

Characterization of a cell wall invertase gene *TaCwi-A1* on common wheat chromosome 2A and development of functional markers

Dongyun Ma · Jun Yan · Zhonghu He ·
Ling Wu · Xianchun Xia

Received: 7 July 2010 / Accepted: 4 October 2010 / Published online: 26 October 2010
© Springer Science+Business Media B.V. 2010

Abstract Cell wall invertase (CWI) is a critical enzyme for sink tissue development and carbon partition, and has a high association with kernel weight. Characterization of *Cwi* genes and development of functional markers are of importance for marker-assisted selection in wheat breeding. In the present study, the full-length genomic DNA sequence of a *Cwi* gene located on wheat chromosome 2A, designated *TaCwi-A1*, was characterized by *in silico* cloning and experimental validation. *TaCwi-A1* comprises seven exons and six introns, with 3,676 bp in total, and an open reading frame (ORF) of 1,767 bp. A pair of complementary dominant markers, CWI21 and CWI22, was developed based on allelic variations at the *TaCwi-A1*

locus. A 404-bp PCR fragment was amplified by CWI21 in varieties with lower kernel weights, whereas a 402-bp fragment was generated by CWI22 in the varieties with higher kernel weights. The markers CWI21 and CWI22 were located on chromosome 2AL using a F_{2:3} population from a cross Doumai/Shi 4185, and a set of Chinese Spring nullisomic–tetrasomic lines. They were linked to the SSR locus *Xbarc15-2AL* with a genetic distance of 10.9 cM. QTL analysis indicated that *TaCwi-A1* could explain 4.8% of phenotypic variance for kernel weight over 2 years. Two sets of Chinese landraces and two sets of commercial wheat varieties were used to validate the association of CWI21 and CWI22 with kernel weight. The results indicated that the functional markers CWI21 and CWI22 were closely related to kernel weight and could be used in wheat breeding for improving grain yield.

Dongyun Ma and Jun Yan contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-010-9524-z) contains supplementary material, which is available to authorized users.

D. Ma · Z. He · L. Wu · X. Xia (✉)
Institute of Crop Science, National Wheat Improvement Center/The National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences (CAAS), 12 Zhongguancun South Street, 100081 Beijing, China
e-mail: xiachun@caas.net.cn

D. Ma
National Wheat Engineering Research Center, Henan Agricultural University, 62 Nongye Road, 450002 Zhengzhou, Henan Province, China

Keywords Allelic variation · *Cwi* gene · Kernel weight · QTL · *Triticum aestivum*

J. Yan
Cotton Research Institute, Chinese Academy of Agricultural Sciences (CAAS), Huanghedadao, 455000 Anyang, Henan, China

Z. He
CIMMYT China Office, c/o CAAS, 12 Zhongguancun South Street, 100081 Beijing, China

L. Wu
Crop Research Institute, Sichuan Academy of Agricultural Sciences, 610066 Chengdu, China

Introduction

Wheat is one of the most important food crops in the world. Increasing grain yield is of great importance to meet the needs of wheat for increasing human populations. Grain yield is closely associated with kernel number per unit area and kernel weight (Kumbhar et al. 1983). In Chinese environments, 140–160 kg per hectare can be added, when the thousand-kernel weight (TKW) is increased by one gram (Tian et al. 2006). Collaku (1989) reported that 20% of the variation in wheat yield can be attributed to TKW. In addition to its contributions to yield per se, TKW is crucial for variety acceptance by farmers in countries such as India and China. Therefore, developing varieties with high kernel weight is an important objective in wheat breeding programs.

In most plants, the assimilated carbon in the form of sugars such as sucrose is transported from the photosynthetic tissues (source) to non-photosynthetic tissues (sink) (Sturm 1999), where the sucrose can be used directly or cleaved into hexoses by sucrose synthase or invertase (Copenald 1990). Invertase is a hydrolase, catalyzing the conversion of sucrose to glucose and fructose. Based on their solubility, localization and pH optima, the invertases in higher plants are of three types, viz. cytoplasmic, vacuolar, and cell wall (Kim et al. 2000). Cell wall invertase (CWI) with an acidic pH optimum and bound to the cell wall ionically catalyzes the hydrolysis of sucrose to glucose and fructose (Sturm 1999).

Many studies indicate that CWI plays an important role in carbon partitioning (Tang et al. 1999) and sink organ development (Sturm 1999; Chourey et al. 2006), and is related to kernel weight (Miller and Chourey 1992; Cheng et al. 1996; Wang et al. 2008). Weber et al. (1995) reported that the amount of CWI in the seed coat was highly related to the level of hexoses in cotyledons and apoplastic endosperm spaces, and proposed that it played a crucial role in establishing sink strength in young seed. Transgenic carrot antisense repression of CWI showed reduced taproot growth and an increased dry weight leaf-to-root ratio (Tang et al. 1999). The seed weight of the maize mutant seed Miniature-1, which lacks of invertase activity, was reduced by 70% relative to the normal weight (Miller and Chourey 1992; Cheng et al. 1996). Hirose et al. (2002) also found that the

expression of the *OsCINI* gene was related to CWI activity and caryopsis development, suggesting that *OsCINI* is important for supplying carbon to the kernels. More recently, the rice *GIFI* gene was cloned. It encodes a cell wall invertase required for carbon partitioning during early grain-filling, and its unique expression pattern confirmed that CWI was highly associated with kernel weight (Wang et al. 2008).

