Title

Some characteristics of crossing over in induced recombination between chromosomes of wheat and rye

Journal

The Plant Journal, 105(6)

ISSN

0960-7412

Authors

Fan, Chaolan
Hao, Ming
Jia, Zhenyu
et al.

Publication Date

2021-03-01

DOI

10.1111/tpj.15140

Peer reviewed

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2021) **105,** 1665–1676 doi: 10.1111/tpj.15140

Some characteristics of crossing over in induced recombination between chromosomes of wheat and rye

Chaolan Fan[1,2], Ming Hao[1], Zhenyu Jia[2], Christian Neri[2], Xue Chen[1], Wenshuai Chen[1], Dengcai Liu[1,3] and Adam J. Lukaszewski[2,*]

[1]*Triticeae Research Institute, Sichuan Agricultural University, Chengdu, Sichuan, China,*

[2]*Department of Botany and Plant Sciences, University of California, Riverside, CA, USA, and*

[3]*State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu, Sichuan, China*

Received 19 October 2020; revised 10 December 2020; accepted 15 December 2020; published online 21 December 2020.
*For correspondence (e-mail adam.lukaszewski@ucr.edu).

SUMMARY

Allopolyploid wheat (*Triticum aestivum* L.) carries three pairs of homoeologous genomes but its meiotic pairing is diploid-like. This is the effect of the *Ph* (pairing homoeologous) system which restricts chromosome pairing to strictly homologous. *Ph1* is the locus with the strongest effect. Disabling *Ph1* permits pairing between homoeologues and is routinely used in chromosome engineering to introgress alien variation into breeding stocks. Whereas the efficiency of *Ph1* and the general pattern of homoeologous crossovers in its absence are quite well known from numerous studies, other characteristics of such crossovers remain unknown. This study analyzed the crossover points in four sets of the *ph1b*-induced recombinants between wheat homologues as well as between three wheat and rye (*Secale cereale*) homoeologous chromosome arms, and compared them to crossovers between homologues in a reference wheat population. The results show the *Ph1* locus also controls crossing over of homologues, and the general patterns of homologous (with *Ph1*) and homoeologous (with *ph1b*) crossing over are the same. In all intervals analyzed, homoeologous crossovers fell within the range of frequency distribution of homologous crossovers among individual families of the reference population. No specific DNA sequence characteristics were identified that could be recognized by the *Ph1* locus; the only difference between homologous and homoeologous crossing over appears to be in frequency. It is concluded that the *Ph1* locus likely recognizes DNA sequence similarity; crossing over is permitted between very similar sequences. In the absence of *Ph1* dissimilarities are ignored, in proportion to the level of the sequence divergence.

Keywords: crossing over, homology, homoeology, pairing control, *Ph1* locus.

INTRODUCTION

Polyploids face a serious challenge when it comes to meiotic chromosome pairing. Only bivalent pairing ensures regular reduction of the chromosome number, and hence efficient production of gametes. In polyploids, the presence of more than two set of chromosomes offers a chance to pair into configurations greater than two, and hence irregular disjunction in anaphase I. In autopolyploids, little can be done about the problem, short of chromosome differentiation, and this is a very slow process (Doyle, 1986). Allopolyploids, on the other hand, are composed of related (homoeologous) genomes and have evolved genetic systems which restrict pairing to strictly homologous, hence guaranteeing diploid-like behavior in the first meiotic division (Jenczewski and Alix, 2004). The best known of these systems, even if not fully understood, is the *Ph* (pairing homoeologous) in polyploid wheats (*Triticum* sp.) (Sears and Okamoto, 1958; Riley and Chapman, 1958).

The *Ph* system in wheat consists of several genes; two major loci are *Ph1* on the long arm of chromosome 5B and *Ph2* on the short arm of chromosome 3D; there are also several other small-effect loci on other chromosomes (Okamoto, 1957; Riley and Chapman, 1958; Mello-Sampayo, 1971; Sears, 1976; Sutton *et al.*, 2003). While the actual mechanism of their operation still remains unclear, when present, they enforce strictly homologous pairing, metaphase I (MI) of meiosis shows only bivalents, and the reduction of the chromosome number in anaphase I is orderly. When absent, homoeologous pairing takes place, with univalents and multivalents in MI, consequent uneven

segregation of chromosomes, and hence aneuploidy, as well as chromatin exchange among the genomes (Sánchez-Moran *et al.*, 2001). Of the two major genes, *Ph1*, has a much stronger effect than *Ph2*; it is this locus that is manipulated/disabled whenever homoeologous crossing over (CO) is desired.

Ph1 is capable of recognizing even subtle differences between chromosomes and may restrict pairing of homologues (Dvorak and McGuire, 1981). Apparently, when *Ph1* is disabled, the stringency of CO is relaxed and divergent homologues as well as homoeologues, even when originating from different genera, can pair and cross over. In chromosome engineering exercises in wheat a mutation of *Ph1*, *ph1b* (in fact it is a deletion of a segment of chromosome 5BL with the locus, Gill *et al.*, 1993), is commonly used but a substitution of chromosome 5B, either by 5D or by some other group-5 chromosome, is as effective. These have been used to induce pairing of wheat chromosomes with their homoeologues from related species, for example, *Hordeum vulgare*, *S. cereale*, *Haynaldia villosa, Agropyron elongatum,* various *Aegilops* species, and others (Sears, 1981; Lukaszewski, 2000; Zhao *et al.*, 2013; Rey *et al.*, 2015; Lukaszewski and Cowger, 2017; Zhang *et al.*, 2018; Danilova *et al.*, 2019). The frequencies of such *ph1b*-induced homoeologous CO vary widely. For chromosomes 4, 5, and 7 of *Hordeum chilense* they ranged from 3.3% to 7.5% (Rey *et al.*, 2015); for the chromosome arms 2RS and 2RL of *S. cereale* matched with chromosome 2B of wheat they were 0.6% and 16.3% (Lukaszewski *et al.*, 2004). Arms 2RS and 2BS are structurally different (not fully syntenic); arms 2BL and 2RL are fully syntenic (Devos *et al.*, 1993).

COs induced by the absence of *Ph1* in wheat pose several interesting questions. Perhaps the most important one is: is there a qualitative difference between homo- and homoeologous COs? If such a difference exists, what are the general DNA landscape features that are recognized by the locus? Apart from the CO frequency (considerably lower between homoeologues than homologues), are the same DNA sequences involved (such as 'hot spots'), or does the absence of *Ph1* permit COs in new/different regions of chromosomes? Is homologous CO in the absence of *Ph1* equivalent to that in the presence of *Ph1*? So far it is only clear that apart from a virtual absence of double COs in a chromosome arm, the general patterns of COs along the chromosome length in *Ph1* and *ph1b* appear to be similar (Lukaszewski, 1995; Lukaszewski *et al.*, 2004; Zhang *et al.*, 2018).

In this study, we examine the effect of *Ph1* on homologous CO frequencies in a defined segment of wheat chromosome 1B, and compare genetic maps created by homologous CO in three wheat chromosome arms with *Ph1* present, with genetic maps of the same chromosome arms recombined with their homoeologues from rye (*S. cereale*), in the *ph1b* background. While the results still do not offer a clear answer as to the nature of the *Ph* system in wheat, they do add some new elements to its characteristics.

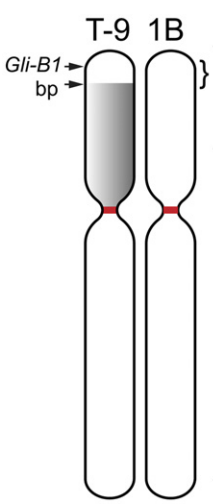
RESULTS

The effect of *Ph1* on homologous crossing over

A total of 22 758 progeny chromosomes were scored for the CO frequency in the terminal approximately 9.6–9.9-Mb wheat segment of the T-9 wheat–rye translocation against corresponding segments in chromosomes 1B from four sources (Figure 1). Of these, 13 677 were in the *Ph1* background (normal pairing control level) and 8081 in the *ph1b* background. With *Ph1* present, the CO rates for the targeted wheat segment in individual chromosomes ranged from 5.2% to 10.9% (average 8.6%) while in *ph1b* they ranged from 18.9% to 25.8% (average 20.0%). In each case, the absence of *Ph1* significantly increased the CO frequency, from 173% for chromosome $1B_{HE}$ to 396% for $1B_{LC}$, with an average increase of 232% (Figure 1). The observed CO rates in the absence of *Ph1* were within the range of frequency expected for the CO rate of identical segments, inferred from the 43% metaphase I pairing frequency of T-9 with 1BS of Pavon.

