Planta (2019) 249:663–675
https://doi.org/10.1007/s00425-018-3033-4

ORIGINAL ARTICLE

Physical location of tandem repeats in the wheat genome and application for chromosome identification

Tao Lang[1] · Guangrong Li[1,2] · Hongjin Wang[1] · Zhihui Yu[1] · Qiheng Chen[1] · Ennian Yang[3] · Shulan Fu[4] · Zongxiang Tang[4] · Zujun Yang[1,2]

Received: 25 June 2018 / Accepted: 20 October 2018 / Published online: 24 October 2018

Abstract

Main conclusion **A general distribution of tandem repeats (TRs) in the wheat genome was predicted and a new web page combined with fluorescence in situ hybridization experiments, and the newly developed Oligo probes will improve the resolution for wheat chromosome identification.**

Comprehensive sequence analysis of tandem repeats (TR) in the wheat reference genome permits discovery and application of TRs for chromosome identification. Genome-wide localization of TRs was identified in the reference sequences of Chinese Spring using Tandem Repeat Finder (TRF). A database of repeats unit size, array number, and physical coverage length of TRs in the wheat genome was built. The distribution of TRs occupied 3–5% of the wheat chromosomes, with non-random dispersal across the A, B, and D genomes. Three classes of TRs surrounding the predicted genes were compared. An optimized computer-assisted website page B2DSC was constructed for the general distribution and chromosomally enriched zones of TR sequences to be displayed graphically. The physical distribution of predicted TRs in the wheat genome by B2DSC matched well with the corresponding hybridization signals obtained with fluorescence in situ hybridization (FISH). We developed 20 oligonucleotide probes representing 20–60 bp lengths of high copy number of TRs and verified by FISH. An integrated physical map of TR-Oligo probes for wheat chromosome identification was constructed. Our results suggest that the combination of both molecular cytogenetics and genomic research will significantly benefit wheat breeding through chromosome manipulation and engineering.

Keywords Chromosome identification · Fluorescence in situ hybridization · Tandem repeats · Wheat genome

Introduction

In humans and other higher organisms, repetitive DNA sequences make up more than 50% of the genome (Charlesworth et al. 1994; Janicki et al. 2011). Based on their distribution modes in linear genomes, the repetitive sequences can be broadly classified into two categories: dispersed and tandem repetitive sequences. Tandem repeats (TR) are organized as 'head-to-tail' arrays, while dispersed repeats are distributed as individual copies in the genome (Jurka et al. 2005). DNA repeats have been previously regarded as genomic 'junk', because, most apparently, they do not encode functional proteins for host cells (Doolittle and Sapienza 1980; Orgel and Crick 1980). However, research studies have continued to unravel the many functions of TRs and have spurred a growing interest in gaining a better understanding of the evolutionary history and genomic composition of TRs (Gemayel et al.

Tao Lang and Guangrong Li have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00425-018-3033-4) contains supplementary material, which is available to authorized users.

✉ Zujun Yang
yangzujun@uestc.edu.cn

[1] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China

[2] Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 611731, China

[3] Crop Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610066, Sichuan, China

[4] Province Key Laboratory of Plant Breeding and Genetics, Sichuan Agricultural University, Chengdu 611130, China

2010). Genomic tandem repeats (TRs) may serve as placeholders for epigenetic signals that govern heterochromatin formation, or may function in the repair of double-strand DNA breaks (Gemayel et al. 2012).

Because of the abundance of published sequence data for wheat, researchers now have an opportunity to apply bioinformatic approaches to reveal the genomic locations of tandem repeats in such large chromosomes (Pavlek et al. 2015). The chromosome complement of bread wheat, *Triticum aestivum* L. ($2n = 6x = 42$), consists of three closely related subgenomes (A, B, and D) totaling about 17 Gb with a high proportion (> 70%) of repetitive DNA (Brenchley et al. 2012; Zimin et al. 2017). DNA sequences, in particular, the tandem repeats in the wheat chromosomes, have been previously isolated based on a combination of genomic and cytogenetic studies (Rayburn and Gill 1986; Anamthawat-Jonsson and Heslop-Harrison 1993; Vershinin et al. 1994; Kishii et al. 2001). Fluorescence in situ hybridization (FISH) with labeled TRs as probes has greatly assisted in chromosome identification procedures (Mukai et al. 1993; Jiang and Gill 2006). More recently, the simplified technique of non-denaturing FISH (ND-FISH), using oligonucleotide probes of repetitive sequences, has been proved to be a simple, cheap, and high-throughput method for identifying the chromosomes of wheat and its relatives (Cuadrado and Jouve 2010; Tang et al. 2014; Fu et al. 2015; Tang et al. 2016; Li et al. 2016; Cuadrado et al. 2017; Du et al. 2017; Xiao et al. 2017). A key prerequisite for successful ND-FISH analysis is the development of suitable oligonucleotide probes. The essentials to the successful development of new multiplex DNA probes for high-resolution FISH on wheat chromosomes are (a) analysis of the length of each potentially useful probe, (b) determination of the number of repeat units, and (c) physical localization of the probes on the A, B, or D genomes (Pavlek et al. 2015; Lang et al. 2018).

In the current study, we present details of a new web server B2DSC which can be used to predict the chromosomal distribution of TRs based on the recently uploaded wheat genomic reference sequences v1.0. The locations of TRs along chromosome arms by the FISH analysis were mostly consistent with the physical localizations predicted by the web server. The predicted TR-Oligo probes were abundantly and widely distributed across the wheat genome, which will be useful in the future for the precise identification of chromosome segments.

Materials and methods

Plant materials

Common wheat line Chinese Spring was originally obtained from the Wheat Genetics Resource Center, Kansas State University, USA, and maintained in our laboratory of Plant Molecular Cytogenetics at the University of Electronic Science and Technology of China. Wheat cultivar Chuanmai 62 (CM62) was developed and provided by Dr. Ennian Yang at the Crop Research Institute, Sichuan Academy of Agricultural Sciences, China.

TR search in wheat genome

The IWGSC RefSeq v1.0 assembly and genomic annotation was downloaded from (https://wheat-urgi.versailles.inra.fr/), and the chromosome sequence was input as a unique FASTA file. The sequence of each chromosome was analyzed using the Tandem Repeats Finder (TRF) algorithm (Benson 1999). All chromosomes were processed with the TRF using alignment parameters of 2, 7, and 7 for match, mismatch, and indels, respectively. A minimum alignment score of 50 was used to identify the tandem repeats in each chromosome (Benson 1999). The tandem repeats annotated by the TRF were divided into three classes according to the size of period distances (< 20, 20–60 and > 60). Results on the distribution of TRs among the A, B, and D chromosomes were analyzed using the SPSS software (version 22.0, SPSS, Chicago, IL).

Optimal web server for TR physical mapping

A web server B2DSC (http://mcgb.uestc.edu.cn/b2dsc) is implemented based on Perl real-time web framework Mojolicious (v6.15). The main steps are shown in Fig. 1. The TR sequence comparison for the wheat genome used the local BLAST program (Camacho et al. 2009). Briefly, each query for DNA sequences over 10 bp is submitted by clicking the 'BLASTN' button (Fig. S1). The '*blastn*' parameters used are customizable. The high-scoring segment pair (HSP) is then produced by combining both parameters of the percentage of identical matches (*pident*) and query coverage per subject (*qcovs*) values no less than 90% (Fig. S2). HSP values lower than those pre-set are discarded. The number of repeats found per Mbp is calculated. The start position, end position, and strand are also collected for each Mbp, and stored in a FASTA format.

After filtering, the counts of repeats per Mbp along each chromosome are used to construct images, which are mainly coded in JavaScript. Three categories of customizable parameters: 'Size', 'Offset', and 'Color' can be set before clicking on the button 'Plot' to draw (Fig. S2). Only the chromosomes having matched HSPs are plotted. By default, the gradient colors are utilized to indicate different levels of counts per Mbp (Fig. S2). The physical distribution of Oligo-CCS1 can be added on the left to show the positions of the centromeres on each chromosome.

