Theoretical and Applied Genetics
https://doi.org/10.1007/s00122-018-3064-z

ORIGINAL ARTICLE

Sequence divergence between spelt and common wheat

Miao Liu[1,2,3] · Qiang Zhao[4] · Feng Qi[4] · Jiri Stiller[1] · Shican Tang[4] · Jiashun Miao[4] · Jan Vrána[5] · Kateřina Holušová[5] · Dengcai Liu[3] · Jaroslav Doležel[5] · John M. Manners[1] · Bin Han[4] · Chunji Liu[1]

Received: 21 August 2017 / Accepted: 23 January 2018

Abstract

Key message **Sequence comparison between spelt and common wheat reveals that the former has huge potential in enriching the genetic variation of the latter.**

Abstract Genetic variation is the foundation of crop improvement. By comparing genome sequences of a *Triticum spelta* accession and one of its derived hexaploid lines with the sequences of the international reference genotype Chinese Spring, we detected variants more than tenfold higher than those present among common wheat (*T. aestivum* L) genotypes. Furthermore, different from the typical 'V-shaped' pattern of variant distribution often observed along wheat chromosomes, the sequence variation detected in this study was more evenly distributed along the 3B chromosome. This was also the case between *T. spelta* and the wild emmer genome. Genetic analysis showed that *T. spelta* and common wheat formed discrete groups. These results showed that, although it is believed that the spelt and common wheat are evolutionarily closely related and belong to the same species, a significant sequence divergence exists between them. Thus, the values of *T. spelta* in enriching the genetic variation of common wheat can be huge.

Introduction

Triticum spelta or spelt wheat is a husked hexaploid wheat closely related to the bread or common wheat (*T. aestivum* L). Available data indicated that the Asian spelt and the European spelta had different origins. Rather than being derived from the spreading of the Asian type, the European spelt was likely originated from hybridization between a tetraploid wheat and a free-threshing hexaploid wheat (Blatter et al. 2004). Although it is now only a minor crop grown in marginal regions in parts of Europe where environmental factors prevent the cultivation of common wheat, spelt wheat is one of the ancient grains with renewed interest. It possesses a unique flavor and a higher vitamin content, and is more nutritious than common wheat (Campbell 1997). Spelt is also one of the plant species that have been proposed to be rich sources of bioactive components and hence suitable for producing high value food products with enhanced health benefits (Shewry and Hey 2015).

In addition to their ability in abiotic tolerance, spelt wheat has also been used to improve biotic resistance including rusts and *Septoria tritici* blotch in modern wheat breeding (Kema 1992; Campbell 1997). Spelt wheat has also been used to improve kernel quality of common wheat. For example, ten waxy alleles not present in common wheat were

Communicated by Dr. Ling.

Miao Liu and Qiang Zhao have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00122-018-3064-z) contains supplementary material, which is available to authorized users.

✉ Bin Han
bhan@ncgr.cn

✉ Chunji Liu
chunji.liu@csiro.au

1 CSIRO Agriculture and Food, 306 Carmody Road, St Lucia, QLD 4067, Australia

2 Crop Research Institute, Sichuan Academy of Agricultural Sciences, 4 Shizishan Road, Jinjiang District, Chengdu 610066, China

3 Triticeae Research Institute, Sichuan Agricultural University, Wenjiang, Chengdu 611130, China

4 National Center for Gene Research, Chinese Academy of Sciences, 500 Caobao Road, Shanghai 200233, China

5 Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, 78371 Olomouc, Czech Republic

Published online: 09 February 2018

detected in spelt and they could affect starch quality thus could be used in modern wheat breeding (Guzmán et al. 2012).

Spelt and common wheat are evolutionarily so closely related that it is now widely accepted that they belong to the same species. Compared with common wheat, spelt is taller, has long and lax ears, a brittle rachis, and tight glumes (Mac 1966; Campbell 1997). The obvious differences on spike architecture are believed to be mainly controlled by a small number of genes. These include the spelt factor (q) on chromosome arm 5AL, the gene for ear compactness on chromosome arm 2DL, and the sphaerococcum factor (s1) which controls seed forms, on 3D (Morris and Sears 1967). However, spelt can be easily separated from common wheat using characteristics not associated with any of the above loci. For example, spelt wheat from Europe and common wheat form distinct groups based on high- and low-molecular-weight glutenin subunit compositions (An et al. 2005).