Genetic mapping with homoeologous and homologous crossing over

All recombinants of homoeologous arms (1BS-1RS, 2BS-2RS, and 2BL-2RL) generated in the absence of *Ph1* were



Pop.	Crossover rate	
	Ph1	*ph1b*
$1B_{HE}$	10.9%(3046)	18.9%(2551)**
$1B_{BE}$	8.0%(5047)	20.1%(2670)**
$1B_{\#55}$	8.3%(4974)	25.8%(526)**
$1B_{LC}$	5.2%(610)	20.6%(2334)**
Total	8.6%(13677)	20.0%(8081)**

Figure 1. The structure of translocation T-9 and the crossover rates of its terminal wheat segment with homologues 1B from four wheats, in *Ph1* and *ph1b*. The shaded area on the short arm of T-9 is occupied by a rye segment; the non-shaded area is from 1BS; it contains the *Gli-B1* locus. The wheat–rye distal translocation breakpoint is located between positions 9.56 and 9.96 Mb from the telomere. $1B_{HE}$, $1B_{BE}$, $1B_{LC}$, and $1B_{\#55}$ stand for chromosomes 1B from cvs. Henika (HE), Begra (BE), Little Club (LC), and an Iranian landrace #55 (#55), respectively. **χ^2 test, $P < 0.01$; bp, breakpoint.

selected cytologically by a change of phase between telomeres and centromeres, and most of them were later confirmed as single breakpoint translocations. They were selected from very large populations, totaling approximately 35 000 individuals. Standard calculations of genetic distances would produce genetic maps of a fraction of a centiMorgan (cM) in length. To avoid this problem, as in the two previous studies (Lukaszewski, 2000; Lukaszewski *et al.*, 2004), relative genetic map distances were used here, with the total map length for each arm fixed at 50 cM, as expected for single COs.

For 1BS-1RS Recs, 128 recombinants were genotyped by 31 914 DArT-seq markers. Homozygous recombinants were processed first; by comparison of Pavon 76, Pavon 1RS.1BL, and Pavon Dt1BL, 1083 markers were identified as specific to 1BS and 999 markers as specific to 1RS. After filtering, 724 1BS and 799 1RS high-quality markers were retained to generate genetic maps. Recombinant T-16 had a high proportion of missing data (32%) and was excluded. A genetic map was generated by MSTMap with 1507 markers sorted into 57 groups (Table S1); 16 markers (one for 1BS and 15 for 1RS) did not link to any other markers and could not be placed on the map. As expected, most recombinant chromosomes indeed showed single translocation breakpoints. However, chromosomes 1B+33, 1B+40, 1B+47, T-23, and T-75 showed multiple breakpoints. Given the known structure of these chromosomes multiple exchanges were unreasonable, and as it was impossible to determine which of these multiple exchange points suggested by markers were real, these six chromosomes were excluded from further analyses.

For 2BS-2RS Recs, 25 recombinants were genotyped. Chromosomes 2BS+12 and 2RS-14 were excluded because DNA markers identified no breakpoints. The 524 markers polymorphic between Pavon and the Pavon 2RS.2BL translocation line sorted themselves into 22 bins (Table S1). With the assumed 50 cM total relative map length, the average bin length was 2.27 cM; the average marker density was 0.09 cM/marker.

For 2BL-2RL Recs, 88 recombinants were genotyped by 9208 DArT markers, 2061 of which were polymorphic between Pavon and Pavon 2BS.2RL. Of these, 605 (257 specific to 2BL and 348 specific to 2RL) were used to generate a genetic map consisting of 41 bins (Table S1). Marker density based on relative genetic distances was 0.08 cM/marker. Nine recombinants (2BL+12, 2BL+33, 2BL+44, 2BL+52, 2RL-2, 2RL-18, 2RL-21, 2RL-38, and 2RL-40) showed no breakpoints and were not included in the map. Another five recombinants (2BL+18, 2RL-3, 2RL-9, 2RL-14, and 2RL-35) were excluded because they appeared to show multiple breakpoints. Recombinant 2RL-12 appeared to be inverted at the end of the arm generating what appeared as multiple breakpoints, and it was also left out. The remaining 72 COs were classified into 40 genetic intervals.

Individual COs were allocated to intervals between two adjacent informative markers. There were 121 COs for 1BS-1RS, 23 for 2BS-2RS, and 72 for 2BL-2RL, located in 56, 21, and 40 genetic intervals, respectively (Table 1). The average number of COs per interval for the three pairs of arms was 2.12, 1.10, and 1.80, respectively. Some recombinant chromosomes 1BS-1RS and 2BL-2RL had up to 10 and 6 COs in a single interval, respectively, while in 2BS-2RS only one interval had two COs (Table S1). This is likely a consequence of a low number of mapped CO points in 2BS-2RS and their wide distribution along the chromosome arm.

Homologous COs in corresponding wheat arms (1BS-1BS, 2BS-2BS, and 2BL-2BL) were identified in the reference nested association mapping (NAM) population (Table S2). A summary of the numbers of recombinants detected for each homologous arm is presented in Table 2. Generally, a third of the total number of members in the population for each arm had no detectable CO events. Arms with single COs were the most frequent (39.4% for 1BS-1BS; 39.0% for 2BS-2BS; 34.5% for 2BL-2BL) followed by arms with double COs and multiple COs. Chromosome arm 1BS had a higher ratio of single COs than the remaining two arms. Again, frequent multiple COs in a chromosome arm in wheat appear unrealistic but for lack of a verification method they were retained in this analysis.

The average number of COs per line for 1BS arms in the NAM population was 0.84, ranging from 0.52 (NAM29, 38 COs/73 lines) to 1.28 (NAM27, 95 COs/74 lines). For 2BS and 2BL, the mean number of COs per line was 1.32 (ranging from 0.75 to 1.63) and 1.21 (ranging from 0.69 to 1.75), respectively. The total numbers of intervals across all analyzed NAM families were 348 for 1BS-1BS, 944 for 2BS-2BS, and 920 for 2BL-2BL, while the average number of COs per interval was 3.23, 2.61, and 2.44, respectively (Table 2, Table S3). The maximum numbers of COs per interval for 1BS, 2BS, and 2BL were 26, 24, and 19.

Physical mapping of homologous and homoeologous crossover intervals

The genetic maps generated here, both by homologous and homoeologous CO, were aligned with the DNA assemblies for the three chromosome arms using the DNA sequence marker information. The lengths of the intervals in which COs occurred were further narrowed down by determining physical locations of redundant markers on DNA assemblies, that is, markers completely linked to the skeleton markers. However, at times the genetic interval of a CO could not be mapped on a DNA assembly because of an absence or evidently incorrect physical location of the skeleton and redundant markers. In such (rare) cases, physical intervals of such COs were incorporated into

Table 1 A summary of the crossover intervals in the populations of homoeologous and homologous recombinants

		Physical intervals*				
		Wheat as reference		Rye as reference		
Recombinants	No. of intervals	Intervals	Mean length	Intervals	Mean length	*P* value*
1BS-1RS	56	37	1.43 Mb (65 bp–14.70 Mb)	43	0.93 Mb (244 bp–9.05 Mb)	0.68
2BS-2RS	21	21	1.63 Mb (42.274 kb–5.27 Mb)	19	1.92 Mb (7.20 kb–7.78 Mb)	0.91
2BL-2RL	40	35	1.44 Mb (18.692 kb–9.36 Mb)	26	2.27 Mb (0.359 kb–7.19 Mb)	0.74
1BS-1BS	348	280	9.06 Mb (2.33 kb–108.30 Mb)	–	–	–
2BS-2BS	944	815	6.20 Mb (46 bp–92.00 Mb)	–	–	–
2BL-2BL	920	823	7.99 Mb (0.63 kb–101.70 Mb)	–	–	–

*Paired sample *t*-tests were done for the lengths of interval sets with wheat and rye sequence assemblies used as references. Adjacent physical intervals that contained different genetic intervals were merged to create a one-to-one match between the wheat and rye genomes.

