Received: 4 May 2020 | Accepted: 8 September 2020

DOI: 10.1002/tpg2.20065

The Plant Genome

ORIGINAL RESEARCH

Sequence analysis of wheat subtelomeres reveals a high polymorphism among homoeologous chromosomes

Miguel Aguilar[1] | **Pilar Prieto**[2]

[1] Área de Fisiología Vegetal. Universidad de Córdoba. Campus de Rabanales, edif. C4, 3ª planta, Córdoba, Spain

[2] Plant Breeding Department, Institute for Sustainable Agriculture, Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC), Alameda del Obispo s/n, Apartado 4084, Córdoba 14080, Spain

Correspondence
Pilar Prieto, Plant Breeding Department, Institute for Sustainable Agriculture, Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC), Alameda del Obispo s/n, Apartado 4084, 14080 Córdoba, Spain.
Email: pilar.prieto@ias.csic.es

Funding information
Spanish Ministerio de Economía y Competitividad (MINECO), Grant number: AGL2015-64833R, and The European Regional Development Fund (FEDER) from the European Union

Abstract

Bread wheat, *Triticum aestivum* L., is one of the most important crops in the world. Understanding its genome organization (allohexaploid; AABBDD; 2n = 6x = 42) is essential for geneticists and plant breeders. Particularly, the knowledge of how homologous chromosomes (equivalent chromosomes from the same genome) specifically recognize each other to pair at the beginning of meiosis, the cellular process to generate gametes in sexually reproducing organisms, is fundamental for plant breeding and has a big influence on the fertility of wheat plants. Initial homologous chromosome interactions contribute to specific recognition and pairing between homologues at the onset of meiosis. Understanding the molecular basis of these critical processes can help to develop genetic tools in a breeding context to promote interspecific chromosome associations in hybrids or interspecific genetic crosses to facilitate the transfer of desirable agronomic traits from related species into a crop like wheat. The terminal regions of chromosomes, which include telomeres and subtelomeres, participate in chromosome recognition and pairing. We present a detailed molecular analysis of subtelomeres of wheat chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS. Results showed a high polymorphism in the subtelomeric region among homoeologues (equivalent chromosomes from related genomes) for all the features analyzed, including genes, transposable elements, repeats, GC content, predicted CpG islands, recombination hotspots and targeted sequence motifs for relevant DNA-binding proteins. These polymorphisms might be the molecular basis for the specificity of homologous recognition and pairing in initial chromosome interactions at the beginning of meiosis in wheat.

1 | INTRODUCTION

Wheat is one of the main food crops in the world. Understanding its genetics and genome organization is of high interest for geneticists and plant breeders. In the framework of breeding, the study of interspecific chromosome associations in hybrids or in interspecific genetic crosses, which are developed to introgress desirable agronomic traits from related species into crops like wheat, is necessary. Interspecific hybridization can allow plant breeders to develop new genetic combinations to be used directly as new crops or indirectly to widen the genetic basis of actual crops. However, unfortunately, the genome stability

of hexaploid wheat, which behaves as a diploid during meiosis, prevents chromosome pairing among wheat and related species carrying desirable traits, having negative effects in a plant breeding framework. Researchers can contribute to allow the manipulation of such chromosome associations for plant breeding purposes by studying the genetic factors controlling chromosome recognition and pairing at the beginning of wheat meiosis.

Although there is a lot of information related to synaptonemal complex formation, recombination and chromosome segregation, how chromosomes specifically identify a partner to properly pair for further recombination and segregation remains to be elucidated. Homologous chromosome recognition must occur at the beginning of meiosis, and in wheat, a polyploid organism, this event must be efficiently controlled, as each chromosome needs to distinguish its homologue from the homoeologous chromosomes of the related genomes. In wheat, the *Ph1* gene suppresses recombination between homoeologous chromosomes (Griffiths et al., 2006; Martín, Rey, Shaw, & Moore, 2017; Okamoto, 1957; Rey et al., 2017; Riley & Chapman, 1958; Riley & Chapman, 1964; Riley & Kempana, 1963; Sears & Okamoto, 1958) although it does not prevent chromosome associations between homoeologues (Calderón, Rey, Martín, & Prieto, 2018).

