the plant journal

SEB Society for Experimental Biology

The Plant Journal (2015) **84,** 733–746 doi: 10.1111/tpj.13036

Exploring the tertiary gene pool of bread wheat: sequence assembly and analysis of chromosome 5M^g of *Aegilops geniculata*

Vijay K. Tiwari[1], Shichen Wang[2], Tatiana Danilova[1], Dal Hoe Koo[1], Jan Vrána[3], Marie Kubaláková[3], Eva Hribova[3], Nidhi Rawat[1], Bhanu Kalia[1], Narinder Singh[1], Bernd Friebe[1], Jaroslav Doležel[3], Eduard Akhunov[2], Jesse Poland[1], Jamal S. M. Sabir[4] and Bikram S. Gill[1,*]

[1]*Department of Plant Pathology, Wheat Genetics Resource Center, Kansas State University Manhattan, Manhattan, KS 66506, USA,*

[2]*Department of Plant Pathology, Kansas State University, Manhattan, KS 66502, USA,*

[3]*Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, CZ 78371, Olomouc, Czech Republic, and*

[4]*Biotechnology Research Group, Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia*

Received 24 July 2015; revised 3 September 2015; accepted 14 September 2015; published online 26 September 2015.
*For correspondence (e-mail bsgill@ksu.edu).

SUMMARY

Next-generation sequencing (NGS) provides a powerful tool for the discovery of important genes and alleles in crop plants and their wild relatives. Despite great advances in NGS technologies, whole-genome shotgun sequencing is cost-prohibitive for species with complex genomes. An attractive option is to reduce genome complexity to a single chromosome prior to sequencing. This work describes a strategy for studying the genomes of distant wild relatives of wheat by isolating single chromosomes from addition or substitution lines, followed by chromosome sorting using flow cytometry and sequencing of chromosomal DNA by NGS technology. We flow-sorted chromosome 5M^g from a wheat/*Aegilops geniculata* disomic substitution line [DS5M^g (5D)] and sequenced it using an Illumina HiSeq 2000 system at approximately 50 × coverage. Paired-end sequences were assembled and used for structural and functional annotation. A total of 4236 genes were annotated on 5M^g, in close agreement with the predicted number of genes on wheat chromosome 5D (4286). Single-gene FISH indicated no major chromosomal rearrangements between chromosomes 5M^g and 5D. Comparing chromosome 5M^g with model grass genomes identified synteny blocks in *Brachypodium distachyon*, rice (*Oryza sativa*), sorghum (*Sorghum bicolor*) and barley (*Hordeum vulgare*). Chromosome 5M^g-specific SNPs and cytogenetic probe-based resources were developed and validated. Deletion bin-mapped and ordered 5M^g SNP markers will be useful to track 5M-specific introgressions and translocations. This study provides a detailed sequence-based analysis of the composition of a chromosome from a distant wild relative of bread wheat, and opens up opportunities to develop genomic resources for wild germplasm to facilitate crop improvement.

Keywords: flow sorting, SNPs, next generation sequencing, single gene FISH, GenomeZipper, synteny.

INTRODUCTION

Next-generation sequencing (NGS) technologies are providing the ability to assess the genomic composition of crop plants and contributing to a massive introduction of molecular and genomics approaches into crop improvement programs (Henry, 2012; Varshney *et al.*, 2014). NGS is suitable for a number of purposes, including marker discovery, genome-wide selection studies and targeted re-sequencing to identify important alleles and genes by comparing the genome of crop species and their wild relatives (Egan *et al.*, 2012; Henry, 2012; Ray and Satya, 2014; Varshney *et al.*, 2014). NGS technologies may also be used to assess gene structure and expression and identify heritable genomic variation related to important agronomic traits (Berkman *et al.*, 2012). Application of NGS is straightforward in species with small genomes and in crops for which reference genome sequences are available (Abe *et al.*, 2012; Liu

© 2015 The Authors
The Plant Journal © 2015 John Wiley & Sons Ltd

et al., 2012; Varshney *et al.*, 2013). However, whole-genome NGS data are more difficult to handle and use in species with large and complex genomes for which reference sequence assemblies are not available. Important species with large, complex genomes are found in the tribe Triticeae, with high repeat content (approximately 85%) and large genome size of 4–5 GB per haploid genome. Many of these species, including bread wheat (*Triticum aestivum*), are polyploid, making the genome even larger (approximately 17 GB) (Choulet *et al.*, 2014; International Wheat Genome Sequencing Consortium, 2014). Exome capture and genotyping-by-sequencing approaches have been used successfully to reduce complexity, but they have their own limitations at the whole-genome level and ignore some potentially important parts of the genome (Winfield *et al.*, 2012; Jordan *et al.*, 2015).

Flow cytometric sorting of individual chromosome(s) and chromosome arm(s) provides an attractive approach to reduce genome complexity. Through sorting, individual chromosomes or chromosome arms are typically isolated at approximately 90–95% purity (Dolezel *et al.*, 2014). High-molecular-weight DNA prepared from flow-sorted chromosomes was used to construct chromosome-specific BAC libraries (Safar *et al.*, 2004, 2010) for preparation of sequence-ready physical maps (Paux *et al.*, 2008; Philippe *et al.*, 2013; Poursarebani *et al.*, 2014) and chromosome reference sequences (Choulet *et al.*, 2014). DNA amplified from a smaller number of these large chromosomes is suitable for NGS technology (Mayer *et al.*, 2011). Although individual chromosomes still represent a considerable target of 500–1000 Mb, sequencing individual chromosomes from flow-sorted DNA enables more robust assembly and permits identification of the majority of genes on a chromosome. Virtual gene order maps may be developed using GenomeZipper (http://pgsb.helmholtz-muenchen.de/plant/barley/gz/index.jsp), which relies on synteny with sequenced model genomes including rice, sorghum and *Brachypodium distachyon* L. (Mayer *et al.*, 2011, Martis *et al.*, 2013; International Wheat Genome Sequencing Consortium, 2014). The gene order maps facilitate structural and comparative genomic studies in cereals and have provided insight into evolutionary aspects of wheat chromosomes (Ma *et al.*, 2010; Hernandez *et al.*, 2012; Jordan *et al.*, 2015). More importantly, DNA amplified from flow-sorted chromosomes has been used for marker discovery (International Wheat Genome Sequencing Consortium, 2014, Tiwari *et al.*, 2014).

The under-utilized tertiary gene pool of wheat (Qi *et al.*, 2007) is a potential gold mine for genetic variability. However, the lack of molecular markers and genomic resources, coupled with hybridization barriers, has led to limited success in utilizing this attractive resource for wheat improvement. Cytogeneticists have developed hundreds of useful stocks whereby whole chromosomes and arms from tertiary gene pool species have been added to the wheat genome (Friebe *et al.*, 1996a,b; Gill *et al.*, 2006; Schneider *et al.*, 2008), but the lack of molecular resources and tools impede their use. We hope that developing genomic resources for these species will result in better understanding of their genic and non-genic constitution, and facilitate precise gene transfer to wheat.

One of the tertiary gene pool species, *Aegilops geniculata* Roth, is an important source of genes for crop improvement (Zaharieva *et al.*, 2001a,b; Dhaliwal *et al.*, 2002; Kuraparthy *et al.*, 2007; Liu *et al.*, 2012). Important germplasm resources have been developed to utilize the important variability from *Ae. geniculata* in wheat breeding (Friebe *et al.*, 2000; Dhaliwal *et al.*, 2002; Kuraparthy *et al.*, 2007; Liu *et al.*, 2012). A number of agronomically important genes have been mapped on *Ae. geniculata* chromosome $5M^g$ from accession TA10437, including *Lr57*, *Yr40* and *Sr53* (Dhaliwal *et al.*, 2002; Kuraparthy *et al.*, 2007; Liu *et al.*, 2012). In addition, chromosome $5M^g$ from accession TA10437 escapes the diploid pairing control of wheat and freely recombines with chromosome 5D of wheat (Liu *et al.*, 2012). Our specific interest in these genes and genetic factors led us to develop alien chromosome-specific SNP markers after Illumina sequencing of the flow-sorted short arm of chromosome $5M^g$ of *Ae. geniculata* (Tiwari *et al.*, 2014). However, due to short sequence reads and low coverage, the sequence data were not suitable for structural and comparative analyses.

To develop genomic resources for targeted genes on $5M^g$, we coupled the powerful NGS approach with flow sorting of a complete *Aegilops* chromosome from its substitution line, and reduced the complexity to an individual chromosome. This study provides useful insight into the functional and structural analysis of an *Aegilops* chromosome from a tertiary wheat gene pool species. We demonstrated the utility of alien chromosome-specific shotgun sequences to study chromosomal rearrangements and its syntenic relationship with wheat and wheat progenitors and model grass genomes. We developed SNP markers and ordered them in homoeologous wheat deletion bins for downstream applications in tracking and mapping introgressions of chromosome 5M across the M-genome species such as *Ae. comosa* Sm. in Sibth. & Sm. (MM), *Ae. crassa* Boiss. ($U^cU^cM^cM^c$), *Ae. biuncialis* Vis. ($U^bU^bM^bM^b$), *Ae. neglecta* Req. ex Bertol. ($U^nU^nM^nM^n$) and *Ae. columnaris* Zhuk ($U^cU^cM^cM^c$). Additionally, we developed repetitive DNA-based cytogenetic markers and database resources for use in molecular cytogenetic analyses.

© 2015 The Authors

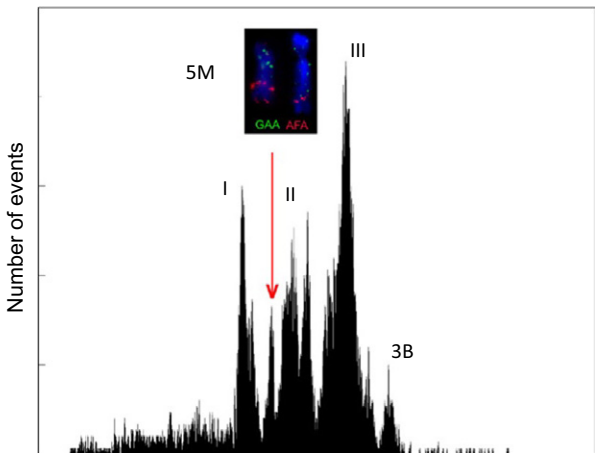


Figure 1. Flow-cytometric analysis of mitotic metaphase chromosomes isolated from a wheat/ *Ae. geniculata* disomic substitution line. Histogram of DAPI fluorescence intensity (flow karyotype) was obtained after analysis of DAPI-stained chromosome suspensions, and comprises three composite peaks representing groups of wheat chromosomes, and a peak representing wheat chromosome 3B. The peak representing chromosome 5M^g is well resolved, and the chromosome was sorted with approximately 95% purity. Inset: Examples of flow-sorted chromosomes 5M^g after FISH using probes for the Afa family (red) and $(GAA)_n$ repeats (yellow/green). The chromosomes were counterstained using DAPI (blue).