Table 2 Distribution of crossovers (COs) per chromosome arm in the NAM population

Chromosome arm	Number of families	Number of individuals	Non-CO* individuals	Single-CO individuals	Double-CO individuals	Multiple-CO individuals	Total Recs
1BS	19	1343	561	529	184	69	785
2BS	27	1865	457	728	442	238	1408
2BL	27	1861	590	642	423	206	1271

*CO = crossover

intervals between the two adjacent COs. As a consequence, the number of physical CO intervals can be different for the two reference assemblies while the total number of COs remains the same (Table 1). The numbers of COs in each genetic and physical interval of the homoeologous recombinants are listed in Table S4. For illustration purposes, the middle position of an interval harboring a CO was assumed as that CO's location.

The characteristics of individual homoeologous CO intervals can be expressed in terms of either the wheat or rye reference DNA sequence assemblies. The lengths of the intervals ranged from a kb to an Mb in length and in most cases, there was a good agreement between the results obtained using the wheat and rye reference assemblies in term of both the order and the interval length (Figure 2; Table 1). However, there were several exceptions: the shortest identified CO interval in the rye assembly (0.24 kb) among 1BS-1RS Recs, with 10 COs, stretched over 7 Mb along the wheat reference assembly. On the other hand, the shortest CO interval (65 bp) of 1BS-1RS Recs on the wheat reference assembly, with one CO, merged with three other COs in the wheat reference assembly and covered 583 kb along the rye reference assembly. Due to inherent ambiguity in the DNA sequence assemblies of both species, the shortest physical intervals harboring COs in homoeologous recombination could not be precisely established.

Physical locations of the homologous CO intervals in each NAM family were determined using the wheat DNA sequence assembly Ref 2.0 as the physical reference (Table S5). For the chromosome arm 1BS, the average interval length was 9062 kb, ranging from 2.33 kb (NAM13) to 108.30 Mb (NAM17). For 2BS, the average interval length was 6197 kb, while the shortest interval was 46 bp (NAM27) located around the 25 Mb point and the longest was 92 004 kb (NAM19) covering from 220 Mb to 312 Mb. For 2BL, the shortest interval was 634 bp (NAM24) in the vicinity of position 550 Mb while the longest was 101.7 Mb (NAM18), from 456 Mb to 558 Mb, with an average length of 7990 kb. The longest intervals with detected COs were in the pericentromeric regions with lower marker coverage and infrequent COs.

To plot the general distribution patterns of COs, chromosome arms were divided into 5 Mb intervals and proportions of COs in any given interval in the total number of COs in the arm were calculated (Figure 3). In homologous recombination in 1BS and 2BL, essentially all COs (96%), occurred in the distal halves of the arms, with approximately 70% of all COs concentrated in the regions stretching from 5 Mb to 50 Mb of the arms' assemblies from the

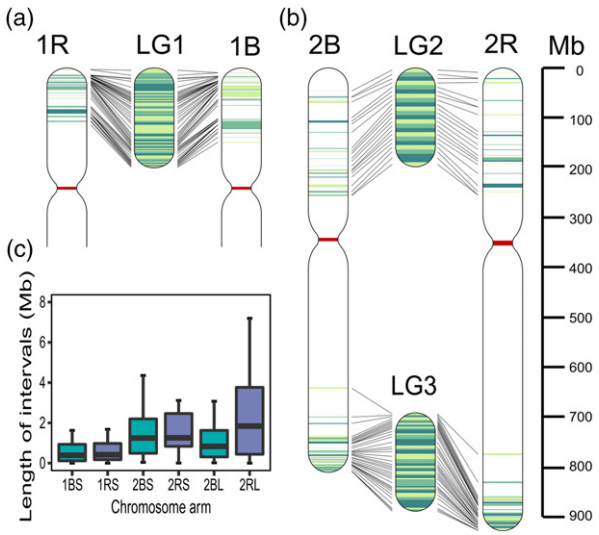


Figure 2. A comparison of genetic intervals with corresponding physical intervals in homoeologous recombination 1BS-1RS, 2BS-2RS, and 2BL-2RL, using wheat (1B, 2B) and rye (1R, 2R) sequence assemblies as references. Linkage groups (LGs) 1, 2, and 3 represent genetic maps of the three pairs of homoeologous arms constructed using relative genetic lengths. (a) 1BS-1RS. (b) 2R-2B. (c) Boxplot of the lengths of physical intervals using wheat and rye assemblies.

telomere. Within this general pattern, there were peaks of higher CO frequencies in individual intervals, and the distribution of such peaks was different in different NAM families (Figure S1). In homoeologous recombination of 1BS with 1RS and 2BL with 2RL the distribution pattern of COs was very similar to that of homologous CO, with some deviations in individual intervals. However, in no case the frequency distribution peaks of homoeologous COs exceed those observed in individual families of NAM (Figure 3).

Chromosome arm 2BS showed a somewhat different pattern of CO distribution: while distal concentration of homologous COs was evident (approximately 50% of all COs occurred within the terminal 50 Mb of the 248 Mb arm) COs occurred along most of the chromosome arm and much closer to the centromere than in the other two arms. In the 2BS-2RS recombination, there was no CO in the terminal 46 Mb of the 2BS arm and all detected COs were distributed more or less evenly along the rest of the arm.

Collinearity of the three pairs of homoeologous chromosome arms

Generally, there was a good fit of DNA sequences between homoeologous chromosome arms, with few notable exceptions (Figure 4). For 1BS versus 1RS, in a 180-Mb portion of the sequence assembly analyzed in detail, the level of collinearity varied. There appears to be an inversion of approximately 10 Mb close to the telomere. A 70-Mb segment immediately adjacent to the inversion showed very high collinearity between 1B and 1R, and this correlated with high CO frequencies. The rest of the two arms (proximal approximately 170 Mb) showed lower levels of collinearity, and low levels of CO, with the exception of a segment between 105 Mb and 120 Mb which was highly collinear, appeared to be inverted, and had a high CO rate.

In the 300-Mb terminal portion of the 2BS-2RS arms analyzed in detail, the arms were highly collinear, with the exception of the terminal approximately 50 Mb with no collinearity at all. This reflects the presence of a 6S segment on rye chromosome 2R (Devos *et al.*, 1993). Perhaps because of high collinearity along the rest of the arms, COs were distributed quite evenly. In the terminal 300 Mb of the 2L arms, some structural variation appeared to be present, with four possible inversions at around approximately 620, approximately 690, approximately 730, and approximately 815 Mb, respectively (Figure 4). The terminal 100 Mb of the 2BL-2RL arms were highly collinear and this is where a majority of COs were located. However, it needs to be pointed out that the most proximal 100 Mb of the analyzed segment were also highly collinear and no COs were formed there.

Characteristics of the crossover intervals

The general features of the CO intervals shorter than 50 kb in length were referenced using the wheat or rye genome assemblies (Table S4). All examined CO intervals for the three arms where located in non-annotated (and presumably non-coding) regions in the wheat genome assembly. In the rye genome assembly almost all these intervals were associated with gene regions, either inside an annotated sequence or containing such a sequence, with two exceptions, one in 1BS-1RS Recs and one in 2BL-2RL Recs, which did not involve annotated sequences. Only two short CO intervals, interval 26/27 of 244 bp in length in 1BS-1RS Recs and interval 10 of 382 bp in 2BL-2RL Recs, were within rye annotated genes, SECCE1Rv1G0006160 and SECCE2Rv1G0131880, respectively, of which neither one has introns (Rabanus-Wallace *et al.*, 2019). More than one highly similar sequence was found for each of these two genes on the corresponding wheat homoeologous arms (1BS, 2BL) using the two interval sequences as queries (Figure 5). For interval 26/27 of 1BS-1RS, two homologous sequences were present, one located within the corresponding 1BS physical interval and unannotated, while the other one was close to the first one and located within an annotated gene TraesCS02G1B02G053700 (Figure 5a). For interval 10 of 2BL-2RL, three homologous sequences were located in the wheat 2BL sequence assembly. One of these sequences was within the corresponding 2BL physical interval and unannotated; the other two were outside the interval, with one of them close to the first one and the other within an annotated gene

