High-Resolution Mapping of Crossover Events in the Hexaploid Wheat Genome Suggests a Universal Recombination Mechanism

Benoit Darrier,* Hélène Rimbart,* François Balfourier,* Lise Pingault,* Ambre-Aurore Josselin,* Bertrand Servin,[†] Julien Navarro,* Frédéric Choulet,* Etienne Paux,* and Pierre Sourdille*¹

*Genetics, Diversity and Ecophysiology of Cereals, Institut National de la Recherche Agronomique, Université Blaise Pascal, 63000 Clermont-Ferrand, France and [†]Génétique Physiologie et Systèmes d'Elevage, Institut National de la Recherche Agronomique, Université de Toulouse, 31320 Castanet-Tolosan, France

ABSTRACT During meiosis, crossovers (COs) create new allele associations by reciprocal exchange of DNA. In bread wheat (*Triticum aestivum* L.), COs are mostly limited to subtelomeric regions of chromosomes, resulting in a substantial loss of breeding efficiency in the proximal regions, though these regions carry ~60–70% of the genes. Identifying sequence and/or chromosome features affecting recombination occurrence is thus relevant to improve and drive recombination. Using the recent release of a reference sequence of chromosome 3B and of the draft assemblies of the 20 other wheat chromosomes, we performed fine-scale mapping of COs and revealed that 82% of COs located in the distal ends of chromosome 3B representing 19% of the chromosome length. We used 774 SNPs to genotype 180 varieties representative of the Asian and European genetic pools and a segregating population of 1270 F₆ lines. We observed a common location for ancestral COs (predicted through linkage disequilibrium) and the COs derived from the segregating population. We delineated 73 small intervals (<26 kb) on chromosome 3B that contained 252 COs. We observed a significant association of COs with genic features (73 and 54% in recombinant and nonrecombinant intervals, respectively) and with those expressed during meiosis (67% in recombinant intervals and 48% in nonrecombinant intervals). Moreover, while the recombinant intervals contained similar amounts of retrotransposons and DNA transposons (42 and 53%), nonrecombinant intervals had a higher level of retrotransposons (63%) and lower levels of DNA transposons (28%). Consistent with this, we observed a higher frequency of a DNA motif specific to the *TIR-Mariner* DNA transposon in recombinant intervals.

KEYWORDS recombination; meiosis; bread wheat; linkage disequilibrium; transposon; hotspot; sequence motif

MEIOTIC recombination is a process that allows reshuffling of diversity by the reciprocal exchange of DNA called a crossover (CO). This phenomenon is conserved in most eukaryotes (for a review see Mercier *et al.* 2015) and follows the formation of a double-strand break (DSB) of DNA generated by the topoisomerase SPO11 complex. However, the number of DSBs is at least 10- to 50-fold greater than the number of COs which rarely exceeds three per bivalent chro-

somosome per meiosis (Mercier *et al.* 2015). This paucity of COs per meiosis with regards to the number of DSBs suggests the existence of a tight control and regulation of recombination in plants that promote DSB repair in a manner that does not lead to COs. For example, the study of the two helicases AtFANCM and RECQ4 (AtRECQ4A and AtRECQ4B) (Crismani *et al.* 2012; Knoll *et al.* 2012; Girard *et al.* 2014; Séguéla-Arnaud *et al.* 2015) revealed three- and sixfold CO frequency increases in the *Atfancm* single mutant and the *Atrecq4a/Atrecq4b* double mutant, respectively, compared to the wild type. These increases result from additional COs from the class-II pathway which suggests that FANCM and RECQ4 prevent CO formation and direct recombination intermediates toward the synthesis-dependent, strand-annealing, non-CO (NCO) pathway. In addition, the AAA-ATPase (unfoldase) FIDGETIN-LIKE-1 (FIGL1) (Girard *et al.* 2015)

Copyright © 2017 by the Genetics Society of America
doi: <https://doi.org/10.1534/genetics.116.196014>

Manuscript received September 19, 2016; accepted for publication May 12, 2017;
published Early Online May 22, 2017.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.196014/-/DC1.

¹Corresponding author: Genetics, Diversity and Ecophysiology of Cereals, Institut National de la Recherche Agronomique, Site de Crouël, 5 Chemin de Beaulieu, 63000 Clermont-Ferrand, France. E-mail: pierre.sourdille@inra.fr

and the RAD51 paralog XRCC2 (Da Ines *et al.* 2013) also prevent or at least limit CO formation, likely by regulating the early invasion step catalyzed by DMC1/RAD51.

At least one CO per chromosome is absolutely required to ensure a correct segregation of chromosomes during meiosis (Mercier *et al.* 2015). In all eukaryotes studied so far, the distribution of COs is not homogeneous along the chromosomes (Lukaszewski and Curtis 1993; Tenaillon *et al.* 2002; Saintenac *et al.* 2009; Mayer *et al.* 2012; Pan *et al.* 2012; Choulet *et al.* 2014; Choi and Henderson 2015; Mercier *et al.* 2015). For cultivated crops, this implies a decrease of the breeding power in regions showing low CO rates (Rodgers-Melnick *et al.* 2015). In human, *Saccharomyces cerevisiae*, *Arabidopsis*, and wheat, >80% of the recombination events occur in less than a quarter of the genome (Myers *et al.* 2005; Chen *et al.* 2008; Mancera *et al.* 2008; Choi *et al.* 2013; International Wheat Genome Sequencing Consortium 2014). This nonhomogeneity of CO distribution is the basis of the definition of recombination hotspots and coldspots (which have significantly high and low CO frequencies, respectively). A general rule is that centromeres are cold regions (Choo 1998) with a few exceptions such as Welsh onion where COs cluster close to centromeres (Jones 1984).

CO location can be determined either through their cytological signature (chiasmata) or through the parental allele switch in progenies (for a review see Baudat *et al.* 2013). This switch can be detected by genotyping the progenies derived from biparental or multi-parental crosses (for a review see Huang *et al.* 2015) but is limited by a constrained number of COs (Darvasi *et al.* 1993). Statistical approaches were also developed to estimate the population recombination parameter, ρ , and to computationally infer genome-wide, population-averaged recombination rates (McVean *et al.* 2002, 2004; Li and Stephens 2003; Slatkin 2008; Hellsten *et al.* 2013; Smukowski Heil *et al.* 2015). These approaches, also termed “coalescent analysis,” rely on linkage disequilibrium (LD) patterns in populations which represent a nonrandom association of alleles at different loci (Lewontin and Kojima 1960).

In plants, coalescent theory was successfully applied to generate maps of CO frequency and hotspots in monkey flower (*Mimulus guttatus*; Hellsten *et al.* 2013) and in *Arabidopsis thaliana*, (Choi *et al.* 2013). In *Mimulus*, a collection of 98 individuals was used to precisely locate 414,734 CO- or gene-conversion events, among which 3235 were highly reliable recombination hotspots and ~13,000 were considered *bona fide* COs. In *Arabidopsis*, megabase-scale variations in CO frequencies along the chromosomes, with an increase from telomere to centromere, were observed (Choi *et al.* 2013; Drouaud *et al.* 2013). Two previously identified hotspots (3a and 3b; Yelina *et al.* 2012) were confirmed by allele-specific amplification from gamete DNA (pollen typing; Drouaud and Mézard 2011). In both cases (*Mimulus* and *Arabidopsis*), CO frequency increases toward transcriptional start sites (TSSs)—and to a lower extent toward transcriptional termination sites (TTSs) in *Arabidopsis* (Choi *et al.*

2013)—and fall off sharply just after, thus exhibiting polarity. In sorghum, a similar correlation was observed with 97–98% of COs occurring in euchromatin regions (Paterson *et al.* 2009). Similar studies aiming at deciphering recombination in maize (Fu *et al.* 2001, 2002; Yao and Schnable 2005) and wheat (Saintenac *et al.* 2011) revealed the presence of small regions exhibiting high recombination rates. These authors evaluated intervals ranging from 380 to 11,600 bp where recombination varied from 22 to 132 cM/Mb compared to average chromosome values of 2.1 and 0.20 cM/Mb for maize and wheat, respectively. Like for *Arabidopsis* and *Mimulus*, CO events were more frequent in the 5' regions of genes confirming that in plants, hotspots are rather localized in promoters as this is the case in yeast (Pan *et al.* 2011), humans, or mice (reviewed in Borde and de Massy 2013 and de Massy 2013).