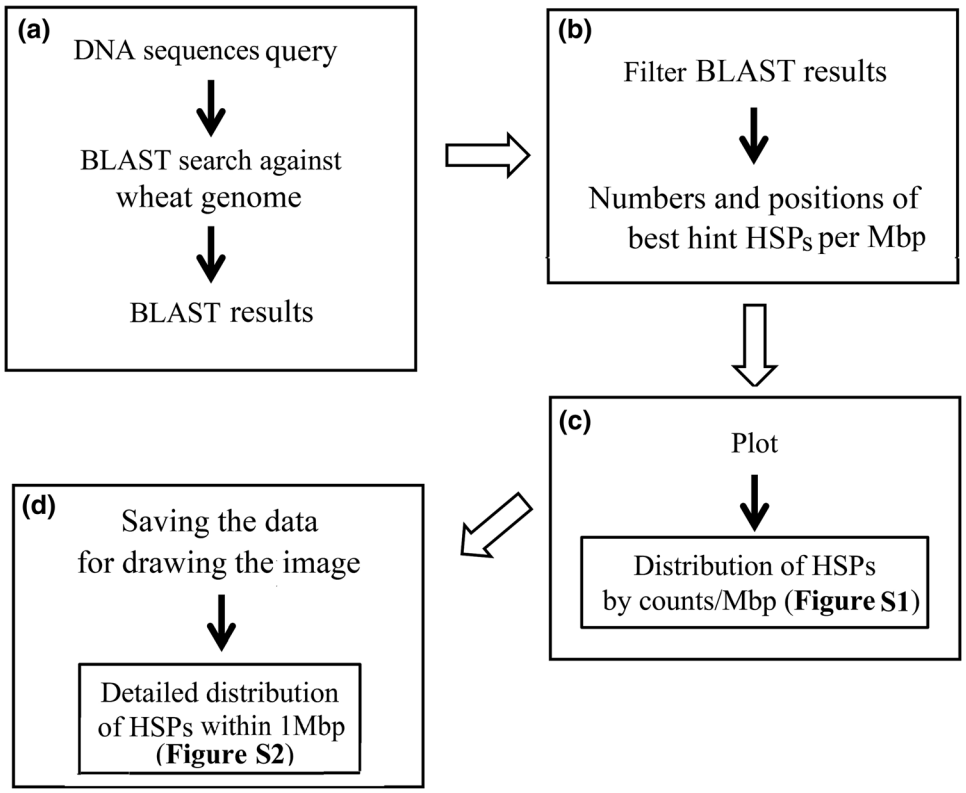

Fig. 1 Schematic illustration of the four steps for using the B2DSC web server. Users are able to scrutinize the results of each step. **a** BLAST search against whole chromosome sequences. **b** Filter HSPs (high-scoring segment pairs) and prepare data for an image plot. **c** Get the distribution chart of the query sequence (per Mbp). **d** Detail distribution in 1 Mbp

The detailed distribution of the homologous sequences is plotted within 1 Mbp. This section is coded in JavaScript and interacts with the server based on AJAX (Asynchronous JavaScript and XML). When a bar or barcode of the query sequence is clicked, the data of the corresponding region are pulled from the server, and then, the number of HSPs per 100 kb, 10 kb, and 1 kb is plotted (Fig. S3).

FISH analysis of TR-derived Oligo probes

Roots of wheat cv. Chinese Spring were collected when they were about 2–3 cm long. Excised root tips were treated with nitrous oxide for 2 h under 1.0 MPa pressure. The treated root tips were fixed in 90% acetic acid for 10 min. After the roots were washed with distilled water, they were digested in 2% cellulase and 1% pectolyase (Yakult Pharmaceutical, Tokyo) for 55 min at 37 °C. The digested root sections were washed and then mashed to form a cell suspension in 100% acetic acid. The cell suspension was dropped onto glass slides for chromosome preparation according to Kato et al. (2004). The TR-Oligo probes were synthesized by Shanghai Invitrogen Biotechnology Co. Ltd. (Shanghai, China), with a 6-carboxyfluorescein (6-FAM) or 6-carboxytetramethylrhodamine (Tamra) attached to the 5′ end. The protocol of ND-FISH using synthesized probes was described by Fu et al. (2015). The chromosome squashes for sequential ND-FISH were washed twice, each for 5 min with 0.1% Tween 20 in 2 × SSC, to remove the hybridization signals. Photomicrographs of FISH chromosomes were taken with an Olympus BX-51 microscope equipped with a DP-70 CCD camera operated with DPmanager software (Olympus).

Integrated physical map of TR probes

The FISH signals were measured using the mask tool to select regions of interest for at least ten mitotic cells for each probe. To integrate FISH results with the predicted data produced by the website B2DSC, the relatively strong hybridization signals and predicted copy number over 20 per Mb scale at the corresponding regions were used for map construction.

Results

Genomic distribution of TRs in wheat subgenomes

To investigate the proportional contribution of the TRs in the wheat genome, sequence data from all 21 assembled individual chromosomes were analyzed using the TRF software. A Perl script allowed us to identify non-redundant TRs from the raw TRF result. The database of all the raw and filtered non-redundant TRs is available at http://mcgb.uestc.edu.cn/tr, where the physical regions of chromosomes 1A–7D

can be used as queries to search. We obtained 2,847,865 non-redundant arrays of TRs comprising a total length of 431.05 Mb, which constitutes 3.06% of the total 14.06 Gb assembled wheat genome (Table 1). The non-redundant TRs were classified into three classes based on their lengths of period distance (PD) of < 20 bp, 20–60 bp, and > 60 bp. It is evident that the array number of TRs of 20–60 bp is more abundant (51.0%) than those with < 20 bp (32.2%) and > 60 bp (16.7%). The lowest proportion of chromosome sequence length contributed by TRs was for chromosome 6A at 2.67%, and highest was for chromosome 4D reaching 4.41% (Fig. 2a). The coverage of TR content varied based on the comparison of overall genomic content among those of the A, B, and D genomes. Although the overall length (in base pairs) of the D genome chromosomes was shorter than the A- and B-genome chromosomes, the percentage of TRs for the D genome was higher than those for the A- and B-genomes. The Pearson coefficient comparison showed that the chromosome length related to the TR contents had a negative correlation ($r = -0.77$, $p < 0.001$). The average TR content among the A, B and D genomes was 2.85%, 3.06% and 4.03%, respectively. The comparison among overall content indicated that the occurrence of tandem repeats in the A, B, and D genomes differed significantly. The distribution of TRs determined across the seven different linkage groups was also compared, and the results showed that a significant variation existed between groups 1 and 3 ($p = 0.0073$), groups 2 and 5 ($p = 0.049$), and groups 4 and 6 ($p = 0.0123$), respectively.

Enrichment of three classes of TRs and their surrounding genes

To provide further evidence for the copy-number enrichment of TRs, the TRs with patterns size ≥ 20 bp, copy number ≥ 50, and Percent match ≥ 80 were clustered. The top 44 TR clusters with total length over 20 kb were selected and listed in Table S1. Two large TR clusters TR-5B-1 (copy number of 4435) and TR-4A-1 (copy number of 1038) were the commonly used repetitive sequences pSc119.2 (McIntyre et al. 1990) and pAs1 (Rayburn and Gill 1986) or pTa535 (Komuro et al. 2013).

The distribution of TRs with high confidential (HC) gene annotation in the functional genomics data was analyzed (Fig. 2). A total of 82.92% predicted HC genes contained TRs among each 6 kb upstream of the transcription start site (TSS) and downstream of transcription terminal sites (TTS). The average distances of the three classes of TRs to TSS and TTS were established across the A, B, and D genomes. All three classes of TRs showed significantly

Table 1 Predicted non-redundant TRs in wheat chromosomes by TRF

Chromosome	TR array < 20 bp		TR array 20–60 bp		TR array > 60 bp		Total TR array	
	No.	Length	No.	Length	No.	Length	No.	Length
1A	37,106	1,515,154	55,880	4,793,166	17,206	9,953,781	110,192	16,262,101
1B	48,302	1,976,323	70,323	6,107,163	23,439	11,605,094	142,064	19,688,580
1D	31,600	1,279,699	57,476	5,141,621	19,711	12,063,287	108,787	18,484,607
2A	49,046	2,014,359	73,201	6,313,247	22,096	11,890,216	144,343	20,217,822
2B	55,229	2,260,823	82,530	7,206,266	27,648	12,158,481	165,407	21,625,570
2D	40,580	1,653,904	75,836	6,828,238	25,709	15,425,397	142,125	23,907,539
3A	47,315	1,929,775	71,566	6,718,087	21,695	11,414,276	140,576	20,062,138
3B	57,322	2,339,074	86,195	7,528,124	28,845	13,298,349	172,362	23,165,547
3D	37,999	1,545,878	70,639	6,327,229	24,199	14,859,668	132,837	22,732,775
4A	47,266	1,938,797	69,063	5,936,842	21,428	12,804,182	137,757	20,679,821
4B	44,716	1,831,990	65,133	5,721,644	22,915	12,057,977	132,764	19,611,611
4D	29,483	1,200,005	54,015	4,833,655	19,735	14,764,295	103,233	20,797,955
5A	44,242	1,800,239	66,924	5,961,863	20,490	11,450,735	131,656	19,212,837
5B	49,685	2,040,222	74,478	6,861,735	24,819	12,977,232	148,982	21,879,189
5D	36,328	1,471,608	63,984	6,971,876	21,985	13,804,499	122,297	22,247,983
6A	38,600	1,580,064	57,919	4,982,875	17,391	9,208,457	113,910	15,771,396
6B	50,128	2,051,456	73,612	6,385,373	24,132	11,814,778	147,872	20,251,607
6D	29,506	1,205,345	53,059	4,741,988	18,573	12,671,334	101,138	18,618,667
7A	48,608	1,996,931	74,223	6,667,751	21,699	12,351,198	144,530	21,015,880
7B	53,359	2,206,824	80,465	6,981,260	26,490	12,379,223	160,314	21,567,307
7D	41,828	1,724,518	76,936	6,863,145	25,955	14,666,230	144,719	23,253,893
Total	918,248	37,562,988	1,453,457	129,873,148	476,160	263,618,689	2,847,865	431,054,825