When searching for novel sources of Fusarium crown rot (CR) resistance, we found that one of the spelt accessions was among the best. Linkage mapping detected that the resistance of this accession (CSCR6) was controlled by a major locus on the long arm of chromosome 3B (Ma et al. 2010). Near-isogenic lines for this locus were generated (Ma et al. 2012) and transcriptomic analysis was conducted against several pairs of the NILs (Ma et al. 2014). Working toward the cloning of gene(s) underlying the 3BL locus, we also sequenced flow-sorted 3B chromosome from a CSCR6-derived hexaploid line CRNIL1A. An initial analysis of the 3B sequences targeting large fragments insertions and deletions identified an extensive dispensable genome component in wheat (Liu et al. 2016). In the study reported here, we further compared the 3B sequences of CRNIL1A with those of CS and detected huge numbers of SNPs, indels, and other types of variation. We then sequenced the whole genome of the *T. spelta* genotype CSCR6 and assessed its sequences against those of CS. Results from these assessments confirmed the rich sequence variation between *T. spelta* and common wheat, and thus the huge potential of exploiting the former in wheat breeding.

Methods

Illumina sequencing of the *T. spelta* genotype CSCR6

CSCR6, a genotype of hexaploid wheat *T. spelta* which was the donor for the CR resistance locus on chromosome arm 3BL in CRNIL1A (Ma et al. 2012), was used for whole-genome sequencing. The genomic DNA was extracted from the fresh leaf tissue of CSCR6. The sequencing library was constructed with an insert size of 400–500 base pairs (bp) for paired-end reads, following the manufacturer's protocol for Illumina. Raw data of 100-bp paired-end reads were generated by an Illumina HiSeq 2000 sequencer. All reads obtained in this work were deposited in the European Nucleotide Archive (ENA) with the accession number of PRJEB23983.

SNP detection between chromosome 3B of CSCR6 and Chinese Spring

The 100-bp paired-end reads were mapped to the wheat reference genome (IWGSC Survey sequence chromosomes version2 ftp://ftpmips.helmholtz-muenchen.de/plants/wheat/IWGSC/genome_assembly/genome_arm_assemblies_CLEANED/) (IWGSC 2014) using the software Bowtie2 version 2.2.3 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) (Langmead and Salzberg 2012). Read pairs aligned on the chromosome 3B were picked up and then mapped to the pseudomolecule of chromosome 3B (https://urgi.versailles.inra.fr/download/wheat/3B/) (Choulet et al. 2014) for precise alignment, using the software Smalt aligner (http://www.sanger.ac.uk/science/tools/smalt-0) with the parameters of a minimum Smith–Waterman (-m) value of 35, a maximum insert size (-i) of 800, and a minimum insert size (-j) of 50.

The single-nucleotide polymorphisms (SNPs) were then called using SSAHA_Pileup (version 0.8; ftp://ftp.sanger.ac.uk/pub/zn1/ssaha_pileup/) with three thresholds to post-filter unreliable SNPs: (1) SSAHA Pileup SNP score > = 20; (2) the ratio of SNP_base/other_base > = 10:1 was used for identifying the heterozygous SNPs following the removal of the low-quality bases; and (3) the minimum distance for adjacent SNPs > = 10 bp. To calculate the distribution of SNPs between the 3B sequences of CSCR6 and CS, a sliding window of 1 Mb along the chromosome sequence was used to analyze the rates of SNPs.

SNP analysis between CSCR6 and the wild emmer genome

To analyze the SNP distribution on other chromosomes, we mapped the 100-bp paired-end reads to the wild emmer reference genome (Avni et al. 2017) https://www.dropbox.com/sh/3dm05grokhl0nbv/AAC3wvlYmAher8fY0srX3gX9a?dl=0) using the software Bowtie2 version 2.2.6 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Samtools version 1.5 (Li et al. 2009) and picard version 2.10.10 (http://broadinstitute.github.io/picard/) were adopted to generate the index file and label the duplicates (REMOVE_DUPLICATES = false). GATK version 4.0 (McKenna et al. 2010) was then used to call variants. Homozygous SNPs were collected to calculate the SNP rate along the pseudomolecules of the wild emmer chromosomes, with a sliding window size of 1 Mb.

Genetic analysis between CSCR6 and common wheat genotypes

Using the same methods as described above, the 100-bp paired-end reads of CSCR6 were mapped to the wheat reference genome (IWGSC Survey sequence chromosomes). Based on the wheat-exome capture (WEC) sequencing data, variants of CSCR6 at the corresponding positions were merged into the matrix of the 62 diverse hexaploid wheat accessions (Jordan et al. 2015). Considering that the sequence depth of CSCR6 was only 12×, many sites may not be covered or read depths may not be sufficient to determine the genotype of the site in variant calling. To deal with this issue, k-nearest neighbor imputation was carried out for rows which had more than 50% entries missing. Missing data were imputed by R package impute [Bioconductor version: Release (3.6)] (https://bioconductor.org/packages/release/bioc/html/impute.html) with parameters: $k = 400$, rowmax = 0.5, colmax = 0.9, maxp = 1500, rng.seed = 362,436,069. With the variant matrix of the 62 common wheat accessions, R package stats (version 3.3) (https://www.r-project.org/) was used to conduct the principal component analysis (PCA) and hierarchical cluster (HCA) analyses.