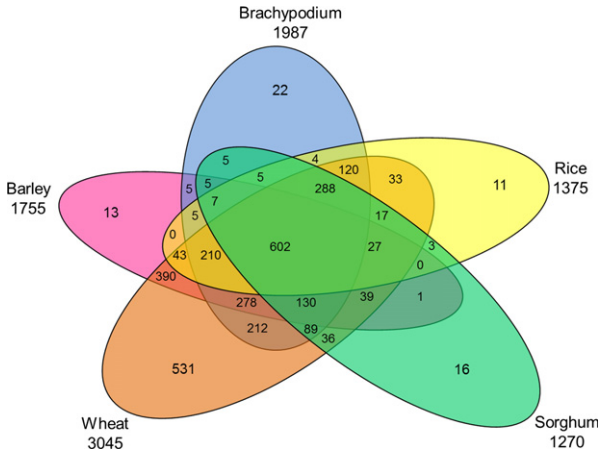


Figure 2. General overview of annotated 5M^g transcripts against five cereal genomes. Numbers represent the total number of annotated genes in the 5M^g assembly compared with various grass genomes, including wheat (homoeologous group 5 chromosomes), barley, *Brachypodium*, rice and sorghum. Numbers given below each species name represent the total number of annotated transcripts in that species.

RESULTS

Flow sorting and sequencing of 5M^g of *Ae. geniculata*

Flow cytometric analysis of DAPI-stained chromosome suspensions prepared from a wheat/ *Ae. geniculata* 5M^g disomic substitution line resulted in a histogram of DAPI fluorescence intensity (flow karyotype) with three composite peaks (I–III) representing groups of wheat chromosomes and a peak for wheat chromosome 3B. Moreover, the peak representing *Ae. geniculata* chromosome 5M^g was well resolved between composite peaks I and II (Figure 1), permitting sorting of chromosome 5M^g with approximately 95% purity as determined by FISH. Sequencing of DNA amplified from flow-sorted chromosome 5M^g in one lane of an Illumina HiSeq 2000 system generated more than 470 154 682 raw reads with a mean read length of 101 bp. The short-paired end reads were filtered and trimmed for quality control. The resulting 369 434 106 reads (79% of the total raw reads) provided approximately approximately 50 × coverage of chromosome 5M^g (estimated size 750 Mb), and were used for further analysis. Assembly of the Illumina paired-end reads was performed using SOAPdenovo (Luo *et al.* 2012) with various *k*-mer sizes (31–79 bp). The assembly with the best N50 (a weighted median statistic such that 50% of the entire assembly is contained in contigs equal to or larger than this value; 79–mer, 400 bp) and the highest number of larger contigs (959 contigs with scaffolds >10 kb) was chosen for further analyses. *De novo* assembly of the 5M^g reads resulted in approximately 246 000 contigs with a length ≥500 bp and a mean coverage depth of 13 ×. Assembled repeat-free sequences were then used for chromosome structure analysis, gene detection and construction of a virtual gene order map.

Estimation of the gene content of chromosome 5M^g, and GenomeZipper analysis

By mapping wheat EST sequences to the repeat-free 5M^g contigs, we predicted the structure of 4236 transcripts (Figure 2). Sequence-based comparisons against the genomes of barley, *Brachypodium*, rice, sorghum and wheat chromosomes were performed. The total numbers of orthologous transcripts found in barley, *Brachypodium*, rice, sorghum and wheat chromosomes were 1755, 1987, 1375, 1270 and 3045, respectively. Among the 4236 predicted non-redundant transcripts, 3147 were found to have orthologs in at least one of the grass genomes (Figure 2). The number of 5M^g genes shared among the species is detailed in Figure 2. In total, approximately 72% (3045) of the annotated genes on 5M^g showed sequence identity (mean 97%) to bread wheat, with high-confidence genes on homoeologous group 5 chromosomes (5A, 5B and 5D). A total of 1882 genes were shared by 5M^g and all group 5 wheat chromosomes (Figure 3a). When genes on chromosome 5M^g were compared with genes/gene fragments in the whole-genome draft assemblies of A and D genome donors of hexaploid wheat, *Triticum urartu* Tumanian ex. Gandilyan and *Ae. tauschii* Coss., we found that 2795 genes/gene fragments were shared by *T. urartu* and *Ae. tauschii* and chromosome 5M^g of *Ae. geniculata* (Figure 3b).

© 2015 The Authors

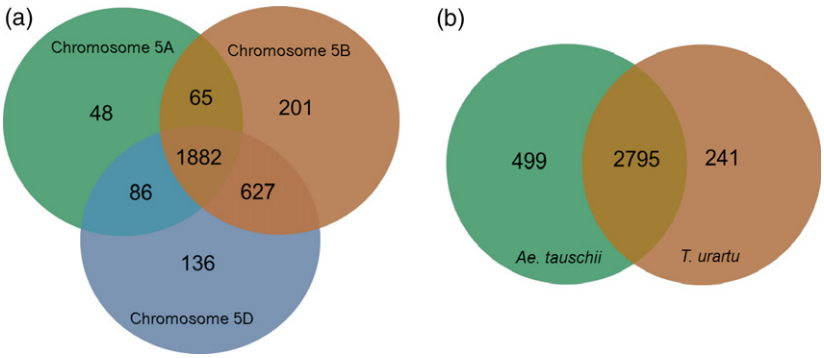


Figure 3. Number of genes annotated on *Ae. geniculata* chromosome 5M^g compared with those on homoeologous groups of wheat, *Ae. tauschii* and *T. urartu*. The Venn diagrams show (a) the number of wheat homoeologous group 5 genes (5A, 5B and 5D) and (b) *Ae. tauschii* (DD) and *T. urartu* (AA) genes shared with chromosome 5M^g of *Ae. geniculata*. Overlaps represent the number of 5M^g genes that have homoeologous genes in both or all three genomes.

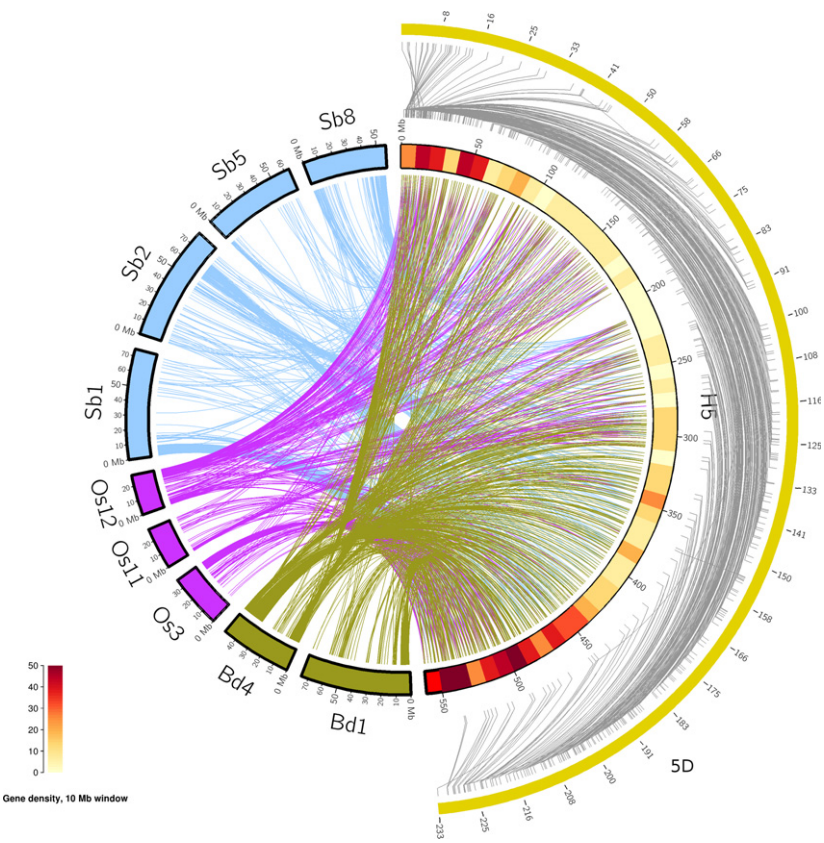


Figure 4. Collinearity of *Ae. geniculata* chromosome 5M^g with sequenced grass genomes. Linear gene order of *Ae. geniculata* GenomeZipper against the sequenced genomes of *Brachypodium*, rice, sorghum and barley. Barley chromosome 5H-specific information was used to virtually anchor and order 5M^g genes. The outermost semi-circle represents the corresponding position of 5M^g genes on a high-density SNP consensus map of wheat chromosome 5D based on chromosome 5H of barley. The scale bar with various colors shows the gene density in a 10 Mb window.

GenomeZipper-based alignment across the sequenced cereal genomes based on syntenic regions generated a virtual gene map of chromosome 5M^g (Figure 4). The 5M^g sequences were compared by stringent BLASTX analysis (Kopecky *et al.* 2013) against genomic sequences of *Brachypodium*, rice, sorghum and barley (Mayer *et al.*, 2011) to identify syntenic blocks or regions (Figure 4). Comparing the 5M^g contigs against sequenced grass genomes identified syntenic regions on *Brachypodium* chromosomes 1 and 4, rice chromosomes 3, 11 and 12, sorghum chromosomes 1, 2, 5 and 10, and barley chromosome 5H (Figure 4). Chromosome-specific information for barley chromosome 5H was then used to anchor and order 5M^g genes and contigs.