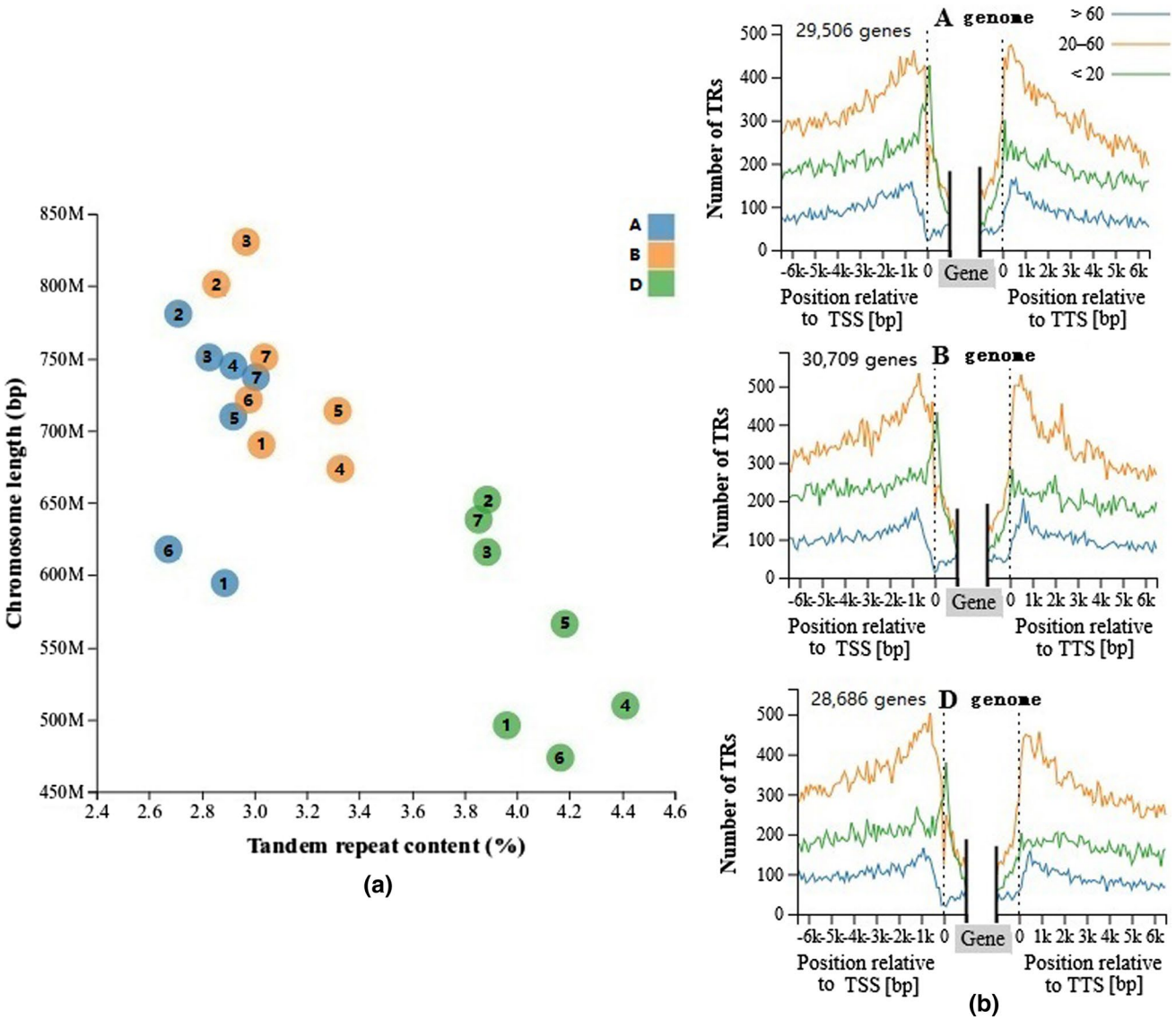

Fig. 2 Genomic distribution of TRs in assembled wheat genome. **a** TR contents of 21 chromosomes showed the D genome of small chromosome size had high TR content; **b** enrichment of TRs in the distance from transcription start sites (TSS) and transcription terminal sites (TTS) by 1 kb windows in the A, B, and D genomes. *, **, and ***refer to $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively

higher ($p < 0.03$) frequencies of occurrence in the regions approximately 1.5 kb upstream of TSS and downstream of TTS in the genome, compared to other comparable regions within a 6 kb distance. An example of 16 predicted HC genes and their surrounding TRs are listed in Table S2. An RNAseq data (NCBI SRR2306547) was used for investigating the expression of the genes; one gene was highly expressed (with RPKM value over 10) and seven genes were unexpressed.

Physical location of known TRs and FISH validation

The results of 11 oligonucleotide probes from Tang et al. (2014) as query sequences were stored in this version of B2DSC using default parameters (*pident* = 85 and *qcovhsp* = 80). The probe Oligo-pSc119.2-1 can replace the role of the original repetitive sequence pSc119.2 (McIntyre et al. 1990) in FISH analysis of wheat, rye, and wheat-rye hybrids (Fu et al. 2015). The core length of the pSc119.2 family repeats is about 120 bp. An example of genomic distribution analysis of Oligo-pSc119.2-1 (green) by B2DSC in B-genome chromosomes is shown in Fig. 3a, and the predicted copy number and the physical positions have been listed in Table S3. The physical map indicates that Oligo-pSc119.2-1 (green) has strong signals on all of the B-genome chromosomes. ND-FISH results showed that the hybridization sites of Oligo-pSc119.2-1 on chromosomes 1B–7B of Chinese Spring were mainly in telomeric

and subtelomeric regions (Fig. 3b). The results showed that the predicted physical locations on the wheat chromosomes (Table S3) match with the FISH hybridization sites (Fig. 3b). Therefore, each of the hybridization positions of Oligo-pSc119.2-1 had been assigned their physical positions on the chromosomes (Fig. S4).

Similarly, to compare the location of the wheat repetitive centromeric DNA sequences predicted by the B2DSC web server (Table S4) with the FISH results, ND-FISH analysis using probe Oligo-Quinta (Table 2) and Oligo-CCS1 (Tang et al. 2014) was conducted and the results showed clear centromeric signals on all the wheat chromosomes (Fig. S5). The hybridization signals of Oligo-Quinta were non-uniform among the centromeres in these subgenomes of CS. The strongest signals were observed on B and A genome chromosomes, while the weakest signals were commonly observed on the D genome chromosomes. The question arises as to whether a correlation between the predicted copy number and the observed strength of Oligo-Quinta hybridization signals exists. The resulting FISH hybridization signal intensities of Oligo-Quinta are clearly correlated to the predicted copy numbers ($r = 0.69$, $p = 0.005$) by the B2DSC web server (Fig. S5), confirming the different centromeric repeat distributions among the chromosomes of A, B, and D genomes.