However, many gene-rich regions show little or no recombination in plants as well (Drouaud *et al.* 2013) while regions enriched in repeats may recombine (Duret *et al.* 2000), questioning whether repeated sequences may play a role in recombination distribution. Transposon abundance, diversity, and activity are highly variable among species and they developed diverse strategies to thrive in the host, such as preferential insertion close to genes (Levin and Moran 2011; Barrón *et al.* 2014). Transposable elements (TEs) contribute to chromosome shape and function (for a review see Slotkin and Martienssen 2007) but contradictory results are observed regarding the impact of TE content on recombination. Repeated sequences and recombination rate are correlated positively in *Caenorhabditis elegans* (Duret *et al.* 2000) and negatively in the *Drosophila melanogaster* genome (Rizzon *et al.* 2002). Even in the same species, in maize, insertion sites of the *Mu* transposon concentrate in the same regions as meiotic recombination (Liu *et al.* 2009), while fine-scale studies at the *bronze* locus revealed a reduction of recombination in regions containing TEs, suggesting that retrotransposons are probably inert with regard to recombination because of methylation and the likely associated condensed chromatin (Dooner and Martínez-Férez 1997; He and Dooner 2009; Rodgers-Melnick *et al.* 2015). In *Arabidopsis*, DNA methylation can silence CO hotspots and plays an essential role in forming domains of meiotic recombination along chromosomes (Yelina *et al.* 2015). Recently, a study in mouse revealed that DNA methylation can prevent transposons from adopting chromatin characteristics amenable to meiotic recombination (Zamudio *et al.* 2015), underlying the importance of DNA conformation and the likely role of epigenetic marks on recombination.

Mouse and human use the same molecular strategies regarding the regulation of recombination. Initially in human, the CCTCCCT motif was found to be overrepresented at hotspots in the *THE1A/B* retrotransposon which is primate specific (Smit 1993; Myers *et al.* 2005, 2008; Pace and Feschotte 2007). Furthermore, Myers *et al.* (2008) suggested that any generic hotspot-promoting motif should operate on diverse genetic backgrounds (such as in different repeat

families) in human, and revealed the presence of a common 13-bp degenerated motif (CCNCCNTNNCCNC) that is recognized by the DNA-binding and chromatin-modifier PRDM9 protein (Baudat *et al.* 2010). PRDM9 has a SET domain that promotes H3K4me3 and DSB sites display a PRDM9-dependent enrichment for H3K4me3, detected before and in the absence of DSBs (Borde and de Massy 2013). In plants, no homolog to PRDM9 subfamily was found (Zhang and Ma 2012). However, most H3K4me3-methylase encoding genes are expressed in meiotic cells in *Arabidopsis* and rice (Zhang and Ma 2012), and this mark is associated with recombinogenic regions in *Arabidopsis* and barley (Choi *et al.* 2013; Aliyeva-Schnorr *et al.* 2015; Shilo *et al.* 2015). This may even be a general eukaryotic regulatory mechanism and may have originated in the ancestor of eukaryotes (Zhang and Ma 2012; Mercier *et al.* 2015).

In plants, several studies have revealed strongly associated repeat motifs with meiotic CO hotspots. Significant associations were found between hotspots and poly-A stretches (Choi *et al.* 2013; Wijnker *et al.* 2013; Shilo *et al.* 2015) which are known to preferentially locate upstream of TSSs, resulting in reduced nucleosome occupancy that facilitates accessibility of the recombination machinery (Wu and Lichten 1994; Berchowitz *et al.* 2009; Segal and Widom 2009; Pan *et al.* 2011). Similarly, CCN- and CTT-repeat sequence motifs were also detected to be associated to recombination hotspots (Choi *et al.* 2013; Shilo *et al.* 2015; Wijnker *et al.* 2013). These motifs are also located close to the TSSs and are similar to the motif targeted by the PRDM9 protein in humans and mouse, which may suggest a similar recognition by the recombination machinery. Poly-A stretches, CCN, and CTT degenerate motifs may thus contribute to the structure of the chromatin in promoter regions in plants with consequences for meiotic recombination. However, no specific motif such as the one targeted by PRDM9 (CCNCCNTNNCCNC) in human and mouse has been found in plants, so far, to explain the presence of recombination hotspots.

Fine-scale analysis of the recombination pattern in bread wheat (*Triticum aestivum* L.) has long been hampered by the size, complexity, and polyploidy of its genome (17 Gb, 80–85% of repeated sequences, $2n = 6x = 42$). Initial studies were based on aneuploid stocks (deletion lines; Endo and Gill 1996) and revealed that the frequency of meiotic recombination is highly biased toward the ends of chromosomes (Paris *et al.* 2000; Paux *et al.* 2008; Saintenac *et al.* 2009). However, this does not rely on the distal position since when the chromosome arms are inverted (telomeres placed at the centromeres and vice versa), COs still occur in the same region whatever its location (Lukaszewski *et al.* 2012). Only one hotspot was reported in the promoter of the *TaHGA3* gene on chromosome 3B (Saintenac *et al.* 2011), and this hotspot was associated to an NCO event covering 453–1098 bp. Recently, the reference sequence of chromosome 3B and the draft assemblies of the 20 other chromosomes were published (Choulet *et al.* 2014; International Wheat Genome Sequencing Consortium 2014), offering the opportunity to

precisely characterize recombination hotspots at the whole-genome scale.

In this work, we used these new resources to obtain a genome-wide overview of recombination and to get better insights into the factors that drive COs in wheat. First, we fine-mapped COs on chromosome 3B using a large segregating population. We showed that these contemporary recombination hotspots are conserved with ancestral recombination breakpoints determined through historical mapping (termed ρ -map herein; Stumpf and McVean 2003) using two collections (European and Asian) representing two different genetic pools of bread wheat. Our analysis of >250 COs located on chromosome 3B revealed that recombination mainly occurs nearby or within genes that are mostly expressed during meiosis. Furthermore, TE composition is different between recombinogenic or nonrecombinogenic intervals and TE-related DNA motifs are associated with high recombination rates, suggesting a possible ancestral mechanism for recombination in eukaryotes.

Materials and Methods

Plant material

The CsRe population was derived from a cross between Chinese Spring (Cs) (international reference) and Renan (Re) (French elite cultivar). The hybrid was self-fertilized and the progeny (1270 individuals) was developed from F₂ plants through self-fertilization of single individuals for each family [single seed descent (SSD)] until the F₆ generation. For chromosome 3B, two fully overlapping sets of 276 and 356 lines were randomly selected to genotype 5778 (Axiom Array) and 1280 (KASPar) SNPs, respectively. The whole population was genotyped with only 96 SNPs to roughly identify the recombinants along the chromosome 3B pseudomolecule (Choulet *et al.* 2014). For whole-genome analysis, 430 lines (including the preceding sets) were randomly selected to genotype 280,226 (Axiom Array) public SNPs, from which 53,287 were used to construct a dense genetic map (Rimbert *et al.* 2017). From these SNPs, 2527 located in 476 recombining scaffolds were used to fine map the COs.

Two contrasted collections were constructed by selecting within the Asian and the North-western European wheat gene pools (which are genetic pools of origin for Cs and Re cultivars, respectively; Balfourier *et al.* 2007); 90 accessions for each collection (Supplemental Material, Table S3), representative of the diversity within their considered genetic pools.

Genotyping data and genetic mapping

SNPs were developed using the Cs survey sequences from International Wheat Genome Sequencing Consortium (2014) as reference to align the reads from Re and make the SNP calling. We identified 192,584 SNPs on chromosome 3B that were polymorphic between Cs and Re. SNPs were assigned *in silico* to the 3B pseudomolecule and unique loci with 100% overlap and 100% identity were selected, leading to 136,458

predicted SNPs. Finally, 22,680 supplementary SNPs derived from other projects (Axiom Array) were added to improve our capacity of SNP development. In addition, 168,211 simple sequence repeats (SSRs) were detected on the pseudomolecule using an in-house pearl script.

In total, 774 SNPs and 61 SSRs from chromosome 3B (Table S4) were used on the CsRe population, from which 466 SNPs were derived from Rimbert *et al.* (2017), 214 from Choulet *et al.* (2014), 88 transferred from Axiom 420K Array (Rimbert *et al.* 2017), and six from previous studies of recombination (Saintenac *et al.* 2011). Then, 540 SNPs and 9 SSRs were retained to refine CO localization. For the two Asian and European collections, only 96 SNPs were used (Table S4). Genotyping data using SNPs were acquired using KASPar technology. Triplex primers were designed manually according to manufacturer's instructions using informatics prediction and context sequences. Two techniques were used to process to genotyping: Fluidigm technology (microfluidics) and Roche LightCycler480. SSR marker polymorphisms were analyzed on 1% agarose gel. Whole-genome genotyping data were obtained through Axiom Array as described in Rimbert *et al.* (2017). Genetic mapping (53,287 SNPs mapped on the 21 chromosomes) was achieved according to Rimbert *et al.* (2017).

