Plant Biotechnology Journal

*Plant Biotechnology Journal* (2017) **15**, pp. 249–256 doi: 10.1111/pbi.12610

# The *in silico* identification and characterization of a bread wheat/*Triticum militinae* introgression line

Michael Abrouk[1], Barbora Balcárková[1], Hana Šimková[1], Eva Komínkova[1], Mihaela M. Martis[2,3], Irena Jakobson[4], Ljudmilla Timofejeva[4], Elodie Rey[1], Jan Vrána[1], Andrzej Kilian[5], Kadri Järve[4], Jaroslav Doležel[1] and Miroslav Valárik[1,*]

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

[2]*Munich Information Center for Protein Sequences/Institute of Bioinformatics and Systems Biology, Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany*

[3]*Division of Cell Biology, Department of Clinical and Experimental Medicine, Bioinformatics Infrastructure for Life Sciences, Linköping University, Linköping, Sweden*

[4]*Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia*

[5]*Diversity Arrays Technology Pty Ltd, Canberra, ACT, Australia*

*Received 29 November 2015;*
*revised 21 July 2016;*
*accepted 8 August 2016.*

**Correspondence* (Tel +420 585 238 714; fax +420 585 238 704; email valarik@ueb.cas.cz)

**Keywords:** GenomeZipper, alien introgression, comparative analysis, chromosome rearrangement, chromosome translocation, linkage drag.

## Summary

The capacity of the bread wheat (*Triticum aestivum*) genome to tolerate introgression from related genomes can be exploited for wheat improvement. A resistance to powdery mildew expressed by a derivative of the cross-bread wheat cv. Tähti × *T. militinae* (*Tm*) is known to be due to the incorporation of a *Tm* segment into the long arm of chromosome 4A. Here, a newly developed *in silico* method termed rearrangement identification and characterization (RICh) has been applied to characterize the introgression. A virtual gene order, assembled using the GenomeZipper approach, was obtained for the native copy of chromosome 4A; it incorporated 570 4A DArTseq markers to produce a zipper comprising 2132 loci. A comparison between the native and introgressed forms of the 4AL chromosome arm showed that the introgressed region is located at the distal part of the arm. The *Tm* segment, derived from chromosome 7G, harbours 131 homoeologs of the 357 genes present on the corresponding region of Chinese Spring 4AL. The estimated number of *Tm* genes transferred along with the disease resistance gene was 169. Characterizing the introgression's position, gene content and internal gene order should not only facilitate gene isolation, but may also be informative with respect to chromatin structure and behaviour studies.

## Introduction

Using interspecific hybridization to widen a crop's gene pool is an attractive strategy for reversing the genetic bottleneck imposed by domestication and for compensating the genetic erosion, which has resulted from intensive selection (Feuillet *et al.*, 2008). Much of the pioneering research in this area has focused on bread wheat (*Triticum aestivum*), in which over 50 related species have been exploited as donors thanks to the plasticity of the recipient's genome (Jiang *et al.*, 1993; Wulff and Moscou, 2014). Typically, introgression events have involved the transfer of a substantially sized donor chromosome segment, which, along with the target, probably bears gene(s), which impact negatively on the host's fitness (a phenomenon also called 'linkage drag') (Gill *et al.*, 2011; Qi *et al.*, 2007; Zamir, 2001). For this reason, very few introgression lines are represented in commercial cultivars (Rey *et al.*, 2015). The prime means of reducing the length of an introgressed segment is to induce recombination with its homoeologous region (Niu *et al.*, 2011). The success of this strategy is highly dependent on the conservation of gene content and order between the donor segment and its wheat equivalent.

The level of resolution with which introgression segments can be characterized has developed over the years along with advances in DNA technology. Large numbers of genetic markers have been identified in many crop species, including wheat (Bellucci *et al.*, 2015; Chapman *et al.*, 2015; Sorrells *et al.*, 2011; Wang *et al.*, 2014). In a recent example, a wheat mapping population has been genotyped with respect to >100 000 markers, but the mapping resolution achieved has only enabled the definition of around 90 mapping bins per chromosome (Chapman *et al.*, 2015). Given that the genomes of most donor species are poorly characterized, marker data at best allow only the position of an introgressed segment to be defined on the basis of the loss of wheat markers; they cannot determine either the size of the introduced segment or analyse its genetic content. The recently developed 'Introgression Browser' (Aflitos *et al.*, 2015) combines genotypic data with phylogenetic inferences to identify the origin of an introgressed segment, but to do so, a high-quality reference sequence of the host genome is needed, along with a large set of donor sequence data. The first of these requirements is being addressed by a concerted effort to acquire a reference sequence for bread wheat (www.wheatgenome.org). So far, only chromosome (3B) has been fully sequenced, and the gene content of each wheat chromosome has been obtained (Choulet *et al.*, 2014; IWGSC, 2014). The so-called GenomeZipper method (Mayer *et al.*, 2011), based on a variety of resources, has been used to predict gene order along each of the 21 bread wheat chromosomes (IWGSC, 2014).

The improved resistance to powdery mildew of an introgressive line 8.1 derived from the cross of bread wheat cv. Tähti (genome formula ABD) and tetraploid *T. militinae* (*Tm*; genome formula $A^tG$) is known to be mainly due to the incorporation of a segment of *Tm* chromatin containing the resistance gene *QPm-tut-4A* into the long arm of chromosome 4A (Jakobson *et al.*, 2006, 2012). Here, a novel *in silico*-based method, termed rearrangement identification and characterization (RICh), has been developed to identify the sequences suitable for generating markers targeting an introgression segment such as the one from *Tm.* The method integrates the GenomeZipper approach with shotgun sequences of chromosome with the introgression. The RICh method was also effective in confirming the identity of the chromosomal rearrangements, which occurred during the evolution of modern wheat.