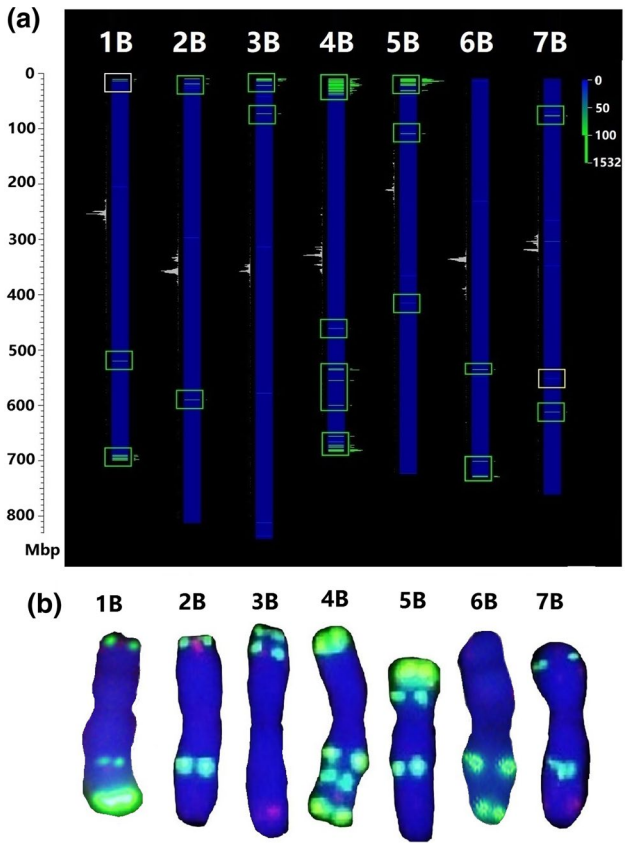

Fig. 3 Physical mapping (**a**) and ND-FISH results (**b**) of Oligo-pSc119.2-1 on B-genome chromosomes of wheat (Chinese Spring) by B2DSC, using default parameters (*pident* = 85, *qcovhsp* = 80) for the blast and filter steps. Gray bars: the distribution of Oligo-CCS1 corresponding to the positions of centromeres. Blue-to-green bars and barcodes: the number of HSPs per Mbp of Oligo-pSc119.2-1. Green box: ND-FISH hybridization sites and numbers of HSPs were well matched; yellow box: ND-FISH hybridization sites with low number of HSPs predicted

Prediction of new TR representative oligo probes

The probes pSc119.2 and pAs1 have been commonly used to identify the wheat chromosomes by FISH (Rayburn and Gill 1986; Mukai et al. 1993). However, their hybridization patterns were poorly distributed across several chromosomes. Our aim is to setup a procedure to enable chromosomal identification across numerous specific regions of wheat, and to achieve this, we need a lot more probes with clearly defined physical positions. In the present study, a total of 20 oligo probes were designed from the above-predicted TR database and the physical distributions and estimation of copy numbers of the 44 TRs clusters (Table S1) identified by TRF were obtained using the B2DSC web server (Table 2). Furthermore, they appeared as distinct and stable hybridization signals on wheat chromosomes as revealed by ND-FISH. ND-FISH results demonstrated that Oligo-3D2 had hybridization site only on the distal region of the short arm of chromosome 3D (Fig. 4a), while Oligo-1B5 produced hybridization signal in the centromeric region of 1B (Fig. 4c). Oligo-4B5 produced multiple hybridization signals on chromosomes 1B, 3B, 4B, 6B, and 7B (Fig. 4e), and the wheat chromosomes were also identifiable using Oligo-pSc119.2 and pTa535 (Fig. 4b, d, f). The hybridization patterns of 12 probes for individual wheat chromosomes and their predicted physical locations in the wheat genome, as predicted by B2DSC, are presented in Fig. 5. The ND-FISH results showed that the predicted TR-locations had distinct signals on wheat chromosomes at those expected sites. As an example, Oligo-5A1 hybridized to subtelomeric regions of chromosomes 5AS, 1BS, 5BS, 1DS, and 5DS, which are the locations of 5S rDNA regions as predicted by the web server B2DSC. The probe Oligo-1BS1 was predicted to hybridize onto 16 wheat chromosomes, and subsequently, all hybridization sites on these 16 chromosomes matched well with expected physical locations of accumulated high copy numbers. However, 12 of the 59 hybridized signals to CS chromosomes occurred despite the lack of any previous predictions. In particular, the TR arrays in pericentromeric regions of chromosome 6BL were displayed by ND-FISH, but the copy number by TR prediction was largely underestimated, which possibly implied that the sequence assembly of the related region needs to be improved.

Table 2 Sequences of Oligo probes and their chromosomal distribution by ND-FISH

Probes	Sequences	ND-FISH hybridization sites on chromosomes
Oligo-1B1	TAATAAAGTATTAGTGGATCACACGGGCTTCGCGGCACC	1B, 6B
Oligo-1B2	CGTTAAACTCGTCTCCGTAGTTGAGAGGGAGCGGCCAAAGC	1B, 5B, 1D
Oligo-1B3	CAATGGGTTCAGAAACTTGTCTGAAAAACAGCGACGAATGATG	1B, 5B, 6B, 7B
Oligo-1B5	CTTTTTTCTTGCGTGTTTTGTTTCAGGCATGATTTTGATTG	1B
Oligo-3D2	TTCTTCCTCGCGGGGGCGCGTCAGCGCACCCGCTGGGTGTAGCCCCCGAGATT	3D
Oligo-1BS1	ACAGGGTGGCATGGTGTCACGTGCTCGTCTCGGCGTCCGTCGTCGC	1A, 4A–7A, 1B–6B, 3D–7D
Oligo-1BS2	GCACCGGATCCCATCAGAACTCCGAAGTTAAGCGTGCTTGGGCGAGAGTAGTA	1A, 4A–7A, 1B, 4B–6B, 1D, 4D–7D
Oligo-1BS4	GTCCTGCTTAAGGCTGAATTTGAGGGTACTCCGGGCCTTTTGGGTCAAGCGGGGA	6A, 1B–7B, 1D, 3D, 4D
Oligo-1BS6	TGCTCTCACTCGGAGTAGTTTTTTAACTTAGGCGAGGACGGGCTCGCGGCTAAGC	1A–7A, 1B–7B, 1D, 4D–7D
Oligo-1BS7	ATAGGAGTATATATATATTTCTCAGGCTGTTTTGATGAGGTGAGGGAGGGATGTG	1B, 2B, 4B, 7B
Oligo-1BS8	ATTTTCCTAGTGTCAACATAGAACAACATGAGTGTTGTGTTAATTTTTT	1A, 4A, 1D–7D
Oligo-4B5	GCTAGCTTCAAAATGACCTATACTTAGAATCTTTTCCTCCAAAATTAG	1B, 3B, 4B, 6B, 7B
Oligo-3A1	AATAATTTTACACTAGAGTTGAACTAGCTCTATAAGCTAGTTCA	3A, 5A, 7A, 5B, 5D
Oligo-3D9	CTTAGCTGCGACTCCGTATCGCCTAAGTTAAAACAACTCCG	5A, 2B, 6B, 7B
Oligo-3D3	TAATAAGCTTAGTTAGCTCCAAAATGACATATTTTC	3B, 4B, 6B, 7B
Oligo-2D17	TTGGGTCCCGATGTGATCCGAATGTTTCGGGAACCCT	3B, 4B, 5B, 6B
Oligo-5A1	CGCATCCGACATGTTACCCCCGGCTTCATCCCTTATGCTTCCCACTCCCACGTC	5A, 1B, 5B, 1D, 5D
Oligo-1D1	CGGAGTCCGTTTTGGCTCCACAAGTAGTCAAAACGTTTGTGACGACCAGATGCT	1D, 2D, 4D–6D
Oligo-4H21	GTAAGCCTCTCTACTTTCTTGCACATGCTATGTGGGTG	1A–7A, 3B, 6B, 7B, 1D–7D
Oligo-2H1	AACTAACTGAACTAAGAGCGGCCGATGTTGGCTTGCAG	3B
Oligo-Quinta	TCAGCCTGAC CACGAAGGTC TATTCCTGCA AGCAATCGAA	1A–7A, 1B–7B, 1D–7D

Integrated physical map for TR-Oligos

After combining the ND-FISH results for the 35 reported Oligo probes (Table S3) from Tang et al. (2014, 2016, 2018) and with another 20 from Lang et al. (2018) and the present study (Table 2), we found that several probes had similar hybridization patterns. For example, probe Oligo-3A1 (Lang et al. 2018) was similar to Oligo-44 (Tang et al. 2018), and Oligo-1BS1 closely resembled Oligo-713 (Tang et al. 2018). A total of 37 non-redundant oligo probes were allocated to 206 predicted chromosome locations with an accumulated copy number over 50 per 1 Mb, which relatively closely matched the physical locations revealed by ND-FISH. An integrated Oligo-based ND-FISH map of Chinese Spring is shown in Fig. 6. The 206 hybridization sites include 59 on the A genome, 68 on the B-genome and 79 on the D-genome chromosomes. About 25 of 37 (70%) oligo probes produced single clear or strong hybridization signals on 13 chromosomes. Each chromosome appeared to have 4–15 hybridization sites; in particular, chromosome 6D had 15 hybridization sites from 10 oligos. There was a lack of hybridization site on chromosome arms 2AL and 4AS. The probes hybridized specifically to the centromeric regions of 1B, 2B, 5B, 6B, 7B, and 6D chromosomes, which can be potentially used to identify the breakpoints in translocations involving these chromosomes. The newly produced TR-Oligos as well as previously reported probes can be used as an oligo 'cocktail' to dissect specific chromosome regions effectively. Our results suggested that 21 distinct hybridization sites involving five probes had higher copy number than predicted. It might suggest that these specific segments rich in TRs remained poorly assembled and underestimated in outputs of wheat reference genome IWGSC RefSeq v1.0.