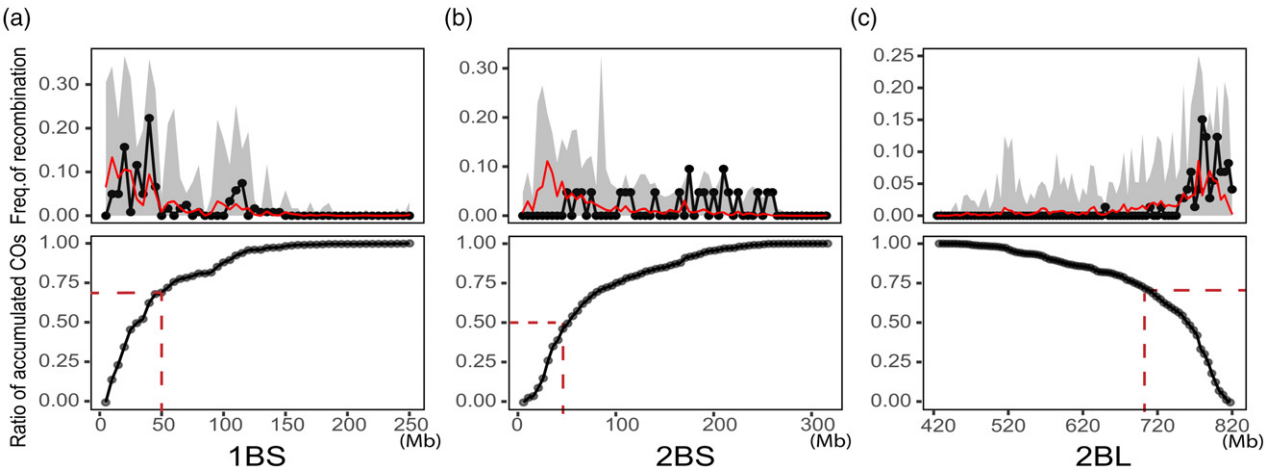


Figure 3. Crossover landscape in three sets of wheat–rye recombinants with *Ph1* absent compared to homologous crossovers (*Ph1* present) in the NAM population. (a) 1BS and 1BS-1RS. (b) 2BS versus 2BS-2RS. (c) 2BL versus 2BL-2RL. *x*-axis: physical length (Mb) of the arm; *y*-axis: crossover frequency/ratio of accumulated COs in top/bottom rows. Top row graphs: Red line, the average crossover distribution in the NAM population; shaded area, the range of crossover distribution in NAM; black line, distribution of homoeologous crossovers. In all instances homoeologous crossovers are within the range of homologous crossovers among different families of NAM. Bottom graphs: The ratios of homologous COs in a given fraction of the arm to the total number of COs in the arm, with the distal concentration of crossovers illustrated.

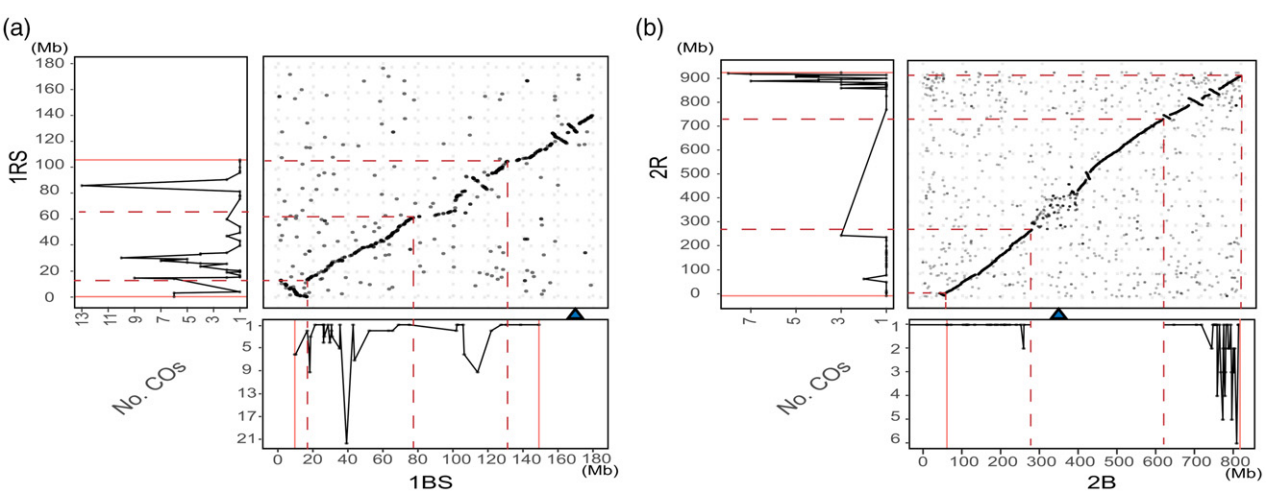


Figure 4. Dot-plot visualization of collinearity of the three pairs of homoeologous chromosome arms in relation to the distribution of the crossover frequency. (a) The distribution of the crossover frequency along the arms 1RS and 1BS (left and bottom panel boxes, respectively). (b) The distribution of the crossover frequency along the arms of chromosomes 2R and 2B (left and bottom panel boxes, respectively). Arrowheads indicate the most likely positions of centromeres, based on the concentration of CENH3 histones (International Wheat Genome Sequencing Consortium *et al.*, 2018).

TraesCS02G2B02G552900 (Figure 5b). All pairs of the wheat–rye homologous sequences appeared to be GC-rich. Thus, homologous 10-kb segments from the wheat and rye genome assemblies were used to calculate the GC content and sequence identity (Figure 5). The sequence similarity peaks and the GC contents co-located within the pairs of homologous sequences with dramatic declines on either side. Moreover, both of the 10-kb wheat segments (1BS, 2BL) contained fully aligned homologous sequences covering the entire lengths of the two annotated rye genes (SECCE1Rv1G0006160 and SECCE2Rv1G0131880) (Figure S2).

DISCUSSION

After some 60 years since the discovery of the *Ph1* locus in wheat, surprisingly little is known about it; even its precise location and the mode of action are still a matter of dispute (Griffiths *et al.*, 2006; Bhullar *et al.*, 2014). It is clear, however, that the locus, in some fashion, controls the stringency of CO. Its absence, either as a deletion of a small fragment of chromosome 5B (Gill *et al.*, 1993) or as a whole chromosome substitution, have been used for decades to induce CO between genetically divergent chromosomes, homoeologues. The original scheme for such

chromosome engineering has been devised by E.R. Sears (Sears, 1981) and is still used to this day.

The general pattern that emerges from these chromosome engineering efforts is that the homoeologous CO frequency in the absence of *Ph1* appears to be proportional to the phylogenetic distance separating the two manipulated chromosomes, and hence to the level of their divergence/differentiation. In closely related homoeologues, such as chromosomes of the A genome of wheat and the A^m genome of *Triticum monococcum*, the absence of *Ph1* produces homologous-like levels of CO and the resulting genetic maps are essentially identical to those produced by CO of homologues (Dubcovsky *et al.*, 1995; Luo *et al.*, 2000). With distantly related chromosomes, such as homoeologues from wheat and barley (*Hordeum vulgare*) (Danilova *et al.*, 2019; Rey *et al.*, 2015) or wheat and rye (Lukaszewski, 2000; Lukaszewski *et al.*, 2004), the CO frequencies are much lower, and are further reduced by any structural differences, to the point where attempts at induced recombination are futile (Lukaszewski *et al.*, 2001). At the same time, even within a single genome, such as that of rye, there can be very wide variation in the frequency of homoeologous pairing of individual chromosome arms with wheat (Naranjo *et al.*, 1988; Naranjo and Fernandez-Rueda, 1996) and consequent widely ranging CO rates. Even for fully syntenic arms, such as 1RS of rye versus 1S chromosome arms of wheat and 2RL of rye versus 2L of wheat, nearly a 40-fold difference in the CO frequencies was observed (Lukaszewski, 2000; Lukaszewski *et al.*, 2004).