Studies of initial chromosome recognition and pairing remain difficult because these are highly dynamic processes, which occur only between some regions of the chromosomes and might not be synchronized from one nucleus to the other (Zickler, 2006). It is necessary to mention here that "pairing" refers to homologous associations before the initiation of the synaptonemal complex, which stabilizes initial chromosome interactions between homologues for synapsis and recombination. Extensive connections involving the early events occurring at the beginning of meiosis (pairing, synapsis and recombination) are significant across phyla, although genome sizes can differ by orders of magnitude, for example if we compare yeast, *Drosophila*, *Caenorhabditis elegans*, plants, mouse and humans.

The telomere-mediated reorganization occurring at the onset of meiosis seems to be a widely conserved first step in the homology search process of chromosomes in higher eukaryotes (Blokhina et al., 2019; Scherthan, 2001; Zickler, 2006). Telomeres are chromosome ends and have a greatly conserved sequence across eukaryotes. In many species, including wheat, telomeres associate and cluster at the nuclear envelope at early stages of meiosis to form a structure, the "bouquet", which brings the chromosome ends close together, facilitating initial interactions between homologues for recognition and pairing to allow further recombination (Bass et al., 2000; Naranjo, 2014; Page & Hawley, 2003; Scherthan, 2001; Zickler & Kleckner, 1998).

Core Ideas

- How homologous chromosomes specifically recognize each other to pair in meiosis is essential for plant breeding.
- The chromosome ends (telomeres and subtelomeres) participate in recognition and pairing.
- A molecular analysis of wheat subtelomeres showed a high polymorphism for all the features studied.
- The subtelomeric polymorphism might contribute to chromosome recognition in wheat meiosis.

Homologues start associating in pairs at the terminal region of chromosomes, where telomeres and subtelomeres (adjacent to telomeres) are (Bass et al., 2000; Prieto, Shaw, & Moore, 2004). Nevertheless, the precise function of these two chromosome regions in initial recognition and pairing is still unclear. In fact, although the telomere bouquet formation seems to assist specific pairing between homologous chromosomes, the molecular mechanism underlying the initial interactions among chromosomes that end up with specific homologous chromosome interactions and pairing remains as a meiosis enigma.

Once chromosomes are in close proximity in the bouquet, other chromosome regions should be considered for chromosome specificity, because the telomeric DNA sequence itself is highly conserved. Subtelomeres are an interesting target to be studied, but the polymorphic nature of subtelomeric regions are technically a challenge. Indeed, subtelomeres are less evolutionary conserved than telomeres and gene-rich, delimited between the telomeric repeats and chromosome-specific sequences and include recombination hot spots (Emden et al., 2019; Linardopoulou et al., 2005; Louis et al., 2005). All these features have complicated the assessment of the potential conserved roles of subtelomeric regions. These DNA segments and the proteins associated with them also play a role in genome stability and replication (Emden et al., 2019; Rietchman, Ambrosini, & Paul, 2005). The distal regions of chromosomes are also important for other essential biological activities such as cell cycle regulation, chromosome dynamics and regulation of subtelomeric genes transcription (Blackburn, 2005; Feuerbach et al., 2002).

Most of the studies related to subtelomeres have been limited to the distal 500 Kb of each chromosome arm in the species studied, including humans, *Arabidopsis* and rice (Kuo, Olsen, & Richards, 2006; Macina et al., 1994; Macina

et al., 1995; Mizuno et al., 2006; Mizuno et al., 2008a). *Arabidopsis thaliana* subtelomeres are simple and small, according to its small genome size and scarcity of repeats (Kuo et al., 2006). Although some blocks of low copy sequences are detected among non-homologous chromosomes (Heacock, Spangler, Riha, Puizina, & Shippen, 2004; Kotani, Hosouchi, & Tsuruoka, 1999), most of the subtelomeric regions do not share similarities among most non-homologous chromosomes in *Arabidopsis*, such as that seen in yeast and humans (Linardopoulou et al., 2005; Louis, Naumova, Lee, Naumov, & Haber, 1994; Mefford & Trask, 2002). The analysis of rice subtelomeres showed high variability of sequences such as tandem repeats, mobile elements, coding sequences or genes (Fan et al., 2008; Mizuno, Wu, & Matsumoto, 2014). In other species such as rye, subtelomeres are mainly organized in large heterochromatic blocks (Evtushenko et al., 2016). In barley, tobacco, tomato and wheat, species-specific subtelomeric repeats have also been identified (Fajkus, Kovarík, Královics, & Bezděk, 1995; Ganal, Lapitan, & Tanksley, 1991; Mao, Devos, Zhu, & Gale, 1997; Prieto, Martin, & Cabrero, 2004; Röder, Lapitan, Sorrells, & Tanksley, 1993; Salina et al., 2009). Nevertheless, and to the best of our knowledge, the molecular organization of subtelomeres is still unclear.