Multiple TR-Oligos for determining the breakpoints of chromosome translocations

Multiple oligo probes can be used to determine the specific locations of the breakpoints on wheat translocated chromosomes (Lang et al. 2018). Above all, the reciprocal non-Robertsonian translocation is particularly difficult to identify by molecular markers only. As an example, a wheat cultivar CM62 contains a pair of 5B–7B reciprocal translocations discovered by ND-FISH with probes Oligo-pSc119.2 (green) and Oligo-pTa535 (Fig. 7a). Oligo-1BS1 produced two distinct signals (located at 86 Mb and 151 Mb) on 5BS and one (279 Mb) on 5BL (Fig. 7b). Another probe Oligo-18 (Tang et al. 2016) had two specific hybridization sites (at positions of 379 Mb and 470 Mb) on 7BL (Fig. 7c). Both Oligo-1BS1 and

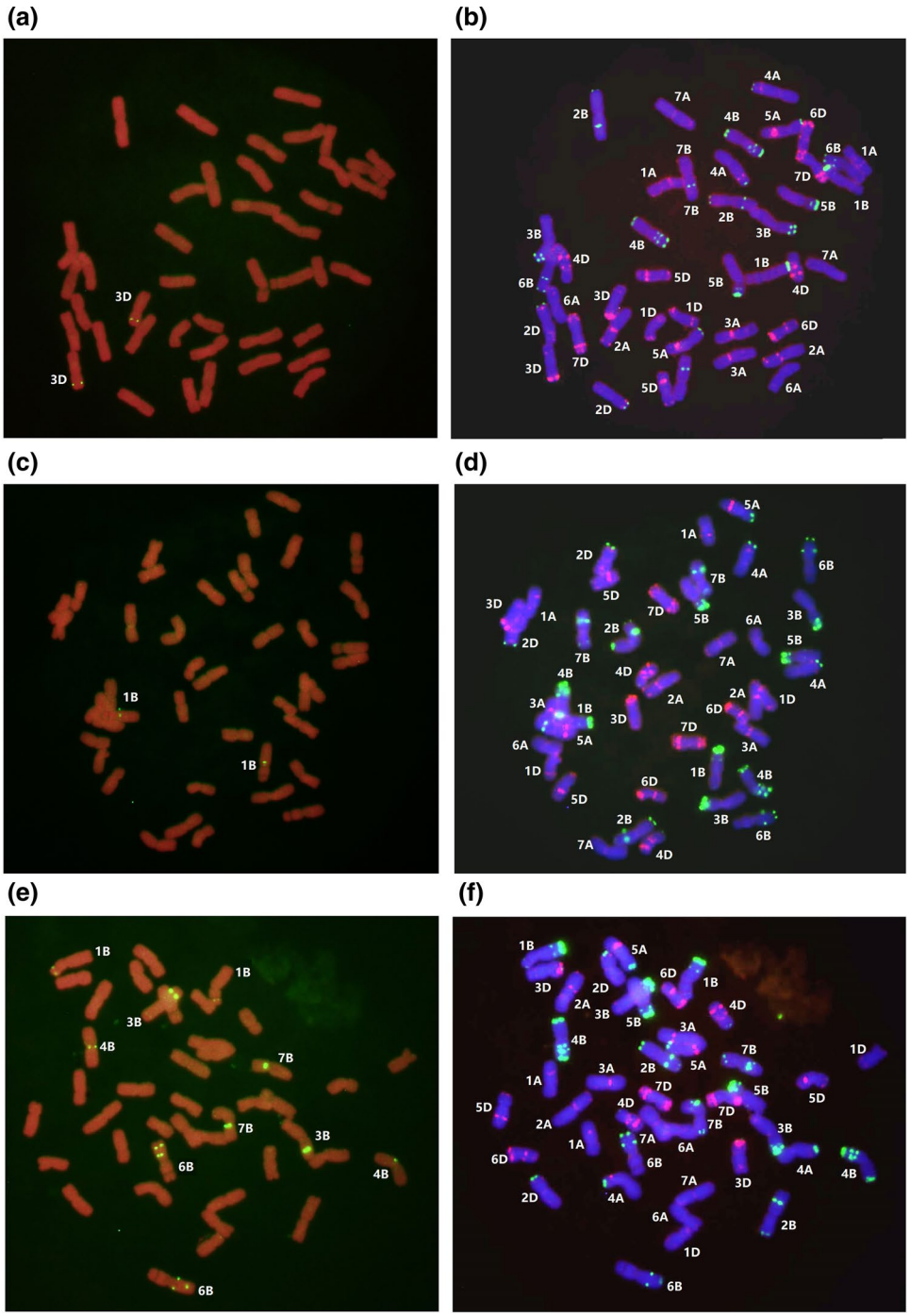

Fig. 4 Sequential ND-FISH analysis with predicted TR-oligo probes on metaphase chromosomes of Chinese Spring. The probes Oligo-3D2 (**a**), Oligo-1B5 (**c**), and Oligo-4B5 (**e**) were hybridized and their chromosomal locations determined. The sequential FISH with the Oligo-pSc119.2 (green) and Oligo-pTa535 (red) were used to identify the individual wheat chromosomes (**b**, **d**, **f**)

Oligo-18 were used to determine the positions of the breakpoints on the 5B–7B translocations. As indicated in Fig. 7d, the breakpoints are between the two hybridization sites between 99 Mb (Oligo-pSc119.2 site) and 151 Mb (Oligo-1BS1) of 5BS, while the hybridization signals of Oligo-18 pinpoint the breakpoint between the centromere (about 300 Mb) and 379 Mb on 7BL. The different physically mapped TR-Oligo probes can be used to localize the chromosome structure variation in wheat effectively.

Discussion

Cytogenetic studies of wheat have revealed that a large number of heterochromatic regions exist in wheat genomes, particularly in the B-genome (Gill et al. 1991). The molecular-based evidence, including DNA blots and FISH studies, has also revealed that TRs comprise a large proportion of the wheat genome (Mukai et al. 1993; Komuro et al. 2013). Due to the expansion of the tandem repeats in the wheat