CO detection

A CO was defined on each line as a switch of parental alleles between two markers of known physical position. According to our strategy, we used scaffold information resulting in mapping/detection of markers along the chromosome 3B pseudomolecule (Choulet *et al.* 2014) to locate and refine the CO position between two markers present on the same scaffold. This strategy allows access to a confident physical position at scaffold scale. Before refinement of CO position and to certify quality and avoid chimeric scaffolds, SNPs were first genetically mapped according to Rimbert *et al.* (2017) using the whole population. Consistency of genotyping data was verified using graphical genotyping (Young and Tanksley 1989) to manually order markers according to their genetic and physical positions. These approaches revealed a few double-allele swaps, *i.e.*, successive transition between alleles of both parents concerning three consecutive markers (ABA or BAB). However, they were explained by only one SNP and the use of an F₆ progeny did not allow us to discriminate between an NCO and two close CO events. These events were thus removed from our analysis and only one allele swap supported by neighboring genotyping data (*e.g.*, ...AAABBB... or ...BBBAAA...) contributed to our CO number.

For the whole-genome analysis, we selected all contigs from the International Wheat Genome Sequencing Consortium (IWGSC) survey sequence which presented an increase of genetic distance between two successive markers. All genotyping information of these contigs and consistency within the global chromosome genetic map were manually verified as described for chromosome 3B contigs with graphical genotyping to exclude chimeric contigs and putative NCOs. We thus confirmed presence of COs supported by neighboring genotyping information and we used physical information available on

the IWGSC survey assembly (International Wheat Genome Sequencing Consortium 2014).

Coalescent analysis of wheat genetic variation

To compare contemporary and ancestral CO frequencies and hotspots in our wheat scaffolds, we applied coalescent theory (Choi *et al.* 2013; Hellsten *et al.* 2013) to a SNP data set generated from our two Asian and European populations. We used PHASE 2.1.1 (Li and Stephens 2003; Crawford *et al.* 2004) to estimate the background recombination rate parameter, ρ , and to infer hotspot position between pairs of SNPs using λ . Presence/absence of polymorphism was encoded using the available multi-allelic function, while other SNPs were encoded as biallelic markers as described in the PHASE user manual. We used the parameter MR = 0 which is the general model for recombination of PHASE 2.1.1 (Li and Stephens 2003; Crawford *et al.* 2004). The MCMC chains were run for 10,000 iterations (-X100 option). From the software output, we extracted in each interval the posterior distribution of λ . We used the median of this posterior distribution as an estimate of the interval specific recombination intensity.

Sequence information and analyses

Sequence information from the IWGSC chromosome survey sequence (CSS) contigs (International Wheat Genome Sequencing Consortium 2014) and the pseudomolecule from chromosome 3B (Choulet *et al.* 2014) were used for detection of correlation between sequence features, recombination, as well as for motif discovery. For chromosome 3B, we used the sequence annotation for genes and TEs described in Choulet *et al.* (2014). For promoter and terminator regions, we used available data when it was already defined or we considered 2-kb upstream or downstream after UTR regions. Concerning IWGSC CSS (International Wheat Genome Sequencing Consortium 2014), we used the Munich Information Center for Protein Sequences (MIPS) annotation as well as new transcript regions identified by RNA-sequencing (RNA-seq) mapping (Lloyd *et al.* 2014; see below). TEs associated to IWGSC CSS contigs were identified using RepeatMasker and ClariTE (Daron *et al.* 2014). We scanned recombinant (Rec), nonrecombinant (NoRec), and control (Overview) intervals (see Figure S1 for details of the design of intervals) for matches with probabilistic matrix provided by MEME Suite 4.10.2 (see below). Sequence analysis was performed on the Unix system using BEDtools, custom command line/script, and R (R Core Team 2014) Bioconductor (Huber *et al.* 2015) Suite with the genomeIntervals package (version 2.17.0 and version 2.25.0) (Quinlan and Hall 2010). Statistical tests of significance were based on logistical regression (GLM function on R) with a Student's *t*-test on the covariate effects.

For motif discovery, we used MEME (version 4.10.2) (Bailey and Elkan 1994) with the following parameters: motifs with nucleotide size ranging between 3 and 20 nt, use of first-order Markov model to take into account both nucleotide and dinucleotide (to capture the observed frequency of dinucleotide repeats across the genome) composition, and

Figure 1 Percentage of COs according to interval size in the Rec data set.

search for motifs on both DNA strands. MEME was also used to generate sequence logos for each discovered motif. Probabilistic matrix computed by MEME and scan_for_matches was used to scan DNA sequences to search motifs on the basis of the probabilistic matrix.

RNA-seq mapping

A total of 656,290,406 RNA-seq reads from four meiotic stages (latent/leptotene, zygotene/pachytene, diplotene/diakinesis, and metaphase I) (Lloyd *et al.* 2014; data publicly available at <http://wheat-urgi.versailles.inra.fr/Seq-Repository/Expression>) with two replicates were mapped on 86,710 CSS contigs from IWGSC (International Wheat Genome Sequencing Consortium 2014) used as a reference with TopHat2 (version 2.0.13) with zero mismatches (-m 0 -N 0 options). Mapped reads were then filtered using SAMTools and only those mapped with a minimum mapping quality of 30 were kept. Duplicates were removed using SAMTools rmdup. Mapping data were then used with Cufflinks (version 2.2.1) to build GTF files of exons for each condition and each replicate with default parameters. Each GTF file was combined with the gene models from MIPS using cuffmerge (Cufflinks package) with default parameters. Quantification of transcripts was done using cuffquant (Cufflinks package) with “rescue method” (-multi-read-correct option) and with the output from cuffmerge as transcript reference for each of the BAM files. Finally, cuffnorm was run with default parameters with the GTF file from cuffmerge as reference annotation and with the CXB files from cuffquant. Chromosome 3B meiosis RNA-seq data were established according to Pingault *et al.* (2015).

Data availability

All data are publicly available either directly in the supplemental materials, in the links above, or on request.

Results

Fine-scale distribution of current recombination on chromosome 3B

We genetically mapped 96 SNP markers from chromosome 3B on 1270 F₆ lines derived through SSD from the cross between Cs and Re cultivars (CsRe). In total, we identified 3031 CO events representing 2.38 COs per individual on average for this chromosome. We confirmed the partitioning of recombi-

nation with 82% of COs located in the distal ends of chromosome 3B representing 19% of the chromosome length (Choulet *et al.* 2014). Using the SNPs available for chromosome 3B (Rimbert *et al.* 2017) on the whole CsRe population, we located 891 COs on 73 scaffolds from the chromosome-3B pseudomolecule. These COs were framed by two markers carried by the same scaffold to avoid calculating the wrong physical distance due to unknown exact gap size between two consecutive scaffolds or error in ordering of scaffolds. The 73 scaffolds followed the same partitioning as recombination with 72% mapped in the two distal regions and an equal contribution of short/long arms (25 vs. 24). The remaining 24 scaffolds were localized in pericentromeric/centromeric regions (Figure S2). The size of the 73 scaffolds ranged from 101 to 2812 kb (Table S1).

By further genotyping only lines which presented COs with additional SNPs on 26 among the 73 scaffolds, we narrowed down the resolution of CO location to 74 intervals of <26 kb (further called “Rec” to refer to the recombinant data set) containing 252 COs. The remaining 639 COs could not be accurately mapped in <26 kb because: (1) the number of COs was found too low in too large features to be selected for additional mapping, and (2) the given region lacked polymorphism. Rec intervals ranged from 108 to 25,848 bp (average: 9344 bp). They contained up to 17 COs while 41% carried a single CO. In total, 62% of COs were located in intervals <10 kb (Figure 1). Among the 74 Rec intervals, 69 intervals (92%) were found in subtelomeric regions while only 5 were part of the central region (Figure S3).

Fitting ancestral and current recombination

Ancestral recombination was estimated by haplotype inference on two populations, each of 90 landraces, representing Asian and European genetic pools to fit best with the CsRe biparental population (see *Materials and Methods*). A recombination background rate was estimated as well as the factor (λ) by which the recombination exceeds the background rate (Li and Stephens 2003; Crawford *et al.* 2004). On the basis of the 74 intervals studied on CsRe, the two largest scaffolds were selected to perform ancestral recombination analysis. The two regions, present on scaffolds v443_0247 and v443_0914 (Figure 2), represent 1.2 and 2.5 Mb, respectively, and are located distally on the short arm (Figure 2A). The mean density of markers in these two scaffolds in the CsRe population was of one SNP per 28 and 39 kb, respectively. Additional SNPs, derived from a complementing analysis (Rimbert *et al.* 2017), were genotyped in the landraces to achieve a better precision in the estimation of ancestral recombination in these two scaffolds.

For v443_0247 (Figure 2B), three intervals (a, b, c) exhibiting COs were identified in CsRe. We observed COs overlapping the same three regions in population data. According to the ρ -map analysis, recombination intensity (λ) was higher in the European pool for intervals (a) and (b), while it was high for both pools for interval (c), resulting in an even stronger deviation from the background rate when pools were merged.