The only indication that *Ph1* may affect homologous chromosome pairing/CO in wheat was reported by Dvorak and McGuire (1981) but the effects were not fully quantified. The quantification is presented here, in populations generated with four different chromosomes 1B tested against a designated segment in a single tester. These data offer several interesting observations. The range of variation in the CO rate with *Ph1* present was far greater than when it was absent (more than twofold), and the chromosome with the lowest CO rate in *Ph1* showed the highest proportional increase when *Ph1* was removed (almost 400%). Differences among individual chromosomes in *ph1b* were small, and in all cases the CO rate was close to that predicted based on pairing frequency of identical segments (that is, the wheat segment in T-9, originating from 1B of Pavon with 1B of Pavon). It needs to be pointed out that in *ph1b* several intralocus and unequal COs involving the multigene *Gli-B1* locus were recovered (Lukaszewski and Brzezinski, 2003). In the NAM population used here as a reference, there was considerable variation in the CO patterns among individual families (Jordan *et al.*, 2018). The NAM population was created from hybrids of 28 wheats with a single tester, very much the same as the set of four 1BS-T-9 populations tested here. The overall CO frequencies in the 1BS arms among different NAM families were not drastically different but the CO distribution patterns were different.

In case of complete chromosomes such as those in the NAM population, any CO event can be allocated to anywhere along the entire recombining portion of the arm. In 1BS versus T-9, CO was restricted to a small fraction (approximately 9.6–9.9 Mb) of the arm. The overall frequencies of CO in 1BS among individual families of NAM did not vary greatly perhaps because sufficiently similar CO

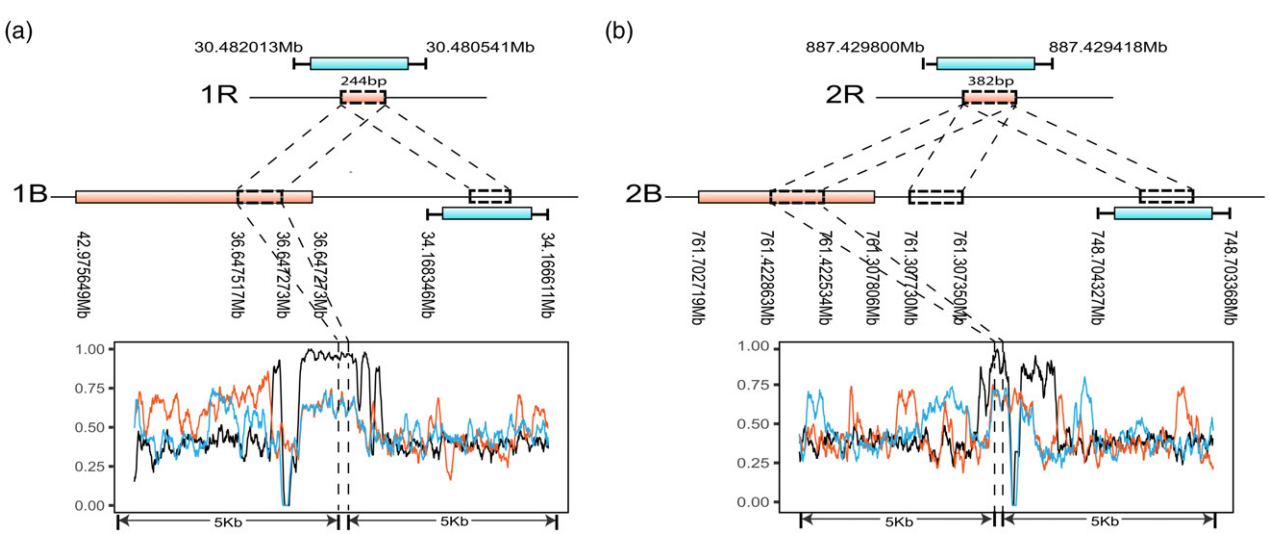


Figure 5. Sequence features of wheat and rye reference genomes defined by two crossover intervals that occurred within an annotated gene on the rye reference genome. (a, b) Interval 26/27 of 1BS-1RS Recs and interval 10 of 2BL-2RL Recs. Pink boxes: Physical crossover intervals. Blue boxes: Annotated genes. Dashed box: Best hits after BLAST. Line plots show the sequence features of the physical intervals on the rye assembly and their best hit sequences of the wheat assembly, all expanded to 10 kb (5 kb on either side). Plots: Yellow line, the GC content of rye sequences; blue line, the GC content of wheat sequences; black line: the similarity level of corresponding wheat and rye sequences.

sites were present somewhere along the recombining portions of the arms; in the small segment of T-9 available for homologous recombination the choices were limited and at times a CO could not be established. However, with *Ph1* removed, the CO frequencies were restored to levels essentially identical to those in perfectly homologous (identical) segments, in very much the same way as between homoeologues from the A genome of wheat and the A^m genome of *T. monococcum* (Dubcovsky *et al*., 1995; Luo *et al*., 2000). This confirms that *Ph1* is also capable of recognizing differences between homologues; hence it regulates their behavior. When *Ph1* is removed, the CO rate of non-identical homologous segments reaches the level of identical segments. Apparently, *Ph1* readily rejects polymorphic DNAs as candidates for CO even if these fall under our definition of homology. In a self-pollinating allopolyploid such as wheat, high CO stringency carries no penalty. Even in spontaneous or artificial hybrids of wheat there apparently is enough perfect DNA sequence homology somewhere along the length of the recombining portions of chromosome arms for successful establishment of chiasmata, and hence regular disjunction in anaphase I. With more divergent chromosomes, such as those from related species (including two of wheat's three diploid progenitors) the choices may be few and the MI pairing frequency drops. In this sense, the distinction between homologues and homoeologues appears too rigid. In fact, it is probably a continuous variation, from identical to completely different DNA sequences. Up to some point on the scale of the chromosome differentiation, sufficient stretches of identical DNA sequences are present to complete a CO. Beyond that point, restrictions apply and the ability to CO start to drop, all the way to zero. *Ph1*, apparently, moves the critical point on that scale to the left, toward fewer differences, or, in other words, imposes far more strict criteria for a CO formation. From what is known about the requirements of the DNA sequence homology and the substrate length for CO (Shen and Huang, 1986, 1989; Datta *et al*., 1997) it appears plausible that the *Ph1* locus may be involved in the intermediate steps of heteroduplex formation, perhaps by rejecting too divergent candidates or by directing its resolution toward conversions.

If the *Ph1* locus recognizes differences between chromosomes destined for CO and the frequency of the actual CO reflects their level of differentiation at the DNA sequence level, it would be of considerable interest to analyze the general DNA landscape features recognized by the locus, ranging from single-nucleotide polymorphisms to such features as the frequencies/proportions or distribution of various classes of DNA in candidate chromosomes. Attempts were made here to compare frequencies/proportions of various classes of DNAs in various intervals with different CO frequencies, but given the interval lengths and low CO numbers, all were inconclusive. The CO saturation rates with homoeologous recombination, in chromosomes of this size, are still too low for meaningful generalizations on the DNA landscape in CO. On the one hand, there is also some ambiguity as to actual location of the CO points. Imprecision in marker scoring lowers the resolution: in most cases in the reference populations of homologous recombinants the actual location of any given exchange point is only a statistical function, as opposed to a biological reality.

The only level of analysis that offers a reasonable level of confidence is the general pattern of distribution of the CO points along chromosome arms. In two of the three chromosome arms studied here, the pattern of the CO distribution in homoeologous pairs (1BS versus 1RS, 2RL versus 2BL) was essentially the same as in homologous recombination, whether detected by mapping of the physical attributes of chromosomes (Lukaszewski and Curtis, 1993) or DNA polymorphisms (Jordan *et al*., 2018). The resolution difference between the two approaches did not in any way affect the outcome; the curves of the CO distribution are the same: there is no CO in the vicinity of the centromere, and all, or almost all, CO is concentrated in the distal halves of the arms, increasing in frequency toward the telomere. In physically long arms, there can be two peaks of the CO frequency, separated by the positive CO interference distance (Lukaszewski and Curtis, 1993). The 2RS-2BS arms deviated from this general pattern: there were no COs in the distal portion of the 2BS arm. This is easily explained by a structural difference (Devos *et al*., 1993). More interesting is an almost even distribution of COs along the rest of the arm. Originally this pattern was attributed to that structural difference in the terminal region which might have altered the pattern of synapsis and CO formation in the rest of the arm (Lukaszewski *et al*., 2004). This study shows that a fairly even distribution of CO is a feature of the 2BS arm itself (no map of the physical distribution of CO in 2RS is yet available) and it appears related to the level of differentiation along the arm (Figure 4). Of the three arms tested, two showed considerably greater differentiation in the proximal than in distal halves, and COs were located distally. In 2BS-2R, the level of collinearity in the proximal half was high and COs were distributed almost evenly along the arms.