## Results

### Chromosome sorting, sequencing and assembly

The flow karyotype derived from the DAPI-stained chromosomes of the DT4AL-TM line included a distinct peak (Figure S1) corresponding to the 4AL telosome (4AL-TM), which enabled it to be sorted to an average purity of 86.2%. The contaminants in the sorted peak comprised a mixture of fragments of various chromosomes and chromatids. DNA of all 45 000 sorted 4AL-TM telosomes was amplified by DNA multiple displacement amplification (MDA). To minimize the risk of representation bias, the products from three independent amplification reactions were pooled. From the resulting 4.5 μg DNA, a total of ~6.2 Gb of sequence was obtained, which was subsequently assembled into 279 077 contigs of individual length >200 bp, with an N50 of 2068 bp (Table 1). When the assembly was aligned with the reference genome sequences of *Brachypodium distachyon* (Vogel *et al.*, 2010), rice (IRGSP, 2005) and sorghum (Paterson *et al.*, 2009), it was apparent that the 4AL-TM telosome shares synteny with segments of *B. distachyon* chromosomes Bd1 and Bd4, rice chromosomes Os3, Os6 and Os11 and sorghum chromosomes Sb1, Sb5 and Sb10 (Figure S2).

### Origin of the introgression segment

The chromosomal origin of the *Tm* introgression segment was established by initially flow sorting the *Tm* chromosome complement. This was achieved by pretreating the chromosomes with fluorescence *in situ* hybridization in suspension (FISHIS) (Giorgi *et al.*, 2013) in which GAA microsatellites were fluorescently labelled by FITC. The resulting DAPI *vs* GAA bivariate flow karyotype succeeded in defining 13 distinct clusters (Figure 1). As the haploid chromosome number of *Tm* is 14, one of the clusters was therefore deemed likely to harbour a mixture of two distinct chromosomes. Two of the clusters (#4 and #8) contained sequences that were amplified by the *Xgwm160* (Roder *et al.*, 1998) and *owm82* primers (these two markers are linked to the *QPm-tut-4A* gene from *Tm* introgression). The dispersed profile of cluster #4 (Figure 1) suggested that it was composed of two different A$^t$ genome chromosomes, because all G chromosomes were identified due to a higher GAA content (Badaeva *et al.*, 2010). The *owm72* marker, also linked to the *QPm-tut-4A* gene, amplified two fragments in *Tm*, one of size 205 bp and the other of size 250 bp; only the former was amplified from 4AL-TM telosome or of cluster #8. The fluorescence *in situ* hybridization (FISH) profile of the chromosomes present in cluster #8 unambiguously identified the introgressed segment as deriving from chromosome 7G.

Table 1 Assembly statistics of chromosome arms 4AL-TM, 4AS-CS and 4AL-CS

| | 4AS-CS | 4AL-CS | 4AL-TM |
|---|---|---|---|
| Sequencing read depth | 241x | 116x | 23x |
| Total contigs | 301 954 | 362 851 | 279 077 |
| Total bases (bp) | 282 335 959 | 361 971 522 | 266 737 930 |
| Assembly coverage* | 0.89x | 0.67x | 0.49x |
| Min contig length (bp) | 200 | 200 | 200 |
| Max contig length (bp) | 70 057 | 129 043 | 28 604 |
| Average contig length (bp) | 935 | 998 | 956 |
| N50 length (bp) | 2782 | 3053 | 2068 |

The data for 4AS-CS and 4AL-CS arms are taken from IWGSC (2014) and data for 4AL-TM were acquired in this study.

*The size of chromosome arms 4AS-CS (318 Mbp) and 4AL-CS (540 Mbp) were taken from Šafář *et al.* (2010). To estimate the assembly coverage of the 4AL-TM arm, the 4AL-CS size was used.

### GenomeZipper improvement

A chromosome 4A zipper was constructed based on Chinese Spring (CS) chromosome specific survey sequences (CSSs) using 1780 specific DArTseq markers ordered in consensus genetic map (Table S1). As DArTseq marker sequences are short (69 nt) and generally nongenic, they were initially anchored to the CSS assembly; this step reduced the number of useful markers to 632 (CSS-DArTseq markers), of which 102 mapped to the short arm and 530 to the long arm. The first version of the zipper comprised a total of 2398 loci. The resulting model for 4AS was collinear with Bd1, Os3 and Sb01, as reported previously (Hernandez *et al.*, 2012). However, the one for 4AL was a mosaic of 15 orthologous blocks (based on the rice genome as the reference), derived from Os11/Bd4/Sb5, Os3/Bd1/Sb1 and Os6/Bd1/Sb10 (Figure S3a). Validation for this complex structure was sought from analysis of the subset of 2638 SNP loci (Wang *et al.*, 2014), which had been assigned a bin locations based on an analysis of a panel of established 4A deletion lines (Endo and Gill, 1996): of these, 750 mapped to five deletion bins on 4AS

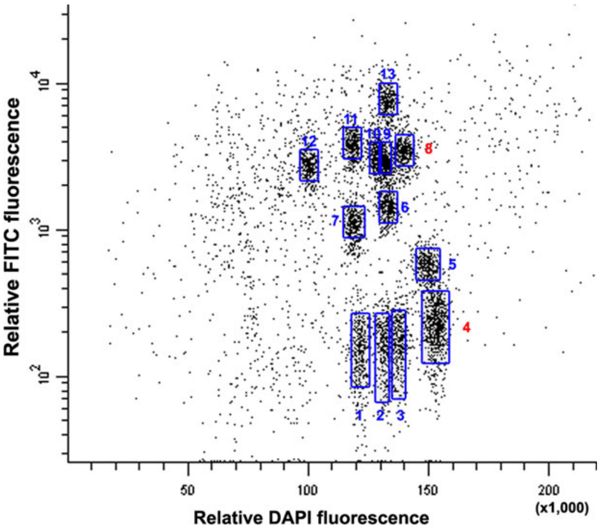

**Figure 1** The bivariate flow karyotype of *T. militinae.* Mitotic chromosomes at metaphase were stained with DAPI and GAA microsatellites were labelled with FITC. A set of 13 distinct clusters were obtained (shown boxed). Cluster #8 harbours the *Tm* chromosome (7G) which was the origin of the introgression segment present in line 8.1. Cluster #4 harbours a putative homoeolog of 7G and based on its width and shape most likely comprises a mixture of two distinct chromosomes.