The cDNA sequence encoding CWI was first cloned from carrot (Sturm and Chrispeels 1990) and was then identified and characterized in a variety of plants, such as tomato (Godt and Roitsch 1997), soybean (Dimou et al. 2005), maize (Cheng et al. 1996), sorghum (Jain et al. 2008), and rice (Hirose et al. 2002). Common features of the amino acid sequences of plant CWI are a β -fructosidase motif (NDPNG/A) and a Cys residue (Sturm 1999). These *Cwi* genes have similar structure, with six to eight exons (Tymowska-Lalanne and Kreis 1998), and most of them contain an extremely small exon (exon 2) coding only three amino acids (DPN) (Sturm 1999). To date, however, no full-length DNA sequence of *Cwi* has been cloned in common wheat.

A large number of quantitative trait loci (QTLs) have been reported to control grain yield and yield components. In particular, QTLs for kernel weight were mapped on almost all 21 wheat chromosomes (Huang et al. 2003; Quarrie et al. 2005; Su et al. 2006). Su et al. (2006) found eight QTLs for kernel weight per spike on chromosomes 2A, 2D, 4B, 5A, 7A and 7B. These QTLs explained each 5.6–16.2% of phenotypic variation. Four QTLs for TKW were detected on 1D, 2A, 5D, and 6A, and their contributions in different environments ranged from 5.9 to 20.1% (Sun et al. 2009). Functional markers developed from polymorphic sites of genes are very important for marker-assisted breeding in wheat (Bagge et al. 2007). Use of functional markers for kernel weight would benefit high-yield wheat breeding.

The objectives of this study were to characterize allelic variations at the *TaCwi-A1* locus on wheat chromosome 2A, develop functional markers for the gene, and evaluate its association with kernel weight in Chinese commercial wheat varieties and landraces, and a F_{2:3} population from a cross Doumai/Shi 4185.

Materials and methods

Plant materials

Two sets of Chinese commercial wheat varieties and two sets of Chinese landraces, viz. Set I, 192 winter wheat varieties from China Wheat Zones I, II, III and IV; Set II, 117 winter wheat varieties from China Wheat Zone IV; Set III, 130 landraces from China Wheat Zones I, II, III, IV, V and VIII; and Set IV, 48 landraces from China Wheat Zones I, II, III, V and VIII, were used for validation of the newly developed STS markers. The China Wheat Zones were described in detail by Zhang et al. (2008). The Chinese Spring (CS) nullisomic-tetrasomic and ditelosomic lines, and a wheat-barley chromosome substitution line N2A–S2H provided by Prof. Xiao Chen, Institute of Crop Science, CAAS, were employed to verify the chromosomal location of the STS markers. Nullisomic-tetrasomic 2A lines are sterile and cannot be maintained. The 2H chromosome is able to restore fertility to CS nullisomic 2A and can be used in identifying genetic effects that are lost in the absence of chromosome 2A. A population of 217 $F_{2:3}$ lines derived from Doumai/Shi 4185 was used to map the STS markers and QTL analysis. Shi 4185 has a low TKW (36.1 g), whereas Doumai has a high TKW (55.0 g).

Field trials

During the 2002–2003 cropping season, the varieties of Set I were grown in randomized complete blocks with three replicates at the Anyang Experimental Station, CAAS, located in Henan Province. Each plot consisted of two 2-m rows spaced 25 cm apart, with 75 plants in each row. During the 2008–2009 cropping season, Sets II and III were grown in randomized complete blocks with three replicates at the Chengdu Experimental Station, Sichuan Academy of Agricultural Sciences, located in Chengdu, Sichuan Province. Set IV varieties were grown in randomized complete blocks with three replicates at the Beijing Experimental Station of CAAS during the 2008–2009 cropping season. The 217 $F_{2:3}$ lines were planted at Beijing Experimental Station of CAAS during the 2008–2009 and 2009–2010 cropping seasons. Test plots were managed according to local practices. All field trials were kept free of weeds and

diseases, with two applications of broad-range herbicides and fungicides, respectively. Plants were harvested by hand when the plants reached physiological maturity. The TKW was evaluated by weighing 200 kernels in duplicate, after the seeds had been dried to a standard moisture content.

Strategies for cloning and sequence analysis of *TaCwi-A1*

We used the cDNA sequence of the rice *GIF1* gene encoding a cell wall invertase (GenBank accession EU095553, Wang et al. 2008) as a probe for a BLAST searching against the wheat EST database in GenBank. All wheat ESTs sharing high similarity with the reference gene were selected as candidate sequences (E value $< e^{-50}$, score > 300 bp and identity $> 70\%$, <http://www.ncbi.nlm.nih.gov>). The putative full-length sequences of wheat *Cwi* genes were constructed by analyzing and assembling the wheat ESTs. Primers were designed based on the constructed sequence using the software Primer Premier 5 (<http://www.premierbiosoft.com>) and synthesized by Shanghai Sangon Biological Engineering Technology & Service Co. Ltd. (<http://www.sangon.com>). Each primer combination was tested with N2A–S2H, N2B–T2A and N2D–T2B to confirm the chromosomal location of the primers and their amplified PCR fragments.