Figure 2 Comparative analysis between ancestral and current recombination. p-map is computed for: top: Asian (blue) and European (red) genetic pools as well as for the compilation of both (black) using λ deviation of 0.75 quantile. Bottom: segregating population (CsRe; green) using the number of COs observed. Physical location (in kilobase pair) along pseudomolecule is shown in abscises. (A) Distribution of recombination rate along the

Similarly, for v443_0914 (Figure 2C) three intervals with COs (d, e, f) were identified on the basis of CsRe population and two of them exhibited high λ deviations, while no λ deviation from the background rate was observed for the largest interval (e; 484 kb). Interestingly, interval (d) was highly saturated with SNPs (Figure 2C), allowing us to reach a very high resolution revealing two historically recombinogenic regions already detected in CsRe (Figure 2D). These two regions (g, h) present strong λ deviation from the background rate similarly to CsRe. However, a high λ value was observed only for the European pool for the first one (g), while for the second one (h) an elevated λ value was only revealed when both pools were analyzed together.

Although some recombinogenic regions may be unstable along the evolution, ancestral and current recombination patterns are mainly conserved. This suggests that sequence features affecting recombination exist in wheat as demonstrated in *Arabidopsis* (Choi and Henderson 2015).

COs occur more frequently in the vicinity of the coding fraction of the chromosomes

For each of the 7264 predicted genes described in Choulet *et al.* (2014), we estimated potential correlations between COs and the presence of genic features (promoter, gene body, terminator) in our 74 Rec intervals (see *Materials and Methods* for details). Because of their resolution and their size, some intervals may have a unique genic feature (promoter or gene body or terminator) while others may have two or three (see Figure S4 for details).

Among the 252 COs mapped, 179 (74%) were comprised of intervals presenting at least one genic feature. To correlate recombination rate with sequence features, a control set of 74 intervals where no CO was detected (NoRec) was defined. To avoid biasing the control set, NoRec intervals were selected in the same scaffolds, in the vicinity of Rec ones and with a similar size (see Figure S1 for details). In addition, we created an additional set of 74 control intervals homogeneously distributed along chromosome 3B (hereafter named “chromosome overview”) by selecting each ~10 Mb, 9344 bp of sequence with no recombination on 406 CsRe lines. The Rec intervals were then compared with the NoRec and chromosome-overview intervals (Figure 3). Genic features were present in 73% of the Rec intervals while they were present in only 54% of the NoRec intervals (Figure 3A), a difference which was significant (P -value < 0.05). In addition, expressed genes (FPKM greater than one during meiosis) represented 67% of the genes located in Rec intervals vs. 48% for NoRec intervals (P -value < 0.1) (Figure 3B). Chromosome-overview intervals showed a lower level of genic features (23%) and similar gene expression (50%) to NoRec intervals.

According to our results, recombination in wheat seems to occur more frequently close to or within the genes as in *Arabidopsis* and *S. cerevisiae* (Pan *et al.* 2011; Tischfield and Keeney 2012; Choi *et al.* 2013). Recombinogenic regions also carry genes which are more prone to being expressed during meiosis, while on the contrary, cold regions carry fewer and less expressed genes. Furthermore, when we achieved a sufficient resolution, eight, five, and nine COs were comprised of intervals where only one promoter, gene body, or terminator was present, respectively; indicating a tendency of recombination to occur more frequently in the promoter and terminator.

Recombinogenic regions have different composition in TEs and associated specific DNA motifs

Rec intervals contained a significantly lower density of TEs (30%) than NoRec intervals (54%) (P -value < 0.05; Figure 4A). Regarding TE composition, Rec intervals carried a similar amount of retrotransposons and DNA transposons (42 and 53%, respectively) contrary to control intervals, which contained two to four times more retrotransposons (63 and 78% for NoRec and chromosome-overview intervals, respectively) than DNA transposons (28 and 20%, respectively), highlighting a particular signature of loci where COs take place. In addition, NoRec intervals showed a significantly higher level of retrotransposons compared to Rec intervals (P -value < 0.05; Figure 4B). At the superfamily level (Figure 4C), *CACTA*, *PIF/HARBINGER*, *MUTATOR*, *MARINER*, *MITE*, and *LINE* were more frequently observed in Rec intervals, while *GYPSY* and *COPIA* were prevalent in NoRec intervals.

Previous studies in *Arabidopsis*, *Drosophila*, and dogs discovered recombination-associated motifs like A-stretch motif, CCT, CCN repeats, and CpG (Comeron *et al.* 2012; Auton *et al.* 2013; Wijnker *et al.* 2013; Shilo *et al.* 2015; Choi *et al.* 2016). To evaluate the existence of recombination-associated motifs in wheat, we searched for DNA motifs over-represented in Rec intervals without *a priori* using MEME Suite and the base frequency of the chromosome (see *Materials and Methods* for details). Subsequently, enrichment in Rec vs. NoRec intervals was evaluated as well as percentage of COs present in sequences which carry these motifs. Distribution along the 3B pseudomolecule was also analyzed. A total of 16 motifs were identified in sequences surrounding COs (Figure S5). We focused our analysis on four of them because they were found between 1.2 and 2.1 times more frequently and showed a higher copy number in Rec compared to NoRec intervals (Figure 5).

Two motifs are simple repeats, [CCG] $_n$ (P -value = 0.187) and poly-A stretches (P -value = 0.145); and two are parts of known TEs, CTCCCTCC (P -value = 0.527) in terminal inverted repeats of the *TC1/Mariner* superfamily (Zhao

pseudomolecule of chromosome 3B (Choulet *et al.* 2014). Position of the two scaffolds is indicated: (B) scaffold v443_0247, (C) scaffold v443_0914, (D) magnification on specific area where high-resolution mapping is available on scaffold v443_0914 for both ρ -maps. Breakpoints (a, b, c, d, e, f, g, h) are delimited by black boxes.

Figure 3 Comparative analysis for gene body and expression between Rec, NoRec, and OverView intervals. Repartition of (A) gene body and their expression during meiosis [(B) percentage of genes expressed; FPKM ≥ 1 (B)] in Rec, NoRec, and OverView (3B chromosome) intervals. Chr, chromosome. * $P < 0.05$, $P < 0.1$.

et al. 2016) and TTAGTCCCGGTT (P -value = 0.362) found in CACTA elements. All four motifs followed the recombination pattern and partitioning of chromosome 3B with higher proportion in the two distal regions (Figure 6; all motifs in Figure S6). Each motif was present on Rec intervals which carry 8.7 to 51.6% of the COs in our data set (Figure 5).

Data from chromosome 3B can be expended to the whole genome

The whole-genome genetic map (Rimbert *et al.* 2017) allowed us to analyze recombination on 406 F_6 lines of CsRe which led to validate 596 COs in 476 intervals of <26 kb (called RecDraft) (Table S2). In total, 90% of COs were delineated in intervals of <10 kb with 81% of intervals presenting only one CO. As for chromosome 3B, we selected control intervals devoid of recombination (NoRecDraft). Similar results as those observed for chromosome 3B were found: more genes in RecDraft intervals (66 vs. 58% in NoRecDraft) and a tendency of these genes to be expressed during meiosis (73 vs. 68%) (Table 1). The percentage of TEs was close to 10% in both intervals of the draft genome (Figure 7A). Nevertheless, the same tendency as for chromosome 3B was observed with more DNA transposons in RecDraft intervals and more Retrotransposons in NoRecDraft intervals (Figure 7B). However, analysis of TE composition revealed some similarity with chromosome 3B, like more *Mutator* elements (+1,4%) in RecDraft intervals, and some difference, like more *LINE* and *COPIA* elements in NoRecDraft intervals (Figure 7C).

Associated recombination motifs were searched on RecDraft intervals with the same approach as the one applied on chromosome 3B. A total of 10 motifs were found, among which 7 were related to the first 20 bp of the *TIR-Mariner* superfamily (Figure S7). [CCG]n was present within intervals of RecDraft where 16% of COs were mapped (Table 2). The motif related to *TIR-Mariner* as well as the A-stretch were present within intervals which carried 26% of COs in RecDraft.

Enrichment and copy numbers were close to those observed for chromosome 3B for the motif related to *TIR-Mariner* and slightly lower for [CCG]n (Table 2) while they were lower for the motif related to *CACTA* and A-stretch.

In light of the results and despite the bias generated by the size and content of the contigs, we confirmed our previous observations from chromosome 3B with the same proportions of genes and TEs in RecDraft and NoRecDraft. We had the same trend regarding gene and gene expression in meiosis, and RecDraft intervals are enriched in DNA transposons while NoRecDraft intervals are enriched in retrotransposons. Concerning the motif discovery, the related-to *TIR-Mariner*, A-stretch, and CCG motifs were confirmed to be putatively related to recombination in the whole-genome analysis, albeit with the lower enrichment factor and copy number.

Discussion

New wheat-genomic tools allow more resolute analysis of recombination pattern

Whole-genome, fine-scale recombination studies in plants have only been conducted in species for which the genome sequence was available (Wu *et al.* 2003; Liu *et al.* 2009; Paape *et al.* 2012; Rodgers-Melnick *et al.* 2015; Shilo *et al.* 2015).