The high polymorphism of subtelomeres has outlined several hypotheses based on their function in chromosome dynamics and stability. For example, in rice, subtelomeres might facilitate recombination and the insertion of transposons (Fan et al., 2008). Interesting enough, in wheat and rye, recombining regions are more frequently involved in chromosome recognition and pairing between homologues than crossover-poor regions (Valenzuela, Perera, & Naranjo, 2013). In fact, recombination occurs widely in the subtelomeric region in all eukaryote kingdoms including plants (Gaut, Wright, Rizzon, Dvorak, & Anderson, 2007; Linardopoulou et al., 2005; Louis et al., 1994; Rudd et al., 2007). In addition, subtelomeric recombination has been shown particularly common in non-coding sequences in protozoa and fungi (Corcoran, Thompson, Walliker, & Kemp, 1988; Glover, Alsford, & Horn, 2013; Rehmeyer et al., 2006), although it also occurs within coding sequences (Fan et al., 2008; Kraemer et al., 2007).

Related to meiosis, the function that subtelomeres might play in specific homologous recognition has been poorly studied, although some stimulating works suggested they might play a key role during meiosis. For example, recent experiments in the Zebra fish *Danio rerio* have suggested the implications of subtelomeric regions in initial homologous recognition at the onset of meiosis, showing presynaptic chromosome co-alignment near the telomeric region, even when telomeres are proximal but not fully associated (Blokhina et al., 2019). The recent model of a large-scale organization and evolutionary dynamics of the subtelomeric regions has gradually been documented in yeast and humans (Anderson, Song, & Langley, 2008). In plants, information concerning subtelomeres and their possible functions during early meiosis is scarce and most of the works are related to *Arabidopsis* and rice. In *Arabidopsis*, the evidences regarding the subtelomeric regions indicate a potential function on chromosome specific interaction between homologues at the beginning of meiosis (Kuo et al., 2006). In wheat, the availability of an extra wild barley pair of homologous chromosomes carrying terminal deletions in this polyploid background (Said, Recio, & Cabrera, 2012) have contributed to shed some light on the implications of subtelomeres in chromosome recognition and pairing (Calderón, Rey, Cabrera, & Prieto, 2014). Fluorescence *in situ* hybridization experiments has revealed that homologous chromosomes failed to recognize each other and did not initiate chromosome pairing when the subtelomeric sequences were absent (Calderón et al., 2014). Furthermore, chromatin remodeling in the chromosome arms without subtelomeres also failed, which implied a delay in pairing and the absence of recombination in these distal chromosome regions (Calderón et al., 2014). The importance of the subtelomeric regions in recombination was also confirmed by other authors in wheat lines carrying a distal deletion of any chromosome arm, which did not recuperate the same level of chiasma frequency as the complete chromosome (Naranjo, 2015).

Another key aspect to understand the nature of wheat subtelomeres and their putative function on initial chromosome associations is how DNA is folded and organized in the nucleus through meiosis, which implies not only the DNA sequence but also the proteins interacting with it. For example, it is well-known that meiotic cohesins are crucial for chromatin compaction (Kleckner, Zhang, Weiner, & Zickler, 2012; Zhu & Wang, 2019). The absence of meiotic cohesins modified the chromosome axis structures and therefore, chromosome associations between homologues failed, suggesting that meiotic cohesion built on chromosome axis structures is essential for the correct association of homologues (Ding et al., 2016). Nevertheless, the impact of these axis structures on initial chromosome interactions for specific recognition and pairing remains to be elucidated. Other architectural proteins like CTCF and Ying Yang 1, which have a critical function connecting higher-order chromatin folding in animals, could also contribute to facilitate initial homologous chromosome interactions (Beagan et al., 2017; Wu, Cheng, Li, Wang, & Liu, 2012) and might play a putative role in meiosis. In fact, CTCF colocalizes with cohesin subunits at many chromosomal positions and form DNA loops (Hirano, 2006; Nasmyth & Haering, 2005; Rubio et al., 2008; Wendt et al.,