and 1888 to 13 deletion bins on 4AL (Figure S3, Balcárková *et al.*, unpublished). The analysis allowed 329 SNP loci (113 on 4AS, 216 on 4AL) to be integrated into the new 4A zipper. Of the 113 4AS SNP loci, just four mapped to an inconsistent locations, demonstrating the model's accuracy; however, six (#3, #6, #8, #12, #14 and #16) of the 15 4AL blocks were inconsistent with respect to the multiple SNP loci allocations. For example, block #12—positioned in the subtelomeric region according to the zipper—included 18 SNP loci assigned to the pericentromeric region. The GenomeZipper was therefore rerun after first removing the 62 CSS-DArTseq markers associated with the misassignment of the blocks (Table S2); of the 570 CSS-DArTseq markers retained (Table S3), 79 were anchored to at least one of the *B. distachyon*, rice or sorghum scaffolds. The set of 2132 loci (745 on 4AS and 1387 on 4AL) revealed just six (rather than 16) blocks (Figure S3b, Table S2). The final structure resembles that described by Hernandez *et al.* (2012). When the model was retested with SNP markers, no further discrepancies were flagged along distal part of chromosome arm 4AL (Figure S3b).

### The *in silico* characterization of the evolutionary chromosome rearrangements on 4AL

The RICh method is based on a stringent identification and density estimation of homoeologs and is validated using a segmentation analysis. To test the approach, the CSS-based scaffolds of chromosome arms 4BS, 4BL, 4DS, 4DL, 5BL, 5DL, 7AS and 7DS (IWGSC, 2014) were compared with that of chromosome 4A, applying as the criteria a 90% level of identity and a minimum alignment length of 100 bp. The numbers of homoeologous loci obtained were, respectively, 719, 762, 636, 877, 850, 673, 602 and 627 (mean 718), but no common distinct blocks allowing for the definition of evolutionary translocations could be identified. A window size of eleven genes was then selected from the 4A zipper for the subsequent segmentation analysis. The ancestral 4AS and 4AL arms had an average density of 0.83, while the remainder of 4AL had a density of only 0.41 (Figure 2a). 4BL and 4DL sequences were homologous to 4AS, and 4BS and 4DS ones to 4AL, confirming the pericentromeric inversion event uncovered before (Devos *et al.*, 1995; Hernandez *et al.*, 2012; Ma *et al.*, 2013; Miftahudin *et al.*, 2004). Immediately following the ancestral 4AL region, the density of homoeologs associated with chromosome group 5 increased (5BL and 5DL: 147 genes, density 0.73), identifying the presence of ancestral 5AL chromatin on this arm (Figure 1b). Finally, the most distal segment of 4AL was associated with an increased density of chromosome group 7 (7AS and 7DS: 557 genes, density 0.45), confirming the ancestral translocation event involving 7BS (Figure 1C).

### Characterization of the *Tm* introgression segment

The RICh approach was then used to characterize the 4AL introgressed *Tm* segment. A direct comparison between the 4AL-TM sequence assembly and the 4A-CS zipper (95% identity, 100 bp minimum alignment length) was then made. For the long arm, the segmentation analysis revealed two distinct regions (Figure 3): the more proximal one had a high density of homologous genes (~0.84, 863 loci), so likely corresponds to a region of the 4AL telosome inherited from bread wheat (Figure 3). However, in the distal part of the arm, the homologous gene density fell to ~0.37, suggesting this as the site of the translocation event (Figure 3). Considering the same number of genes in the homologous regions of CS DT4AL chromosome arm (4AL-CS) and 4AL-TM, the comparison between these proximal segments revealed that 16% of homologous genes (167 of 1030) in the 4AL-TM assembly were not identified and may be accounted to the sequencing and assembly imperfection. If this rate of imperfection is applied to the regions including the introgressed segment (357 CS genes *vs* 131 *Tm* homologous genes), the presence of 169 CS nonhomologous genes in the introgression segment could be estimated. The number of such genes represents the size of linkage drag (neglecting allelic variation of the homologous genes).

## Discussion

Introgression from related species provides many opportunities to broaden the genetic base of wheat, but its impact on wheat improvement has been limited by a combination of imperfect homology between donor and recipient chromatin, the loss of key recipient genes, the suppression of recombination and linkage drag effect. Thus, obtaining an accurate understanding of the size, homology, orientation and position of an introgressed segment could help to determine which introgression events are more likely to avoid incurring a performance penalty. Such knowledge would also be informative in the context of isolating a valuable gene introduced via an introgression event. Gaining this information requires saturating the target region with molecular markers. In an effort to clone of *QPm-tut-4A* gene introgressed to the wheat 4A chromosome from *T. militinae,* we developed new method for chromosome rearrangements and introgressions identification and characterization.

The presence of ancient intra- and interchromosomal rearrangements is a known complicating issue in the polyploid wheat genome, and the 4AL chromosome arm, which is one of the site of the introgression event selected in line 8.1, has a particularly complex structure. The composition of the proximal segment of the 4AL telosomes carried by DT4AL-TM and the standard CS DT4AL stock was largely identical, as expected. However, distal part of the telosomes differs in presence of *Tm* introgressive segment (Jakobson *et al.*, 2012), but no difference by synteny blocks could be detected. In hybrids between the tetraploid forms *T. turgidum* and *T. timopheevi*, Gill and Chen (1987) noted that while the latter's G genome chromosomes paired most frequently with those from the B genome, chromosome 4A was occasionally involved in pairing with chromosome 7G, presumably as a result of the presence of the 7BS segment on the *T. turgidum* 4AL arm. The likelihood is therefore that the *Tm* chromosome 7G segment, which has contributed the 4A-based powdery mildew resistance of line 8.1, was introduced via homologous recombination with the segment of 4AL carrying 7BS chromatin.

