* region where *TaCWI-5D* is located. The lack of positive selection for either *Hap-4A-T* or *Hap-4A-C* at *TaCWI-4A* might be due to inconsistent effects on agronomic traits. We found differences between haplotypes at the same locus influence selection time and intensity (Hou et al. 2014; Qin et al. 2014). An alternative explanation is that *Hap-4A-T* increases TKW, but *Hap-4A-C* increases SN or GN (Tables 4, 5, 6), thus there is a balancing effect between the two haplotypes in terms of yield. It seemed that irrigation or rainfed conditions also influence selection of either haplotype. Wheat production in Australia, Canada and USA is mainly rainfed. The frequency of *Hap-4A-C* was higher in these countries than in China, Europe, Russia and Mexico, where rainfall is higher or production is dependent on irrigation. In the rainfed production areas, seed setting was more important than TKW for yield achievement as evidenced by the extreme preference of *TaGWI-4A-C* in Australian breeding programs. In irrigated areas, there is positive selection for large spikes with high TKW and GN (Zhou et al. 2007; Hou et al. 2014).

Based on Ji et al. (2010) anther sink strength at the pre-anthesis stage is very important for maintaining grain number. They found the starch sink is almost exhausted in drought-sensitive cultivars, but maintained at normal levels in drought-tolerant cultivars after drought stress during the stage of anther development. This causes partial sterility in

sensitive cultivars and leads to a significant decline in seed setting. As an essential gene for starch synthesis in anthers *TaCWIs* have an important role in drought tolerance in wheat. This is consistent with our results on the global distribution of *Hap-4A-C* and *Hap-4A-T*, i.e., most of released cultivars in rainfed area are *Hap-4A-C* (Fig. 7).

Association of *TaGWI -5D* to plant height and tiller number might be caused by its tight linkage gene *TaGA20ox1*

The influence of *TaCWI* gene on plant height and tiller number might be caused by linkage to other important genes. In a search of the wheat genomics database (http://www.wheatgenome.org/Projects/IWGSC-Bread-Wheat-Projects), we found gene (Traes-5DL-3E77D28A6.1, *TaGA20ox1*), that is active in gibberellin metabolism. Gibberellins are hormones that control plant height, tillering and heading date (Peng et al. 1999). This gene is 2.34 cM from *TaCWI-5D*. We therefore cannot ignore the possibility of a selection sweep that encompasses both *TaCWI* and *TaGA20ox1*, because the average LD in modern wheat cultivars is 5–10 cM (Hao et al. 2011). Another possibility is that breeding targeted both genes, because short stem has been a continual breeding objective since the late 1960s. The genetic distance of *TaCWI-4A* and *TaGA20ox1* (Traes_4AL_FABDF4EDA.1) on chromosome 4A is 35.66 cM. This could be the reason that *TaCWI-4A* was not associated with PH and tiller number.

Extreme selection occurred at *TaGWI-5B*, and-*4A* during tetraploidization of wheat

Diversity at the *TaCWI-4A* locus was dramatically lower following tetraploidization (Figs. 8, 9, S7, S8). The diversity of *TaCWI-5D* was also reduced during hexaploidization (Fig. S7c), but much less than for *TaCWI-4A* and *TaCWI-5B*. The reason may be that evolutionary history of the D genome in polyploid wheat is shorter than A and B genomes. The F_{ST} of *Ae. tauschii* with *TaCWI-5D* further supported this view. There were reductions in haplotype numbers from 9 to 2, and 8 to 1 at *TaCWI-4A* and *TaCWI-5B*, respectively, from tetraploid to common wheat, supporting the commonly held view that natural hybridization between tetraploid wheat and *Ae. tauschii* was a rare event (Fig. 9a, b). This was further supported by the presence of 14 vs. 2 *TaCWI-5D* haplotypes in *Ae. tauschii* and common wheat (Fig. 9c). Dramatic reductions in diversity at the *TaSus1* loci were also detected during polyploidization (Hou et al. 2014). All of these examples indicate that strong bottlenecks occurred during polyploidization of wheat.

In summary, *TaCWIs* were cloned and mapped on wheat chromosomes 4A, 5B and 5D. *TaCWI-4A* is likely located in the 5A segment of the 4A-5A-7B translocation which now characterizes chromosome 4A in polyploid wheat (Liu et al. 1992). Haplotype *Hap-5D-C* at *TaCWI-5D* was strongly selected in global wheat breeding. In rainfed production regions, *Hap-4A-C* was favored because it brought more seeds, but in well irrigated conditions, *Hap-4A-T* was favored in modern breeding because of higher TKW. We assume that changes in haplotype frequency are related to genetic gain; if a gain is positive, a haplotype will be kept in breeding, otherwise declines in frequency.