2008). Although CTCF has not been found in plants (Ong & Corces, 2014), other orthologous plant protein(s) might be playing a similar function. Other proteins such as the high mobility group proteins HMG were also suggested as meiotic actors in initial steps of homologous chromosome pairing through their presumed interaction at AT-rich sites (Subirana & Messeguer, 2011). Thus, it seems realistic to study the presence and distribution of the putative binding sites of these proteins in wheat subtelomeric regions.

Recently, an annotated reference sequence has been published, covering the genome of hexaploid bread wheat in the form of 21 chromosome-like sequence assemblies and allowing the study of 107,891 high-confidence genes, which also comprise their genomic context of regulatory sequences (Alaux et al., 2018; IWGSC, 2014; IWGSC, 2018). In contrast, very little information about bread wheat subtelomeres is available, even in these relevant publications. Thus, the identification and molecular characterization of the full subtelomeric sequences in all wheat chromosome arms remains a challenge and further work on the annotation of the wheat genome sequence should be carried out.

It seems clear that subtelomeres might be the target of a deep molecular analysis to shed some light on their structure and any peculiar feature that could potentially be related to initial chromosome recognition and pairing. Here we have identified and characterized bread wheat subtelomeric sequences in those wheat chromosome arms in which the telomeric sequences were annotated (1AS, 4AS, 7AS, 7BS and 7DS), allowing the determination of the exact beginning of the subtelomeric sequence. Different features such as gene content, transposable elements, repeats, hot and cold recombination spots, among others, were initially analyzed within the 500 Kb distal subtelomeric regions and extended to a larger chromosome region (5 Mb) in those chromosomes in which telomeric sequences were identified. Results presented in this work help to describe the molecular organization of wheat subtelomeric regions and suggest that the marked differences in the pattern of different DNA repetitive sequences and DNA protein-binding sequences might contribute to chromosome specificity, which is needed when initial chromosome interactions take place at the beginning of meiosis and allow homologous chromosome associations and recombination in hexaploid wheat.

2 | MATERIALS AND METHODS

2.1 | DNA sequences

All the sequences analyzed in this study were obtained from the International Wheat Genome Sequencing Consortium (IWGSC, RefSeq v1.1, https://www.wheatgenome.org). Bread wheat variety (*Triticum aestivum* cv. Chinese Spring) was used for generating IWGSC RefSeq v1.0 assembly.

2.2 | DNA sequence analysis and prediction tools

Prediction of coding genes, non-coding RNAs, transposable elements, repetitive sequences was done as follows. RNA genes were detected by filtering alignments of Rfam (release 12.2) covariance models, genes annotated with high confidence by IWGSC, and RNA genes produced by filtering predictions from tRNAscan-SE v1.23. Retrotransposons (Long terminal repeats, LTR), type I LINE transposon, type I SINE transposons, type II transposons were detected using the MIPS Repeat Database (REdat) with RepeatMasker. Satellite repeats are sequences with multiple copies of the same base sequence on a DNA sequence. The repeated pattern can vary in length from a single base to several thousand bases long. Satellite repeats were also detected using the MIPS Repeat Database (REdat) with RepeatMasker. Low complexity DNA (Poly-purine or poly-pyrimidine stretches, or regions of extremely high AT or GC content) was detected by Dust. Tandem repeats (duplications of more complex sequences of hundreds of bases) were identified by Tandem Repeats Finder. Emboss CpG plot was used for CG content calculation and CpG island prediction (Rice, Longden, & Bleasby, 2000). Parameters for the analysis were set as follows: window Size (100), minimum length of a reported island (200), minimum observed/expected before a CpG island is reported (0,6), minimum average percentage of C plus G in a set of 10 windows that are required before a CpG island is reported (50).

Hot and cold recombination spots were predicted with iRSpot-EL (Liu, Wang, Long, & Chou, 2016). Size of sliding window (in Kb) and step size parameters were set at 2 and 200, respectively.