To increase resolution of the analysis, the GenomeZipper method (Mayer *et al.*, 2011), combining genetic maps, data from chromosome shotgun sequencing, and synteny information with sequenced model genomes has been adopted. The method has been useful for developing virtual gene orders in both wheat and barley chromosomes (IWGSC, 2014; Mayer *et al.*, 2011). The most crucial data set is a reliable genetic map, which serves as backbone to integrate and orient the identified syntenic blocks. Two zippers for chromosome 4A have been published to date. The first was based on relatively low coverage sequencing of the chromosome, employing as its backbone a barley linkage map formed from expressed sequence tags distributed over the

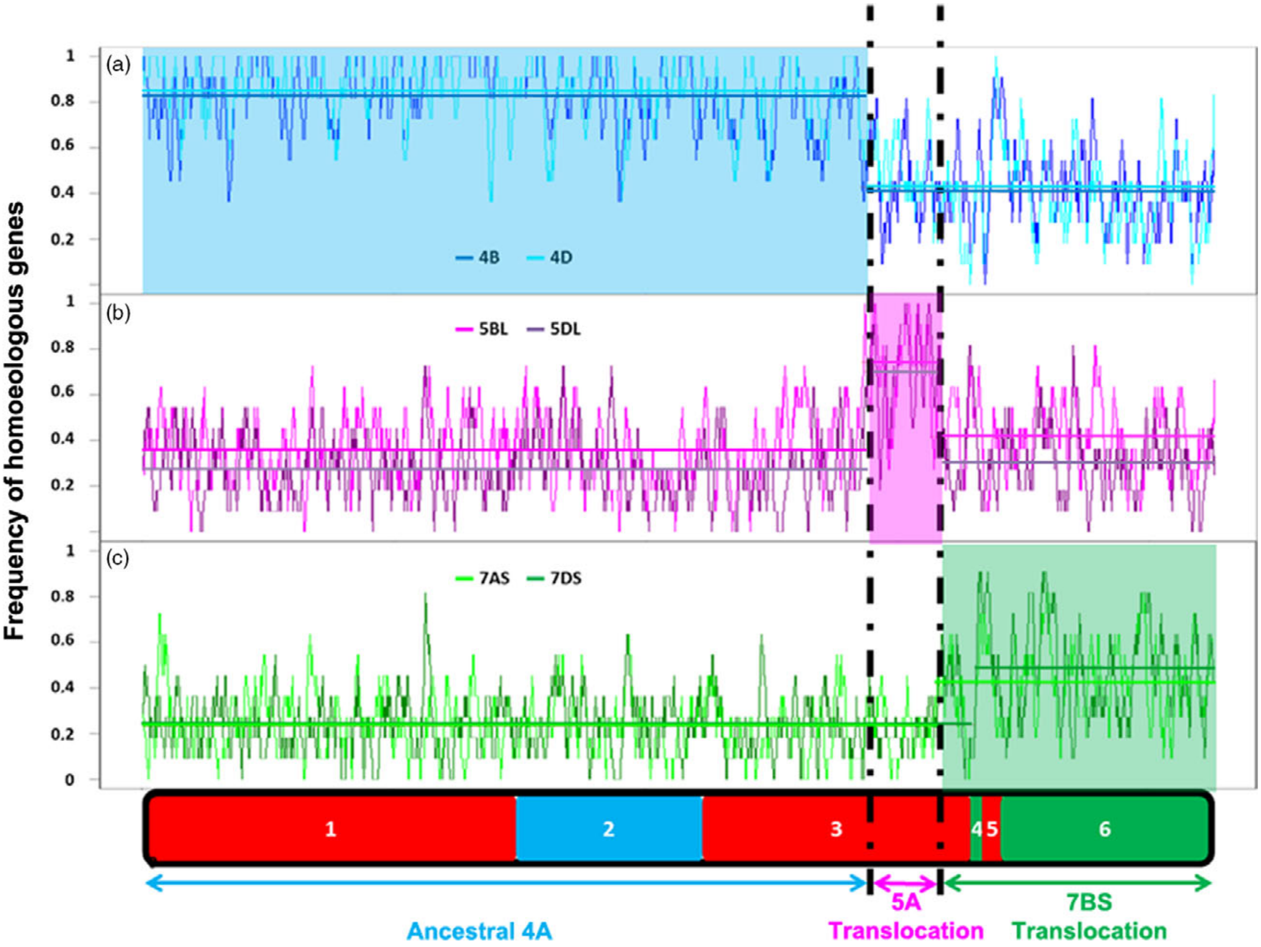

**Figure 2** Variation in homoeologous gene density along the various 4A-CS chromosome segments compared to their homoeologous chromosomes. The structure of native 4A-CS chromosome is represented at the bottom with syntenic blocks with rice genome shown in different colours (red = Os3; blue = OS11; green = Os6). (a) The 4A homoeologous gene density compared to 4B and 4D chromosomes, (b) comparison with the 5BL and 5DL chromosome arms and (c) comparison with chromosome arms 7AS and 7DS is shown as homoeologous genes frequency histogram. Homoeologous regions are characterized by a high average frequency (denoted by the horizontal lines). The lower average frequency shown by the group 7 chromosomes reflects a significantly lower sequencing coverage.