Genomic DNA was extracted from seeds using a method modified from Lagudah et al. (1991). PCRs were performed in an MJ Research PTC-200 thermal cycler in a total volume of 20 μ L, including 20 mM of Tris–HCl (pH 8.4), 20 mM of KCl, 100 μ M of each of dNTPs, 1.5 mM of $MgCl_2$, 5 pmol of each primer, 1 unit of *Taq* DNA polymerase (TIANGEN Biotech Co., Ltd., Beijing, <http://www.tiangen.com>) and 50 ng of template DNA. Reaction conditions were 95°C for 5 min, followed by 40 cycles of 95°C for 45 s, 58°C for 1 min 15 s, and 72°C for 2 min, with a final extension of 72°C for 8 min. The PCR products were separated by electrophoresis on a 1.5% agarose gel stained with ethidium bromide and visualized using UV light. Targeted fragments with expected sizes (Table 1) were recovered and cloned into the pMD18-T vector, and sequenced by Shanghai Sangon Biological Engineering Technology and Service Co. Ltd. (<http://www.sangon.com>).

Three wheat varieties with higher and three with lower kernel weights were used for cloning the

Table 1 Three primer sets used for cloning *TaCwi-A1*, and two functional markers developed for the gene

Primer set or marker	Primer sequence (5'–3')	Amplified region ^a	Size of PCR fragments (bp)
C1	Forward: GCCCCAGTGTGAGAGCAATGGTAGT Reverse: CACCTCGTAGTTGATGTCGC	1–1,361	1,361
C2	Forward: CAGCCCAGCATCAAGTCCGA Reverse: CCGAGCAAAGAAGCCCAACA	1,242–2,996	1,755
C3	Forward: GCCGGTGAGTCTCAAGGACC Reverse: TTCCTACGGGCAAAATGAT	2,164–3,676	1,513
CWI21	Forward: GTGGTGATGAGTTCATGGTTAAG Reverse: AGAAGCCCAACATTAAATCAAC	2,705–3,108	404
CWI22	Forward: GGTGATGAGTTCATGGTTAAT Reverse: AGAAGCCCAACATTAAATCAAC	2,707–3,108	402

^aAmplified region was counted from the genomic DNA sequence of *TaCwi-A1* cloned in this study

full-length sequence of *TaCwi-A1* on chromosome 2A to identify allelic variations at the locus. The PCR of each primer set was repeated three times for each DNA sample, and each PCR product was sequenced at least three times to ensure the accuracy of the nucleotide sequences. The genomic DNA sequences of *TaCwi-A1* were aligned by the software DNAMAN (<http://www.lynnon.com>) to detect the allelic variants among different varieties. The intron positions were also determined by DNAMAN through alignment of the amplified genomic DNA sequence and the constructed cDNA sequences.

Development of functional markers for *TaCwi-A1*

Three wheat varieties with higher and three with lower TKW were used for developing new STS markers. Based on the sequence divergence of the allelic variants at the *TaCwi-A1* locus, two complementary dominant markers, CWI21 and CWI22, were designed to identify allelic variants at the *TaCwi-A1* locus. These were validated with the two sets of Chinese landraces and two sets of commercial varieties. PCRs for CWI21 and CWI22 were 95°C for 5 min, followed by 38 cycles of 95°C for 45 s, 58°C (CWI21) or 60°C (CWI22) for 45 s, and 72°C for 60 s, with a final extension of 72°C for 8 min.

Statistical analyses

Analysis of variance (ANOVA) was conducted by the Analyze-Compare Means-Independent-Samples t-Test in SPSS10.0 with genotype classes as a categorical variable to derive the mean TKW for each

class and to test the significance of the differences between the two classes. Linkage analysis was performed with the software Map Manager QTX (Manly et al. 2001) and the Kosambi function was used for calculating genetic distances (Kosambi 1944). QTL analysis was conducted using the software QTL Cartographer V2.5 (Wang et al. 2005), and a LOD score of 3.0 was used as the threshold for declaration of linkage and QTL detection.

Results

Cloning and characterization of *TaCwi-A1*

Using the cDNA sequence of rice *GIF1* gene (GenBank accession EU095553) as a probe for BLAST searching against the wheat EST database in GenBank, four wheat ESTs, viz. CK206103, CD927722, CJ673133, and CD927242, covering the full coding sequence of the reference gene, were identified and assembled to construct the assumed full-length cDNA of the *Cwi* gene. Three primer sets, namely, C1, C2 and C3, were generated to amplify the full-length genomic DNA sequence of *TaCwi-A1* based on the assumed *Cwi* sequence. Each primer set was tested with CS nullisomic–tetrasomic lines N2B–T2A and N2D–T2B, and barley chromosome substitution line N2A–S2H, for confirming its chromosomal location (Sup-fig. 1). The PCR products amplified from these primer sets were recovered and cloned, and eight clones from each PCR product were sequenced. Thus the sequences obtained in the clones from C1, C2, and C3 were assembled to construct the full-length

DNA sequence of *TaCwi-A1* (Table 1), which was then aligned with the corresponding sequence assembled with cDNAs by DNAMAN to determine the exons and introns.