Repeat landscape of chromosome 5M^g

To identify the main repetitive DNA elements of chromosome 5M^g, *de novo* repeat reconstruction was performed on Illumina sequence reads using the RepeatExplorer pipeline (Novak *et al.*, 2013). A variety of repetitive DNA elements were identified based on all-to-all sequence read comparisons using rigorous threshold similarities (90% identity over 55% of the sequence length). Graph-based clustering resulted in 128 093 clusters containing 2 755 143 sequence reads. Two hundred clusters, corresponding to 0.01% of the Illumina sequence reads, were manually annotated using a combination of approaches based on similarity searches against sequence databases (TREP,

© 2015 The Authors

http://wheat.pw.usda.gov/ITMI/Repeats/; GIRI, http://www.girinst.org/) and graph layout analysis. Annotation of the largest clusters led to classification of 72.50% of chromosome 5M^g sequences, and provided information about the content of various types of repetitive elements (Figure 5). The largest clusters contained mainly *Ty3*/*gypsy*-like elements (Figure 5), which were found to be the most abundant (33.46% of the total repeat content). *Ty1*/*copia*-like retroelements represented 12.82% of the repeats, whereas 10.49% of the repeats represented putative non-annotated LTR retroelements, which did not contain any coding domain that allowed their classification (Figure 5 and Table S1). On the other hand, this group of putative LTR elements shared high similarity with previously annotated LTR sequences and/or contained putative LTR sequences (Table S1). Sequences similar to DNA transposons represented 7.88% of the repeats; only 0.37% of the repeats were found to be homologous to LINE elements. We were not able to classify sequences within 51 clusters that accounted for 6.45% of the analyzed repeats (Figure 5). These sequences may represent unknown DNA repeats or artifacts originating during the whole-genome amplification procedure.

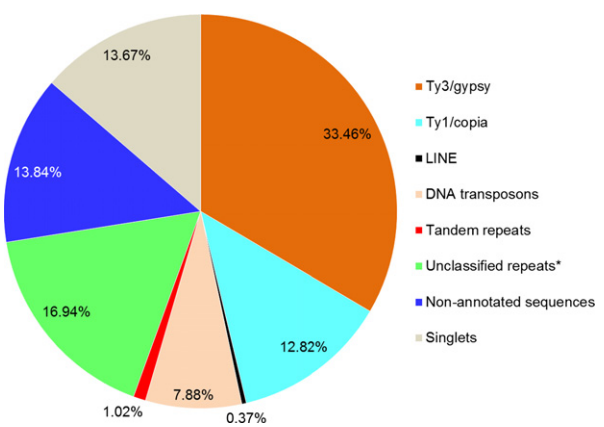
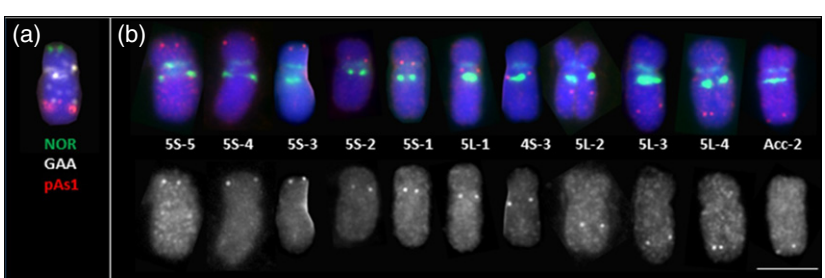


Figure 5. Repetitive landscape of chromosome 5M^g of *Ae. geniculata*. Composition of various repetitive elements in sequenced reads of chromosome 5M^g of *Ae. geniculata*.

Based on the graphic layouts used by SeqGrapheR (https://cran.r-project.org/web/packages/SeqGrapheR/index.html), contigs derived from clusters containing putative tandem repeats were identified using Dotter analysis (Sonnhammer and Durbin, 1995) (Figure S1). Dotter analysis revealed the presence of tandem organized repeats in 12 clusters, which together represented 1.02% of the sequence reads. Within these clusters, known satellite/tandem sequences, as well as newly identified tandem repeats, were found.

To develop a useful, repetitive DNA-based resource for chromosome 5M^g, a sequence database (http://olomouc.ueb.cas.cz/projects/Aegilops_geniculata_5M) for various types or families of DNA repeats was prepared by merging reads from clusters with the same annotation. This database is useful for fast annotation of previously assembled genomic sequences by mapping the reads using various alignment programs. Novak *et al.* (2013) showed that collections of reads gathered from clustering analysis provide better sensitivity for detecting divergent repetitive elements because they better represent their sequence variability. Databases of Illumina reads and assembled contigs specific to the various types of repetitive elements identified in the data for chromosome 5M^g were established and are publicly available at http://olomouc.ueb.cas.cz/projects/Aegilops_geniculata_5M.



Chromosome arm	FISH site name	FLcDNA KOMUGI database	Average distance from the centromere		
			5M^g-TA6675	5M^g-TA10437	5D
5S	5S-5	tplb0027f03	0.69±0.05	0.73±0.05	0.70±0.04
	5S-4	tplb0016e11	0.71±0.07	0.70±0.04	0.69±0.04
	5S-3	tplb0006h03	0.71±0.07	0.67±0.07	0.63±0.03
	5S-2	tplb0002p18	0.22±0.01	0.19±0.02	0.10±0.02
	5S-1	tplb0016k09	0.12±0.01	0.08±0.03	0.18±0.03
5L	5L-1	tplb0014l23	0.09±0.03	0.10±0.02	0.07±0.02
	4S-3	tplb0013i03	0.17±0.03	0.15±0.02	0.16±0.03
	GA		0.22±0.02	0.22±0.01	
	5L-2	AK331808	0.50±0.03	0.48±0.02	0.45±0.03
	5L-3	AK334748	0.69±0.03	0.70±0.04	0.61±0.03
	5L-4	tplb0043p15	0.72±0.02	0.74±0.03	0.70±0.02
	Acc-2		0.86±0.02	0.82±0.02	0.81±0.03

Figure 6. Application of wheat tandem repeats and full-length cDNA FISH probes to chromosome 5M^g of *Ae. geniculata*.
(a) FISH pattern of tandem repeats $(GAA)_n$, pAs1 and NOR on chromosome 5M^g. Repeat $(GAA)_n$ is shown in white, pAs1 is shown in red and NOR is shown in green; chromosomes counterstained with DAPI are blue.
(b) Physical mapping of group 5 wheat FLcDNA FISH probes on chromosome 5M^g from substitution line TA6675. The wheat FLcDNA signals are red, $(GA)_n$ and centromeric repeats are green; chromosomes counterstained with DAPI are blue. Each chromosome is presented as a merged image and as a separate red-channel image showing the FLcDNA sites. Scale bar = 5 μm. The positions of cDNA FISH probes and blocks of tandem repeats on chromosome 5M^g from the substitution line TA6675, *Ae. geniculata* accession TA10437 and wheat chromosome 5D are shown in the table. Chromosome measurement data for chromosome 5D were taken from Danilova *et al.* (2014).

© 2015 The Authors

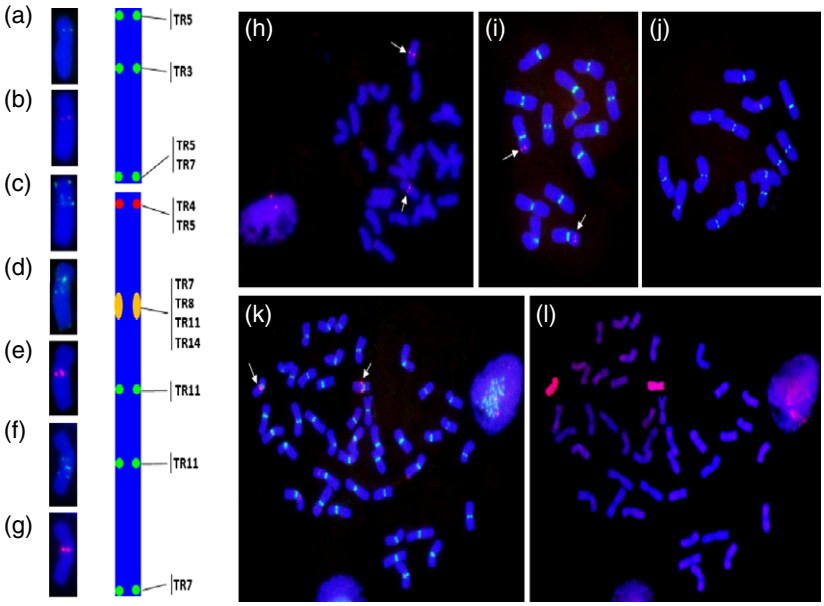


Figure 7. Chromosome-specific localization of simple sequence repeats (SSRs).
(a–g) FISH patterns of individual SSRs on chromosome 5M^g and a corresponding ideogram of localization patterns.
(h,i) The FISH pattern of TR14 was specific to the proximal region of the long arm of chromosome 5M^g (h), and was also specific to short-arm subterminal region of *Ae. comosa* subsp. *subventricosa*, which is the M-genome donor species of *Ae. geniculata* (i).
(j) However, no hybridization signal was detected in another sub-species of *Ae. comosa* subsp. *comosa*.
(k) This probe was also used for detection of chromosome 5M^g in a wheat/ *Ae. geniculata* addition.
(l) The presence of chromosome 5M was further confirmed in the disomic addition line using genomic in situ hybridization (GISH). The FISH signals are indicated by arrows, and the transposable element Quinta, a centromere-specific probe, was used for detection of centromere position. A description of the probes is given in Table S1.

Single-gene FISH

Tandem repeats GAA, GA and pAs1, the LTR of retroelement Quinta, and minor NOR sites were detected on chromosome 5M^g (Figure 6). Wheat chromosomes 3B, 4B, 5B and 6B have large blocks of GA– repeats (Cuadrado *et al.*, 2008a), but their pattern differs from that of 5M^g. Therefore, in the following experiments, the GA– repeat was used to identify chromosome 5M^g. FISH using all 11 wheat Full Length cDNA probes produced distinct signals on chromosome 5M^g of *Ae. geniculata* (Figure 6). The order and relative positions of the probes on chromosome 5M^g were similar in substitution line TA6675, *Ae. geniculata* TA10437 (Figure 6) and chromosome 5D of *Triticum aestivum* cv. Chinese Spring, indicating no major rearrangements except for the positions of the two short-arm probes 5S-1 and 5S-2. These probes are located very close to each other in the pericentromeric region. Their order is opposite on chromosome arm 5DS and the short arms of chromosomes 5A and 5B of wheat (Danilova *et al.*, 2014) and 5M^{g}S of *Ae. geniculata*. This region of chromosome 5D of wheat cv. Chinese Spring is probably inverted; however, additional work is required to confirm this assumption. The results of this analysis indicate that chromosome 5D-specific mapping information may be used for ordering homoeologous 5M^g genes.