Author contributions Jiang Q isolated the *TaCWI*. Jiang Y carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. Zhang X and Chen X supervised the study and revised the manuscript critically. Hou J, Hao C, Wang L, Zhang H and Zhang S participated in the agronomic data collection and analysis.

Acknowledgments We thank Dr. D.C. Liu, Institute of Genetics and Developmental Biology, Chinese Academy of Science and Dr. SL Xue, Nanjing Agricultural University, for their help in mapping of the *TaCWI-4A* and *TaCWI-5D*, respectively. We also gratefully acknowledge help from Prof. Robert A McIntosh, University of Sydney in English editing. This research was supported by Chinese Ministry of Science and Technology (grant no. 2010CB125900), the Animal and Plant Transgenic Project (2013ZX08009-001) and CAAS Innovation Project.

Conflict of interest The authors declare that they have no conflict of interest.

References

Bate N, Twell D (1998) Functional architecture of a late pollen promoter: pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. Plant Mol Biol 37:859–869

Bordes J, Ravel C, Jaubertie JP, Duperrier B, Gardet O, Heumez E, Pissavy AL, Charmet G, Le Gouis J, Balfourier F (2013) Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. Theor Appl Genet 126:805–822

Boter M, Ruíz-Rivero O, Abdeen A, Prat S (2004) Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. Genes Dev 18:1577–1591

Chen DH, Ronalds PC (1999) Rapid DNA mini preparation method suitable for AFLP and other PCR applications. Plant Mol Biol Rep 17:53–57

Cheng WH, Taliercio EW, Chourey PS (1996) The *miniature1* seed locus of maize encodes a cell wall invertase required for normal development of endosperm and maternal cells in the pedicel. Plant Cell 8:971–983

Cho J, Kyu S, Lee SK (2005) Molecular cloning and expression analysis of the cell-wall invertase gene family in rice (*Oryza sativa* L.). Plant Cell Rep 24:225–236

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1659

Copenald L (1990) Enzyme of sucrose metabolism. Methods Plant Biochem 3:73–85

Dong YS, Cao YS, Zhang XY, Liu SC, Wang LF, You GX, Pang BS, Li LH, Jia JZ (2003) Development of candidate core collections in Chinese common wheat germplasm. J Plant Genet Resources 4:1–8 (in Chinese)

Dorion S, Lalonde S, Saini HS (1996) Induction of male sterility in wheat by meiotic-stage water deficit is preceded by a decline in invertase activity and changes in carbohydrate metabolism in anthers. Plant Physiol 111:137–145

Hanocq E, Laperche A, Jaminon O, Lainé A, Le Gouis J (2007) Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. Theor Appl Genet 114:569–584

Hao CY, Dong YC, Wang LF, You GX, Zhang HN, Ge HM, Jia JZ, Zhang XY (2008) Genetic diversity and construction of core collection in Chinese wheat genetic resources. Chinese Sci Bull 53:1518–1526

Hao CY, Wang LF, Ge HM, Dong YC, Zhang XY (2011) Genetic diversity and linkage disequilibrium in Chinese bread wheat (*Triticum aestivum* L.) revealed by SSR markers. PLoS One 6:e17279

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27:297–300

Hou J, Jiang QY, Hao CY, Wang YQ, Zhang HN, Zhang XY (2014) Global selection on sucrose synthase haplotypes during a century of wheat breeding. Plant Physiol 164:1918–1929

Ji XM, Shiran B, Wan JL, Lewis DC, Jenkins CLD, Condon AG, Richard AR, Dolferus R (2010) Importance of pre-anthesis anther sink strength for maintenance of grain number during reproductive stage water stress in wheat. Plant Cell Environ 33:926–942

Jia JZ, Zhao SC, Kong XY, Li YR, Zhao GY et al (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95

Kang BH, Xiong YQ, Williams DS, Pozueta-Romero D, Chourey PS (2009) Miniature1-encoded cell wall invertase is essential for assembly and function of wall-in-growth in the maize endosperm transfer cell. Plant Physiol 151:1366–1376

Kim JY, Mahéa A, Guya S, Brangeona J, Rochea O, Choureyb PS, Prioula JL (2000) Characterization of two members of the maize gene family, *Incw3* and *Incw4*, encoding cell-wall invertases. Gene 245:89–102