Mapping reads of flow-sorted chromosome 3B from CRNIL1A to the 3B pseudomolecule of CS

The flow-sorted chromosome 3B sequences from the hexaploid wheat line CRNIL1A were obtained from a previous study (Liu et al. 2016). This line was developed from a backcross of Janz*2/CSCR6. Janz is a commercial bread wheat variety (Ma et al. 2012). After trimming, high-quality reads of CRNIL1A 3B were mapped to the CS 3B pseudomolecule (https://urgi.versailles.inra.fr/download/wheat/3B/) (Choulet et al. 2014) using CLC Genomics Workbench 8.0 (CLC Bio, Cambridge, MA, USA). Reads were aligned using the following settings: mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.9, and similarity fraction = 0.9.

Variant analysis between CRNIL1A 3B and CS 3B pseudomolecule

Sequence variant detection was performed using the Basic Variant Detection function from the CLC Genomics Workbench 8.0, with settings specifying a ploidy level of 6. To minimize the rate of false-positive, the following strict filters were applied: ignored the broken pairs, minimum coverage = 4, minimum count = 2, minimum frequency = 90%, neighborhood radius = 5, minimum central quality = 20, and minimum neighborhood quality = 15. A sliding window of 1 Mb along the chromosome sequence was used to analyze the rates of SNPs and other variants including MNVs (multinucleotide variants), insertions, deletions, and replacements.

Validation of single-nucleotide polymorphisms

To evaluate the authenticity of the detected SNPs, 20 sequences were randomly selected from CS 3B pseudomolecule for validation. One pair of primers was designed for each of these sequences, which was specific to 3B chromosome. These sequences contain a total of 85 SNPs between CSCR6 and CS 3B, and 116 between CRNIL1A and CS 3B (Table S1).

Genomic DNA was isolated from 20-day-old seedlings of CSCR6 and CRNIL1A using the method of hexadecyl-trimethyl-ammonium bromide (CTAB) (Murray and Thompson 1980). PCR amplification was performed in 10-µl reaction mixtures containing 50 ng of genomic DNA, 200 µM of each dNTP, 0.2 µM of each primer, and 0.5 units of Taq DNA polymerase. The cycling conditions for the reaction were 94 °C for 5 min to pre-denature, which was followed by 35 cycles of 94 °C for 1 min, 30 s at the appropriate annealing temperature (ranging from 55 to 60 °C depending on the primers, see Table S1 for details), 72 °C for 1 min, and a final extension at 72 °C for 10 min. PCR products were separated on 1.0% agarose gels.

PCR products were purified using the QIAquick Extraction Kit (QIAGEN). The recovered PCR fragments were inserted into the pGEM-T easy vector (Promega) and transformed into *Escherichia coli*. At least three independent clones for each fragment were Sanger-sequenced in both directions. Allele sequences from CSCR6 and CRNIL1A were aligned and compared with CS the 3B sequences using the DNAman software package, and each of the anticipated SNPs was visually examined.

SSR detection from the 3B sequences of CRNIL1A

SSR detection was carried out using the Tandem Repeat Finder (TRF) software (Benson 1999) and Tandem Repeats Analysis Program (TRAP) (Sobreira et al. 2006). A highly stringent criterion was used to identify both perfect and imperfect SSRs with 1–5 nucleotide cores: (1) for mono-nucleotide or di-nucleotide SSRs, the minimum length was 16 bp; for tri-nucleotide SSRs, the minimum length was 18 bp; for tetra-nucleotide or penta-nucleotide SSRs, the minimum length was 20 bp; (2) For imperfect SSRs, less than 10% mismatches or gaps relative to a perfect SSR of the same length and motif were included.

Results

SNPs detected on chromosome 3B between the *T. spelta* genotype CSCR6 and the common wheat genotype CS

Six illumina Hiseq 2000 lanes of CSCR6 containing approximately 217.2 Gb of 100-bp paired-end reads were obtained. As the genome size of hexaploid common wheat was about 17 Gb (IWGSC 2014), these data represented a sequencing depth of 12.8-fold across the whole genome. After removing low-quality reads, the remaining reads were first mapped to the IWGSC's whole-genome survey sequences of CS. This allowed a total of 14.0-Gb sequences located on the 3B chromosome with an average depth of 15.8-fold. The CSCR6 3B reads were then aligned to the pseudomolecule of CS 3B for precise position. Following SNP calling and subsequent filtering, 1388,078 SNPs between CSCR6 and CS along the chromosome 3B were detected. Interestingly, the significantly lower density of variation commonly reported at the centromere and pericentromeric region in wheat were not found in this analysis. The density of SNPs in the centromeric and pericentromeric region was not any lower than those in other regions along the chromosome (Fig. 1). Based on the annotation features of 7264-known protein-coding genes and their positions on this chromosome, only 10,183 or 0.7% of the SNPs were found in coding regions of 2938 genes, which indicated that most SNPs happened on untranslated regions.