Physical localization of repetitive DNA-based probes

We identified repetitive DNA elements, especially simple sequence repeats (SSRs), from sequences of flow-sorted chromosome 5M^g (Table S2). A subset of these were used for developing FISH probes to identify chromosome 5M^g. Figure 7 shows FISH patterns of individual SSRs on chromosome 5M^g together with an ideogram of localization patterns. The FISH pattern of TR14 was specific to the proximal region of the long arm of chromosome 5M^g and the short-arm sub-terminal region of *Ae. comosa* subsp. *subventricosa* (Figure 7h,i). However, no hybridization signal was detected in another sub-species of *Ae. comosa* subsp. *comosa*, implying that potential M-genome donor species may have arisen from hybridization between *Ae. comosa* subsp. *subventricosa* and one of the U-genome donor species (Figure 7i,j). This probe was also used to detect chromosome 5M^g in the wheat/ *Ae. geniculata* chromosome addition line (Figure 7k,l). Thus, a combination of this probe and others may be used to identify the 5M^g alien chromosome in a wheat background.

Anchoring and ordering of 5M^g NGS contigs

Our results with single-gene FISH revealed no evidence of a major structural rearrangement between chromosomes 5D and 5M^g, which indicated that the 5D data may be used to map 5M^g contigs and genes. We used barley-specific GenomeZipper information to anchor 5M^g genes (Table S3) and 5M^g contigs. Because chromosomes 5M^g and 5D show a close relationship, chromosome 5D-specific high-density SNP, POPSEQ and 5D physical map data (Wang *et al.* 2014, Chapman *et al.*, 2015; Luo *et al.*, 2013) were used to anchor 5M^g genes and repeat-free NGS contigs (Figure 4 and Table S3). Approximately 2600 5M^g genes or gene fragments and 5M^g contigs were anchored and ordered based on GenomeZipper, POPSEQ and *Ae. tauschii* map-based datasets. The chromosome 5D marker sequences (POPSEQ and *Ae. tauschii* physical map) were used for BLAST-based searches against 5M^g

© 2015 The Authors

contigs, and information about the closest markers, based on sequence similarity, was provided with respect to the GenomeZipper data. Because an *Ae. geniculata* -specific genetic map was not used in this study, our ability to resolve 5M^g contigs was limited, but single-gene FISH data indicated that 5D-specific data may be transferred to 5M^g for comparative genomics studies, including anchoring and ordering of 5M^g genes and contigs.

Development of SNP markers

Chromosome 5M^g is an important source for agronomically important genes (such as *Lr57*, *Yr40* and *Sr53*). To develop a marker resource for identification of 5M-specific introgressions across the various M-genome species, we focused on developing 5M-specific SNP markers. To develop short arm- and long arm-specific SNPs, 5M^g reads from *Ae. geniculata* were mapped onto homoeologous group 5 wheat chromosomes (5A, 5B and 5D) and short-arm and long-arm assemblies of hexaploid wheat cv. Chinese Spring. After this step, we focused on finding chromosome 5M-specific SNPs that had the same allele on all group 5 wheat chromosomes but a different allele on 5M^g. We observed that 4538 SNPs had the same alleles in 5A, 5B and 5D chromosomes and different alleles in 5M^g. By mapping SNP sequences to arm-specific chromosome survey sequences (CSS) assemblies of Chinese Spring group 5 chromosomes, arm-specific 5M SNPs were identified (Table S4). A total of 364 SNPs were identified on the short arm and 1462 SNPs were identified on the long arm of chromosome 5M^g (Table S4). To find SNPs in genic regions, we aligned these SNPs against wheat ESTs mapped on chromosomes 5A, 5B and 5D. The method described by Tiwari *et al.* (2014) was then used to confirm that the flanking sequences of 5M^g-specific SNPs matched the 5M^g assemblies. Flanking sequences of 5M^g-specific SNPs were BLAST-searched against the 5M^g assemblies so that the primers designed based on the flanking sequences would work for chromosomes 5A, 5B, 5D and 5M^g. To develop specific SNP assays, interference from variations that may be located in the flanking sequences was eliminated, and only 235 SNPs that showed no variation in the 100 bp flanking sequences of SNPs between the 5A and 5M^g contigs were selected as described by Tiwari *et al.* (2014). To locate these genic SNPs, 5M^g-specific sequences were then mapped on deletion bin-mapped wheat ESTs (5A, 5B and 5D), and SNPs that showed at least 99% similarity over 100 bp sequences were selected. Because 5M^g is similar to chromosome 5D of wheat (based on DNA sequence composition and single-gene FISH) and a physical map of 5D is available, these ESTs were ordered based on the 5D physical map. A total of 135 ordered 5M^g-specific SNPs (Table S5) were placed in seven homoeologous deletion bins of chromosome 5D, which provides an excellent resource for screening for 5M^g-specific introgressions from M-genome species (http://wheatgenetics.org/download/category/19-5 m-geniculata).

Identification of 5M^g contigs specific to the *Lr57* and *Yr40* region

Chromosome 5M^g of *Ae. geniculata* TA10437 is a source of important genes including *Lr57*, *Yr40* (Kuraparthy *et al.*, 2007) and *Sr53* (Liu *et al.* 2011). Although the *Lr57* and *Yr40* genes were mapped in the distal region of chromosome 5D with a very small distal segment from chromosome 5M^g, fine mapping and identification of the putative candidates for these genes was not possible because of the limited number of markers for *Ae. geniculata*. We have developed germplasm and genomic resources to narrow down the genetic region containing these two important genes. Kuraparthy *et al.* (2007) mapped *Lr57* and *Yr40* genes and identified the breakpoint between 5M^g of *Ae. geniculata* and 5D of hexaploid wheat using 5D-specific EST-derived sequenced-tagged sites markers from wheat deletion bin 5DS2-0.78-1.00. By combining an NGS approach with rice synteny and the *Ae. tauschii* physical map, the *Lr57* and *Yr40* region may be narrowed to a small genetic interval (Table S3). We used all the EST sequences mapped by Kuraparthy *et al.* (2007) for a BLASTN search against the *Ae. tauschii* physical map database (http://probes.pw.usda.gov/WheatDMarker/phpblast/blast.php). By combining both results, we determined the virtual map position of flanking ESTs of the *Lr57* and *Yr40* region on *Ae. tauschii* chromosome 5D (Tables S3 and S5). Our results with single-gene FISH do not indicate any major rearrangement between chromosomes 5D and 5M^g. The 5D physical map information may be used to anchor and order 5M^g-specific contigs for the *Lr57* and *Yr40* region (Table S5).

DISCUSSION

Complexity reduction and sequencing of chromosome 5M^g

Like other members of the tribe Triticeae, *Aegilops* species have large and complex genomes, and their ploidy varies from diploid to hexaploid. *Ae. geniculata* is allotetraploid, and its 1C genome size of 10.06 Gb (Eilam *et al.*, 2008) is similar to the 12.38 Gb genome size for tetraploid durum wheat (*T. turgidum* subsp. *durum*) (Boyko *et al.*, 1984). To work efficiently on such a complex genome, reducing genome complexity, which involves high repetitive DNA content (approximately 85%) and homoeologous genes, is useful. Several approaches to reduce genome complexity are useful, such as sequencing complementary DNA for generation of ESTs, RNA-Seq, exome capture, genotyping-by-sequencing and chromosome flow sorting. The first two strategies require developmental stage specificity of RNA samples (Martienssen *et al.*, 2005; Kopecky *et al.*, 2013). Exome capture is specific to the functional and

© 2015 The Authors

hence smaller part of the genome, and genotyping-by-sequencing may be used only for mapping and low-end genetic diversity studies due to the low coverage of very small sequence reads. On the other hand, the flow cytometric chromosome sorting approach reduces genome complexity (Dolezel *et al.*, 2014) and plays an important role in dissecting large and complex cereal genomes such as wheat, rye and barley (Mayer *et al.*, 2011, Martis *et al.*, 2013; International Wheat Genome Sequencing Consortium, 2014). Chromosome flow sorting in wild relatives of wheat has also been reported (Molnár *et al.*, 2011; Molnár *et al.*, 2013; Grosso *et al.*, 2012).

However, in some species, not every chromosome may be discriminated and flow sorted. Thus, chromosome $5M^g$ was not separated in flow karyotypes of *Ae. geniculata*, and was only sorted together with a group of other five chromosomes ($2M^g$, $4M^g$, $2U^g$, $2U^g$ and $4U^g$) (Molnár *et al.*, 2011). Our results demonstrate that one solution may be to sort the chromosome of interest from alien chromosome addition and substitution lines. This is possible in cases where the substituting or added chromosome differs in relative size (measured in our case by DAPI fluorescence) from the remaining chromosomes in the karyotype. Based on the observations by Vrána *et al.* (2015), who mapped distributions of chromosomes within the three composite peaks on the flow karyotype of hexaploid wheat, and chromosome sizes (Safar *et al.*, 2010), our results indicate that the peak of chromosome $5M^g$ was resolved from the left shoulder of composite peak II due to the absence of wheat chromosome 5D, which has almost the same molecular size as $5M^g$ (748 Mb versus 750 Mb). In the flow karyotype of the substitution line lacking 5D, chromosome $5M^g$ was resolved clearly between the right shoulder of chromosome 4D (648 Mb) on composite peak I and the left shoulder of chromosomes 1A (799 Mb) and 6A (705 Mb) on composite peak II. The apparent discrepancies between the estimated chromosome sizes and their peak position are most probably due to the fact that chromosomes were classified by flow cytometry after staining by DAPI, which binds preferentially to AT-rich regions of DNA. Thus, DAPI fluorescence does not correspond to the DNA amount (Doležel *et al.*, 1992). This observation indicates differences in AT/GC ratio between chromosome $5M^g$ and wheat 5D.