The results of this study, as well as those of Zhang *et al*. (2018), demonstrate clearly that the change in the *Ph1* status (*Ph1* versus *ph1b)* does not affect the CO pattern in a chromosome arm; only its frequency. In all cases here, the distribution of homoeologous CO, while at times differing in individual intervals from the average for the entire reference wheat population, never strayed beyond the range of variation for homologous COs in *Ph1* among different families of NAM (Figure 3, Figure S1). The general pattern was always the same, whether for homologues, such as in the

NAM population, for closely related homoeologues such as 2SS of *Aegilops speltoides* and 2BS of wheat (Zhang *et al.*, 2018), or distantly related such as chromosomes of rye and wheat here, CO was concentrated in the distal regions of chromosome arms and absent in the vicinity of the centromere.

This study was not particularly successful in detailed characterization of the DNA sequences involved in homoeologous CO. While the starting numbers of DNA polymorphisms were high, considerable ambiguity inherent is their scoring reduced their numbers to relatively few. Consequently, in a majority of cases individual COs could be allocated only to sizable bins on the DNA reference assemblies. These bins ranged in length between 65 bp and 14.7 Mb. The same problem was faced by Zhang *et al.* (2018), with COs allocated to bins ranging in length from 0.04 to 157 Mb of the wheat reference assembly. Unlike Zhang *et al.* (2018), this study had access to the reference rye DNA assembly, but this only generated additional complication of interpretation when in some segments the two assemblies, wheat and rye, were non-collinear. As the CO intervals were quite long a detailed analysis of their DNA features was possible in only a few cases. Interestingly, when the wheat DNA assembly was used as the reference, all COs appear to have occurred in non-coding (or, rather, non-annotated) sequences. When the rye assembly served as the reference, a majority of the few analyzed COs occurred either in coding (annotated) DNA sequences or in intervals which included such sequences. Unfortunately, with so few intervals analyzed in detail, no generalization appears plausible.

The highest resolution so far in allocation of a homoeologous CO event to a specific DNA sequence was reported for 5BS-5DS recombinants induced by the *ph1b* mutation in wheat (Ibba *et al.*, 2019). Four independent CO events were narrowed down to a 39-bp fragment identical between the two donors. The length of the segment is surprising given that CO, at least in yeast, is strongly affected by mismatches and if such are present, conversions are a far more likely outcome (Datta *et al.*, 1997). However, as speculated above, in the absence of *Ph1* such differences are probably overlooked, and a CO probably progresses all the way to its conclusion. Far more surprising was that four independent events were recovered in a single stretch of 39 bp, in chromosome arms that are hundreds Mb in length. However, in this study, the shortest interval harboring a homologous CO was 46 bp and for a homoeologous one it was 65 bp.

In most cases the patterns of the CO distribution generated by referencing the two DNA assemblies were similar. This indicates that the CO points on wheat and rye chromosomes were in corresponding positions and there appear to be no independent hotspots on either wheat or rye chromosomes.

The density of polymorphic markers on the 2B arm in this study was similar to that in the study of Zhang *et al.* (2018), who reported the physical size of the intervals ranging from 0.04 to 150 Mb, which is significantly larger than any set of recombinants in this study (Table S4). Physical sizes of the CO intervals in the NAM population were at the same orders of magnitude as the Recs population here (Table 1, Table S5), although the NAM population was genotyped by sequence-based genotyping coupled with wheat 90K SNPs (Jordan *et al.*, 2018). Additionally, all wheat–rye recombinants here were verified cytologically even if some of them were not identified by DNA markers. Within the resolution in this study, we were able to explore the general distribution of CO under the effects of non-functional *Ph1* but not the DNA sequence features inside CO intervals.

In conclusion, the absence of *Ph1* is likely the wild-type condition where regular pairing and CO between divergent homologues is permitted (and advantageous). Pairing of homoeologues is constrained by the level of their divergence. *Ph1* appears to react to the DNA sequence divergence and restricts COs to identical or very similar sequences. This is highly beneficial in polyploids. The locus does not appear to affect to any detectable extent the general CO landscape; the distribution of COs involving homoeologues induced in the absence of *Ph1* is the same as that of homologous CO generated with *Ph1* present. Clearly, the same chromosome regions recombine in both cases. The only change is in frequency, and that appears to depend primarily on the level of chromosomal divergence.

EXPERIMENTAL PROCEDURES

Plant material

All stocks used in this study were generated in *T. aestivum* L. cv. 'Pavon 76', a spring white hard wheat from the breeding program at the Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Mexico. For a study of homologous CO, substitutions of chromosomes 1B from four wheats, cvs. Henika (abbreviated HE), Begra (BE), Little Club (LC), and an Iranian landrace #55 (#55) in Pavon were combined with Pavon *ph1b*. These lines, with substituted chromosomes 1B and with *Ph1* or *ph1b*, were crossed to two lines of Pavon T-9 with *Ph1* or *ph1b*. T-9 is a wheat–rye translocation (Lukaszewski, 2000) with the breakpoint located between DNA markers at positions 9.56 and 9.96 Mb, and hence the wheat segment is no longer than 9.96 Mb and is terminal; the rest of the arm is from rye chromosome arm 1RS (Figure 1). Heterozygotes T-9 with each of the four substituted chromosomes 1B, homozygous *Ph1*, or *ph1b*, were backcrossed as male to Pavon Dt1BL lines (lines missing chromosome arm 1BS), and recombined chromosomes were selected. All selection was performed by C-banding, utilizing polymorphism for the terminal C-band between T-9 and each of the four chromosomes 1B and the presence/absence of the rye segment.

Three sets of wheat–rye homoeologous recombinants (1BS-1RS, 2BS-2RS, and 2BL-2RL) were developed previously by

induced recombination between wheat and rye homoeologues, in the absence of the *Ph1* locus (Lukaszewski, 2000; Lukaszewski *et al.*, 2004). Additional recombinants 1BS-1RS and 2BL-2RL were selected from the leftover populations in the earlier studies. Selection was performed by C-banding, utilizing C-band polymorphism between wheat and rye chromosome arms. Each chromosome is a single breakpoint translocation, composed of a rye and a wheat segment. As before, recombinant chromosomes were classified into two types based on the configuration of the wheat and rye segments: chromosomes with designation T- have proximal rye and terminal wheat segments; designation B+ represents chromosomes with proximal wheat and terminal rye segments.

The set of 1BS-1RS recombinants consisted of 128 chromosomes, of which 108 were available as homozygotes and 20 were hemizygous (present in a single copy with chromosome arm 1BL). Including the newly selected recombinants, the 2BL-2RL set consisted of 88 chromosomes, all monosomic. The set of 2BS-2RS recombinants consisted of 25 chromosomes, all homozygous. The three sets of recombinants are referred to as 1BS-1RS Recs, 2BS-2RS Recs, and 2BL-2RL Recs, respectively.

Genotyping

Recombinants 1BS-1RS and 2BL-2RL, along with their parental lines, were genotyped by DArT-seq at the Diversity Arrays Technology (DArT, Canberra, Australia). Genomic DNA from each line carrying a recombinant chromosome was extracted from young leaves using the DNA extraction method recommended by the Diversity Arrays Technology. Recombinants 2BS-2RS and their parental lines were genotyped by the wheat 90K Illumina iSelect SNP arrays (https://www.illumina.com/), courtesy of Dr. X. Chen, USDA-ARS, Fargo ND, USA.

Construction of genetic maps

Genetic maps of 1BS-1RS Recs, 2BS-2RS Recs, and 2BL-2RL Recs were constructed for each pair of arms. Each recombinant chromosome arm contains a segment from a wheat arm of Pavon 76 and a segment from a rye chromosome. Wheat- and rye-specific DNA markers were identified by comparing the genotypes of the parental lines: Pavon 76 versus the Pavon translocation line from which recombinants were generated (1RS.1BL, 2RS.2BL, 2BS.2RL, respectively). Wheat-specific markers were present in Pavon 76 and absent in the respective translocation line, and vice versa for the rye-specific markers.