The complete genomic DNA sequence of *TaCwi-A1* was 3,676 bp, with seven exons and six introns, including the 5' and 3' flanking sequences. The exon–intron structure of *TaCwi-A1* was very similar to those of the maize and rice genes encoding cell wall invertase. As also for those genes the second exon of wheat *TaCwi-A1* was 9 bp, encoding part of a highly conserved region, which is present in all known invertase proteins. The third exon was the longest among the seven exons.

The cDNA sequence of the cloned *TaCwi-A1* was 1,954 bp, containing an open reading frame (ORF) of 1,767 bp, a 17-bp 5' untranslated region (UTR) and a 170-bp 3' UTR. The deduced pre-mature peptide sequence of *TaCwi-A1* consisted of 588 amino acid residues, with a predicted molecular weight of 65.4 kDa. The putative cleavage site of the deduced pre-mature peptide was located in the same position as that from other *Cwi* genes, based on the alignment of the deduced amino acids of *TaCwi-A1* with those in *Cwi* genes in other plants (Sup-fig. 2). The

cleavage would generate a mature protein of 564 amino acid residues with a molecular weight of 63 kDa and a transit peptide of 24 amino acids with a molecular weight of 2.4 kDa. A basic pI of 8.9 was calculated using the DNAMAN sequence analysis tool, suggesting that in the cell wall environment, where the pH is acidic, *TaCwi-A1* would be ionically bound to the cell wall components.

Identification and validation of functional markers for *TaCwi-A1*

The amplified genomic DNA sequences of *TaCwi-A1* from wheat varieties with different kernel weights showed two allelic variants, *TaCwi-A1a* and *TaCwi-A1b* (Fig. 1). Nine single-nucleotide polymorphisms (SNPs), four in introns and five in exons, were detected in the DNA sequences from varieties with lower and higher TKW. Two complementary dominant markers, CWI21 and CWI22, were designed based on the SNP at the position of the 2,727th base of *TaCwi-A1*. CWI22 yielded a 402-bp PCR fragment in varieties with allele *TaCwi-A1a*, and CWI21 yielded a 404-bp fragment in those with allele *TaCwi-A1b* (Fig. 2). In the test of 192 varieties in

Fig. 1 Part of DNA sequences alignment of the alleles *TaCwi-A1a* and *TaCwi-A1b*. The DNA sequences of exons are underlined. The shaded bases are SNPs between the two alleles. The sequences of the forward primers and complementary sequences of reverse primers for CWI21 and CWI22 are boxed

<i>TaCwi-A1a</i>	ACAGCTTAGAAAACTAGACTCGTCGTCACGCGCACATTGGCATTTGGCCTTGACGCGCACAC	2610
<i>TaCwi-A1b</i>	ACAGCTTAGAAAACTAGACTCGTCGTCACGCGCACATTGGCATTTGGCCTTGACGCGCACAC	2610
<i>TaCwi-A1a</i>	CTCCAATCTAGCTAGGCTCCAACCTGAACTGGTGCAACCTAACAATAGATCGTTTAAATTT	2670
<i>TaCwi-A1b</i>	CTCCAATCTAGCTAGGCTCCAACCTGAACTGGTGCAACCTAACAATAGATCGTTTAAATTT	2670
	CWI22 Forward primer	
<i>TaCwi-A1a</i>	GACCAACATAATGTCTGTTCCTTTTGTAAAGCGTGTGGTGATGAGTTTCATGGTTAAATAC	2730
<i>TaCwi-A1b</i>	GACCAACATAATGTCTGTTCCTTTTGTAAAGCGTGTGGTGATGAGTTTCATGGTTAAATAC	2730
	CWI21 Forward primer	
<i>TaCwi-A1a</i>	CGGACTAAGTTGTTTCATCAGGCTGACGTCGAGGTGAGCTTTGAGGTGCCGAGCCTGGAG	2790
<i>TaCwi-A1b</i>	CGGACTAAGTTGTTTCATCAGGCTGACGTCGAGGTGAGCTTTGAGGTGCCGAGCCTGGAG	2790
<i>TaCwi-A1a</i>	GGAGCGGAGGCGCTGGACCCGGCGCTAGCCAACGACGCCCAGAAGCTGTGCAGCTTGAGG	2850
<i>TaCwi-A1b</i>	GGAGCGGAGGCGCTGGACCCGGCGCTAGCCAACGACGCCCAGAAGCTGTGCAGCTTGAGG	2850
<i>TaCwi-A1a</i>	GGTGCCGACGTGGAAGGCGGCGTGGGCCCCCTTCGGTCTGTGGGTGCTCGCCTCGTCCAAG	2910
<i>TaCwi-A1b</i>	GGTGCCGACGTGGAAGGCGGCGTGGGCCCCCTTCGGTCTGTGGGTGCTCGCCTCGTCCAAG	2910
<i>TaCwi-A1a</i>	CTGGAGGAGAAGACGGCGGTCTTCTTCCAGGTGTTCAAGGCCGCGCGCAACATCAACAGC	2970
<i>TaCwi-A1b</i>	CTGGAGGAGAAGACGGCGGTCTTCTTCCAGGTGTTCAAGGCCGCGCGCAACATCAACAGC	2970
<i>TaCwi-A1a</i>	ACCAAGCCGGTGGTCTCATGTGCTCCGACCCACCACGTAATGACATCTCCTGATCCTC	3030
<i>TaCwi-A1b</i>	ACCAAGCCGGTGGTCTCATGTGCTCCGACCCACCACGTAATGACATCTCCTGATCCTC	3030
<i>TaCwi-A1a</i>	ATTTCTCCCTTAAATTCATTATTTCTCAGCATGTGACATGCACACATTAGCTAGGTTG	3090
<i>TaCwi-A1b</i>	ATTTCTCCCTTAAATTCATTATTTCTCAGCATGTGACATGCACACATTAGCTAGGTTG	3090
	CWI22 Reverse primer	
<i>TaCwi-A1a</i>	ATTTAATGTTGGGCTTCTTTGCTCGGTGCAGGTATCATCTTTGAACCCGAACCTCTACAAGC	3150
<i>TaCwi-A1b</i>	ATTTAATGTTGGGCTTCTTTGCTCGGTGCAGGTATCATCTTTGAACCCGAACCTCTACAAGC	3150
	CWI21 Reverse primer	