chromosome arms 4HS (117 loci), 4HL (16 loci), 5HL (36 loci) and 7HS (36 loci) (Hernandez *et al.*, 2012). The second was based on the 4A CSS and wheat SNP map and consisted of 167 markers on 4AS ordered into 56 mapping bins and 200 (92 mapping bins) markers on 4AL; these were combined with a linkage map developed from a mapping population bred from a cross between bread wheat cv. Opata and a synthetic wheat (Sorrells *et al.*, 2011). Neither of these two zippers was able to provide a sufficient level of resolution to identify the *Tm* introgression into 4AL chromosome arm. The present new zipper was based on consensus DArT map derived from crosses with CS and comprised 55% more markers and 25% more mapping bins than the latter one, which approximately doubled the number of ordered genes/loci (2132 *vs* 1004), and was informative with respect to the *Tm* introgression. When this improved zipper was used in conjunction with the RICh method, it was also possible to recognize the three evolutionary rearrangements, which have long been known to have generated the structure of the modern chromosome arm 4AL (Figure 2) (Devos *et al.*, 1995; Hernandez *et al.*, 2012; Ma *et al.*, 2013; Miftahudin *et al.*, 2004). Similarly, it was able to identify that a lower density of homologous genes obtained at the distal end of the 4AL-TM telosome (Figure 3) is representing the region harbouring the segment introgressed from *Tm*. The *Tm* introgression overlaps with almost the entire chromosome 7BS segment now present on 4AL (Figure 3, Table S2), while the proximal region of the 4AL-CS and 4AL-TM telosomes is essentially of bread wheat origin. The number of wheat loci retained in this latter region did, however, differ by 16% in gene content (4AL-CS—1030 and 4AL-TM—863 genes). This difference may be result of lower sequencing coverage of the 4AL-TM (30x compared to 116x of the 4AL-CS (IWGSC, 2014)) and thus lower representation of the 4AL-TM sequence assembly. If we assume the similar gene density in homologous chromosomes of relative species, as reported before by Tiwari *et al.* (2015), and if the same rate of missing genes as above due to sequencing and assembly imperfections is assumed, estimated 169 CS nonhomologous genes were carried by the introgression in linkage drag. Knowledgeable selection of parental lines that have relatively high frequency of homologous genes in the region of interest (e.g. QTL for resistance in the *Tm* introgression, Figure 3) may

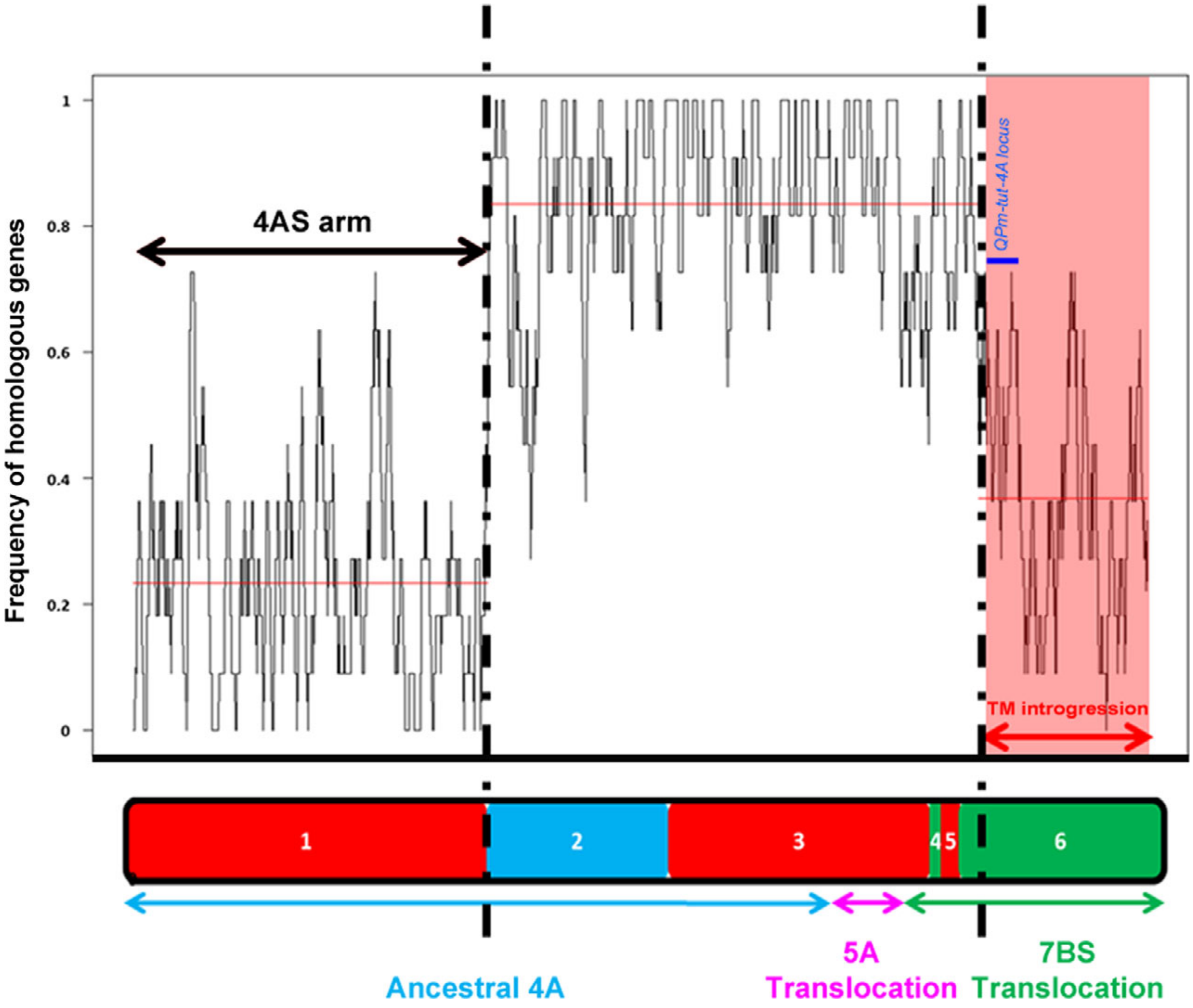

**Figure 3** Variation in homologous gene density between 4A-CS chromosome and 4AL-TM telosome. The structure of native 4A-CS chromosome is represented at the bottom with syntenic blocks with rice genome shown in different colours (red = Os3; blue = OS11; green = Os6). The homologous gene density along the 4A-CS zipper compare to 4AL-TM assembly is shown with the black line. The segment of the *Tm* introgression overlaps the 7BS translocation in 4AL (red highlight). The equivalent region on the 4AL-CS telosome harbours 357 genes, only 131 have homologous genes on the *Tm* segment. The dark blue bar represents approximate localization of the *QPm-tut-4A* locus.