SNP analysis between CSCR6 and the wild emmer genome

The number of SNPs detected between CSCR6 and the wild emmer genome varied across different chromosomes. The highest number was observed on chromosome 3B (3,826,870), while lowest on 6B (2,049,741) (Table S2). The SNP frequency also varied among the chromosomes with 4B and 6B having the highest (5418.5) and lowest (2914.8) density (2914.8) per 1-Mb interval, respectively (Table S2). The distribution of SNPs within a chromosome was also not even (Fig. 2). The density of SNPs in the centromeric and pericentromeric region was also not significantly lower than those in other regions in all but chromosome 2 (Fig. 2).

Genetic analysis between spelt (CSCR6) and common wheat

After removing the rows which contain NA records, only 28,838 of 1,341,350 rows were kept for PCA (PC1: 11.22% variance, PC2: 7.513% variance) and HCA analyses between CSCR6 and the 62 common wheat accessions. The results showed that CSCR6 and common wheat accessions were clearly separated in a PCA along both first and second axes (Fig. 3). The HCA results also showed that CSCR6 was not closely related with the common wheat accessions (Fig. S1). In addition, 1,151,351 complete rows were obtained following the k-nearest neighbor imputation and used for PCA (PC1: 10.60% variance, PC2: 5.71% variance) and HCA analyses. Both analyses showed that CSCR6 and the 62 common wheat accessions formed discrete groups (Figs. S2, S3). These results all indicated that a significant sequence divergence exists between the spelt and common wheat genotypes.

Variants detected on chromosome 3B between CRNIL1A and CS

As described before (Liu et al. 2016), about 65.0 million of CRNIL1A 3B reads were successfully mapped to the 3B pseudomolecule of CS and they covered a total of 692.7 Mb of the CS 3B pseudomolecule. A total of 2,264,952 sequence variants were detected on this chromosome between CRNIL1A and CS. Of these, most were SNPs, followed by MNVs, deletions, insertions, and replacements.

More transition types of change were detected compared with those of transversions among the 2,125,304 SNPs

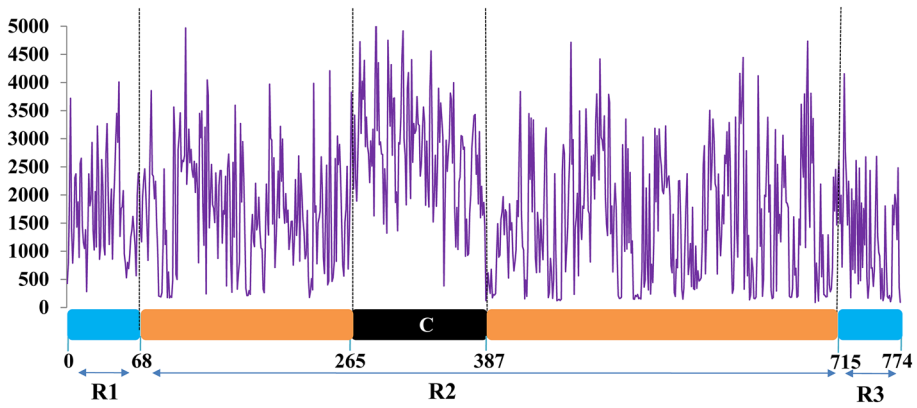

Fig. 1 Distribution of SNPs on chromosome 3B between the *T. spelta* genotype CSCR6 and CS. The x-axis represents the physical distance along whole CS 3B chromosome, split into 1-Mb windows. The y-axis indicates the number of SNPs. R1 and R3 (blue): distal regions; R2 (yellow and black): proximal region; and C (black): the centromeric and pericentromeric regions (Choulet et al. 2014) (color figure online)