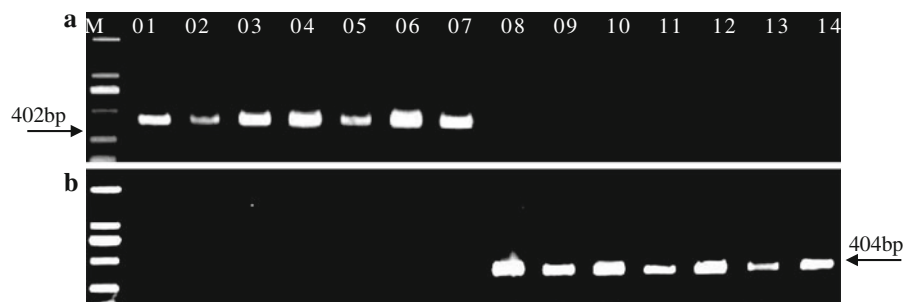


Fig. 2 PCR fragments amplified with the markers CWI22 (a) and CWI21 (b) in 14 Chinese winter wheat varieties. M DNA ladder DL2000; 01 Doumai (TKW 55.0 g); 02 Zhongmai 9 (53.2 g); 03 Chuanmai 39 (55.0 g); 04 SW18390 (50.5 g); 05 Neimai 8 (52.6 g); 06 Shannong 664 (52.2 g); 07

Chuanmai 43 (54.2 g); 08 Fengmai 24 (37.8 g); 09 Shi 4185 (36.1 g); 10 Linhan 917 (38.0 g); 11 Yumai 21 (32.2 g); 12 Mianyang 26 (39.3 g); 13 Chuannong 16 (34.9 g); 14 Yangmai 10 (27.2 g)

Set I using CWI21 and CWI22, the 404-bp fragment was detected in 46 varieties (*TaCwi-A1b*), with an averaged TKW of 42.6 g, whereas the 402-bp fragment was found in 146 varieties (*TaCwi-A1a*), with a mean TKW of 45.0 g, indicating a significant difference in TKW between genotypes *TaCwi-A1a* and *TaCwi-A1b* ($P < 0.05$) (Table 2). Among the 117 varieties in Set II, the 402-bp fragment was found in 102 varieties (*TaCwi-A1a*), with a mean TKW of 45.2 g, whereas the 404-bp fragment was detected in 15 genotypes (*TaCwi-A1b*), with an averaged TKW of 42.8 g. In the test of the varieties in set III using CWI21 and CWI22, the 404-bp fragment was detected in 125 genotypes (*TaCwi-A1b*), with a mean TKW of 28.3 g, whereas the 402-bp fragment was found in 5 varieties (*TaCwi-A1a*), with an averaged TKW of 37.8 g, showing a significant difference of TKW between the genotypes *TaCwi-A1a* and *TaCwi-A1b* ($P < 0.05$). Similar results were also obtained with Set IV varieties, in which the 402-bp fragment was found in 5 varieties (*TaCwi-A1a*), and the 404-bp fragment was detected in 43 genotypes (*TaCwi-A1b*),

with average TKW of 35.7 and 28.6 g, respectively. These results confirmed the association of *TaCwi-A1* with TKW in Chinese wheat germplasms. Among the 309 Chinese commercial varieties in Sets I and II, the frequencies of *TaCwi-A1a* and *TaCwi-A1b* were 80.3 and 19.7%, respectively, whereas the frequencies of *TaCwi-A1a* and *TaCwi-A1b* were 5.6 and 94.4%, respectively, in the 178 Chinese landrace wheats (Sets III and IV).