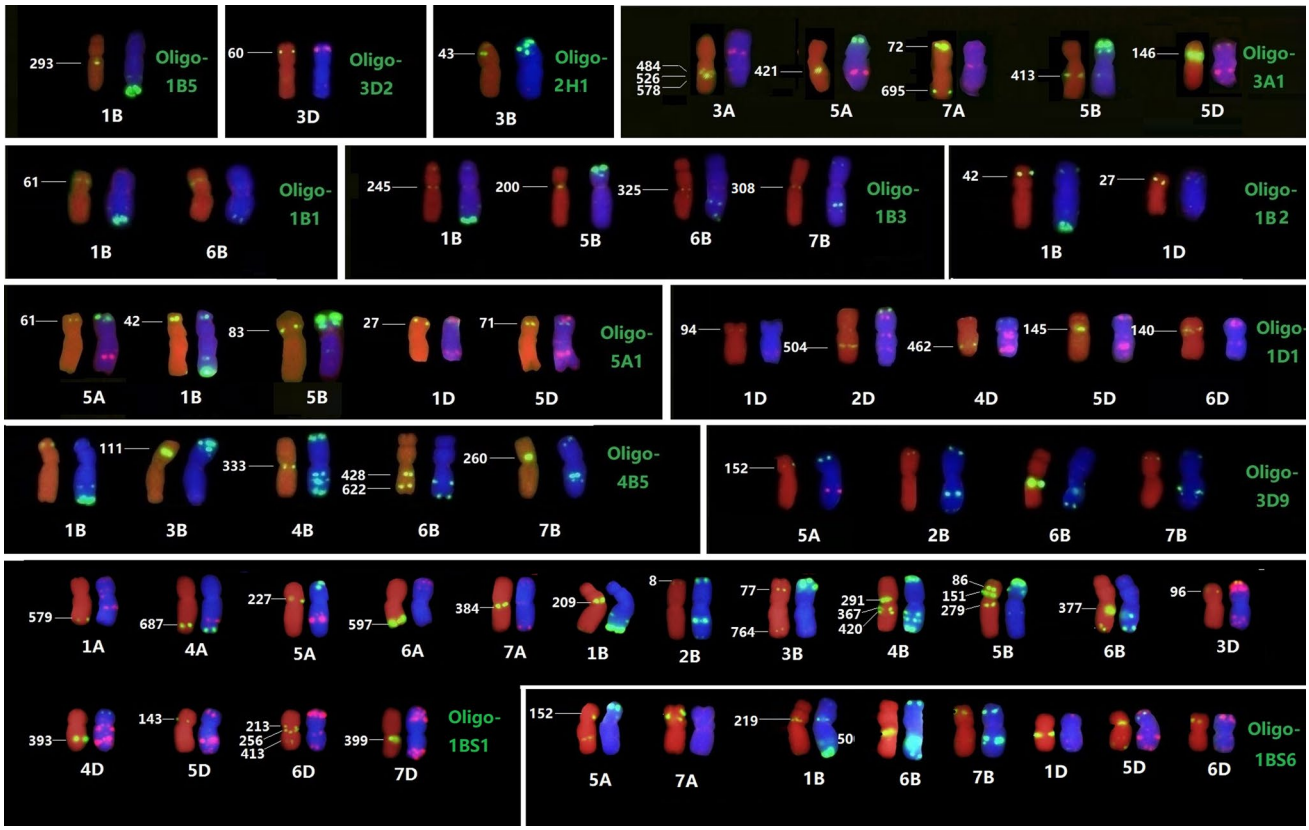

Fig. 5 Hybridization sites of 13 oligo probes on metaphase chromosomes of Chinese Spring, which produced signals (green) with reference to the patterns of Oligo-pSc119.2 (green) and Oligo-pTa535 (red) on the right. The numbers on the left side of each chromosome indicate the mid points of their positions as predicted by IWGSC v1.0 reference genome using the B2DSC web server

genome, the presence of massive blocks of heterochromatin have been found, and the size of this heterochromatin is correlated with the copy number of the tandem arrays (Vershinin et al. 1995; Evtushenko et al. 2016). However, the repeats, including the high tandem repeats, have largely been ignored. Recently, the physical maps of wheat chromosomes such as 1BS (Raats et al. 2013), 1BL (Philippe et al. 2013), 5DS (Akpinar et al. 2015), and 5BS (Salina et al. 2018) were constructed for analyzing the individual chromosome arms. The availability of a whole-genome assembly of Chinese Spring wheat provides opportunities to examine the distribution of TRs at a whole-genomic level. The present study performed a genome-wide survey of TRs using TRF analysis, and built a database of TRs arrays with repeats available at http://mcgb.uestc.edu.cn/tr, which may be useful for the further analysis of structure or possible function of TRs in the specific regions of wheat genome. The total amount of detected TRs constitutes 3–5% of the 17 Gb assembled wheat genome by TRF prediction. Arrays are not uniformly distributed across the A, B, and D genome chromosomes, and TRs showed a higher density on D than A- or B-genome chromosomes (Fig. 2). The present IWGSC reference genome v1.0 contained 3.1% of sequences with an unassigned genome, and about 1.9% of the assembled genome with undefined nucleotides. The TR contents by TRF may increase to 5–6% of the entire genome, which is in agreement with a previous report of overall TR contents in grass genomes (Bilinski et al. 2017). Additional improvements will thus be needed to generate complete assembly of the wheat genomes. The essential to achieve this will be data derived from FISH studies using different types of probes, which can effectively validate the genome assembly quality (Zhao et al. 2017; Lu et al. 2018). Therefore, the bioinformatic analysis and validation by FISH are important for improving annotation of genome projects. A genome-wide survey of TRs has revealed that, for the genome regions that are about 1.5 kb upstream of Transcription Start Sites (TSS) and downstream of Transcription Terminal Sites (TTS), these TRs were significantly accumulated (Fig. 2). The differential gene expression level of selected 16 genes surrounding TRs was demonstrated in Table S2. It is thus worthwhile to have the gene family and TRs clusters data, as well as the stress-induced transcriptomic data to reveal the possibilities of TRs involved in gene regulation in wheat.

The comprehensive combination of molecular cytogenetic and genomic approaches will facilitate the precise

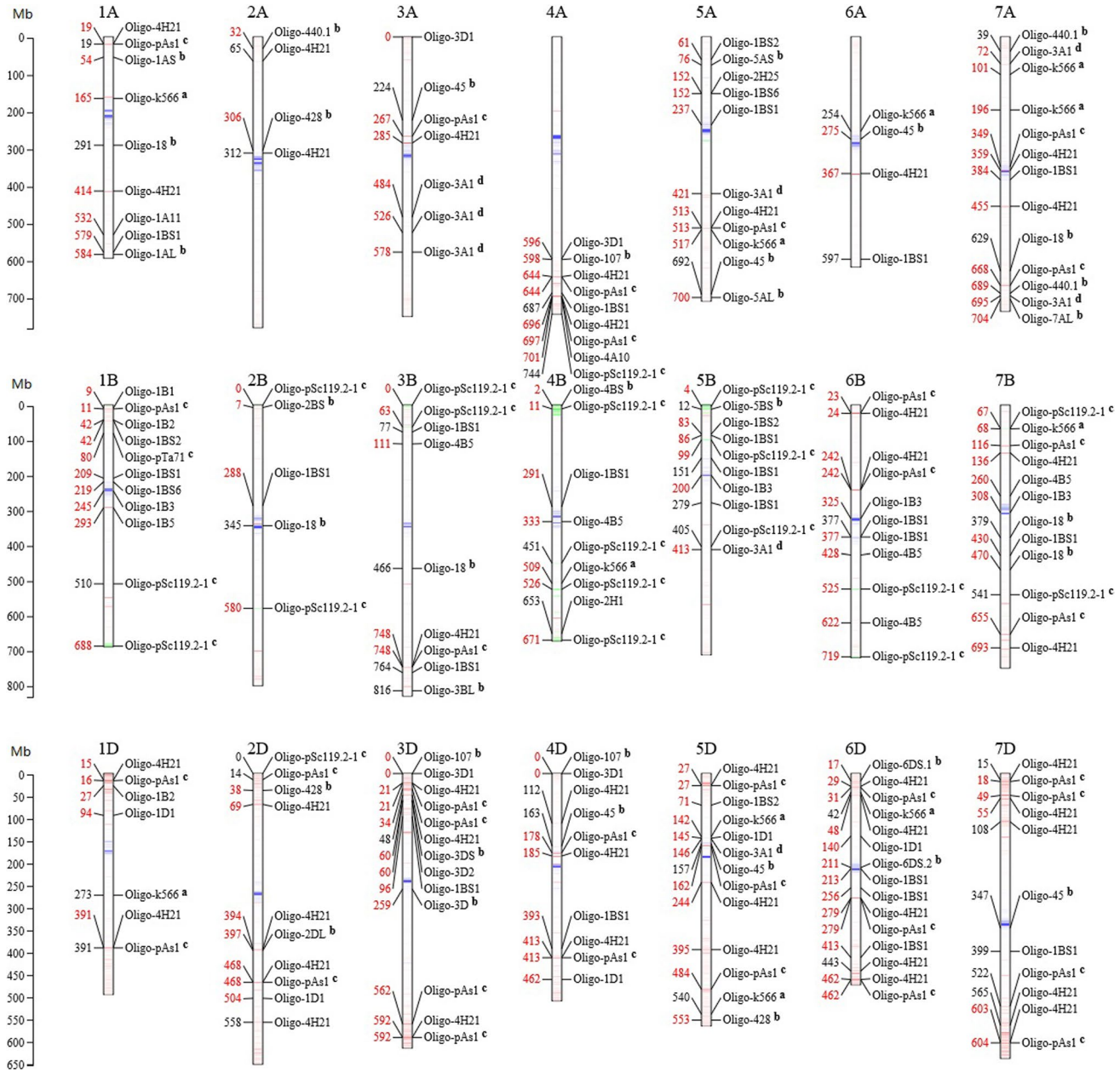

Fig. 6 Integrated physical map of TR-oligos predicted and validated by ND-FISH. Numbers on the left represent the physical position with unit of Mb. The copy number predicted over 50 is marked in red. The superscripts a, b, c, and d refer to the publications of Tang et al. (2016), (2018), (2014) and Lang et al. (2018), respectively. The probes without superscript are from the present study. The blue lines in the chromosomes were Oligo-CCS1 hybridization sites indicating the locations of centromeric regions

identification of chromosome structure variation in plants by FISH (Kato et al. 2005; Jiang and Gill 2006; Lyslak et al. 2006). ND-FISH technology, based on repetitive probes, has provided a low-cost, high-throughput technique to characterize the karyotypes of wheat (Cuadrado et al. 2009; Danilova et al. 2012; Tang et al. 2014, 2016) and its related species such as *Aegilops* (Mirzaghaderi et al. 2014), rye (Cuadrado and Schwarzacher 1998; Fu et al. 2015), *Dasypyrum* (Li et al. 2016; Xiao et al. 2017), *Thinopyrum* (Du et al. 2017; Li et al. 2017), and *Hordeum* (Cuadrado et al. 2013; Tang et al. 2016). However, the positions of the signals generated with the previously reported oligo probes on wheat chromosomes were mostly estimated by the relative fraction length on the chromosomes. The previously developed FISH probes from TRs lacked definite physical locations in the wheat genomes (Tang et al. 2018). The present study has established a user-friendly B2DSC web server which provides an easily visible and user-friendly means for viewing the general distribution