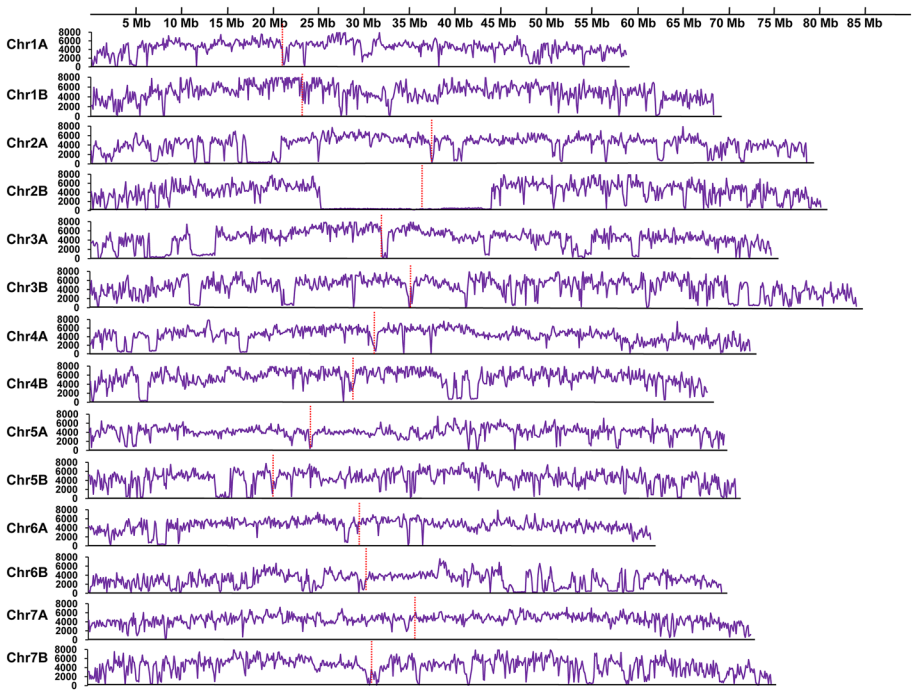

Fig. 2 Distribution of SNPs between CSCR6 and emmer wheat on each of the 14 chromosomes. The x-axis represents the physical distance (split into 1-Mb windows) along chromosomes. The y-axis indicates the number of SNPs, and the vertical red dotted lines indicate centromere locations as reported by Avni et al. (2017) (color figure online)

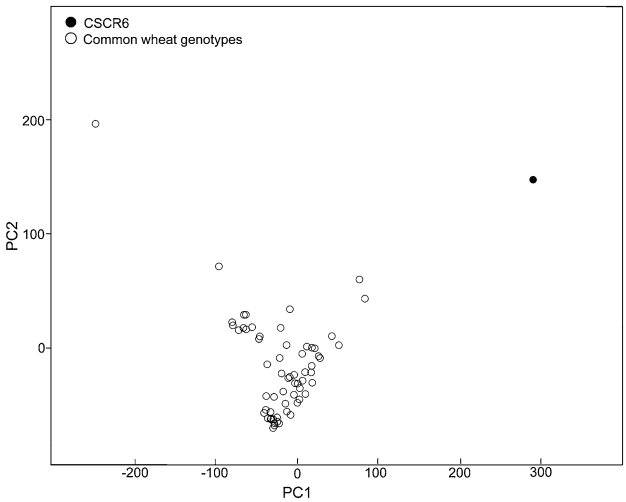

Fig. 3 Principal component analysis (PCA) between CSCR6 and 62 common wheat accessions. The analysis was carried out with the SNP variants between CSCR6 and 62 common wheat accessions following the removal of the rows which contained NA records in the variant matrix

detected. The numbers of C/T and G/A transitions were similar, and the difference between G/T and A/C transversions was also not significant. However, the total number of transitions was about 2.27-fold of that the transversions (Table 1). In addition, according to the positions of SNPs between CRNIL1A and CS on chromosome 3B, we found 696,493 SNPs overlapping with those detected between CS and CSCR6.

Table 1 Classification of nucleotide substitutions in SNPs on chromosome 3B between CRNIL1A and CS

Substitutions	No.	Percentage
Transitions (Ts)	1,475,281	–
G/A	737,599	50
C/T	737,682	50
Transversions (Tv)	650,023	–
G/T	185,998	28.6
A/C	186,376	28.7
C/G	164,668	25.3
T/A	112,981	17.4

The lengths for the 43,870 insertions detected varied from 1 to 22 bp and those for the 45,007 deletions varied from 1 to 27 bp. Mono-nucleotide variations accounted for majority of the insertions (83.7%) and deletions (79.2%) (Fig. 4). In fact, the numbers or proportions of the insertions or deletions showed strong negative correlations with the length of the variants. Those insertions and deletions with 5 bp or longer accounted only for 2.7 and 4.6%, respectively, of the totals detected. Similarly, those 2 bp variations accounted for 95.5% of the 47,428 MNVs detected (Fig. 4), and most of the replacements detected were between mono-nucleotide and di-nucleotides (71.7%).