QTL mapping and chromosomal location of *TaCwi-A1*

Linkage analysis with the $F_{2:3}$ population of Doumai/Shi 4185 showed that CWI21 and CWI22 were linked to *Xbarc15* on chromosome 2AL, with a genetic distance of 10.9 cM, indicating *TaCwi-A1* is located on chromosome 2AL (Fig. 3). A major QTL for TKW on chromosome 2A co-segregated with CWI21 and CWI22, and explained 4.8% of the phenotypic variance in the $F_{2:3}$ populations over 2 years, with a LOD score of 3.3 (Fig. 3). The locations of CWI21

Table 2 Association between allelic variants of *TaCwi-A1* and TKW in Chinese commercial wheat varieties and landraces

Plant material	Genotype	Number of accessions	Mean TKW (g) ^a	Standard deviation	Range
Set I	<i>TaCwia-A1a</i>	146	45.0a	5.1	33.4–56.7
	<i>TaCwia-A1b</i>	46	42.6b	3.6	32.4–50.6
Set II	<i>TaCwia-A1a</i>	102	45.2a	5.9	25.8–55.0
	<i>TaCwia-A1b</i>	15	42.8a	6.0	27.3–51.5
Set III	<i>TaCwia-A1a</i>	5	37.8a	6.9	31.0–47.0
	<i>TaCwia-A1b</i>	125	28.3b	5.7	13.5–45.5
Set IV	<i>TaCwia-A1a</i>	5	35.7a	6.4	27.0–46.0
	<i>TaCwia-A1b</i>	43	28.6b	4.0	18.0–38.0

^a Different letters following the mean TKW indicate significant differences between the two genotypes ($P < 0.05$)

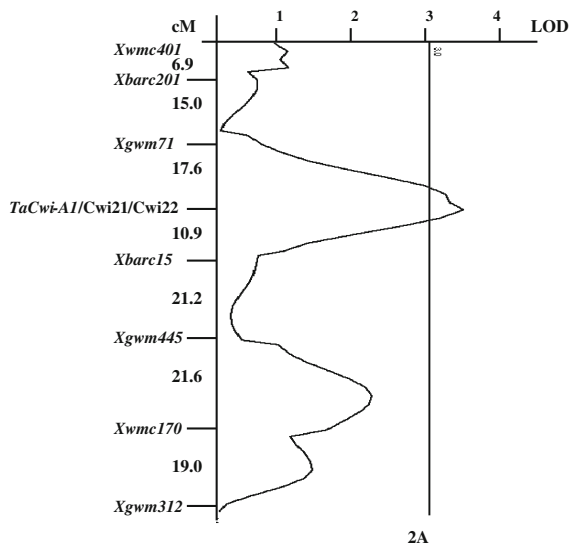


Fig. 3 Logarithm of odds (LOD) contours obtained by composite interval mapping (CIM) for quantitative trait loci (QTL) on TKW in the $F_{2:3}$ population of Doumai/Shi4185. The data of TKW were averaged over 2 years. LOD thresholds, 3.0

and CWI22 on chromosome 2A were also confirmed by amplifying genomic DNA from a set of CS nullisomic–tetrasomic lines, and wheat–barley chromosome substitution line N2A–S2H (Sup-fig. 3). No PCR product was amplified in N2A–S2H.

Discussion

Cloning and characterization of *TaCwi-A1*

Common wheat is an allohexaploid with an enormous genome size. It is essential to make sure that primer sets used for cloning genes are chromosome-specific. In the present study, two primer sets, CW1 and CW2, initially designed for amplification of the full-length *Cwi* genes did not show genome specificity in PCR amplification. The PCR products amplified from the CS nullisomic–tetrasomic lines, N2B–T2A and N2D–T2B, with primer set CW1 were cloned and sequenced. Two divergent sequences were obtained in the clones from N2B–T2A, and one of them was the same as that in the clones from N2D–T2B, implying that the common sequence was on chromosome 2A. Similarly, the same sequence obtained in clones from N2B–T2A and N2D–T2B with primer sets CW2 was the downstream sequence of *TaCwi-A1*

gene on chromosome 2A. Three specific primer sets, C1, C2 and C3, were then designed based on the chromosome 2A-specific sequences to amplify the full-length sequence of *TaCwi-A1* on 2A and to identify sequence polymorphisms of different varieties. The three primer sets were also tested with CS nullisomic–tetrasomic lines N2B–T2A and N2D–T2B, and the wheat–barley chromosome substitution line N2A–S2H, to confirm their locations. Through this strategy, the complete genomic DNA sequence of *TaCwi-A1* on chromosome 2A was constructed.

There was a high degree of identity between the deduced amino acid sequence of *TaCwi-A1* and those from other plants. The deduced amino acid sequence of *TaCwi-A1* shared sequence identities of 79.3% with rice *GIF1* (Wang et al. 2008), 75.2% with maize *Incw2* (Carlson et al. 2000), and 67.8% with sorghum *Incw1* (Jain et al. 2008). The deduced amino acid sequences predicted from wheat *TaCwi-A1* cDNA showed the presence of highly conserved β -fructosidase motifs and cysteine catalytic sites. *TaCwi-A1* has a proline in the fifth residue of the cysteine catalytic site, indicating the typical structure of a *Cwi* gene. All known cell wall invertase genes have a proline residue in this sequence motif, whereas vacuolar invertase genes possess a valine residue (Goetz and Roitsh 1999; Tymowska-Lalanne and Kreis 1998).

Molecular mechanism for the association between allelic variations of *TaCwi-A1* and TKW

The exons are the regions encoding proteins in the open reading frame (ORF) of genes. In this study, five SNPs between *TaCwi-A1a* and *TaCwi-A1b* were found in the exons. However, four of them were silent mutations not producing changes in the amino acid sequences. The point mutation in the fifth exon at position 2,845th base pair in *TaCwi-A1* is a missense mutation resulting in the amino acid transition from Leu to Val at codon 442. This may account in part for the TKW variance between genotypes with the two alleles.