increase chances of unobstructed recombination as was observed in the *QPm-tut-4A* locus (Jakobson *et al.*, 2012). So, reducing the length of the introgression segment by inducing further rounds of recombination can lessen (or even eliminate) any negative effects of linkage drag. Application of the RICh approach should prove informative regarding the order or frequency of homologous genes of any such selections. Overall, the RICh method offers a robust means of both characterizing chromosome rearrangements and of predicting the gene content of a specific chromosomal region. Recent advances in high-throughput genotyping permits the elaboration of ever higher density linkage maps (Bellucci *et al.*, 2015; Chapman *et al.*, 2015; Sorrells *et al.*, 2011; Wang *et al.*, 2014). The status of chromosome flow sorting is such that almost any wheat chromosome (Tsõmbalova *et al.*, 2016) and also chromosomes in many crops (Doležel *et al.*, 2014) can now be isolated to a reasonable purity, while the advances in NGS sequencing make RICh widely affordable. These developments should facilitate the preparation of materials needed for applying the RICh approach, thereby offering novel opportunities for a wide range of prebreeding activities, positional cloning, chromatin hybridization and structural studies.

## Experimental procedures

### Plant materials

Grains of the bread wheat ditelosomic CS DT4AL line were provided by Dr. Bikram Gill (KSU, Manhattan, KS), those of the two nullisomic–tetrasomic lines N4AT4B and N4AT4D (Sears and Sears, 1978) by the National BioResource Centre (Kyoto, Japan), those of *Tm* ($2n = 4x = 28$, genome formula $A^tA^tGG$) accession K-46007 by the N.I. Vavilov Institute of Plant Industry (St. Petersburg, Russia). The line denoted DT4AL-TM was generated from the cross CS DT4AL × 8.1: the line carries 40 bread wheat chromosomes and a pair of 4AL telosomes with the *Tm* introgression (4AL-TM) and is resistant to powdery mildew (Jakobson *et al.*, 2012).

### Flow sorting and amplification of the 4AL telosome carried by 4AL-TM

Liquid suspensions of mitotic chromosomes were prepared from root tips of 4AL-TM seedlings as described by (Vrána *et al.*, 2000). The telosomes were separated from the rest of the genome by flow sorting, using a FACSAria II SORP flow cytometer and sorter (BD Biosciences, San Jose, CA). The level of contamination within a sorted peak was determined using FISH, based on probes detecting telomeric repeats, the Afa repeat and $(GAA)_n$, following the methods described by Kubaláková *et al.* (2003). The flow-sorted 4AL-TM telosomes were treated with proteinase, after which DNA was extracted using a Millipore Microcon YM-100 column (www.millipore.-com). Chromosomal DNA was MDA amplified using the Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare) as described by Šimková *et al.* (2008).

### Identifying the origin of the introgression segment on the 4AL-TM telosome

Chromosomes of *T. militinae* were flow sorted as described above. However, prior to flow cytometry, GAA microsatellites on chromosomes were labelled by FITC using FISHIS protocol (Giorgi *et al.*, 2013). Bivariate analysis (DNA content/DAPI *vs* GAA/FITC) enabled discrimination of 13 of 14 chromosomes of *T. militinae*. Individual chromosome fractions were sorted into tubes for PCR amplification and onto microscopic slides for identification of sorted chromosomes by FISH. Three markers linked to the *Tm* powdery mildew resistance gene *QPm-tut-4A* were used for the selection of the critical cluster: these were the microsatellite *Xgwm160* (Roder *et al.*, 1998) and two unpublished, one (*owm72*) amplified by the primer pair 5′-TGCTTGCTTGTA GATTGTGCA/5′-CCAGTAAGCTTTGCCGTGTG) and the other (*owm82*) by 5′-GGGAGAGACGAAAGCAGGTA/5′-CTTGCATG CACGCCAGAATA. Each 20 μL PCR contained 0.01% (w/v) o-cresol sulphonephtalein, 1.5% (w/v) sucrose, 0.2 mM dNTP, 0.6 U Taq DNA polymerase and 1 μM of each primer in 10 mM Tris-HCl/50 mM KCl/1.5 mM $MgCl_2$/0.1% (v/v) Triton X-100. The template comprised about 500 sorted chromosomes. Test reactions were seeded with either 20 ng genomic DNA extracted from CS, *Tm*, N4AT4B or N4AT4D, or with 50 pg of MDA amplified DNA from 4AL-CS and 4AL-TM telosomes. The reactions were subjected to an initial denaturation (95 °C/5 min), followed by 40 cycles of 95 °C/30 s, 55 °C/30 s and 72 °C/30 s, and completed with an elongation of 72 °C/10 min. The products were electrophoretically separated through 4% nondenaturing polyacrylamide gels and visualized by EtBr staining. The markers were mapped using a $F_2$ population bred from the cross CS × 8.1 (Jakobson *et al.*, 2012).

### Sequencing of the 4AL telosome

A CSS assembly of CS chromosome arms 4AS (4AS-CS) and 4AL (4AL-CS) were acquired from Internation Wheat Genome Sequencing Consortium (IWGSC, 2014). Two sequencing libraries of DNA amplified from the 4AL-TM telosome were constructed using a Nextera kit (Illumina, San Diego, CA) with the insert size adjusted to 500 and 1000 bp. The resulting clones were sequenced as paired-end reads by IGA (Udine, Italy) using a HiSeq 2000 device (Illumina). The 4AL-TM reads were assembled with SOAPdenovo2 software, applying a range of k-mers (54–99, with a step size of 3) to select the assembly with the highest coverage and the largest N50. Assembled scaffolds (k-mer of 69, minimum length 200 bp) were chosen for further analysis (Table 1).

### DArTseq and SNP maps for GenomeZipper construction and validation

A DArTseq consensus map, based on four crosses involving cv. Chinese Spring as a parent has been provided by DArT PL (www.diversityarrays.com). Individual maps were created using DArT PL's OCD MAPPING program (Petroli *et al.*, 2012) to order DArTseq and array-based DArTs. DArT PL's consensus mapping software (Raman *et al.*, 2014) was applied to create a consensus map using similar strategy as described in Li *et al.* (2015). Version 3.0 of consensus map with approximately 70 000 markers was used in this study.