Based on the annotation features of 7264-known protein-coding genes and their positions on this chromosome, in total, only 14,693 (or 0.6% of total variants) of them were located in the coding regions (Table S3), which also showed

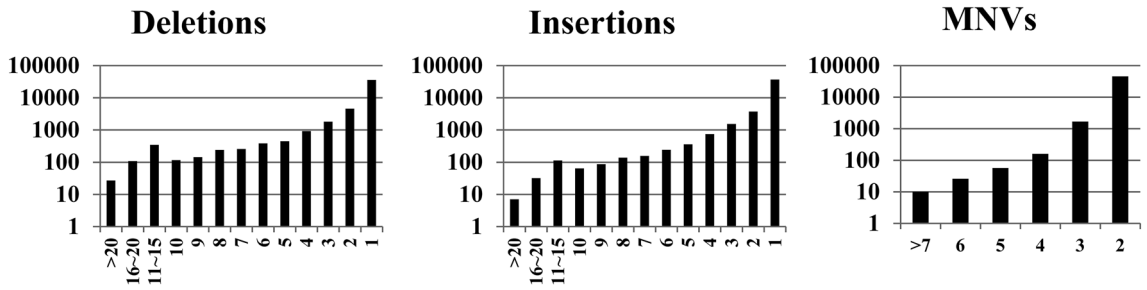

Fig. 4 Distribution of insertions, deletions, and MNVs between the CRNIL1A 3B sequence and CS 3B pseudomolecule based on their lengths. The x-axis represents the lengths (in base pairs) of variants and the y-axis indicates the number of variants

the sequence variation mainly occurred in the non-genic regions.

Distribution of various types of variants detected along chromosome 3B between CRNIL1A and CS

Based on the size of the CS chromosome 3B pseudomolecule, the average densities of the variants detected per 1-Mb window were 2744.5 for SNPs, 61.2 for MNVs, 56.7 for insertions, 58.1 for deletions, and 4.3 for replacements.

The variants detected between the CRNIL1A 3B and CS 3B pseudomolecule were not evenly distributed along the chromosome. There were some intervals with significant variations in SNPs frequencies. We found that 16 intervals (1 Mb each) with significantly higher SNPs frequencies ($\geq$ 6000 SNPs each interval), and 16 intervals with prominently lower SNPs frequencies ($\leq$ SNPs 200 SNPs each interval). Similar to the SNP distribution between the CSCR6 and CS, the densities of variants in centromere/pericentromeric region were also not significantly lower (Fig. S4).

Validation of single-nucleotide polymorphisms

Two of the twenty primers designed for evaluating the authenticity of the SNPs did not produce clean and specific PCR products (Table S1), and thus, SNPs in these two sequences were ignored in further analysis. Sequences of the other 18 PCR products were determined by direct sequencing. They contained 77 SNPs between CSCR6 and CS 3B, and 106 between CRNIL1A and CS 3B. Among them, 69 (89.6%) were the same as those detected by sequence comparison between CSCR6 and CS 3B and 102 (96.2%) were the same for the comparison between CRNIL1A and CS 3B, respectively (Table S1). The low false discovery rate (10.4 and 3.8%, respectively) could be due to several reasons including sequencing method, sequence coverage, mapping method, or criterion used for SNPs detection.

SSRs detected from the 3B sequences of CRNIL1A

In total, 21,542 potential SSRs were identified from CRNIL1A 3B sequence, which means, on average, that there is one SSR for every 30.41 kb. The most abundant type of repeat units was di-nucleotide (61.3%), followed by tri-nucleotide (25.7%), penta-nucleotide (6.2%), tetra-nucleotide (5.76%), and mono-nucleotide (1.0%) repeat motifs (Table S4). The SSRs detected contained a total of 143 motif types of which mono-, di-, tri-, tetra-, and penta-nucleotide repeats had 2, 4, 10, 33, and 94 types, respectively (Table S4). In all repeat motifs, AG/TC (34.4%) was the most frequent, followed by AT/TA, AC/TG, AAG/TTC, AGG/TCC, AAAT/TTTA, AAAAG/TTTTC, and so on (Fig. 5, Table S4).