Introns have very important gene-regulatory functions. They can influence DNA transcription by variable splicing (Fedorova and Fedorov 2003). One way of alternative splicing is frame shifting, which alters the splicing position in the intron causing the DNA to be read differently (Bergman 2001). GT-AG

is a typical structure for intron splicing in higher plants. In the present case, all intron sequences have typical 5'-GT and 3'-AG boundaries. The point mutation in the fourth intron at bp 2727 used to design the STS marker, had changed from AT to AG, a 3'-splice site in introns. This implies that alternative splicing might be involved in the post-transcriptional modification of *TaCwi-A1b*, resulting in an altered transcription of the fifth exon and generating a different mRNA encoding a different polypeptide.

Evaluation of the functional markers CWI21 and CWI22

Function markers are ideal tools for marker-assisted selection in wheat breeding. In the present study, two complementary dominant functional markers CWI21 and CWI22 for *TaCwi-A1* were developed and validated. The markers showed a high association with TKW and thus can be used in the improvement of wheat yield. Previously, Sun et al. (2009) found two QTL clusters related to kernel width, TKW, and test weight on chromosomes 2A and 5D. Wang et al. (2009) detected QTL *QTgw.nfcrl-2A* for TKW on chromosome 2A, in the interval *Xbarc1165-Xbarc124*. In the present study, a major QTL for TKW on chromosome 2A co-segregating with CWI21 and CWI22 was detected. CWI21 and CWI22 (*TaCwi-A1*) were linked with *Xbarc15-2AL* with a genetic distance of 10.9 cM, and it might correspond to the QTL for kernel weight detected by Wang et al. (2009), which was located on chromosome 2A, in the interval *Xwmc522-Xcfa2263* near to *Xbarc15-2AL*.

There was considerable overlapping in the TKW distributions of varieties with the different fragments produced by CWI21 and CWI22. Some varieties with the 404-bp fragment amplified by CWI21 showed higher TKW, and some with 402-bp fragment showed lower TKW. This can be attributed to other genes affecting TKW. In previous studies, many QTLs for TKW were reported and located on almost all 21 wheat chromosomes (Li et al. 2007; Kumar et al. 2006). *TaCwi-A1* is only one of the genes associated with kernel weight, and obviously cannot explain all the variance of kernel weight. More research is needed to identify and clone other major genes or QTLs responsible for TKW and to develop function markers for wheat breeding.

Frequencies of allelic variants at the *TaCwi-A1* locus in Chinese commercial wheat varieties and landraces

Two allelic variants, *TaCwi-A1a* and *TaCwi-A1b*, were detected in Chinese commercial wheat varieties and landraces. The *TaCwi-A1a* allele associated with higher TKW was detected in 80.3% of Chinese commercial wheat varieties, but was present in only 5.6% of Chinese landraces. Increased TKW has been a major objective in Chinese wheat breeding programs during the past 60 years. Zhou et al. (2007a) noted that TKW had remarkably increased with annual genetic gains of 1.3% in the Northern China Plain over 30 years and that simultaneous increases in kernel number per spike and TKW had also occurred in the Yellow and Huai Valley Wheat Zone. The genetic improvement of grain yield primarily attributed to the increased TKW and kernel weight per spike were also observed in the Southern China winter wheat region from 1949 to 2000 (Zhou et al. 2007b). Wang et al. (2008) also reported that the *GIF1* gene in rice might have been a target for artificial selection because grain-filling directly contributes to kernel weight. It seems that selection for TKW in wheat breeding has resulted in higher frequencies of *TaCwi-A1a* associated with higher TKW in Chinese commercial wheat varieties. Although selection did not target *TaCwi-A1a*, the present work indicates that *TaCwi-A1* is an important gene related to wheat kernel weight, and ultimately, yield.

Acknowledgments The authors are very grateful to Prof. Robert McIntosh, Plant Breeding Institute, University of Sydney, for reviewing this manuscript. This study was supported by the National Basic Research Program (2009 CB118300), the National Science Foundation of China (30830072), National 863 Program (2006AA10Z1A7 and 2006AA100102), International Collaboration Project from the Ministry of Agriculture (2006-G2), and an earmarked fund for the Modern Agro-industry Technology Research System.

References

- Bagge M, Xia XC, Lübberstedt T (2007) Functional markers in wheat. *Curr Opin Plant Biol* 10:211–216
- Bergman J (2001) The functions of introns: from junk DNA to designed DNA. *Perspect Sci Chris Faith* 53:170–178