King G, Nienhuis J, Hussey C (1993) Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphisms. Theor Appl Genet 86:1028–1032

Koonjul PK, Minhas JS, Nunes C, Sheoran IS, Saini HS (2005) Selective transcriptional down-regulation of anther invertases precedes the failure of pollen development in water-stressed wheat. J Exp Bot 56:179–190

Kumar S, Nei M, Joel D, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9:299–306

Kumbhar MB, Larik AS, Hafiz HMI, Rind MJ (1983) Interrelationship of polygenic traits affecting grain yield in *Triticum aestivum* L. Wheat Information Service 57:42–45

Liu CJ, Atkinson MD, Chinoy CN, Devos KM, Gale MD (1992) Nonhomoeologous translocations between group 4, 5 and 7 chromosomes within wheat and rye. Theor Appl Genet 83:305–312

Ma DY, Yan J, He ZH, Wu L, Xia XC (2012) Characterization of a cell wall invertase gene *TaCwi-A1* on common wheat chromosome 2A and development of functional markers. Mol Breeding 29:43–52

Miller ME, Chourey PS (1992) The maize invertase-deficient *miniature-1* seed mutation is associated with aberrant pedicel and endosperm development. Plant Cell 4:297–305

Oliver SN, Van Dongen JT, Alfred SC, Mamun EA, Zhao XC, Saini HS, Fernandes SF, Blanchard CL, Sutton BG, Geigenberger P, Dennis ES, Dolferus R (2005) Cold-induced repression of the rice anther-specific cell wall invertase gene *OSINV4* is correlated with sucrose accumulation and pollen sterility. Plant Cell Environ 28:1534–1551

Peng JR, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256–261

Qin L, Hao CY, Hou J, Wang YQ, Wang YQ, Li T, Wang LF, Ma ZQ, Zhang XY (2014) Homologous haplotypes, expression, genetic effects and geographic distribution of the wheat yield gene *TaGW2*. BMC Plant Biol 14:107

Rogers HJ, Bate N, Combe J, Sullivan J, Sweetman J, Swan C, Lonsdale DM, Twell D (2001) Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene g10. Plant Mol Biol 45:577–585

Sturm A (1996) Molecular characterization and functional analysis of sucrose-cleaving enzymes in carrot (*Daucus carota* L.). J Exp Bot 47:1187–1192

Sturm A (1999) Invertases, primary structures, functions, and roles in plant development and sucrose partitioning. Plant Physiol 121:1–7

Sturm A, Tang GQ (1999) The sucrose-cleaving enzymes of plants are crucial for development, growth and carbon partitioning. Trends Plant Sci 4:401–407

Tang GQ, Luscher M, Sturm A (1999) Antisense repression of vacuolar and cell wall invertase in transgenic carrot alters early plant development and sucrose partitioning. Plant Cell 11:177–189

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tian JC, Deng ZY, Hu RB, Wang YX (2006) Yield components of super wheat cultivars with different types and the path coefficient analysis on grain yield. Acta Agron Sin 32:1699–1705

Tymowska-Lalanne Z, Kreis M (1998) The plant invertases: physiology, biochemistry and molecular biology. Adv Bot Res 28:70–117

Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Res 31:3688–3691

Wang ET, Wang JJ, Zhu XD, Hao W, Wang LY, Li Q, Zhang LX, He W, Lu BR, Lin HX, Ma H, Zhang GQ, He ZH (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat Genet 40:1370–1374

Webster H, Keeble G, Dell B, Fosu-Nyarko J, Mukai Y, Moolhuijzen P, Bellgard M, Jia JZ, Kong XY, Feuillet C, Frédéric C, Appels R (2012) Genome-level identification of cell wall invertase genes in wheat for the study of drought tolerance. Funct Integr Genomics 39:569–579

Zhang XY, Li CW, Wang LF, Wang HM, You GX, Dong YS (2002) An estimation of the minimum number of SSR alleles needed to reveal genetic relationships in wheat varieties information from large-scale planted varieties and cornerstone breeding parents in Chinese wheat improvement and production. Theor Appl Genet 106:112–117

Zhou Y, He ZH, Sui XX, Xia XC, Zhang XK, Zhang GS (2007) Genetic improvement of grain yield and associated traits in the Northern China winter wheat region from 1960–2000. Crop Sci 47:245–253

Zhuang QS (2003) Chinese wheat improvement and pedigree analysis. Agricultural Press Beijing (in Chinese)