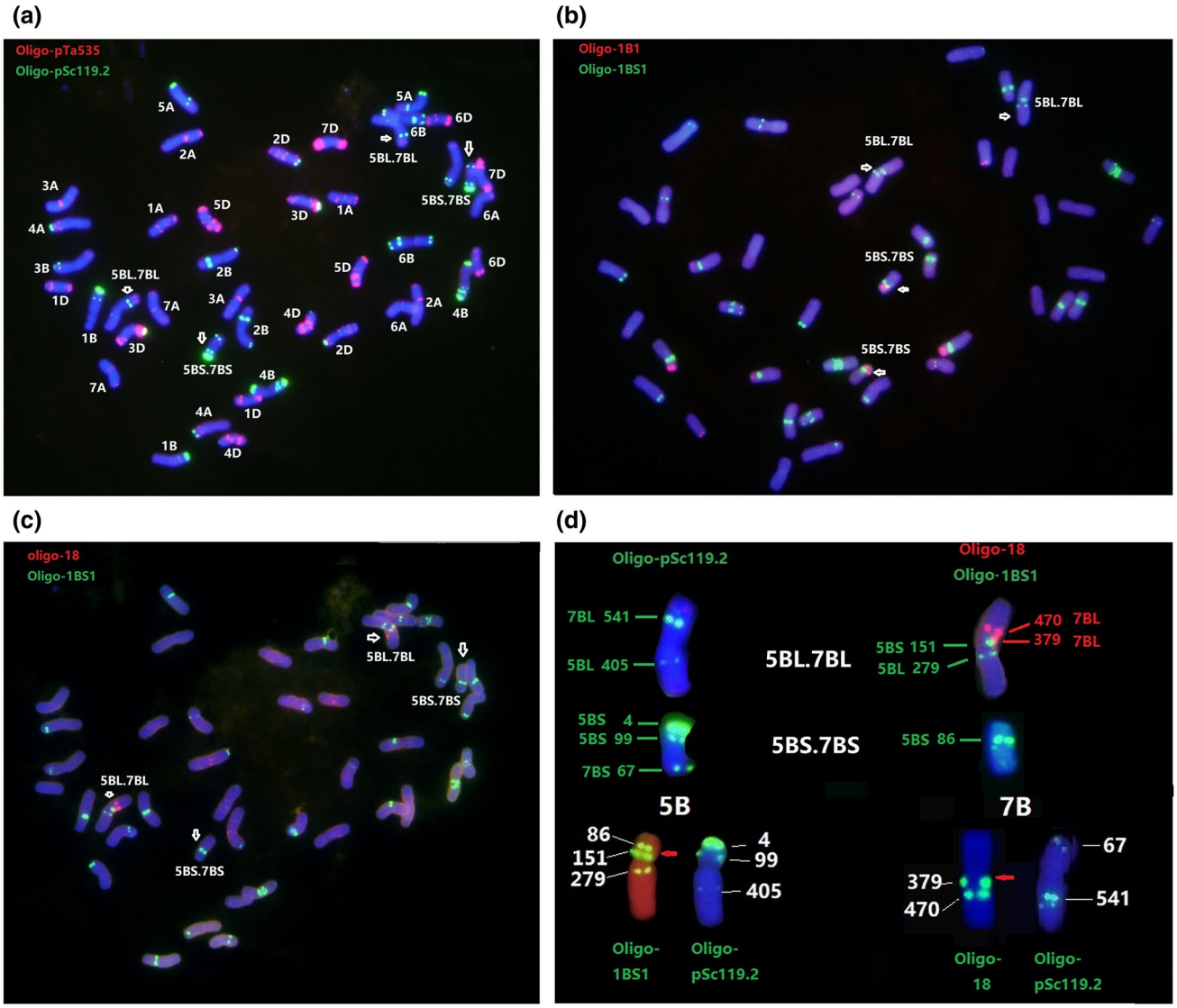

Fig. 7 Multiple ND-FISH for determining the breakpoint of the 5B–7B chromosome translocation in wheat cv. CM62. Oligo-pSc119.2 (green) and Oligo-pTa535 (red) were used to identify individual wheat chromosomes (**a**). The probes Oligo-1B1 (**b**), Oligo-18 (**c**), and Oligo-1BS1 (**b**, **c**) were used to identify the 5B–7B translocations. **d** Structures of translocation chromosomes T5BS.7BS and T5BL.7BL (upper) compared to the normal 5B and 7B chromosomes (lower), showing the physical locations of the translocation breakpoints (red arrowed) on 5B (99–151 Mb) and 7B (centromere to 379 Mb) of cv. CM62

patterns of TRs and copy-number calculations of repeated sequences in wheat chromosomes, and the output data can be viewed through drawing a chromosome plot. The present copy-number prediction of Oligo-CCS1 by B2DSC was confirmed by ND-FISH, which was useful for defining the centromere positions of wheat chromosomes (Table S3, Fig. S5). Therefore, the B2DSC web server may become a widely used and convenient resource for searching for new tandem repeats of common wheat and designing new oligo probes for ND-FISH. The present study located the 5B–7B non-Robertsonian translocation using a combination 'cocktail' of TR-oligos, namely Oligo-1BS1 and Oligo-18 (Fig. 7). In another example, probe Oligo-3A1 produced three hybridization sites on the long arm of chromosome 3A by ND-FISH, a result which was unobtainable using the other earlier-developed FISH probes. Probe Oligo-3A1 was used to precisely determine that a breakpoint on chromosome arm 3AL was between the two hybridization sites in a wheat-*Th. intermedium* translocation chromosome (Lang et al. 2018). The breakpoint was predicted by B2DSC and was also confirmed by molecular markers. A database of an integrated physical map of Oligo probes (Fig. 6) will be a useful tool for precise identification of wheat chromosome translocations in specific physical regions with largely

increased resolution using the combinations of oligo probes. We have also found that, using an integrated physical map of the new TR based Oligo probes, we can successfully identify specific chromosomal regions of related genera of wheat, such as *Secale*, *Dasypyrum,* and *Thinopyrum* (Yang et al. unpublished data). Our B2DSC web server can also be available to search for the TRs in the other grass species including maize, barley, *Aegilops tauschii*, *Brachypodium* genomes. Our studies that integrate the cytogenetic and genomic research approaches may facilitate not only gene isolation, but also chromatin structure and chromosome behavior studies of wheat and its related genomes (Abrouk et al. 2017).

Author contribution statement ZuY and GL designed the experiments. TL, GL, HW, ZhY, and QC performed the experiments, EY, ZuY, ZT, and SF analyzed the data, ZY and GL wrote the paper. All authors read and approved the manuscript.

Acknowledgements We are thankful to Dr. Ian Dundas at the University of Adelaide, Australia, for reviewing the manuscript, and the National Key Research and Development Program of China (2016YFD0102000), Applied and Basic Project (2016JY0075) from Science and Technology Department of Sichuan, China, and National Natural Science Foundation of China (No. 31171542) for their financial support.