The strategy used here will be extremely useful to develop genomic resources for targeted chromosomes from wild wheat relatives. Combining NGS technology with flow cytogenetics provides a powerful approach for studying complex genomes at the individual chromosome level or even chromosome arm level (Dolezel *et al.*, 2012; Choulet *et al.*, 2014, International Wheat Genome Sequencing Consortium, 2014; Tiwari *et al.*, 2014). These advances have opened up opportunities for developing genomic resources for the wheat gene pool.

Gene content analysis

Comparative genomics approaches have been extensively used in wheat and other cereals, and the inter-relationship of syntenic blocks among the various cereal genomes has been established (Devos and Gale, 2000; Salse *et al.*, 2008; Molnár *et al.*, 2013). Gene content-based analysis predicted the presence of approximately 4236 genes on chromosome $5M^g$, which is very similar to the corresponding homoeologous group 5 chromosomes of hexaploid wheat, with similarities of 96%, 95% and 98% for chromosome 5A, 5B and 5D, respectively. Lucas *et al.* (2014) reported the presence of 4286 genes on 5D, which supports our estimated gene number. Approximately 72% of the annotated $5M^g$ genes showed sequence identity to the wheat homoeologous genes on chromosomes 5A, 5B and 5D. If all annotated genes (including low-confidence genes) are included, then 3261 (approximately 77%) of the $5M^g$ genes had homoeologs on the group 5 wheat chromosomes. The smaller and highly fragmented $5M^g$ assembly was the major limitation in identification of complete or partial genes. We were able to identify 1118 genes that had complete open reading frames, but we were unable to differentiate complete or partial genes in the rest of the cases. The reference CSS assembly used for identification of homoeologous genes for chromosome $5M^g$ had approximately 90% completeness in terms of representation of total genes on the wheat genome (International Wheat Genome Sequencing Consortium, 2014). Our flow-sorting results were confirmed with 95% purity for chromosome $5M^g$. This indicates that there is a possibility of approximately 5% contamination from other chromosomes. Compounded with the highly fragmented $5M^g$ assembly, use of very stringent sequence identity criteria, and partially incomplete CSS reference assembly, this may explain the existence of a subset of $5M^g$ genes without any wheat homoeologs. Similar results were observed for the number of genes on chromosome $5M^g$ and homoeologous chromosomes $5A^u$ and $5D^t$ from *T. urartu* and *Ae. tauschii*, the diploid A- and D-genome donors of hexaploid wheat (Figure 3b). A close similarity between the sizes of chromosome $5M^g$ (750 Mb) and wheat chromosome 5D (748 Mb) (Lucas *et al.*, 2014) was also observed. These similarities indicate that 5D may be used as a reference to study chromosome $5M^g$. The annotated sequences identified here for chromosome $5M^g$ represent an important genomic resource for wheat improvement.

We used GenomeZipper to develop a virtual gene order map of *Ae. geniculata* chromosome $5M^g$. Various studies on crop plants have established that GenomeZipper is a useful tool for gene-based anchoring and ordering of physical maps, especially in crops where a reference sequence is not available (Mayer *et al.*, 2011, Kopecky *et al.*, 2013; Martis *et al.*, 2013; Philippe *et al.*, 2013; International

© 2015 The Authors

Wheat Genome Sequencing Consortium, 2014). Mapping of 5M^g genes showed collinear order with barley chromosome 5H, in agreement with various published reports, further highlighting the conservation of large-scale co-linearity among species of Triticeae (Mayer *et al.*, 2011; Lucas *et al.*, 2014). Chromosome 5M^g was collinear with three rice chromosomes (3Os, 11Os and 12Os), two *Brachypodium* chromosomes (Bd1 and Bd5) and four sorghum chromosomes (Sb1, Sb2, Sb5 and Sb8). Our findings were also supported by previous reports that indicated a syntenic relationship of group 5 wheat chromosomes (5A and 5D) with the *Brachypodium*, rice, sorghum and barley genomes (Vitulo *et al.*, 2011; Lucas *et al.*, 2014).

Determination of homoeologous relationships of 5M^g using single-gene FISH

Single-gene FISH is an important technique that utilizes highly conserved genic sequences among Triticeae species to develop homoeolog-specific wheat and Triticeae-specific FISH markers (Lamb *et al.*, 2007; Danilova and Birchler, 2008; Ma *et al.*, 2010; Karafiátová *et al.*, 2013). Danilova *et al.* (2014) used single-gene FISH to analyze homoeologous relationships and chromosomal rearrangements within the Triticeae. Our study with single-gene FISH, using 11 FL-cDNA probes, indicated no major rearrangement between 5M^g of *Ae. geniculata* and chromosome 5D of hexaploid wheat (Figure 6). Low-resolution, single-gene FISH provided the gene order and allowed us to use comparative gene mapping based on synteny conservation between chromosome 5D and 5M^g.

Limited information on evolutionary and cytogenetic relationships between wheat and its wild relatives hampers transfer of alien genes to wheat (Friebe *et al.*, 1996b; Ceoloni, 2006). Information about chromosomal rearrangements of a given chromosome is important for the success of gene transfer by induced homoeologous recombination (Riley *et al.*, 1968; Sears, 1977), because it depends on genome affinity and chromosome co-linearity (Danilova *et al.*, 2014). The extent of meiotic pairing of alien chromosomes or chromosome arms with wheat homoeologs corresponds to the level of structural rearrangements (Ceoloni *et al.*, 1988; Devos *et al.*, 1993; Cuadrado *et al.*, 1997; Lukaszewski *et al.*, 2004). Chromosomes 5M^g (this study) and 5D (Lucas *et al.*, 2014) are of similar size (approximately 750 Mb), and the number of predicted genes is also similar.

Chromosome 5M^g did not show any major structural rearrangements relative to the group 5 wheat chromosomes (Figure 6), and therefore gene transfer from 5M^g may be facilitated by induced homoeologous pairing between 5M^g and the group 5 wheat chromosomes. Several important genes have been transferred from *Ae. geniculata* accession TA10437 to wheat (Kuraparthy *et al.*, 2007; Liu *et al.*, 2011). Liu *et al.* (2011) reported a higher than expected chromosomal meiotic metaphase I pairing between chromosomes 5D and 5M^g (from the same accession of *Ae. geniculata* used in this study), suggesting no major rearrangements in chromosome 5M^g.

The annotated 5M genes and contigs were anchored using GenomeZipper. We used the barley physical map for chromosome 5H to deduce a virtual gene order for 5M^g. Due to the absence of an *Ae. geniculata* -specific genetic map, we are unable to resolve 5M chromosomal events. However, single-gene FISH data indicated that wheat chromosome 5D data may be transferred to chromosome 5M^g, and 5D-specific markers may be used to anchor 5M^g contigs and to deduce gene order. We used 5D high-density SNP marker data (Figure 4), POPSEQ data and *Ae. tauschii* 5D map data (Table S3) to anchor 5M^g contigs (Luo *et al.*, 2013; Wang *et al.*, 2014; Chapman *et al.*, 2015). To avoid any discrepancies between marker datasets, we used the 5M^g virtual order deduced by GenomeZipper (based on barley 5H) and provided corresponding POPSEQ and *Ae. tauschii* 5D map positions (Table S3).

These ordered contigs may be used as a resource for developing more markers to target 5M^g specific-genes such as *Lr57*, *Yr40* and *Sr53* (Table S5). Using a comprehensive *in silico* approach, we narrowed the *Lr57* and *Yr40* region and corresponding sequenced contigs of chromosome 5M^g (Tables S3 and S5). SNPs identified in this research will be useful in the eventual map-based cloning of genes *Lr57* and *Yr40*.

5M^g repeat content: analysis and visualization

The high repeat content of 5M^g (74.0%) is as expected from a member of the tribe Triticeae (Wicker *et al.*, 2005, 2007; Daron *et al.*, 2014). Daron *et al.* (2014) reported 85% transposable elements in the 833 Mb chromosome 3B, and 77% and 74% in the diploid A- and D-genome donors *T. urartu* and *Ae. tauschii*, respectively. Class I transposable elements were major contributors to the repeat content of 5M^g (Figure 5), with *Ty3*/*gypsy*-like elements being the most abundant (33.46%), followed by *Ty1*/*copia*-like retroelements (12.82%) and putative non-annotated LTR retroelements (10.49%). DNA transposons represented only approximately 8% of the reads, with 0.4% of reads being homologous to LINE elements. Retroelements that move via an RNA intermediate are known to be the major component of plant genomes (Sabot *et al.*, 2005; Wicker *et al.*, 2007; Charles *et al.*, 2008; Tenaillon *et al.*, 2010). Daron *et al.* (2014) analyzed the transposable element composition of wheat chromosome 3B, and also found that *Ty3*/*gypsy*-like elements accounted for the highest proportion (47%) of the total transposable element fraction, in agreement with our results.

Repeat-based FISH probes were developed and used to characterize chromosome 5M^g. The chromosomal localization of various repetitive DNA sequences, including SSRs such as $(GAA)_n$, tandem repeat sequences (pSc119.2 and the *Afa* family) and ribosomal genes have been used to

© 2015 The Authors
The Plant Journal © 2015 John Wiley & Sons Ltd,

identify the chromosomes of wheat and its wild relatives (Jiang *et al.*, 1994; Pedersen and Langridge, 1997; Cuadrado *et al.*, 2008b; Sepsi *et al.*, 2008). *De novo* identification of repetitive DNA elements by NGS of whole-genome or flow-sorted chromosomes has proven a useful resource for developing cytogenetic FISH markers in several plant species (Macas *et al.*, 2007, 2011; Torres *et al.*, 2011; Kopecky *et al.*, 2013). Thus, the probes developed in our study may be used to identify the 5M^g alien chromosome in a wheat background.

Development of 5M^g-specific SNPs

Despite the rapid advancement in NGS technology, very limited marker information is available for alien chromosomes, especially from the tertiary gene pool of wheat. We developed 4538 5M^g-specific SNPs, of which 235 high-confidence SNPs (Table S6) were located in the chromosome 5D deletion bins of bread wheat and ordered by the 5D physical map. These bin-mapped and ordered SNPs (upon validation) may be used to identify large to cryptic translocations between wheat and 5M chromosomes from M-genome species such as *Ae. comosa* (MM), *Ae. crassa* (UUMM), *Ae. biuncialis* (UUMM), *Ae. neglecta* (UUMM) and *Ae. columnaris* (UUMM), which will reduce the extensive genomic in situ hybridization - and FISH-based screening of progenies required to select targeted translocations.