Table 1 Comparative analysis between 3B pseudomolecule and draft genome

	Rec on 3B	Draft genome (21 chromosomes)
COs	252	596
Intervals	74	476
Genes in Rec (%)	66.22	66.38
Genes in NoRec (%)	41.45	58.50
Expression in Rec (%)	67.35	72.78
Expression in Norec (%)	47.62	67.98

Figure 4 Percentages of TEs in Rec, NoRec, and OverView intervals for chromosome 3B. Percentages of (A) TEs, of (B) TEs according to their classes [DNA transposons (DNA) and retrotransposons (Retro)], and of (C) TEs according to superfamilies. Data are extracted from chromosome 3B pseudomolecule: recombinant (Rec; red), not recombinant (NoRec; blue), and chromosome overview (OverView; pink) intervals, respectively. *P*-values are 0.05 (*) and 0.01 (**).

However, this mainly restricted such analyses to species with small and compact genomes with low rates of TEs. Only maize had a large genome and data showed a correlation between *Mu* TE insertion sites and CO localization (Liu *et al.* 2009). Bread wheat has a six-times-larger genome than maize and recombination at the chromosome scale is low and occurs preferentially in distal regions. It is thus positively correlated with gene density and negatively with TE content (Choulet *et al.* 2014).

We exploited the pseudomolecule of chromosome 3B (Choulet *et al.* 2014) as well as high-throughput detection of SNPs (Rimbert *et al.* 2017) to map 252 CO events in intervals of <26 kb using genotyping data. This resolution has never been reached in wheat before and was close to the one observed for the four hotspots described in rice (Wu *et al.* 2003). In the collared flycatcher, a songbird species, 279 COs were localized within intervals of <10 kb (Smeds *et al.* 2016) which was only 1.8 times more compared to our analysis (156 COs) while the wheat genome is 17 times larger. Similarly in human, genome-wide sperm sequencing allowed mapping of 13–45% of COs in 30-kb intervals (Lu *et al.* 2012; Wang *et al.* 2012) showing that even in a highly documented genome sixfold smaller than that of wheat, reaching a resolution lower than 30 kb is difficult. Such a deep analysis has never been conducted in wheat before due to the large genome size (~17 Gb) and the high proportion (~85%) of repeated elements. Analyzing 252 COs seems few compared to the 3031 detected in the whole population on chromosome 3B. However, the remaining

2779 were mainly present between or on large scaffolds, preventing their resolution despite development of numerous SNPs.

Ancestral and current recombination follow the same pattern

In avian (Singhal *et al.* 2015; Smeds *et al.* 2016) and in *Saccharomyces* species (Lam and Keeney 2015), recombination patterns are highly conserved during evolution. This led us to assay using another approach to reveal historical recombination using LD and coalescent theory (ρ -map) (Li and Stephens 2003; Crawford *et al.* 2004) in wheat. Our results revealed that the hotspots observed in collections are present in the F_6 population, confirming their stability across many generations. In our study, the choice of the genotypes could be discussed because it presented the maximum divergence between the genetic pools and selection pressure can be present, especially in the European genetic pool. Then, lack of resolution in some regions could influence the estimation of recombination intensity (λ) and the presence/absence of SNP markers can suggest some structural polymorphisms which could also influence the estimation of historical recombination rates.

However, intervals with common COs were found between the two genetic pools and CsRe population. This demonstrated that this strategy could be used with the existing data to define complete recombination-hotspot maps of bread wheat as this was done in human (McVean *et al.* 2002, 2004; Slatkin 2008), *Arabidopsis* (Choi *et al.* 2013), *Mimulus*

Figure 5 Motif detection on Rec sequence using MEME Suite. Only the four more frequent motifs are mentioned. Enrich Seq, ratio between Rec and NoRec intervals (number intervals:number intervals); Enrich Nb, ratio between Rec and NoRec intervals (number motifs:number motifs); %CO, percentage COs associated to the motif; Related, annotation of the motif; Logo, 5'–3' sequence; LogoRC, reverse complement sequence.

(Hellsten *et al.* 2013), or *Drosophila* (Smukowski Heil *et al.* 2015). Specific recombination intensities were different depending on the collection used which could be attributed to the difference in LD patterns at fine scale in the two populations, or differences arising from the complex evolutionary histories of the two genetic pools we used.

This first approach offers opportunity to use recent high-throughput genotyping data (Rimbert *et al.* 2017), but in the future massive data such as those generated by genotyping by sequencing (Deschamps *et al.* 2012) can lead to the complete recombination map in wheat, allowing the same approach of parameter detection as realized in human and *Arabidopsis* (Myers *et al.* 2005, 2008; Choi *et al.* 2013). We could thus identify recombination breakpoints located in pericentromeric regions which could be enhanced to improve recombination in these poorly recombinogenic regions.

Retrotransposons associate with reduced recombination rate

We found that NoRec intervals have a significantly higher TE content (53%) and more retro-transposons (*Gypsy* and *COPIA* elements; 63.7%) than Rec intervals. Interestingly, the percentage of *COPIA* element is even higher in NoRec intervals than in the central part of chromosome 3B (called R2 in Choulet *et al.* 2014) which is made of 70% of retro-transposons and exhibits a very low recombination rate (0.05 cM/Mb; Choulet *et al.* 2014; Daron *et al.* 2014).

Retrotransposons constitute a large part of the wheat genome (Choulet *et al.* 2014; Daron *et al.* 2014) and contributes in presence/absence variations leading to the disparity of the pattern of dispersion of genes between lines (Glover *et al.* 2015). In maize, examination of the recombination rates across the *Bronze* (*Bz1*) locus in three haplotypes differing by the presence and absence of a 26-kb intergenic retrotransposon cluster revealed that the genetic distance between the markers was twofold smaller in the presence of the retrotransposon cluster (Dooner and He 2008; He and Dooner 2009). This suggests that reduction in recombination in regions bearing a high density of retroelements can be due

to structural variations created by retrotransposon insertions or deletions.

Duret *et al.* (2000) showed that the amount of retroelements (LTR and non-LTR retrotransposons) correlates negatively with recombination rate in the *C. elegans* genome. The negative impact of LTR transposons on recombination in wheat could also be attributed to their associated epigenetic patterns which condense DNA and lock the regions, preventing their accessibility to the recombination machinery. As an example, in mouse, methylation leads to DNA conformation which prevents transposons from adopting a permissive chromatin structure for meiotic recombination (Zamudio *et al.* 2015). Finally, in barley, pericentromeric regions with low recombination rates bear LTR retrotransposons that carry the epigenetic landmarks H3K9me2 and H3K27me1 which establish a constitutive heterochromatic state (Baker *et al.* 2015). These results support the hypothesis that COs do not occur, or only very rarely, in chromosomal regions where chromatin is condensed and poorly accessible to the recombination mechanism.

It was shown that LTR retrotransposons are more frequent in pericentromeric regions of the host genomes (for a review see Li *et al.* 2013) and that they could play important roles in maintaining chromatin structures and centromere functions (Zhao and Ma 2013), or reshuffling the structure of the centromeric sequences (Wei *et al.* 2013). In wheat, a *Cereba*-like element called centromeric retrotransposon in wheat (*CRW*) represents the main component of the centromeres (Li *et al.* 2013). Meiotic recombination is almost completely suppressed at the centromeric and pericentromeric regions (Gore *et al.* 2009). We can thus speculate that centromere regions are suppressed in recombination, probably because they contain a high density of LTR retrotransposons.

Recombination and DNA transposons are located in the same regions

Analysis of data at the chromosome 3B and whole-genome scales showed that recombination more often occurs in or close to genes that tend to be expressed in meiosis. In total, 75% of the COs from chromosome 3B fall into an interval

Figure 6 Distribution of the four most frequent motifs along 3B pseudomolecule (in base pairs). Distribution is estimated along chromosome 3B pseudomolecule (774 Mb) using a sliding window of 10 Mb and a step of 1 Mb.

containing at least one gene feature (promoter, gene body, or terminator) confirming thus the results observed in yeast (Mancera *et al.* 2008). At the whole-genome level, similar results were found with 2/3 of COs overlapping genes, as this is the case for most COs in plants (reviewed in Mercier *et al.* 2015). Our COs were found in the gene body and slightly more frequently in promoters and terminators, confirming the results observed in *Arabidopsis* (Choi *et al.* 2013) as well as in avian (Singhal *et al.* 2015; Smeds *et al.* 2016) and in *Saccharomyces* species (Lam and Keeney 2015), where recombination pattern is highly conserved and localized to the promoter (TSS) and terminator (TTS) regions.