Discussion

Genetic variation is the foundation of plant breeding. By comparing 3B sequences from a single accession of *T. spelta*-derived line with the international reference genotype of hexaploid wheat CS, 2,125,304 SNPs were detected. Assuming that a similar level of variants could be detected across the wheat genome, the total number of SNPs between these two genotypes would be as high as 46.6 million based on the proportion of chromosome 3B in the wheat genome (Choulet et al. 2014). Compared with those reported among hexaploid wheat cultivars (Montenegro et al. 2017), the number of SNPs detected between spelt and common wheat genotypes was much higher. The genetic analysis based on sequence differences also found that *T. spelta* and common wheat accessions formed discrete groups. Clearly, although it is believed that spelt and common wheat are evolutionarily very closely related and belong to the same species, significant sequence divergence exists between them. As there are no known barriers in homologous recombination between spelt and common wheat chromosomes, it should be highly feasible to exploit the novel variants in the former

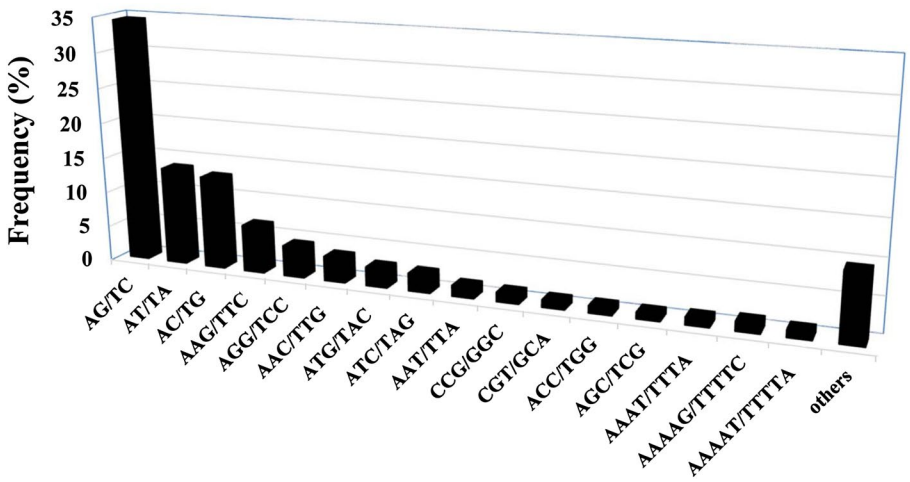

Fig. 5 Frequency distribution of SSRs based on motif types

to improve the genetic variation of the latter through the traditional hybridization.

To confirm the huge amount of sequences detected between the *T. spelta*-derived line CRNIL1A and CS, we sequenced and assessed the whole genome of the *T. spelta* accession CSCR6. Surprisingly, the number of SNPs detected between the *T. spelta* accession and CS was lower than that between its hexaploid derivative and CS. Two possible factors might have contributed to the difference. One of them is the quality and the other quantity of the genome sequences used. The flow-sorted chromosome 3B of the *T. spelta*-derived line CRNIL1A was sequenced on Illumina MiSeq instrument and pair end reads with length of 2 × 300 bp and sequences representing a depth of 41 times for this chromosome were obtained (Liu et al. 2016). The whole genome of the *T. spelta* accession CSCR6 was sequenced on Illumina HiSeq 2000 sequencer and the 217.2-Gb raw sequences of 100-bp paired-end reads represented only a sequence depth of 12.8 times. It seems that both sequence quality and depth affect the numbers of variants detected. Furthermore, the recurrent parent of CRNIL1A, 'Janz', is a commercial bread wheat variety; it also has various sequence diversity with CS. Therefore, the variants detected between CRNIL1A and CS could contain both sequence variations between CSCR6 with CS and Janz with CS. This could be another reason that the number of SNPs detected between the *T. spelta* accession (CSCR6) and CS was lower than that between its hexaploid derivative (CRNIL1A) and CS. On the other hand, we found that 696,493 SNPs between CRNIL1A and CS were overlapping with those between CS and CSCR6. These SNPs could be more useful for the cloning of CR resistance gene(s) underlying the 3BL locus.

In addition to the SNPs, this study also detected large numbers of other types of variants including MNVs, insertions, deletions, replacements, and SSRs from the flow-sorted chromosome 3B of the *T. spelta* derivative. The numbers of such variants were much smaller than the SNPs detected. However, due to the presence of the three homoeologous groups, primers designed based on single SNPs often fail to amplify a specific fragment for a given subgenome. When two or more SNPs or indels were used in designing primers; however, the success rates of the derived markers to amplify locus-specific markers could be dramatically increased (Ma et al. 2015). Thus, these non-SNP variations could be very useful for designing locus-specific markers.

Results from previous studies showed that genetic variants were not evenly distributed along chromosomes. Both sequence variation and recombination frequency are lower around the centromeric regions in comparison with those at either ends of a chromosome. This 'V'-shaped distribution of variation and recombination frequency is more pronounced in cereals (Akhunov et al. 2010; Choulet et al. 2014; Goettel et al. 2014; Avni et al. 2017). However, this pattern was not observed in the distribution of the variants along the 3B chromosome detected between the *T. spelta* genotype and CS. Rather, the variants detected in this study are more evenly distributed along the chromosomes (Figs. 1, 2, S4). The possible reasons for this difference between the current study and those earlier ones are not clear, but the results suggest that *T. spelta* can be particularly efficient in improving genetic variation near the centromeric regions of common wheat chromosomes.