CONCLUSIONS

The production of an Illumina sequence assembly of chromosome 5M^g provides useful insight into the genome organization of a wild wheat relative and Triticeae genome evolution. We characterized the synteny between chromosome 5M^g and its sequenced grass genome relatives, and our analysis indicated no major chromosomal rearrangements between chromosomes 5M^g and 5D of wheat. The NGS data were used for constructing the virtual gene order of 5M^g, anchoring and mapping of contigs on a physical map, and studying structural rearrangements in chromosome 5M^g. The data were used to develop SNPs and repetitive DNA-based cytogenetic markers. The flow-sorted chromosome approach presented here is a powerful strategy for developing knowledge and molecular tools for wild relatives of wheat, and marks an important step forward for effective utilization of the tertiary gene pool for wheat improvement.

EXPERIMENTAL PROCEDURES

Plant material

Plant materials included *Ae. geniculata* (accession number TA10437) and the bread wheat (*T. aestivum* L. cv. Chinese Spring)/ *Ae. geniculata* 5M^g chromosome disomic substitution line TA6675 [DS5M^g(5D)]. The TA6675 line contains 20 pairs of wheat chromosomes and chromosome pair 5M^g from *Ae. geniculata*, substituting for chromosome 5D of wheat. For chromosome sorting, seeds of disomic substitution line TA6675 were used. The plant materials used in this study were procured, developed and maintained by the Wheat Genetics Resource Center at Kansas State University (http://www.k-state.edu/wgrc/).

Chromosome analysis, sorting and sequencing

Aqueous suspensions of mitotic metaphase chromosomes were prepared from wheat/ *Ae. geniculata* disomic substitution line TA6675 as described by Vrána *et al.* (2000). Samples were stained using DAPI and analyzed using a FACSAria II SORP flow cytometer and sorter (BD Biosciences, http://www.bdbiosciences.com/us/home?cc=US&WT.srch=1&gclid=CNiqvu3CwsgCFQiraQod3ygJSg). Three independent samples of 33 000 chromosomes (equivalent of 3 x 50 ng DNA) were sorted into 40 μl of sterile deionized water in a 0.5 ml PCR tube. The contamination of sorted fractions by other chromosomes was assessed for each sort run by analyzing 1000 chromosomes on a microscopic slide. The chromosomes were identified by FISH using probes for the Afa family and $[GAA]_n$ repeats. The protocol described by Šimková *et al.* (2008) was used to amplify DNA of flow-sorted chromosome 5M^g. The sorted fractions were treated using proteinase and purified using Microcon YM-100 columns (Millipore, http://www.emdmillipore.com/). The 5M^g DNA was amplified using an Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare, http://www3.gehealthcare.com/en/global_gateway). The three samples of amplified DNA were then pooled to reduce possible amplification bias. The amplified chromosomal DNA was sequenced using a HiSeq 2000 system (Illumina, https://www.illumina.com/). The pooled multiple displacement amplification-amplified DNA (5.0 μg) was used to create a shotgun DNA-seq library. DNA sequencing libraries were prepared using an Illumina TruSeq DNA kit, and agarose gel size selection was used for library preparation steps. The mean fragment size for the sequencing library was 504 bp. Because the size of the adapter sequences was 120 bp, the final insert size for sequencing was 384 bp. The 5M^g-specific library was sequenced on one lane of an Illumina HiSeq 2000 system at the Center for Genome Research & Biocomputing, Oregon State University (Corvallis, OR).

Assembly of chromosome 5M^g sequences

Sequence data were checked for read quality, and, after initial filtering and trimming, Illumina paired-end reads were assembled using SOAPdenovo (Luo *et al.* 2012), and four *k*-mer-based parameters were used (31, 57, 75 and 79) to achieve optimum assembly. The raw 5M^g reads were deposited in the National Center for Biotechnology Information Sequence Read Archive under accession number SRX1167449. The program was run on the Beocat high-performance computing clusters at Kansas State University, and the assembly with the best N50 was selected for further analyses. Table S6 provides detailed information about 5M^g assemblies generated using various *k*-mers.

The repetitive sequences from the assembly were masked by BLAST searching against the GIRI plant repeats database (http://www.girinst.org/). The 75-mer-based run provided the assembly with the best sequence coverage and N50. The repetitive DNA contents in chromosome 5M^g sequences were identified using Vmatch (http://www.vmatch.de) against the MIPS-REdat Poaceae version 8.6.2 repeat library (http://pgsb.helmholtz-muenchen.de/plant/recat/search.jsp).

Gene estimation and GenomeZipper analysis

Approximately one million cDNA sequences of hexaploid wheat were downloaded from the National Center for Biotechnology

© 2015 The Authors

Information Sequence Read Archive and mapped to 5M^g assemblies to predict genes and structures using the PASA pipeline with default parameters (Haas *et al.*, 2003). In total, approximately 318 166 cDNA sequences were aligned to 5M^g assemblies. After filtering with a minimum of 90% similarity and correct splice sites, 52 203 cDNA sequence alignments were used for gene content prediction. The predicted 5M^g transcripts were BLAST-searched against other sequenced grass species including *Brachypodium*, rice, sorghum and barley transcript sequences to detect possible orthologous genes. The best hits with an e-value <1e-30 and for which ≥60% of the overall length of the coding sequence was aligned were selected as the corresponding orthologous genes.

Genes on *Ae. geniculata* chromosome 5M^g and its sequence contigs were then ordered based on syntenic relationships of the orthologous genes. The repeat-masked 5M^g contigs were also compared to a bread wheat chromosome 5D shotgun sequence assembly (International Wheat Genome Sequencing Consortium, 2014) using the National Center for Biotechnology Information BLASTN tool with a cut-off e-value of 1e-30. The 5D consensus map (1009 markers) constructed based on a 90K genotyping assay of eight populations (Wang 2014) and physical mapping data of gene markers from an integrated physical map of chromosome 5D of *Ae. tauschii* was then used to order 5M^g genes and NGS contigs. A high-density POPSEQ map (Chapman *et al.*, 2015) of bread wheat chromosome 5D was also used to order 5M^g genes and contigs. The corresponding relationships of the ordered genes with various grass genomes and wheat 5D consensus maps were visualized using Circos plot (Krzywinski *et al.*, 2009). To order the 5Mg contigs, barley-based GenomeZipper data were used. The corresponding mapping information of 5D POPSEQ and the 5D physical map (*Ae. tauschii*) are provided against the GenomeZipper-based ordered 5M virtual map for downstream applications.

Repeat reconstruction and annotation

A random dataset, representing 0.4 × coverage of chromosome 5M^g, was retrieved from the Illumina data and used for reconstruction of repetitive elements using a graph-based method (Novak *et al.*, 2010) using the Louvain clustering algorithm (Blondel *et al.*, 2008). Resulting clusters were manually annotated using several sources, including similarity searches using RepeatMasker (http://www.repeatmasker.org), and BLASTX and BLASTN tools were used for similarity searches against public databases and also against the database of domains derived from plant mobile elements (http://repeatexplorer.umbr.cas.cz). Dotter analysis (Sonnhammer and Durbin, 1995) was used to confirm the presence of tandem organized repeats.

FISH with single-gene and repetitive DNA element-based probes

Somatic chromosome preparations, direct probe labeling by nick translation and FISH were performed as described by Kato *et al.* (2006) and Kato *et al.* (2004), with modifications as described by Danilova *et al.* (2012). To study the structure of the 5M^g chromosome, 11 cDNA FISH probes (Table S3) mapped on group 5 wheat chromosomes were used, and their positions relative to the centromere were measured on five to eight 5M^g chromosomes from the substitution line TA6675 and *Ae. geniculata* TA10437 (Danilova *et al.*, 2014).

Chromosome preparation and repetitive DNA element-based FISH were performed as described by Danilova *et al.* (2014) and Koo *et al.* (2010). The oligonucleotide probes CY5-$(AAG)_7$, 6-FAM-$(GA)_{12}$ and TEX615-TCAGAGTTCATTTGAAATGCTTTTCA, labeling the D-genome-specific Afa family repeat (Nagaki *et al.*, 1995), were synthesized by Integrated DNA Technologies (https://www.idtdna.com/site) and used to identify chromosome 5M^g on chromosomal spreads of *Ae. geniculata* and the substitution line TA6675. Clone *pTa71* containing 8S, 5.8S and 26S rRNA wheat genes (Gerlach and Bedbrook, 1979) labeled with fluorescein-12-dUTP (Perkin Elmer, http://www.perkinelmer.com/lab-solutions/default.xhtml) was used to detect NOR sites. A probe for the LTR of transposable element Quinta (Li *et al.*, 2013) was labeled with fluorescein-12-dUTP and used to detect centromeres. Other SSRs used as FISH probes (Table S4) were amplified in the absence of template DNA (Cox *et al.*, 1993). Wheat chromosomes 3B, 4B, 5B and 6B have large blocks of (GA) repeats (Cuadrado *et al.*, 2008a), but their pattern is different from that of 5M^g. Images were captured using an Axioplan 2 microscope (Zeiss, http://zeiss-campus.magnet.fsu.edu/) and a CoolSNAP HQ2 camera (Photometrics, http://www.photometrics.com/) and AxioVision 4.8 software. Images were processed using Adobe Photoshop software (Adobe Systems, http://www.adobe.com/).

SNP discovery and chromosome 5M^g-specific markers for short and long arms

To identify *Ae. geniculata* chromosome 5M^g-specific SNPs, we used the approach described by Tiwari *et al.* (2014). Before aligning reads to references, all reads were trimmed to remove low-quality bases (Phred score ≤15, http://www.phrap.org/phredphrap consed.html) at the ends; reads of more than 30 bp after quality trimming were then filtered based on overall quality (80% of bases must have quality ≥15). CSS assemblies from wheat chromosome arms 5AS, 5AL, 5BS, 5BL, 5DS and 5DL (International Wheat Genome Sequencing Consortium, 2014; http://www.wheatgenome.org/) were used as reference to map 5M^g reads, and variants were called in the second step. Bowtie software (version 1.3) (http://bowtie-bio.sourceforge.net/index.
html) was used to map quality filtered reads of 5M^g to the Chinese Spring 5A, 5B and 5D contigs using mapping quality control parameters as: –k 1 –best –v 3. The 5M^g reads that mapped on the short arm of group 5 assemblies were considered as 5M^g short-arm SNPs, whereas those that mapped on the long arm were considered as 5M^g long-arm SNPs. The alignment results were saved in SAM format files using SAMtools (version 1.8) (http://www.htslib.org), and used to generate pileup files, which were then fed to BCFtools (http://www.htslib.org) to call raw variations using default parameters. To develop 5M^g-specific marker assays and avoid any other variation in the flanking SNP region, the program BLAT (Kent, 2002;) was used to align the 100 bp flanking sequences of 5M^g-specific SNPs against the reference contigs of homologous wheat chromosome arms (5AS, 5AL, 5BS, 5BL, 5DS and 5DL). The alleles of SNPs with 100% similarity on the flanking sequences in 5A, 5B and 5D were then compared to select potential 5M^g -specific SNPs.