We also found a higher rate of DNA transposons, albeit not significant, in Rec intervals. A similar result was observed in the *C. elegans* genome where the amount of transposons (DNA-based elements) correlates positively with recombination rate (Duret *et al.* 2000). This could be explained by a higher survival rate of DNA transposons inserted close to genes and a remaining opportunist and strategic processes to maintain, or even under certain conditions to increase, the number of DNA transposons. An example in maize was described at the *a1* locus where *MuDR*, autonomous element of *Mutator* family which encodes a transposase, supported the

capacity of DNA transposons to influence recombination (Yandeau-Nelson *et al.* 2005).

Open chromatin structure is known to play an important role in CO formation in both yeast (Wu and Lichten 1994; Berchowitz *et al.* 2009; Borde *et al.* 2009; Pan *et al.* 2011) and mammals (Buard *et al.* 2009; Berg *et al.* 2010; Grey *et al.* 2011), especially through the trimethylation of lysine 4 of histone H3 (H3K4me3 landmark). DNA methylation also represses CO formation (Maloisel and Rossignol 1998) and is sufficient to silence CO hotspots in *Arabidopsis* (Yelina *et al.* 2015). CO distribution is largely modified in *Arabidopsis met1* and *ddm1* mutants, which show an increase of proximal recombination events and a simultaneous decrease in pericentromeric and distal regions (Colome-Tatche *et al.* 2012; Melamed-Bessudo and Levy 2012; Mirouze *et al.* 2012; Yelina *et al.* 2012). Since H3K4me3 and H2A.Z landmarks are important for promotion of gene transcription (Liu *et al.* 2009; Yelina *et al.* 2012; Choi *et al.* 2013) and are associated with high recombination rates in *Arabidopsis* (Yelina *et al.* 2012) as well as in barley (Aliyeva-Schnorr *et al.* 2015; Baker *et al.* 2015), this suggests that both recombination and DNA-transposon insertion benefit from low compaction of DNA to occur within the genome.

Figure 7 Percentages of TEs in RecDraft and NoRecDraft intervals for the whole genome. Percentages of (A) TEs, of (B) TEs according to their classes [DNA transposons (DNA) and retrotransposons (Retro)] and of (C) TEs according to superfamilies. Data are extracted from the whole genome: recombinant (RecDraft; red) and not recombinant (NoRecDraft; blue) intervals, respectively.

TE-related motifs are thus present in recombinogenic windows

We revealed four motifs on chromosome 3B and at the whole-genome level that were overrepresented in Rec intervals: two simple motifs (A-stretch and [CCG]) and two related to TEs (*TIR-Mariner* and *CACTA*). Research along the pseudomolecule of chromosome 3B for the various motifs showed a similar partitioning as recombination and preferential location of these four motifs in distal regions of the chromosome.

These two different categories (specific TE-associated motif and simple repeats) were found in previous analyses in other species (Comeron *et al.* 2012; Auton *et al.* 2013; Choi *et al.* 2013; Wijnker *et al.* 2013; Shilo *et al.* 2015). The A-stretch motif was already found in *Arabidopsis*, human, and *Drosophila* to be associated with recombination in ~2-kb areas around CO events, suggesting that A-stretches do not directly induce recombination but that they contribute to its occurrence as a result of DNA decompaction (Myers *et al.* 2008; Comeron *et al.* 2012; Shilo *et al.* 2015). A-stretches were described as likely stiff to avoid folding around the nucleosome and are thus a strong inhibitor to

nucleosome formation. Presence of A-rich motifs close to genes could regulate their expression (Struhl and Segal 2013) and may thus reflect a higher level of expression and less condensed DNA.

The second motif found in wheat is CCG that could be related to CCN found in canids, *Arabidopsis*, and *Drosophila* (Comeron *et al.* 2012; Auton *et al.* 2013; Shilo *et al.* 2015). Our complete motif was CCGCCGNCGCCGCCGCCGCC. It shares similarity with the CCNCCNTNNCCNC motif found in human hotspots corresponding to the PRDM9 site (Myers *et al.* 2008; Baudat *et al.* 2010). The CCN motif is associated with the H3K4me3 epigenetic landmark and the gene body of the genes in *Arabidopsis*, and could thus influence chromatin compaction (Shilo *et al.* 2015).

We found that related *TIR-Mariner* motifs are putatively associated with recombination. The *TC1-Mariner* element is the most diverse group and widespread superfamily in eukaryotes and it was extensively studied in animals (Feschotte and Wessler 2002). Its insertion site is in the TA sequence and, like DNA transposons, it generally locates in genic regions (Zhao *et al.* 2016). In our case, most of the related *TIR-*

Table 2 Comparative analysis of motifs on 3B pseudomolecule and on draft genome

Motif	Related	3B		Draft genome		CO (%)
		Enrich sequence	Enrich no.	Enrich sequence	Enrich no.	
1	<i>TIR-Mariner</i>	1.2	1.3	1.2	1.2	25.8
2	<i>CACTA</i>	1.9	2	1.3	1.1	0.8
3	CCG	2.1	2.3	1.8	1.8	15.9
4	A-stretch	1.3	1.8	1.1	1.1	25.8

Mariner motifs detected that are present in Rec intervals are found close to gene features (73% for chromosome 3B).

***TIR-Mariner* motif is similar to those described for PRDM9**

The related *TIR-Mariner* superfamily motif tends to be more present in the Rec intervals that we defined on chromosome 3B and at the whole-genome scale. Interestingly, part of our related *TIR-Mariner* motif (TCCCTCC) is similar to the PRDM9 motif (CCTCCCT) driving DSB location in human (Pratto *et al.* 2014) and contains five common bases (TCCCT). The CCTCCCT motif is more frequent in human hotspots and is overrepresented (five- to sixfold) in the primate-specific retrotransposon *THE1A/B* consensus (Myers *et al.* 2005, 2008). PRDM9 leads recombination by ensuring H3K4me3 methylation (Baudat *et al.* 2013). It seems to be very specific to a few species and is rather rare among eukaryotes. PRDM9 is absent in plants (Zhang and Ma 2012), birds (Muñoz-Fuentes *et al.* 2011), *Drosophila* (Heil and Noor 2012), and even in some mammals such as the canids (Muñoz-Fuentes *et al.* 2011; Auton *et al.* 2013). On the contrary, recombination is essentially universal.

In canids where PRDM9 is absent, recombination occurs in CpG-rich regions around promoters with little association with H3K4me3 marks (Auton *et al.* 2013). In mouse *prdm9* mutants, recombination hotspots are reverted to gene promoters as it is observed in yeast, plants, and in our analysis, suggesting a functional gain to specify location of recombination (Brick *et al.* 2012).

Conclusion

Until now, no study regarding recombination hotspots was done at this scale and resolution in bread wheat. With a large segregating population (1270 individuals), we were able to map ~900 CO events using the latest genomic sequences available. Availability of the full genome sequence and use of another population will allow determining more precisely hot and cold regions in bread wheat because, in light of previous studies, we obviously did not capture all the recombination breakpoints and ancestral recombination could be an excellent alternative since we showed that the correlation with current recombination is elevated, especially when the resolution is high.

Recombination in wheat occurs more frequently in regions associated with genes, especially with promoters or terminators. These regions are also enriched in DNA transposons and especially from the *Mariner* family. We identified a correlation between recombination events and a motif associated with the *TIR* sequence of *Mariner* similar to the PRDM9 motif which is responsible for most COs in mammals. Despite the fact that there is yet no known PRDM9 homolog in plants, the association we found with these motifs suggests that both systems could be functionally related.

Acknowledgments

The authors are grateful to Alain Loussert who performed all the genotyping experiments as well as to personnel from

Génotypage et Séquençage en Auvergne (GENTYANE) for technical help. The authors thank Christine Mézard and Ian Henderson for help during redaction and for discussions. The seeds of both Asian and European accessions were provided by the Biological Resources Center for small grain cereals (Institut National de la Recherche Agronomique Clermont-Ferrand). The BREEDWHEAT project and the International Wheat Genome Sequencing Consortium are acknowledged for providing SNP (Axiom) genotyping data and wheat genome sequences, respectively. B.D. was supported by Institut National de la Recherche Agronomique (meta-program SELGEN) and Region Auvergne.