Markers were disregarded in the following situations: (i) missing frequency above 20%; (ii) allele frequency above 60%; (iii) allele frequency below 40%. Genetic maps were generated by the MSTMap online (http://www.mstmap.org/) with default settings. For the 1BS-1RS Recs, due to the dosage difference between homozygotes and heterozygotes, the genetic map was first generated for homozygotes only; hemizygotes were fitted into the map manually.

As a control for homoeologous CO genetic maps for the wheat NAM population were used, with access to data (http://wheatgenomics.plantpath.ksu.edu/nam/), and some data filtering kindly provided by Drs. E. Akhunov and K. Jordan, Kansas State University, Manhattan, KS, USA. The NAM population consists of 28 recombinant inbred line (RIL) families created by self-pollination for seven generations of hybrids of 28 spring wheat accessions with one common landrace 'Berkut'. Since all RILs share one parent, the population is ideal for studying CO of any specific chromosome arm with different homologous arms. This NAM population was genotyped by the sequence-based genotyping (Jordan *et al.*, 2018). The skeleton map of each chromosome in each line contains skeleton markers, those with the least missing data for each genetic bin, and these markers were used in this study to identify chromosome intervals containing homologous CO points. Skeleton markers for specific chromosome arms (1BS, 2BS, and 2BL) were identified by locating them on the CS DNA sequence assembly 2.0 kindly provided by Drs. M.C. Luo and T.T. Zhu, University of California Davis, CA, USA. Co-segregation of markers with no more than 15% missing data were then used to narrow down the physical intervals of COs. Skeleton markers on specific arms were selected by mapping their physical locations according to the chromosome partitioning in International Wheat Genome Sequencing Consortium *et al.* (2018) which define the 1BS arm as 0–172 Mb, the 2BS arm as 0–248 Mb, and the 2BL arm as 433–820 Mb.

Crossover detection among the Recs and NAM populations

Because of the mode of selection (change in phase between the telomere and the centromere), all Recs chromosomes are single breakpoint translocations, and hence, each has a single change of phase, from wheat to rye. In most cases, this was confirmed by *in situ* probing with total genomic DNA of rye. No multiple COs per arm were expected and none were detected. With wheat- and rye-specific markers for each pair of arms, each change of phase (a CO) could be allocated to an interval between two adjacent markers, both wheat, both rye, or one of each.

The 28 families of the NAM population contain a total of 2100 RILs. The numbers of COs per chromosome arm in the NAM populations were not known in advance and no *a priori* assumptions were made. Genetic maps of the NAM population (Jordan *et al.*, 2018) were used to identify homologous CO points in chromosome arms 1BS, 2BS, and 2BL, with *Ph1* present. Precise identification of these CO points was seriously complicated by large numbers of the so-called singletons, mis-scored markers, and absent data. Given low chiasmata frequencies in wheat and a very strong positive chiasma interference, the probability of two COs flanking a single DNA marker appears inestimably low. These were ignored. For the sake of this study, a phase change involving two consecutive markers was interpreted as a double CO, even though it is obvious that with a very large number of singletons, many doubletons represent two consecutive mis-scored markers. For the 1BS arm, only 19 families with 1343 RILs had sufficient numbers of skeleton markers (17–32 per arm in a family), and hence only 19 genetic maps were created. The remaining nine families probably include one parent with the 1RS.1BL translocation and were useless in the current study.

For chromosome 2B, the NAM20 family appeared to carry a rye insert in 2BL, and it was excluded from analyses for both 2B arms. Therefore, for chromosome 2B, 27 families were used as reference maps.

Physical localization of crossover intervals

DNA sequence-based markers used in the construction of the genetic maps were aligned against the Chinese Spring DNA sequence assembly 2.0 by blastn-2.7.1 (https://www.ncbi.nlm.nih.gov/books/NBK279680). The best hits using default settings were retained. This created a physical map for each chromosome. Markers were ordered by genetic distance. Redundant markers (that is, markers that were non-informative in the map construction) were used to narrow down the CO intervals. Pre-publication access to the rye DNA assembly and annotation was kindly provided by Drs. N. Stein and M. Wallace of the Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany. TBtools v0.66831 (Chen *et al.*, 2020) was used for sequence information extraction.

Ideally, all CO intervals along chromosome arms should appear in the same order on genetic and physical maps. However, in some cases these orders were different, for several reasons, including possible differences between chromosomes used in genetic mapping versus those used in sequencing, and vagaries of the algorithms used to create the maps and DNA assemblies. For the purpose of this study, the genetic order of markers was taken as standard. Within blocks of redundant markers, those that did not fit into a physical region with the majority of other redundant markers were discarded.

Collinearity of homoeologous chromosome arms

The DNA sequence collinearity of the three pairs of homoeologous chromosome arms was compared by MuMmer 3.0 (http://mummer.sourceforge.net/) (Kurtz *et al.*, 2004).

DNA sequence features of the crossover intervals

The general genic features of the CO intervals with physical lengths below 50 kb were identified using the IWGSC RefSeq v1.1 and the rye annotation. Intervals inside a gene region were further analyzed for their sequence features. Sequences of intervals and their host genes were blasted against the homoeologous genome assembly databases. The distribution of the GC content was measured by a sliding window of 100-bp and 10-bp steps with the Python script. Wheat and rye sequence similarities were also calculated by a sliding window of 100-bp and 10-bp steps with the Python script after alignment with DNAMAN v8.0. Graphical representations were constructed using the R package ggplot2 (v.2.2.1) (Wickham, 2016).

Statistics

Proportions of recombined homologous segments in the T-9 translocation in the total numbers of chromosomes tested in *Ph1* and *ph1b* were compared by the χ^2 test. R3.6.1 (https://cran.rstudio.com/) was used to carry out the paired-sample *t*-test analysis of paired sets of physical length of intervals of three studied arms referring to wheat and rye assemblies.

ACKNOWLEDGMENTS

We acknowledge and appreciate the assistance provided by Drs. E. Akhunov and K. Jordan, Kansas State University, Drs. N. Stein and M. Wallace of the Leibniz Institute of Plant Genetics and Crop Plant Research, and Drs. M.C. Luo and T.T. Zhu, University of California, Davis, who kindly offered access to their data and provided valuable pointers along the way. This project was supported in part by funding from USDA-NIFA (#CA-R-BPS-5411-H) to AJL and from NIH (T3GM062756) to CN and in part by the Chinese National Key Research and Development Program (#2016YFD0102000).

AUTHOR CONTRIBUTIONS

AJL conceived the project and created the recombinants; CF selected additional recombinants, carried out the experiments, and created the first draft of the manuscript; ZJ, MH, CN, XC, and WC assisted in data analysis; MH and DL assisted in writing the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data relevant to the contents of this article are presented in the main text and supplementary tables and figures. All plant materials can be obtained from the corresponding author upon request.

SUPPORTING INFORMATION

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

Figure S1. Variation in the crossover distribution along chromosome arm 1BS among 19 families in the NAM population versus the 1BS-1RS crossover distribution.

Figure S2. Alignment of the 10-kb sequences of wheat and rye assembly with the sequence of the corresponding annotated gene of wheat and rye.

Table S1. Genetic maps of the 1BS-1RS, 2BS-2RS, and 2BL-2RL recombinants.

Table S2. A summary of the number of recombined and non-recombined chromosome arms 1BS and of chromosome 2B in each family of the NAM population.

Table S3. A summary of the number of crossovers in each genetic interval of the 1BS chromosome arm and chromosome 2B in each family of the NAM population.

Table S4. A summary of genetic and physical intervals among the recombinants 1BS-1RS, 2BS-2RS, and 2BL-2RL.

Table S5. A summary of the physical lengths of the crossover intervals on the chromosome arm 1BS and on chromosome 2B in each family of the NAM population.

REFERENCES

Bhullar, R., Nagarajan, R., Bennypaul, H., Sidhu, G.K., Sidhu, G., Rustigi, S., Wettstein, V.D. and Gill, K.S. (2014) Silencing of a metaphase I-specific gene results in a phenotype similar to that of the Pairing homoeologous 1 (*Ph1*) gene mutations. *Proc. Natl Acad. Sci. USA*, **111**, 14187–14192.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. and Xia, R. (2020) TBtools - an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant.* **13**, 1194–1202.