To develop 5M^g SNPs from genic regions and localize 5M^g SNP sequences on wheat deletion bins, 5M^g sequences were aligned against the wheat EST database (http://wheat.pw.usda.gov/wEST/). SNP sequences with a high similarity percentage were shortlisted and assigned to deletion bins of homologous group 5 wheat chromosomes. The alleles of SNPs with 100% similarity on the flanking sequences in 5A, 5B and 5D chromosome-specific ESTs were then compared to select potential 5M^g-specific SNPs. Because the majority of 5D ESTs were ordered based on the *Ae. tauschii* physical map, we used this information to order 5M^g-specific SNPs

© 2015 The Authors

(based on the order of the aligned EST with that particular SNP). Additionally, we used a consensus genetic map and a POPSEQ map of chromosome 5D to order ESTs corresponding to 5M contigs. Only sequences with a 99% match based on BLAST search were used to assign 5D deletion bin positions.

ACKNOWLEDGMENTS

Part of the research in the Gill laboratory was supported by grants from the Kansas Wheat Commission, Heartland Plant Innovations (Manhattan, KS) and King Abdulaziz University. The computing for this project was performed on the Beocat high-performance computing cluster at Kansas State University, which is funded in part by US National Science Foundation grants CNS-1006860, EPS-1006860 and EPS-0919443. J.D., E.H., J.V. and M.K. were supported by the National Program of Sustainability (award number LO 2014) and the Czech Science Foundation (award number P501-12-G090); E.A. and S.W. are supported by US Department of Agriculture grants 2011-68002-30029 and 2012-67013-19401. This is contribution number 15-431-J from Kansas Agricultural Experiment Station, Kansas State University, Manhattan, KS. The work was performed under the auspices of Wheat Genetics Resource Center-I/UCRC partially funded by US National Science Foundation grant contract IIP-1338897.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Examples of graph layouts visualized using SeqGrapheR.

Table S1. Distribution of repetitive elements.

Table S2. List of SSRs used in this study as FISH probes.

Table S3. GenomeZipper analysis and virtual gene map of chromosome 5M^g.

Table S4. Short arm- and long arm-specific SNPs.

Table S5. 5M-specific SNPs ordered in 5D deletion bins.

Table S6. Characteristics of 5M^g sequence assemblies.

REFERENCES

Abe, A., Kosugi, S., Yoshida, K. *et al.* (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* **30**, 174–178.

Berkman, P.J., Lai, K., Lorenc, M.T. and Edwards, D. (2012) Next-generation sequencing applications for wheat crop improvement. *Am. J. Bot.* **99**, 365–371.

Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. (2008) Fast unfolding of communities in large networks. *J. Stat. Mech.* **10**, 10008.

Boyko, E.V., Badaev, N.S., Maximov, N.G. and Zelenin, A.V. (1984) Does DNA content change in the course of Triticale breeding. *Cereal Res. Commun.* **12**, 99–100.

Ceoloni, C.A.J.P.P. (2006) Chromosome engineering of the durum wheat genome: strategies and applications of potential breeding value. In *Genetic Resources, Chromosome Engineering, and Crop Improvement: Cereals* (Singh, R.J. and Jauhar, P.P., eds). Boca Raton, FL: CRC Press, pp. 27–59.

Ceoloni, C., Del Signore, G., Pasquini, M. and Testa, A. (1988) Transfer of mildew resistance from *Triticum longissimum* into wheat by ph1-induced homoeologous recombination. In *7th International Wheat Genetics Symposium* (Miller, T.E. and Koebner, R.M.D., eds). Cambridge, UK: Institute of Plant Science Research, pp. 221–226.

Chapman, J.A., Mascher, M., Buluc, A. *et al.* (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.* **16**, 26.

Charles, M., Belcram, H., Just, J. *et al.* (2008) Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics*, **180**, 1071–1086.

Choulet, F., Alberti, A., Theil, S. *et al.* (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science*, **345**, 1249721.

Cox, A.V., Bennett, S.T., Parokonny, A.S., Kenton, A., Callimassia, M.A. and Bennett, M.D. (1993) Comparison of plant telomere locations using a PCR-generated synthetic probe. *Ann. Bot.* **72**, 239–247.

Cuadrado, A., Vitellozzi, F., Jouve, N. and Ceoloni, C. (1997) Fluorescence *in situ* hybridization with multiple repeated DNA probes applied to the analysis of wheat–rye chromosome pairing. *Theor. Appl. Genet.* **94**, 347–355.

Cuadrado, A., Cardoso, M. and Jouve, N. (2008a) Increasing the physical markers of wheat chromosomes using SSRs as FISH probes. *Genome*, **51**, 809–815.

Cuadrado, A., Cardoso, M. and Jouve, N. (2008b) Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. *Cytogenet. Genome Res.* **120**, 210–219.

Danilova, T.V. and Birchler, J.A. (2008) Integrated cytogenetic map of mitotic metaphase chromosome 9 of maize: resolution, sensitivity, and banding paint development. *Chromosoma*, **117**, 345–356.

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

Danilova, T.V., Friebe, B. and Gill, B.S. (2014) Development of a wheat single gene FISH map for analyzing homoeologous relationship and chromosomal rearrangements within the Triticeae. *Theor. Appl. Genet.* **127**, 715–730.

Daron, J., Glover, N., Pingault, L. *et al.* (2014) Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biol.* **15**, 546.

Devos, K.M. and Gale, M.D. (2000) Genome relationships: the grass model in current research. *Plant Cell*, **12**, 637–646.

Devos, K.M., Millan, T. and Gale, M.D. (1993) Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye and barley. *Theor. Appl. Genet.* **85**, 784–792.

Dhaliwal, H.S., Singh, H. and William, M. (2002) Transfer of rust resistance from *Aegilops ovata* into bread wheat (*Triticum aestivum* L.) and molecular characterisation of resistant derivatives. *Euphytica*, **126**, 153–159.

Dolezel, J., Vrana, J., Safar, J., Bartos, J., Kubalakova, M. and Simkova, H. (2012) Chromosomes in the flow to simplify genome analysis. *Funct. Integr. Genomics* **12**, 397–416.

Dolezel, J., Vrana, J., Capal, P., Kubalakova, M., Buresova, V. and Simkova, H. (2014) Advances in plant chromosome genomics. *Biotechnol. Adv.* **32**, 122–136.

Doležel, J., Sgorbati, S. and Lucretti, S. (1992) Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plant.* **85**, 625–631.

Egan, A.N., Schlueter, J. and Spooner, D.M. (2012) Applications of next-generation sequencing in plant biology. *Am. J. Bot.* **99**, 175–185.

Eilam, T., Anikster, Y., Millet, E., Manisterski, J. and Feldman, M. (2008) Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum*. *Genome*, **51**, 616–627.

Friebe, B., Badaeva, E.D., Kammer, K. and Gill, B.S. (1996a) Standard karyotypes of *Aegilops uniaristata*, *Ae. mutica*, *Ae. comosa* subspecies *comosa* and *heldreichii* (Poaceae). *Plant Syst. Evol.* **202**, 199–210.

Friebe, B., Jiang, J., Raupp, W.J., McIntosh, R.A. and Gill, B.S. (1996b) Characterization of wheat–alien translocations conferring resistance to diseases and pests: current status. *Euphytica*, **91**, 59–87.

Friebe, B., Qi, L.L., Nasuda, S., Zhang, P., Tuleen, N.A. and Gill, B.S. (2000) Development of a complete set of *Triticum aestivum–Aegilops speltoides* chromosome addition lines. *Theor. Appl. Genet.* **101**, 51–58.

Gerlach, W.L. and Bedbrook, J.R. (1979) Cloning and characterization of ribosomal-RNA genes from wheat and barley. *Nucleic Acids Res.* **7**, 1869–1885.

Gill, B.S., Friebe, B., Raupp, W.J., Wilson, D.L., Cox, T.S., Sears, R.G., Brown-Guedira, G.L. and Fritz, A.K. (2006) Wheat Genetics Resource Center: the first 25 years. *Adv. Agron.* **89**, 73–136.

Grosso, V., Farina, A., Gennaro, A., Giorgi, D. and Lucretti, S. (2012) Flow sorting and molecular cytogenetic identification of individual chromo-

© 2015 The Authors

somes of *Dasypyrum villosum* L. (*H. villosa*) by a single DNA probe. *PLoS ONE*, **7**, e50151.

Haas, B.J., Delcher, A.L., Mount, S.M. *et al.* (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666.

Henry, R.J. (2012) Next-generation sequencing for understanding and accelerating crop domestication. *Brief Funct. Genomics* **11**, 51–56.

Hernandez, P., Martis, M., Dorado, G. *et al.* (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J.* **69**, 377–386.

International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, **345**, 1251788.

Jiang, J., Friebe, B. and Gill, B.S. (1994) Recent advances in alien gene transfer in wheat. *Euphytica*, **73**, 199–212.

Jordan, K.W., Wang, S., Lun, Y. *et al.* (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol.* **16**, 48.

Karafiátová, M., Bartoš, J., Kopecký, D., Ma, L., Sato, K., Houben, A., Stein, N. and Doležel, J. (2013) Mapping non recombining regions in barley using multicolor FISH. *Chromosome Res.* **21**, 739–751.

Kato, A., Lamb, J.C. and Birchler, J.A. (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., Albert, P.S., Vega, J.M. and Birchler, J.A. (2006) Sensitive FISH signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. *Biotech. Histochem.* **81**, 71–78.

Kent, W.J. (2002) BLAT – the BLAST-like alignment tool. *Genome Res.* **12**, 656–664.