Literature Cited

- Aliyeva-Schnorr, L., S. Beier, M. Karafiátová, T. Schmutzer, U. Scholz *et al.*, 2015 Cytogenetic mapping with centromeric bacterial artificial chromosomes contigs shows that this recombination-poor region comprises more than half of barley chromosome 3H. *Plant J.* 84: 385–394.
- Auton, A., Y. Rui Li, J. Kidd, K. Oliveira, J. Nadel *et al.*, 2013 Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet.* 9: e1003984.
- Bailey, T. L., and C. Elkan, 1994 Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2: 28–36.
- Baker, K., T. Dhillon, I. Colas, N. Cook, I. Milne *et al.*, 2015 Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me1 and H3K27me3 abundance. *Plant J.* 84: 111–124.
- Balfourier, F., V. Roussel, P. Strelchenko, F. Exbrayat-Vinson, P. Sourdille *et al.*, 2007 A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor. Appl. Genet.* 114: 1265–1275.
- Barrón, M. G., A.-S. Fiston-Lavier, D. A. Petrov, and J. González, 2014 Population genomics of transposable elements in *Drosophila*. *Annu. Rev. Genet.* 48: 561–581.
- Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober *et al.*, 2010 PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327: 836–840.
- Baudat, F., Y. Imai, and B. de Massy, 2013 Meiotic recombination in mammals: localization and regulation. *Nat. Rev. Genet.* 14: 794–806.
- Berchowitz, L. E., S. E. Hanlon, J. D. Lieb, and G. P. Copenhaver, 2009 A positive but complex association between meiotic double-strand break hotspots and open chromatin in *Saccharomyces cerevisiae*. *Genome Res.* 19: 2245–2257.
- Berg, I. L., R. Neumann, K.-W. G. Lam, S. Sarbajna, L. Odenthal-Hesse *et al.*, 2010 PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.* 42: 859–863.
- Borde, V., and B. de Massy, 2013 Programmed induction of DNA double strand breaks during meiosis: setting up communication between DNA and the chromosome structure. *Curr. Opin. Genet. Dev.* 23: 147–155.
- Borde, V., N. Robine, W. Lin, S. Bonfils, V. Géli *et al.*, 2009 Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* 28: 99–111.
- Brick, K., F. Smagulova, P. Khil, R. D. Camerini-Otero, and G. V. Petukhova, 2012 Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485: 642–645.
- Buard, J., P. Barthès, C. Grey, and B. de Massy, 2009 Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. *EMBO J.* 28: 2616–2624.

- Chen, S. Y., T. Tsubouchi, B. Rockmill, J. S. Sandler, D. R. Richards *et al.*, 2008 Global analysis of the meiotic crossover landscape. *Dev. Cell* 15: 401–415.
- Choi, K., and I. R. Henderson, 2015 Meiotic recombination hotspots - a comparative view. *Plant J.* 83: 52–61.
- Choi, K., X. Zhao, K. A. Kelly, O. Venn, J. D. Higgins *et al.*, 2013 *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat. Genet.* 45: 1327–1336.
- Choi, K., C. Reinhard, H. Serra, P. A. Ziolkowski, C. J. Underwood *et al.*, 2016 Recombination rate heterogeneity within *Arabidopsis* disease resistance genes. *PLoS Genet.* 12: e1006179.
- Choo, K. H., 1998 Why is the centromere so cold? *Genome Res.* 8: 81–82.
- Choulet, F., A. Alberti, S. Theil, N. Glover, V. Barbe *et al.*, 2014 Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345: 1249721.
- Colome-Tatche, M., S. Cortijo, R. Wardenaar, L. Morgado, B. Lahouze *et al.*, 2012 Features of the *Arabidopsis* recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proc. Natl. Acad. Sci. USA* 109: 16240–16245.
- Cameron, J. M., R. Ratnapan, and S. Bailin, 2012 The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet.* 8: 33–35.
- Crawford, D. C., T. Bhangale, N. Li, G. Hellenthal, M. J. Rieder *et al.*, 2004 Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* 36: 700–706.
- Crismani, W., C. Girard, N. Froger, M. Pradillo, J. L. Santos *et al.*, 2012 FANCM limits meiotic crossovers. *Science* 336: 1588–1590.
- Da Ines, O., F. Degroote, S. Amiard, C. Goubely, M. E. Gallego *et al.*, 2013 Effects of *XRCC2* and *RAD51B* mutations on somatic and meiotic recombination in *Arabidopsis thaliana*. *Plant J.* 74: 959–970.
- Daron, J., N. Glover, L. Pingault, S. Theil, V. Jamilloux *et al.*, 2014 Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biol.* 15: 546–560.
- Darvasi, A., A. Weinreb, V. Minke, J. I. Weller, and M. Soller, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134: 943–951.
- de Massy, B., 2013 Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. *Annu. Rev. Genet.* 47: 563–599.
- Deschamps, S., V. Llaca, and G. D. May, 2012 Genotyping-by-sequencing in plants. *Biology (Basel)* 1: 460–483.
- Dooner, H. K., and L. He, 2008 Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. *Plant Cell* 20: 249–258.
- Dooner, H. K., and I. M. Martínez-Férez, 1997 Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. *Plant Cell* 9: 1633–1646.
- Drouaud, J., and C. Mézard, 2011 Characterization of meiotic crossovers in pollen from *Arabidopsis thaliana*. *Methods Mol. Biol.* 745: 223–249.
- Drouaud, J., H. Khademian, L. Giraut, V. Zanni, S. Bellalou *et al.*, 2013 Contrasted patterns of crossover and non-crossover at *Arabidopsis thaliana* meiotic recombination hotspots. *PLoS Genet.* 9: e1003922.
- Duret, L., G. Marais, and C. Biéumont, 2000 Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. *Genetics* 156: 1661–1669.
- Endo, T. R., and B. S. Gill, 1996 The deletion stocks of common wheat. *J. Hered.* 87: 295–307.
- Feschotte, C., and S. R. Wessler, 2002 Mariner-like transposases are widespread and diverse in flowering plants. *Proc. Natl. Acad. Sci. USA* 99: 280–285.
- Fu, H., W. Park, X. Yan, Z. Zheng, B. Shen *et al.*, 2001 The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. *Proc. Natl. Acad. Sci. USA* 98: 8903–8908.
- Fu, H., Z. Zheng, and H. K. Dooner, 2002 Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. *Proc. Natl. Acad. Sci. USA* 99: 1082–1087.
- Girard, C., W. Crismani, N. Froger, J. Mazel, A. Lemhemdi *et al.*, 2014 FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers. *Nucleic Acids Res.* 42: 9087–9095.
- Girard, C., L. Chelysheva, S. Choinard, N. Froger, N. Macaisne *et al.*, 2015 AAA-ATPase FIDGETIN-LIKE 1 and helicase FANCM antagonize meiotic crossovers by distinct mechanisms. *PLoS Genet.* 11: e1005369 (erratum: *PLoS Genet.* 11: e1005448).
- Glover, N., J. Daron, L. Pingault, K. Vandepoele, E. Paux *et al.*, 2015 Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. *Genome Biol.* 16: 188–200.
- Gore, M. A., J.-M. Chia, R. J. Elshire, Q. Sun, E. S. Ersoz *et al.*, 2009 A first-generation haplotype map of maize. *Science* 326: 1115–1117.
- Grey, C., P. Barthès, G. Chauveau-Le Fric, F. Langa, F. Baudat *et al.*, 2011 Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol.* 9: e1001176.
- He, L., and H. K. Dooner, 2009 Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for *Helitron* and retrotransposon insertions. *Proc. Natl. Acad. Sci. USA* 106: 8410–8416.
- Heil, C. S. S., and M. A. F. Noor, 2012 Zinc finger binding motifs do not explain recombination rate variation within or between species of *Drosophila*. *PLoS One* 7: e45055.
- Hellsten, U., K. M. Wright, J. Jenkins, S. Shu, Y. Yuan *et al.*, 2013 Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proc. Natl. Acad. Sci. USA* 110: 19478–19482.
- Huang, B. E., K. L. Verbyla, A. P. Verbyla, C. Raghavan, V. K. Singh *et al.*, 2015 MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet.* 128: 999–1017.
- Huber, W., V. J. Carey, R. Gentleman, S. Anders, M. Carlson *et al.*, 2015 Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12: 115–121.
- International Wheat Genome Sequencing Consortium, 2014 A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345: 1251788.
- Jones, G. H., 1984 The control of chiasma distribution. *Symp. Soc. Exp. Biol.* 38: 293–320.
- Knoll, A., J. D. Higgins, K. Seeliger, S. J. Reha, N. J. Dangel *et al.*, 2012 The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in *Arabidopsis*. *Plant Cell* 24: 1448–1464.
- Lam, I., and S. Keeney, 2015 Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. *Science* 350: 932–937.
- Levin, H. L., and J. V. Moran, 2011 Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12: 615–627.
- Lewontin, R. C., and K. Kojima, 1960 The evolutionary dynamics of complex polymorphisms. *Evolution* 14: 458–472.
- Li, B., F. Choulet, Y. Heng, W. Hao, E. Paux *et al.*, 2013 Wheat centromeric retrotransposons: the new ones take a major role in centromeric structure. *Plant J.* 73: 952–965.