Danilova, T.V., Poland, J. and Friebe, B. (2019) Production of a complete set of wheat–barley group-7 chromosome recombinants with increased grain β-glucan content. *Theor. Appl. Genet.* **132**, 3129–3141.

Datta, A., Hendrix, M., Lipstich, M. and Jinks-Robertson, S. (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl Acad. Sci. USA*, **94**(18), 9757–9762.

Devos, K.M., Atkinson, M.D., Chinoy, C.N., Francis, H.A., Harcourt, R.L., Koebner, R.M.D., Liu, C.J., Masojc, P., Xie, D.X. and Gale, M.D. (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. *Theor. Appl. Genet.* **85**, 673–680.

Doyle, G.G. (1986) The allotetraploidization of maize: 4. Cytological and genetic evidence indicative of substantial progress. *Theor. Appl. Genet.* **71**, 585–594.

Dubcovsky, J., Luo, M.C. and Dvorak, J. (1995) Differentiation between homoeologous chromosomes 1A of wheat and $1A^m$ of *Triticum monococcum* and its recognition by the wheat *Ph1* locus. *Proc. Natl Acad. Sci. USA*, **92**, 6645–6649.

Dvorak, J. and McGuire, P. (1981) Nonstructural chromosome differentiation among wheat cultivars with special reference to differentiation of chromosomes in related species. *Genetics*, **97**, 391–414.

Gill, K.S., Gill, B.S., Endo, T.R. and Mukai, Y. (1993) Fine physical mapping of *Ph1*, a chromosome pairing regulator gene in polyploid wheat. *Genetics*, **134**: 1231–1236.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I. and Moore, G. (2006) Molecular characterization of Ph1 as a major

chromosome pairing locus in polyploid wheat. *Nature*, **439**(7077), 749–752.

Ibba, M.I., Zhang, M., Cai, X. and Morris, C.F. (2019) Identification of a conserved *ph1b*-mediated 5DS–5BS crossing over site in soft-kernel durum wheat (*Triticum turgidum* subsp. *durum*) lines. *Euphytica*, **215**: 200.

International Wheat Genome Sequencing Consortium, Appels, R. and Eversole, K. *et al.* (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*; **361**: eaar7191.

Jenczewski, E. and Alix, K. (2004) From diploids to alloploids: the emergence of efficient pairing control genes in plants. *Critical. Rev. In. Plant. Sci.* **23**, 21–45.

Jordan, K.W., Wang, S., He, F. *et al.* (2018) The genetic architecture of genome wide recombination rate variation in allopolyploid wheat revealed by nested association mapping. *Plant J.* **95**(6), 1039–1054.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. and Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biol.* **5**: R12. https://doi.org/10.1186/gb-2004-5-2-r12

Lukaszewski, A.J. (1995) Physical distribution of translocation breakpoints in homoeologous recombinants induced by the absence of the *Ph1* gene in wheat and triticale. *Theor. Appl. Genet.*, **90**(5), 714–719.

Lukaszewski, A.J. (2000) Manipulation of the 1RS.1BL translocation in wheat by induced homoeologous recombination. *Crop Sci.* **40**(1), 216–225.

Lukaszewski, A.J., Brzezinski, W. *et al.* (2003) Novel alleles at the *Gli-B1* locus generated by homologous recombination in the absence of the *Ph1* locus. In *Proc 10th Int Wheat Genet Symp, Sept 1–6, 2003*. (Pogna, N.E., ed). Italy: Paestum, pp. 607–609.

Lukaszewski, A.J. and Cowger, C. (2017) Re-Engineering of the *Pm21* transfer from *Haynaldia villosa* to bread wheat by induced homoeologous recombination. *Crop Sci.* **57**, 2590–2594.

Lukaszewski, A.J. and Curtis, C.A. (1993) Physical distribution of recombination in B-genome chromosomes of tetraploid wheat. *Theor. Appl. Genet.* **86**(1), 121–127.

Lukaszewski, A.J., Porter, D.R., Baker, C.A., Rybka, K. and Lapinski, B. (2001) Attempts to transfer Russian wheat aphid resistance from a rye chromosome in Russian triticales to wheat. *Crop Sci.* **41**(6), 1743–1749.

Lukaszewski, A.J., Rybka, K., Korzun, V., Malyshev, S.V., Lapinski, B. and Whitkus, R. (2004) Genetic and physical mapping of homoeologous recombination points involving wheat chromosome 2B and rye chromosome 2R. *Genome*, **47**(1), 36–45.

Luo, M.C., Yang, Z.L., Kota, R.S. and Dvořák, J. (2000) Recombination of chromosomes $3A^m$ and $5A^m$ of *Triticum monococcum* with homoeologous chromosomes 3A and 5A of wheat: The distribution of recombination across chromosomes. *Genetics*, **154**, 1301–1308.

Mello-Sampayo, T. (1971) Genetic regulation of meiotic chromosome pairing by chromosome 3D of *Triticum aestivum*. *Nat. New Biol.* **230**(9), 22–23.

Naranjo, J., Roca, R., Giraldez, P. and Giocoechea, P.G. (1988) Chromosome pairing in hybrids of *phIb* mutant wheat with rye. *Genome*, **30**, 639–646.

Naranjo, T. and Fernandez-Rueda, P. (1996) Pairing and recombination between individual chromosomes of wheat and rye in hybrids carrying the *ph1b* mutation. *Theor. Appl. Genet.* **93**(1-2), 242–248.

Okamoto, M. (1957) Asynaptic effect of chromosome V. *Wheat Inform. Serv.* **5**, 6.

Rabanus-Wallace, M.T., Hackauf, B. and Mascher, M. (2019) Chromosome-scale genome assembly provides insights into rye biology, evolution, and agronomic potential. *bioRxiv* 2019.12.11.869693. https://doi.org/10.1101/2019.12.11.869693

Rey, M.D., Calderón, M.C. and Prieto, P. (2015) The use of the *ph1b* mutant to induce recombination between the chromosomes of wheat and barley. *Front. Plant. Sci.* **6**, 160.

Riley, R. and Chapman, V. (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. *Nature*, **182**(4637), 713–715.

Sánchez-Moran, E., Benavente, E. and Orellana, J. (2001) Analysis of karyotypic stability of homoeologous-pairing (*ph*) mutants in allopolyploid wheats. *Chromosoma*, **110**(5), 371–377.

Sears, E.R. (1976) Genetic control of chromosome pairing in wheat. *Ann. Rev. Genet.* **10**(1), 31–51.

Sears, E.R. (1981) Transfer of alien genetic material to wheat. In *Wheat Science – Today and Tomorrow* (Evans Lt, W.J. and Peacock, eds). Cambridge, UK: Cambridge University Press, pp. 75–89.

Sears, E.R. and Okamoto, M. (1958) Intergenomic chromosome relationships in hexaploid wheat. Proc 10th Int Congr Genet. *Montreal*, **2**, 258–259.

Shen, P. and Huang, H.V. (1986) Homologous recombination in Escherichia coli: dependence on substrate length and homology. *Genetics*, **112**: 441–457.

Shen, P. and Huang, H.V. (1989) Effect of base pair mismatches on recombination via the RecBCD pathway. *Mol. Gen. Genet.* **218**(2), 358–360.

Sutton, T., Whitford, R., Baumann, U., Dong, C.M., Able, J.A. and Langridge, P. (2003) The *Ph2* pairing homoeologous locus of wheat (*Triticum aestivum*): identification of candidate meiotic genes using a comparative genetics approach. *Plant J.* **36**: 443–456.

Wickham, H. (2016) *ggplot2. Elegant Graphics for Data Analysis*, 2nd edn. Springer International Publishing. ISBN 978-3-319-24277-4.

Zhao, R.H., Wang, H.Y., Xiao, J. *et al.* (2013) Induction of 4VS chromosome recombinants using the *CSph1b* mutant and mapping of the wheat yellow mosaic virus resistance gene from *Haynaldia villosa*. *Theor. Appl. Genet.* **126**(12), 2921–2930.

Zhang, W., Zhu, X., Zhang, M., Chao, S., Xu, S. and Cai, X. (2018) Meiotic homoeologous recombination-based mapping of wheat chromosome 2B and its homoeologues in *Aegilops speltoides* and *Thinopyrum elongatum*. *Theor. Appl. Genet.* **131**(11), 2381–2395.