A SNP deletion map (Balcárková *et al.*, unpublished) was used for validation. Genomic DNAs of a set of 15 chromosome 4A deletion lines (Endo and Gill, 1996) and DNAs amplified from 4AL-CS and 4AS-CS chromosome arms as controls were genotyped at USDA-ARS (Fargo, ND) using a iSelect 90k SNP array (Wang *et al.*, 2014) on Infinium platform (Illumina). The raw genotypic data were manually analysed using GenomeStudio V2011.1 software (Illumina).

### Comparative analysis and GenomeZipper analysis

Synteny between related genomic segments was assessed using ChromoWIZ software (Nussbaumer *et al.*, 2014). The number of conserved genes present within a series of 0.5-Mbp genomic windows (window shift 0.1 Mbp) was determined. The consensus chromosome 4A linkage map used as the backbone for the GenomeZipper analysis comprised 1780 DArTseq markers (Table S1). As these sequences are mostly short (69 nt) and few identify coding sequence, they were first aligned to the set of 4A CSS contigs, preserving only those contigs that matched the entire DArTseq marker sequence at a level of at least 98% identity. The retained CSS contigs ('CSS-DArTseq markers') were used for the construction of the zipper, which was subsequently validated against the SNP deletion map (2706 SNPs). Similarly as above, only those 4A CSS contigs that aligned with SNP loci along their entire length (98% identity threshold) were retained. Ordering of the CSS-DArTseq markers was compared with that ordered by SNPs from the deletion bin map and CSS-DArTseq markers which do not follow the SNP order were eliminated, and a second version of the zipper was generated using the remaining markers (Table S3). This version was revalidated against the SNP deletion map.

### The RICh approach

To identify introgressed/translocated regions, the final 4A zipper was compared to the complete set of CSS sequences obtained from chromosome arms 4BS, 4BL, 4DS, 4DL, 5BL, 5DL, 7AS and 7DS (IWGSC, 2014). Alignments were performed using the BLAST algorithm (Altschul *et al.*, 1990). The BLAST outputs were filtered by applying the following criteria: a minimum identity of either 90% (translocation analysis) or 95% (introgression analysis) and a minimum alignment length of 100 bp. For each comparison, the density of homologous genes was evaluated using a sliding window of eleven genes (five upstream and five downstream), and a segmentation analysis was performed using the R package changepoint v1.1 (Killick and Eckley, 2014), applying the parameter segment neighbourhoods method with a BIC penalty on the mean change. The method allows a statistical detection of gene density changes along the chromosome, corresponding to an increase or decrease in the level of synteny. For translocation events, an increase in synteny level with one group of homoeologs is required, while for an introgression, a loss of orthology is anticipated.

## Acknowledgements

We thank Z. Dubská and R. Šperková for their technical assistance with flow sorting and M. Kubaláková for her FISH-based identification of sorted chromosomes. L. Pingault and E. Paux are acknowledged for their contribution to the statistical analyses, S. Chao for the deletion lines genotyping and J. Bartoš for his critical reading of the manuscript and for his helpful suggestions. This research was supported by grants awarded by the National Program of Sustainability I (No. LO1204), by the Czech Science Foundation (No. 14-07164S) and by the Estonian Ministry of Agriculture.

## Conflict of interests

Dr. A Kilian is head of Diversity Arrays Technology Pty Ltd.

## References

Aflitos, S.A., Sanchez-Perez, G., de Ridder, D., Fransz, P., Schranz, M.E., de Jong, H. and Peters, S.A. (2015) Introgression browser: high-throughput whole-genome SNP visualization. *Plant J.* **82**, 174–182.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Badaeva, E.D., Budashkina, E.B., Bilinskaya, E.N. and Pukhalskiy, V.A. (2010) Intergenomic chromosome substitutions in wheat interspecific hybrids and their use in the development of a genetic nomenclature of Triticum timopheevii chromosomes. *Russian Journal of Genetics*, **46**, 769–785.

Bellucci, A., Torp, A.M., Bruun, S., Magid, J., Andersen, S.B. and Rasmussen, S.K. (2015) Association mapping in Scandinavian winter wheat for yield, plant height, and traits important for second-generation bioethanol production. *Front. Plant Sci.* **6**, 1046.

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

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

Devos, K.M., Dubcovsky, J., Dvořák, J., Chinoy, C.N. and Gale, M.D. (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theor. Appl. Genet.* **91**, 282–288.

Doležel, J., Vrána, J., Cápal, P., Kubaláková, M., Burešová, V. and Simková, H. (2014) Advances in plant chromosome genomics. *Biotechnol. Adv.* **32**, 122–136.

Endo, T. and Gill, B. (1996) The deletion stocks of common wheat. *J. Hered.* **87**, 295–307.

Feuillet, C., Langridge, P. and Waugh, R. (2008) Cereal breeding takes a walk on the wild side. *Trends Genet.* **24**, 24–32.

Gill, B.S. and Chen, P.D. (1987) Role of cytoplasm-specific introgression in the evolution of the polyploid wheats. *Proc. Natl Acad. Sci. USA*, **84**, 6800–6804.

Gill, B.S., Friebe, B.R. and White, F.F. (2011) Alien introgressions represent a rich source of genes for crop improvement. *Proc. Natl Acad. Sci. USA*, **108**, 7657–7658.

Giorgi, D., Farina, A., Grosso, V., Gennaro, A., Ceoloni, C. and Lucretti, S. (2013) FISHIS: fluorescence in situ hybridization in suspension and chromosome flow sorting made easy. *PLoS ONE*, **8**, e57994.

Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Gálvez, S., Schaaf, S., Jouve, N. *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.

IRGSP. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

IWGSC. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, **345**, 1251788.

Jakobson, I., Peusha, H., Timofejeva, L. and Järve, K. (2006) Adult plant and seedling resistance to powdery mildew in a *Triticum aestivum* x *Triticum militinae* hybrid line. *Theor. Appl. Genet.* **112**, 760–769.