- Li, N., and M. Stephens, 2003 Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165: 2213–2233.
- Liu, S., C.-T. Yeh, T. Ji, K. Ying, H. Wu *et al.*, 2009 *Mu* transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet.* 5:e1000733.
- Lloyd, A. H., M. Ranoux, S. Vautrin, N. Glover, J. Fourment *et al.*, 2014 Meiotic gene evolution: can you teach a new dog new tricks? *Mol. Biol. Evol.* 31: 1724–1727.
- Lu, S., C. Zong, W. Fan, M. Yang, J. Li *et al.*, 2012 Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* 338: 1627–1630.
- Lukaszewski, A. J., and C. A. Curtis, 1993 Physical distribution of recombination in B-genome chromosomes of tetraploid wheat. *Theor. Appl. Genet.* 86: 121–127.
- Lukaszewski, A. J., D. Kopecky, and G. Linc, 2012 Inversions of chromosome arms 4AL and 2BS in wheat invert the patterns of chiasma distribution. *Chromosoma* 121: 201–208.
- Maloisel, L., and J. L. Rossignol, 1998 Suppression of crossing-over by DNA methylation in *Ascobolus*. *Genes Dev.* 12: 1381–1389.
- Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454: 479–485.
- Mayer, K. F. X., R. Waugh, P. Langridge, T. J. Close, R. P. Wise *et al.*, 2012 A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491: 711–716.
- McVean, G., P. Awadalla, and P. Fearnhead, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160: 1231–1241.
- McVean, G. A., S. R. Myers, S. Hunt, P. Deloukas, D. R. Bentley *et al.*, 2004 The fine-scale structure of recombination rate variation in the human genome. *Science* 304: 581–584.
- Melamed-Bessudo, C., and A. A. Levy, 2012 Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 109: E981–E988.
- Mercier, R., C. Mézard, E. Jenczewski, N. Macaisne, and M. Grelon, 2015 The molecular biology of meiosis in plants. *Annu. Rev. Plant Biol.* 66: 297–327.
- Mirouze, M., M. Lieberman-Lazarovich, R. Aversano, E. Bucher, J. Nicolet *et al.*, 2012 Loss of DNA methylation affects the recombination landscape in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 109: 5880–5885.
- Muñoz-Fuentes, V., A. Di Rienzo, and C. Vilà, 2011 *Prdm9*, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes. *PLoS One* 6: e25498.
- Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310: 321–324.
- Myers, S., C. Freeman, A. Auton, P. Donnelly, and G. McVean, 2008 A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.* 40: 1124–1129.
- Paape, T., P. Zhou, A. Branca, R. Briskine, N. Young *et al.*, 2012 Fine-scale population recombination rates, hotspots, and correlates of recombination in the *Medicago truncatula* genome. *Genome Biol. Evol.* 4: 726–737.
- Pace, J. K., and C. Feschotte, 2007 The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res.* 17: 422–432.
- Pan, J., M. Sasaki, R. Kniewel, H. Murakami, H. G. Blitzblau *et al.*, 2011 A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* 144: 719–731.
- Pan, Q., F. Ali, X. Yang, J. Li, and J. Yan, 2012 Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. *PLoS One* 7: 1–9.
- Paris, J. D., K. M. Haen, and B. S. Gill, 2000 Saturation mapping of a gene-rich recombination hot spot region in wheat. *Genetics* 154: 823–835.
- Paterson, A. H., J. E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood *et al.*, 2009 The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551–556.
- Paux, E., P. Sourdille, J. Salse, C. Sainetnac, F. Choulet *et al.*, 2008 A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* 322: 101–104.
- Pingault, L., F. Choulet, A. Alberti, N. Glover, P. Wincker *et al.*, 2015 Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. *Genome Biol.* 16: 29–43.
- Pratto, F., K. Brick, P. Khil, F. Smagulova, G. V. Petukhova *et al.*, 2014 DNA recombination: recombination initiation maps of individual human genomes. *Science* 346: 1256442.
- Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- R Core Team, 2014 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rimbert, H., B. Darrier, J. Navarro, J. Kitt, F. Choulet *et al.*, 2017 High throughput SNP discovery and genotyping in hexaploid wheat. *PLoS One* (in press).
- Rizzon, C., G. Marais, M. Gouy, and C. Biemont, 2002 Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Res.* 12: 400–407.
- Rodgers-Melnick, E., P. J. Bradbury, R. J. Elshire, J. C. Glaubitz, C. B. Acharya *et al.*, 2015 Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc. Natl. Acad. Sci. USA* 112: 3823–3828.
- Saintenac, C., M. Falque, O. C. Martin, E. Paux, C. Feuillet *et al.*, 2009 Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). *Genetics* 181: 393–403.
- Saintenac, C., S. Faure, A. Remay, F. Choulet, C. Ravel *et al.*, 2011 Variation in crossover rates across a 3-Mb contig of bread wheat (*Triticum aestivum*) reveals the presence of a meiotic recombination hotspot. *Chromosoma* 120: 185–198.
- Segal, E., and J. Widom, 2009 Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* 19: 65–71.
- Séguéla-Arnaud, M., W. Crismani, C. Larchevêque, J. Mazel, N. Froger *et al.*, 2015 Multiple mechanisms limit meiotic crossovers: TOP3 α and two BLM homologs antagonize crossovers in parallel to FANCM. *Proc. Natl. Acad. Sci. USA* 112: 4713–4718.
- Shilo, S., C. Melamed-Bessudo, Y. Dorone, N. Barkai, and A. A. Levy, 2015 DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in *Arabidopsis*. *Plant Cell* 27: 2427–2436.
- Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn *et al.*, 2015 Stable recombination hotspots in birds. *Science* 350: 928–932.
- Slatkin, M., 2008 Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* 9: 477–485.
- Slotkin, R. K., and R. Martienssen, 2007 Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8: 272–285.
- Smeds, L., C. F. Mugal, A. Qvarnström, and H. Ellegren, 2016 High-resolution mapping of crossover and non-crossover

- recombination events by whole-genome re-sequencing of an avian pedigree. *PLoS Genet.* 12: e1006044.
- Smit, A. F. A., 1993 Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* 21: 1863–1872.
- Smukowski Heil, C. S., C. Ellison, M. Dubin, and M. A. F. Noor, 2015 Recombining without hotspots: a comprehensive evolutionary portrait of recombination in two closely related species of *Drosophila*. *Genome Biol. Evol.* 7: 2829–2842.
- Struhl, K., and E. Segal, 2013 Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20: 267–273.
- Stumpf, M. P. H., and G. A. T. McVean, 2003 Estimating recombination rates from population-genetic data. *Nat. Rev. Genet.* 4: 959–968.
- Tenaillon, M. I., M. C. Sawkins, L. K. Anderson, S. M. Stack, J. Doebley *et al.*, 2002 Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics* 162: 1401–1413.
- Tischfield, S. E., and S. Keeney, 2012 Scale matters: the spatial correlation of yeast meiotic DNA breaks with histone H3 trimethylation is driven largely by independent. *Cell Cycle* 11: 1496–1503.
- Wang, J., H. C. Fan, B. Behr, and S. R. Quake, 2012 Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150: 402–412.
- Wei, L., M. Xiao, Z. An, B. Ma, A. S. Mason *et al.*, 2013 New insights into nested long terminal repeat retrotransposons in *Brassica* species. *Mol. Plant* 6: 470–482.
- Wijnker, E., G. Velikkakam James, J. Ding, F. Becker, J. R. Klasen *et al.*, 2013 The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife* 2: e01426.
- Wu, J., H. Mizuno, M. Hayashi-Tsugane, Y. Ito, Y. Chiden *et al.*, 2003 Physical maps and recombination frequency of six rice chromosomes. *Plant J.* 36: 720–730.
- Wu, T. C., and M. Lichten, 1994 Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science* 263: 515–518.
- Yandea-Nelson, M. D., Q. Zhou, H. Yao, X. Xu, B. J. Nikolau *et al.*, 2005 *MuDR* transposase increases the frequency of meiotic crossovers in the vicinity of a *Mu* insertion in the maize *a1* gene. *Genetics* 169: 917–929.
- Yao, H., and P. S. Schnable, 2005 *Cis*-effects on meiotic recombination across distinct *a1-sh2* intervals in a common *zea* genetic background. *Genetics* 170: 1929–1944.
- Yelina, N. E., K. Choi, L. Chelysheva, M. Macaulay, B. de Snoo *et al.*, 2012 Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. *PLoS Genet.* 8: e1002844.
- Yelina, N. E., C. Lambing, T. J. Hardcastle, X. Zhao, B. Santos *et al.*, 2015 DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in *Arabidopsis*. *Genes Dev.* 29: 2183–2202.
- Young, N. D., and S. D. Tanksley, 1989 Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theor. Appl. Genet.* 77: 95–101.
- Zamudio, N., J. Barau, A. Teissandier, M. Walter, M. Borsos *et al.*, 2015 DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev.* 29: 1256–1270.
- Zhang, L., and H. Ma, 2012 Complex evolutionary history and diverse domain organization of SET proteins suggest divergent regulatory interactions. *New Phytol.* 195: 248–263.
- Zhao, D., A. A. Ferguson, and N. Jiang, 2016 What makes up plant genomes: the vanishing line between transposable elements and genes. *Biochem. Biophys. Acta.* 1859: 366–380.
- Zhao, M., and J. Ma, 2013 Co-evolution of plant LTR-retrotransposons and their host genomes. *Protein Cell* 4: 493–501.

Communicating editor: A. Houben