Koo, D.H., Nam, Y.W., Choi, D., Bang, J.W., de Jong, H. and Hur, Y. (2010) Molecular cytogenetic mapping of *Cucumis sativus* and *C. melo* using highly repetitive DNA sequences. *Chromosome Res.* **18**, 325–336.

Kopecky, D., Martis, M., Cihalikova, J. *et al.* (2013) Flow sorting and sequencing meadow fescue chromosome 4F. *Plant Physiol.* **163**, 1323–1337.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645.

Kuraparthy, V., Chhuneja, P., Dhaliwal, H.S., Kaur, S., Bowden, R.L. and Gill, B.S. (2007) Characterization and mapping of cryptic alien introgression from *Aegilops geniculata* with new leaf rust and stripe rust resistance genes *Lr57* and *Yr40* in wheat. *Theor. Appl. Genet.* **114**, 1379–1389.

Lamb, J.C., Danilova, T., Bauer, M.J., Meyer, J.M., Holland, J.J., Jensen, M.D. and Birchler, J.A. (2007) Single-gene detection and karyotyping using small-target fluorescence *in situ* hybridization on maize somatic chromosomes. *Genetics*, **175**, 1047–1058.

Li, B., Choulet, F., Heng, Y., Hao, W., Paux, E., Liu, Z., Yue, W., Jin, W., Feuillet, C. and Zhang, X. (2013) Wheat centromeric retrotransposons: the new ones take a major role in centromeric structure. *Plant J.* **73**, 952–965.

Liu, W., Rouse, M., Friebe, B., Jin, Y., Gill, B.S. and Pumphrey, M.O. (2011) Discovery and molecular mapping of a new gene conferring resitance to stem rust, *Sr53*, derived from *Aegilops geniculata* and characterization of spontaneous translocation stocks with reduced alien chromatin. *Chromosome Res.* **19**, 669–682.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M. (2012) Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 251364.

Lucas, S.J., Akpinar, B.A., Simkova, H., Kubalakova, M., Dolezel, J. and Budak, H. (2014) Next-generation sequencing of flow-sorted wheat chromosome 5D reveals lineage-specific translocations and widespread gene duplications. *BMC Genom.* **15**, 1080.

Lukaszewski, A., Lapinski, B. and Rybka, K. (2004) Limitations of *in situ* hybridization with total genomic DNA in routine screening for alien introgressions in wheat. *Cytogenet. Genome Res.* **109**, 373–377.

Luo, R., Liu, B., Xie, Y. *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, **1**, 18. doi:10.1186/2047-217X-1-18.

Luo, M.C., Gu, Y.Q., You, F.M. *et al.* (2013) A 4–gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. *Proc. Natl Acad. Sci. USA* **110**, 7940–7945.

Ma, L., Vu, G.T., Schubert, V., Watanabe, K., Stein, N., Houben, A. and Schubert, I. (2010) Synteny between *Brachypodium distachyon* and *Hordeum vulgare* as revealed by FISH. *Chromosome Res.* **18**, 841–850.

Macas, J., Neumann, P. and Navratilova, A. (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genom.* **8**, 427.

Macas, J., Kejnovsky, E., Neumann, P., Novak, P., Koblizkova, A. and Vyskot, B. (2011) Next generation sequencing-based analysis of repetitive DNA in the model dioceous plant *Silene latifolia*. *PLoS ONE*, **6**, e27335.

Martienssen, R.A., Zaratiegui, M. and Goto, D.B. (2005) RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. *Trends Genet.* **21**, 450–456.

Martis, M.M., Zhou, R., Haseneyer, G. *et al.* (2013) Reticulate evolution of the rye genome. *Plant Cell*, **25**, 3685–3698.

Mayer, K.F., Martis, M., Hedley, P.E. *et al.* (2011) Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell*, **23**, 1249–1263.

Molnár, I., Kubaláková, M., Šimková, H., Cseh, A., Molnár-Láng, M. and Doležel, J. (2011) Chromosome isolation by flow sorting in *Aegilops umbellulata* and *Ae. comosa* and their allotetraploid hybrids *Ae. biuncialis* and *Ae. geniculata*. *PLoS ONE*, **6**, e27708.

Molnár, I., Šimková, H., Leverington-Waite, M., Goram, R., Cseh, A., Vrána, J., Farkas, A., Doležel, J., Molnár-Láng, M. and Griffiths, S. (2013) Syntenic relationships between the U and M genomes of *Aegilops*, wheat and the model species *Brachypodium* and rice as revealed by COS markers. *PLoS ONE*, **8**, e70844.

Nagaki, K., Tsujimoto, H., Isono, K. and Sasakuma, T. (1995) Molecular characterization of a tandem repeat, Afa family, and its distribution among Triticeae. *Genome*, **38**, 479–486.

Novak, P., Neumann, P. and Macas, J. (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics*, **11**, 378.

Novak, P., Neumann, P., Pech, J., Steinhaisl, J. and Macas, J. (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, **29**, 792–793.

Paux, E., Sourdille, P., Salse, J. *et al.* (2008) A physical map of the 1–gigabase bread wheat chromosome 3B. *Science*, **322**, 101–104.

Pedersen, C. and Langridge, P. (1997) Identification of the entire chromosome complement of bread wheat by two-colour FISH. *Genome*, **40**, 589–593.

Philippe, R., Paux, E., Bertin, I. *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.

Poursarebani, N., Nussbaumer, T., Simkova, H., Safar, J., Witsenboer, H., van Oeveren, J., Dolezel, J., Mayer, K.F., Stein, N. and Schnurbusch, T. (2014) Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A. *Plant J.* **79**, 334–347.

Qi, L., Friebe, B., Zhang, P. and Gill, B.S. (2007) Homoeologous recombination, chromosome engineering and crop improvement. *Chromosome Res.* **15**, 3–19.

Ray, S. and Satya, P. (2014) Next generation sequencing technologies for next generation plant breeding. *Front Plant Sci.* **5**, 367.

Riley, R., Chapman, V. and Johnson, R. (1968) Incorporation of alien disease resistance in wheat by genetic interference with regulation of meiotic chromosome synapsis. *Genet. Res.* **12**, 199–219.

Sabot, F., Guyot, R., Wicker, T., Chantret, N., Laubin, B., Chalhoub, B., Leroy, P., Sourdille, P. and Bernard, M. (2005) Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. *Mol. Genet. Genomics* **274**, 119–130.

Safar, J., Bartos, J., Janda, J. *et al.* (2004) Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *Plant J.* **39**, 960–968.

Safar, J., Simkova, H., Kubalakova, M., Cihalikova, J., Suchankova, P., Bartos, J. and Dolezel, J. (2010) Development of chromosome-specific BAC

© 2015 The Authors
The Plant Journal © 2015 John Wiley & Sons Ltd,

resources for genomics of bread wheat. *Cytogenet. Genome Res.* **129**, 211–223.

Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M. and Feuillet, C. (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell*, **20**, 11–24.

Schneider, A., Molnar, I. and Molnar-Lang, M. (2008) Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat. *Euphytica*, **163**, 1–19.

Sears, E.R. (1977) An induced mutant with homoeologous pairing in common wheat. *Can. J. Gen. Cytol.* **19**, 585–593.

Sepsi, A., Molnár, I., Szalay, D. and Molnár-Láng, M. (2008) Characterization of a leaf rust-resistant wheat–*Thinopyrum ponticum* partial amphiploid BE–1, using sequential multicolor GISH and FISH. *Theor. Appl. Genet.* **116**, 825–834.

Šimková, H., Svensson, J.T., Condamine, P., Hribova, E., Suchankova, P., Bhat, P.R., Bartos, J., Safar, J., Close, T.J. and Dolezel, J. (2008) Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. *BMC Genom.* **9**, 294.

Sonnhammer, E.L. and Durbin, R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, GC1–GC10.

Tenaillon, M.I., Hollister, J.D. and Gaut, B.S. (2010) A triptych of the evolution of plant transposable elements. *Trends Plant Sci.* **15**, 471–478.

Tiwari, V.K., Wang, S., Sehgal, S. *et al.* (2014) SNP discovery for mapping alien introgressions in wheat. *BMC Genom.* **15**, 273.

Torres, G.A., Gong, Z., Iovene, M., Hirsch, C.D., Buell, C.R., Bryan, G.J., Novak, P., Macas, J. and Jiang, J. (2011) Organization and evolution of subtelomeric satellite repeats in the potato genome. *G3*, **1**, 85–92.

Varshney, R.K., Song, C., Saxena, R.K. and Azam, S. (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* **31**, 240–246.

Varshney, R.K., Terauchi, R. and McCouch, S.R. (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* **12**, e1001883.

Vitulo, N., Albiero, A., Forcato, C. *et al.* (2011) First survey of the wheat chromosome 5A composition through a next generation sequencing approach. *PLoS ONE*, **6**, e26421.

Vrána, J., Kubalakova, M., Simkova, H., Cihalikova, J., Lysak, M.A. and Dolezel, J. (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). *Genetics*, **156**, 2033–2041.

Vrána, J., Kubaláková, M., Číhalíková, J., Valárik, M. and Doležel, J. (2015) Preparation of sub-genomic fractions enriched for particular chromosomes in polyploid wheat. *Biol. Plant.* **59**, 445–455.

Wang, S., Wong, D., Forrest, K. *et al.* (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* **12**, 787–796.

Wicker, T., Zimmermann, W., Perovic, D., Paterson, A.H., Ganal, M., Graner, A. and Stein, N. (2005) A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv–eIF4E locus: recombination, rearrangements and repeats. *Plant J.* **41**, 184–194.

Wicker, T., Sabot, F., Hua-Van, A. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982.

Winfield, M.O., Wilkinson, P.A., Allen, A.M. *et al.* (2012) Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnol. J.* **10**, 733–742.

Zaharieva, M., Gaulin, E., Havaux, M., Acevedo, E. and Monneveux, P. (2001a) Drought and heat responses in the wild wheat relative *Aegilops geniculata* Roth: potential interest for wheat improvement. *Crop Sci.* **41**, 1321–1329.

Zaharieva, M., Monneveux, P., Henry, M., Rivoal, R., Valkoun, J. and Nachit, M.M. (2001b) Evaluation of a collection of wild wheat relative *Aegilops geniculata* Roth and identification of potential sources for useful traits. *Euphytica*, **119**, 33–38.

© 2015 The Authors