Jakobson, I., Reis, D., Tiidema, A., Peusha, H., Timofejeva, L., Valárik, M., Kladivová, M. *et al.* (2012) Fine mapping, phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat-*Triticum militinae* introgression line. *Theor. Appl. Genet.* **125**, 609–623.

Jiang, J., Friebe, B. and Gill, B. (1993) Recent advances in alien gene transfer in wheat. *Euphytica*, **73**, 199–212.

Killick, R. and Eckley, I.A. (2014) changepoint: an R package for changepoint analysis. *J. Stat. Softw.* **58**, 1–19.

Kubaláková, M., Valárik, M., Bartoš, J., Vrána, J., Číhalíková, J., Molnár-Láng, M. and Doležel, J. (2003) Analysis and sorting of rye (*Secale cereale* L.) chromosomes using flow cytometry. *Genome*, **46**, 893–905.

Li, H., Vikram, P., Singh, R., Kilian, A., Carling, J., Song, J., Burgueno-Ferreira, J. *et al.* (2015) A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genom.* **16**, 216.

Ma, J., Stiller, J., Berkman, P.J., Wei, Y., Rogers, J., Feuillet, C., Doležel, J. *et al.* (2013) Sequence-based analysis of translocations and inversions in bread wheat (*Triticum aestivum* L.). *PLoS ONE*, **8**, e79329.

Mayer, K., Martis, M., Hedley, P., Simkova, H., Liu, H. and Morris, J. (2011) Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell*, **23**, 1249–1263.

Miftahudin, Ross, K., Ma, X.-F., Mahmoud, A.A., Layton, J., Milla, M.A.R., Chikmawati, T. *et al.* (2004) Analysis of expressed sequence tag loci on wheat chromosome group 4. *Genetics*, **168**, 651–663.

Niu, Z., Klindworth, D.L., Friesen, T.L., Chao, S., Jin, Y., Cai, X. and Xu, S.S. (2011) Targeted introgression of a wheat stem rust resistance gene by DNA marker-assisted chromosome engineering. *Genetics*, **187**, 1011–1021.

Nussbaumer, T., Kugler, K.G., Schweiger, W., Bader, K.C., Gundlach, H., Spannagl, M., Poursarebani, N. *et al.* (2014) chromoWIZ: a web tool to query and visualize chromosome-anchored genes from cereal and model genomes. *BMC Plant Biol.* **14**, 348.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G. *et al.* (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature*, **457**, 551–556.

Petroli, C., Sansaloni, C., Carling, J., Steane, D., Vaillancourt, R. and Myburg, A. (2012) Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the eucalyptus genome. *PLoS ONE*, **7**, e44684.

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

Raman, H., Raman, R., Kilian, A., Detering, F., Carling, J. and Coombes, N. (2014) Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PLoS ONE*, **9**, e101673.

Rey, E., Molnár, I. and Doležel, J. (2015) Genomics of wild relatives and alien introgressions. In *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics* (Molnár-Láng, M., Ceoloni, C. and Doležel, J., eds), pp. 347–381. Cham: Springer International Publishing.

Roder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.H., Leroy, P. and Ganal, M.W. (1998) A microsatellite map of wheat. *Genetics*, **149**, 2007–2023.

Šafář, J., Šimková, H., Kubaláková, M., Číhalíková, J., Suchánková, P., Bartoš, J. and Doležel, J. (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet Genome Res.* **129**, 211–223.

Sears, E.R. and Sears, L.M.S. (1978) The telocentric chromosomes of common wheat. In *Proc 5th Int Wheat Genet Symp* (Ramanujam, S., ed), pp. 389–407. New Delhi: Indian Soc of Genet Plant Breed.

Šimková, H., Svensson, J.T., Condamine, P., Hřibová, E., Suchánková, P., Bhat, P.R., Bartoš, J. *et al.* (2008) Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. *BMC Genom.* **9**, 294.

Sorrells, M.E., Gustafson, J.P., Somers, D., Chao, S.M., Benscher, D., Guedira-Brown, G., Huttner, E. *et al.* (2011) Reconstruction of the synthetic W7984 x Opata M85 wheat reference population. *Genome*, **54**, 875–882.

Tiwari, V.K., Wang, S., Danilova, T., Koo, D.H., Vrana, J., Kubalakova, M., Hribova, E. *et al.* (2015) Exploring the tertiary gene pool of bread wheat: sequence assembly and analysis of chromosome 5M(g) of *Aegilops geniculata*. *Plant J.* **84**, 733–746.

Tsõmbalova, J., Karafiátová, M., Vrána, J., Kubaláková, M., Peuša, H., Jakobson, I., Järve, M. *et al.* (2016) A haplotype specific to North European wheat (*Triticum aestivum* L.). *Genet. Resour. Crop Evol.* DOI 10.1007/s10722-016-0389-9

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K. *et al.* (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.

Vrána, J., Kubaláková, M., Simková, H., Číhalíková, J., Lysák, M.A. and Doležel, J. (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). *Genetics*, **156**, 2033–2041.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M. *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.

Wulff, B.B.H. and Moscou, M.J. (2014) Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Front. Plant Sci.* **5**, 692. doi: 10.3389/fpls.2014.00692

Zamir, D. (2001) Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* **2**, 983–989.

## Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

**Figure S1** The flow karyotype of DT4AL-TM, a bread wheat line ditelosomic for 4AL, the distal portion of which includes a segment translocated from *T. militinae*.

**Figure S2** A comparative analysis of the telosomes 4AL-CS and 4AL-TM with the *B. distachyon*, rice and sorghum genomes.

**Figure S3** Refining the robustness of the 4A zipper.

**Table S1** Consensus 4A DArTseq map, based on four independent populations, each involving CS as one parent.

**Table S2** The new 4A zipper, composed of 2132 loci, constructed using the CS-based 4A specific DArTseq map and validated by reference to SNPs mapped using a panel of deletion lines.

**Table S3** The set of CSS-DArTseq markers used to construct the new zipper.