References

Abrouk M, Balcárková B, Šimková H, Komínkova E, Martis MM, Jakobson I, Timofejeva L, Rey E, Vrána J, Kilian A, Järve K, Doležel J, Valárik M (2017) The in silico identification and characterization of a bread wheat/*Triticum militinae* introgression line. Plant Biotechnol J 15:249–256

Akpinar BA, Magni F, Yuce M, Lucas SJ, Šimková H, Šafář J, Vautrin S, Bergès H, Cattonaro F, Doležel J, Budak H (2015) The physical map of wheat chromosome 5DS revealed gene duplications and small rearrangements. BMC Genom 16:453

Anamthawat-Jonsson K, Heslop-Harrison JS (1993) Isolation and characterization of genome-specific DNA sequences in Triticeae species. Mol Gen Genet 240:151–158

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Bilinski P, Han Y, Hufford MB, Lorant A, Zhang P, Estep MC, Jiang J, Ross-Ibarra J (2017) Genomic abundance is not predictive of tandem repeat localization in grass genomes. PLoS One 12:e0177896

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM et al (2012) Analysis of the bread wheat genome using whole–genome shotgun sequencing. Nature 491:705–710

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinform 10:421

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Cuadrado Á, Jouve N (2010) Chromosomal detection of simple sequence repeats (SSRs) using nondenaturing FISH (ND–FISH). Chromosoma 119:495–503

Cuadrado A, Schwarzacher T (1998) The chromosomal organization of simple sequence repeats in wheat and rye genome. Chromosoma 107:587–594

Cuadrado Á, Golczyk H, Jouve N (2009) A novel, simple and rapid nondenaturing FISH (ND–FISH) technique for the detection of plant telomeres. Potential used and possible target structures detected. Chromosome Res 17:755–762

Cuadrado Á, Carmona A, Jouve N (2013) Chromosomal characterization of the three subgenomes in the polyploids of *Hordeum murinum* L.: new insight into the evolution of this complex. PLoS One 8:e81385

Cuadrado A, Bustos A, Jouve N (2017) On the allopolyploid origin and genome structure of the closely related species *Hordeum secalinum* and *Hordeum capense* inferred by molecular karyotyping. Ann Bot 120:245–255

Danilova TV, Friebe B, Gill BS (2012) Single–copy gene fluorescence in situ hybridization and genome analysis: Acc–2 loci mark evolutionary chromosomal rearrangements in wheat. Chromosoma 121:597–611

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Du P, Zhuang L, Wang Y, Yuan L, Wang Q, Wang D, Dawadondup Tan L, Shen J, Xu H, Zhao H, Chu C, Qi Z (2017) Development of oligonucleotides and multiplex probes for quick and accurate identification of wheat and *Thinopyrum bessarabicum* chromosomes. Genome 60:93–103

Evtushenko EV, Levitsky VG, Elisafenko EA, Gunbin KV, Belousov AI, Šafář J, Doležel J, Vershinin AV (2016) The expansion of heterochromatin blocks in rye reflects the co–amplification of tandem repeats and adjacent transposable elements. BMC Genom 17:337

Fu SL, Chen L, Wang YY, Li M, Yang ZJ, Qiu L, Yan BJ, Ren ZL, Tang ZX (2015) Oligonucleotide probes for ND–FISH analysis to identify rye and wheat chromosomes. Sci Rep 5:10552

Gemayel R, Vinces MD, Legendre M, Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu Rev Genet 44:445–477

Gemayel R, Cho J, Boeynaems S, Verstrepen KJ (2012) Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. Genes 3:461–480

Gill BS, Friebe B, Endo TR (1991) Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). Genome 34:830–839

Janicki M, Rooke R, Yang G (2011) Bioinformatics and genomic analysis of transposable elements in eukaryotic genomes. Chromosome Res 19:787–808

Jiang J, Gill BS (2006) Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome 49:1057–1068

Jurka J, Kapitonov V, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467

Kato A, Lamb JC, Birchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci USA 101:13554–13559

Kato A, Vega JM, Han FP, Lamb JC, Birchler JA (2005) Advances in plant chromosome identification and cytogenetic techniques. Curr Opin Plant Biol 8:148–154

Kishii M, Nagaki K, Tsujimoto H (2001) A tandem repetitive sequence located in the centromeric region of common wheat (*Triticum aestivum*) chromosomes. Chromosome Res 9:417–428

Komuro S, Endo R, Shikata K, Kato A (2013) Genomic and chromosomal distribution patterns of various repeated DNA sequences

in wheat (*Triticum aestivum* L.) revealed by a fluorescence in situ hybridization procedure. Genome 56:131–137

Lang T, La S, Li B, Yu Z, Chen Q, Li J, Yang E, Li G, Yang Z (2018) Precise identification of wheat—*Thinopyrum intermedium* translocation chromosomes carrying resistance to wheat stripe rust in line Z4 and its derived progenies. Genome 61:177–185

Li GR, Gao D, Zhang H, Li JB, Wang H, La S, Ma J, Yang Z (2016) Molecular cytogenetic characterization of *Dasypyrum breviaristatum* chromosomes in wheat background revealing the genomic divergence between *Dasypyrum* species. Mol Cytogenet 9:6

Li J, Lang T, Li B, Yu Z, Wang H, Li G, Yang E, Yang Z (2017) Introduction of *Thinopyrum intermedium* ssp. *trichophorum* chromosomes to wheat by trigeneric hybridization involving *Triticum, Secale* and *Thinopyrum* genera. Planta 245:1121–1135

Lu FH, McKenzie N, Kettleborough G, Heavens D, Clark MD, Bevan MW (2018) Independent assessment and improvement of wheat genome sequence assemblies using fosill jumping libraries. Gigascience 7:1–10. https://doi.org/10.1093/gigascience/giy053

Lyslak M, Fransz P, Schubert I (2006) Cytogenetic analyses of arabidopsis. Methods Mol Biol 323:173–186

McIntyre CL, Pereira S, Moran LB, Appels R (1990) New *Secale cereale* (rye) DNA derivatives for the detection of rye chromosome segments in wheat. Genome 33:635–640

Mirzaghaderi G, Houben A, Badaeva E (2014) Molecular–cytogenetic analysis of *Aegilops triuncialis* and identification of its chromosomes in the background of wheat. Mol Cytogenet 7:91

Mukai Y, Nakahara Y, Yamamoto M (1993) Simultaneous discrimination of the three genomes in hexaploid wheat by multicolor fluorescence in situ hybridization using total genomic and highly repeated DNA probes. Genome 36:489–494

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607

Pavlek M, Gelfand Y, Plohl M, Meštrović N (2015) Genome–wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. DNA Res 22:387–401

Philippe R, Paux E, Bertin I, Sourdille P, Choulet F, Laugier C et al (2013) A high density physical map of chromosome 1BL supports evolutionary studies, map–based cloning and sequencing in wheat. Genome Biol 14:R64

Raats D, Frenkel Z, Krugman T, Dodek I, Sela H, Simková H et al (2013) The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. Genome Biol 14:R138

Rayburn AL, Gill BS (1986) Isolation of a D-genome specific repeated DNA sequence from *Aegilops squarrosa*. Plant Mol Biol Rep 4:102–109

Salina EA, Nesterov MA, Frenkel Z, Kiseleva AA, Timonova EM, Magni F, Vrána J, Šafář J, Šimková H, Doležel J, Korol A, Sergeeva EM (2018) Features of the organization of bread wheat chromosome 5BS based on physical mapping. BMC Genom 19(Suppl 3):80

Tang ZX, Yang ZJ, Fu SL (2014) Oligonucleotides replacing the roles of repetitive sequences pAs1, pSc119.2, pTa–535, pTa71, CCS1, and pAWRC.1 for FISH analysis. J Appl Genet 55:313–318

Tang SY, Qiu L, Xiao ZQ, Fu SL, Tang ZX (2016) New oligonucleotide probes for ND–FISH analysis to identify barley chromosomes and to investigate polymorphisms of wheat chromosomes. Genes 7:118

Tang S, Tang Z, Qiu L, Yang Z, Li G, Lang T, Zhu W, Zhang J, Fu S (2018) Developing new Oligo probes to distinguish specific chromosomal segments and the A, B, D genomes of wheat (*Triticum aestivum* L.) using ND-FISH. Front Plant Sci 9:1104

Vershinin AV, Svitashev S, Gummesson PO, Salomon B, Bothmer R, Bryngelsson T (1994) Characterization of a family of tandemly repeated DNA sequences in Triticeae. Theor Appl Genet 89:217–225

Vershinin AV, Schwarzacher T, Heslop-Harrison JS (1995) The large–scale organization of repetitive DNA families at the telomeres of rye chromosomes. Plant Cell 7:1823–1833

Xiao Z, Tang S, Qiu L, Tang Z, Fu S (2017) Oligonucleotides and ND–FISH displaying different arrangements of tandem repeats and identification of *Dasypyrum villosum* chromosomes in wheat backgrounds. Molecules 22:973

Zhao G, Zou C, Li K, Wang K, Li T, Gao L, Zhang X, Wang H, Yang Z, Liu X, Jiang W, Mao L, Kong X, Jiao Y, Jia J (2017) The *Aegilops tauschii* genome reveals multiple impacts of transposons. Nat Plants 3:946–955

Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL (2017) The first near–complete assembly of the hexaploid bread wheat genome. Gigascience 6:1–7