- Carlson SJ, Shanker S, Chourey PS (2000) A point mutation at the *Miniature1* seed locus reduces levels of the encoded protein, but not its mRNA, in maize. *Mol Gen Genet* 263:367–373
- 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
- Chourey PS, Jain M, Li QB, Carlson SJ (2006) Genetic control of cell wall invertases in developing endosperm of maize. *Planta* 223:159–167
- Collaku A (1989) Analysis of the structure of correlations between yield and some quantitative traits in bread wheat. *Buletini i Shkencave Bujqësore* 28:137–144
- Copeland L (1990) Enzyme of sucrose metabolism. *Methods Plant Biochem* 3:73–85
- Dimou M, Flemetakis E, Delis C, Aivalakis G, Spyropoulos KG, Katinakis P (2005) Genes coding for a putative cell-wall invertase and two putative monosaccharide/H⁺ transporters are expressed in roots of etiolated *glycine max* seedlings. *Plant Sci* 169:798–804
- Fedorova L, Fedorov A (2003) Introns in gene evolution. *Genetica* 118:123–131
- Godt DE, Roitsch T (1997) Regulation and tissue-specific distribution of mRNAs for three extracellular invertase isoenzymes of tomato suggests an important function in establishing and maintaining sink metabolism. *Plant Physiol* 115:273–282
- Goetz M, Roitsch T (1999) The different pH optima and substrate specificities of extracellular and vacuolar invertases from plants are determined by a single amino-acid substitution. *Plant J* 20:707–711
- Hirose T, Takano M, Terao T (2002) Cell wall invertase in the developing rice caryopsis: molecular cloning of *OsCIN1* and analysis of its expression in relation to its role in grain filling. *Plant Cell Physiol* 43:452–459
- Huang XQ, Cöster H, Ganai MW, Röder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theor Appl Genet* 106:1379–1389
- Jain M, Chourey PS, Li QB, Pring DR (2008) Expression of cell wall invertase and several other genes of sugar metabolism in relation to seed development in sorghum (*Sorghum bicolor*). *J Plant Physiol* 165:331–344
- Kim JY, Mahe A, Guy S, Brangeon J, Roche O, Chourey PS, Prioul JL (2000) Characterization of two members of the maize gene family, *Incw3* and *Incw4*, encoding cell-wall invertases. *Gene* 245:89–102
- Kosambi DD (1944) The estimation of map distance from recombination values. *Ann Eugen* 12:172–175
- Kumar N, Kulwal PL, Gaur A, Tyagi AK, Khurana JP, Khurana P, Balyan HS, Gupta PK (2006) QTL analysis for grain weight in common wheat. *Euphytica* 151:135–144
- Kumbhar MB, Larik AS, Hafiz HMI, Rind MJ (1983) Inter-relationship of polygenic traits affecting grain yield in *Triticum aestivum* L. *Wheat Inf Serv* 57:42–45
- Lagudah ES, Appels R, Brown AHD, McNeil D (1991) The *Nor-D3* locus of *Triticum tauschii*: natural variation and linkage to markers in chromosome 5. *Genome* 34:387–395
- Li SS, Jia JZ, Wei XY, Zhang XC, Li LZ, Chen HM, Fan YD, Sun HY, Zhao XH, Lei TD, Xu YF, Jiang FS, Wang HG, Li LH (2007) A intervarietal genetic map and QTL analysis for yield traits in wheat. *Mol Breed* 20:167–178
- Manly KF, Cudmore RH, Meer JM (2001) Map manager QTX, cross-platform software for genetic mapping. *Mamm Genome* 12:930–932
- 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
- Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C (2005) A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its uses to compare QTLs for grain yield across a range of environments. *Theor Appl Genet* 110:865–880
- Sturm A (1999) Invertase. Primary structures, functions, and roles in plant development and sucrose partitioning. *Plant Physiol* 121:1–8
- Sturm A, Chrispeels MJ (1990) cDNA cloning of carrot extracellular β -fructodisase and its expression in response to wounding and bacterial infection. *Plant Cell* 2:1107–1119
- Su JY, Tong YP, Liu QY, Li B, Jing RL, Li JY, Li ZS (2006) Mapping quantitative trait loci for post-anthesis dry matter accumulation in wheat. *J Integra Plant Biol* 48:938–944
- Sun XY, Wu K, Zhao Y, Kong FM, Han GZ, Jiang HM, Huang XJ, Li RJ, Wang HG, Li SS (2009) QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica* 165:615–624
- Tang GQ, Lüscher 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
- 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) Expression of the *Arabidopsis thaliana* invertase gene family. *Planta* 207:259–265
- Wang S, Basten CJ, Zeng ZB (2005) Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University. <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>
- Wang E, Wang JJ, Zhu XD, Wei H, 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
- Wang RX, Hai L, Zhang XY, You GX, Yan CS, Xiao SH (2009) QTL mapping for grain filling rate and yield-related traits in RILs of the Chinese winter wheat population Heshangmai × Yu8679. *Theor Appl Genet* 118:313–325
- Weber H, Borisjuk L, Heim U, Buchner P, Wobus U (1995) Seed coat-associated invertase of fava bean control both unloading and storage functions: cloning of cDNAs and cell type-specific expression. *Plant Cell* 7:1835–1846
- Zhang XK, Xiao YG, Zhang Y, Xia XC, Dubcovsky J, He ZH (2008) Allelic variation at the vernalization genes *Vrn-A1*,

- Vrn-B1*, *Vrn-D1*, and *Vrn-B3* in Chinese wheat cultivars and their association with growth habit. *Crop Sci* 48:458–470
- Zhou Y, He ZH, Sui XX, Xia XC, Zhang XK, Zhang GS (2007a) Genetic improvement of grain yield and associated traits in the Northern China winter wheat region from 1960 to 2000. *Crop Sci* 47:245–253
- Zhou Y, Zhu HZ, Cai SB, He ZH, Zhang XK, Xia XC, Zhang GS (2007b) Genetic improvement of grain yield and associated traits in the Southern China winter wheat region: 1949 to 2000. *Euphytica* 157:465–473