Author contribution statement CL and DL conceived the study; BH, JMM, CL, QZ, and FQ obtained the sequences of the *T. spelt* genotype CSCR6; JD, JV, and KH sorted and sequenced chromosome 3B of the genotype CRNIL1A; ML, JS, QZ, FQ, ST, and JM conducted data analysis with supervision from CL and BH; ML, QZ, and CL prepared the manuscript with contribution from JD.

Acknowledgements Work reported here was supported by the Commonwealth Scientific and Industrial Organization (CSIRO), Australia (Project code: R-10191-01), and a Chinese Academy of Sciences-CSIRO Joint Project (Project No. R-1910-1). ML is grateful to the Sichuan Agricultural University and the China Scholarship Council for funding his visit to CSIRO. JV, KH, and JD were supported by the Czech Ministry of Education, Youth and Sports (Award LO 1204 from the National Program of Sustainability I).

Compliance with ethical standards

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

References

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N et al (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11(1):702

An X, Li Q, Yan Y, Xiao Y, Hsam SLK et al (2005) Genetic diversity of European spelt wheat (*Triticum aestivum* ssp. *spelta* L. em. Thell.) revealed by glutenin subunit variations at the *Glu-1* and *Glu-3* loci. Euphytica 146(3):193–201

Avni R, Nave M, Barad O, Baruch K, Twardziok SO et al (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357:93–97

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27(2):573

Blatter RHE, Jacomet S, Schlumbaum A (2004) About the origin of European spelt (*Triticum spelta* L.): allelic differentiation of the HMW glutenin B1-1 and A1-2 subunit genes. Theor Appl Genet 108:360–367

Campbell KG (1997) Spelt: Agronomy, genetics, and breeding. Plant Breed Rev 15:187–213

Choulet F, Alberti A, Theil S, Glover N, Barbe V et al (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science 345(80):1249721

Goettel W, Xia E, Upchurch R, Wang ML, Chen P et al (2014) Identification and characterization of transcript polymorphisms in soybean lines varying in oil composition and content. BMC Genomics 15(1):1

Guzmán C, Caballero L, Martín LM, Alvarez JB (2012) Waxy genes from spelt wheat: new alleles for modern wheat breeding and new phylogenetic inferences about the origin of this species. Ann Bot 110(6):1161–1171

Jordan KW, Wang S, Lun Y, Gardiner LJ, MacLachlan R et al (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biol 16(1):48

Kema GHJ (1992) Resistance in spelt wheat to yellow rust. III. Phytogenetical considerations. Euphytica 63:225–231

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al (2009) The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25(16):2078–2079

Liu M, Stiller J, Holušová K, Vrána J, Liu D et al (2016) Chromosome-specific sequencing reveals an extensive dispensable genome component in wheat. Sci Rep 6:36398

Ma J, Li HB, Zhang CY, Yang XM, Liu YX et al (2010) Identification and validation of a major QTL conferring crown rot resistance in hexaploid wheat. Theor Appl Genet 120(6):1119–1128

Ma J, Yan GJ, Liu CJ (2012) Development of near-isogenic lines for a major QTL on 3BL conferring *Fusarium* crown rot resistance in hexaploid wheat. Euphytica 183(2):147–152

Ma J, Stiller J, Zhao Q, Feng Q, Cavanagh C et al (2014) Transcriptome and allele specificity associated with a 3bl locus for fusarium crown rot resistance in bread wheat. PLoS ONE 9(11):e113309

Ma J, Stiller J, Zheng Z, Liu YX, Wei Y et al (2015) A high-throughput pipeline for detecting locus-specific polymorphism in hexaploid wheat (*Triticum aestivum* L.). Plant Methods 11(1):39

MacKey J (1966) Species relationship in Triticum. Hereditas 2:237–276

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing nextgeneration DNA sequencing data. Genome Res 20:1297–1303

Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H et al (2017) The pangenome of hexaploid bread wheat. Plant J 90(5):1007–1013

Morris R. Sears ER (1967) The cytogenetics of wheat and its relatives. In: Quisenberry KS, Reitz LP (eds) Wheat and wheat improvement. Madison, pp 19–87

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326

Shewry PR, Hey S (2015) Do "ancient" wheat species differ from modern bread wheat in their contents of bioactive components? J Cereal Sci 65:236–243

Sobreira TJP, Durham AM, Gruber A (2006) TRAP: automated classification, quantification and annotation of tandemly repeated sequences. Bioinformatics 22(3):361–362