Sequences associated with hot recombination spots (Darrier et al., 2017) were identified (simple repeat: CCGCCGCCG, and sequences associated with transposable elements: CTCCCTCC, TTAGTCCCGGTT). These sequences were localized and displayed by means of MAST (MEME Suite 5.0.5) (Bailey & Gribskov, 1998). Parameters were set as follows: direct and reverse complement sequences were analyzed, and results combined; E-value ≤10 (MAST displays all sequences, exact or degenerate, matching query with E-values below the given specified threshold); the p-value of a hit must be less than 0.0001 to be shown in the output.

Distribution of predicted DNA-binding sites of putative wheat proteins homologous to SMC1β cohesin (CCACCAGGTGGC), YY1 (GGGGGCAGTGG) and HMG

proteins ($[AT]_{n>5}$) was obtained by means of MAST (MEME Suite 5.0.5) (Bailey & Gribskov, 1998). Parameters were set as follows: direct and reverse complement sequences were analyzed and results combined; E-value ≤10 (MAST displays all sequences, exact or degenerate, matching query with E-values below the given specified threshold); the p-value of a hit must be less than 0.0001 to be shown in the output.

Exact matches of double strand breaks (DSB) sites were obtained with CONREPP (Subirana & Messeguer, 2010).

3 | RESULTS

3.1 | Structure of wheat chromosome ends

The sequences of all wheat chromosome ends were analyzed based on the available sequences at the IWGSC (RefSeq v1.1, https://www.wheatgenome.org). Table 1 shows the distal sequences of short and long arms of all 21 chromosomes including A, B and D genomes. We found that the current assemblies of most chromosome ends do not contain the plant terminal telomeric repeat (5′-TTTAGGG-3′, or 3′-AAATCCC-5′ on the complementary strand). For this reason, we decided to use only the short arms ends of chromosomes 1A, 4A, 7A, 7B and 7D in the present study. Fortunately, this allowed us to compare the features of subtelomeric regions of different chromosomes of the same genome and the three homoeologues 7AS, 7BS and 7DS.

The first goal of this study was to analyze the distal region to the canonical telomere sequence. As shown in Table 1, a short stretch (10–30 bp) of variable degenerate telomere repeats can be found. A preliminary analysis of the region in wheat revealed a structure depicted in Figure 1. We could not detect a consistent sequence pattern among the different chromosomes that could indicate a clear border between the proper telomere and subtelomere, nor between the distal subtelomere region and the rest of the chromosome. Within the 3 Kb ends, we did not find a feature that would be shared by all five chromosome arms analyzed, except for the plant telomeric repeat. We saw a high diversity of features, including coding genes, retrotransposons, transposons, tandem repeats and low complexity DNA, right adjacent to the telomere repeats and degenerate telomere repeats, among the five chromosome ends analyzed.

We characterized the distal subtelomeric region (500 Kb) of five chromosome arms, focusing on different features that could be related to chromosome recognition and pairing within the subtelomeric region: genes, transposable elements, repeat sequences, GC content and CpG islands, distribution of binding sites of proteins putatively involved in chromosome pairing and recombination, as well as predicted hot/cold recombination spots. For some of the most relevant features, the analysis was extended to a larger 5 Mb sequence.

3.2 | Coding and non-coding genes

Figure 2 shows the localization of gene sequences (both protein coding and RNA genes, direct and reverse complementary sequences were considered) within the distal subtelomeres of chromosomes 1A, 4A, 7A, 7B and 7D short arms. A total of 264 genes were predicted in these five regions altogether within the 500 Kb distal subtelomere. As shown in Table 2, gene density varies among the 5 wheat chromosome ends studied. Chromosome 1A short arm contains an array of ribosomal RNA genes that starts right after the telomere repeats. Chromosome arm 1AS subtelomere also harbors another set of 31 genes that remain uncharacterized. With 141 genes within the first 500 Kb, the distal subtelomere of 1AS shows the highest gene density among the 5 chromosome ends studied. Gene density is different among chromosomes but about the same order of magnitude: 1 gene per tens of Kb. The five chromosome arms analyzed have a different and specific pattern of gene distribution. Great differences can be found both among nonhomologous chromosomes of the same genome (1AS, 4AS and 7AS) and even among homoeologues 7AS, 7BS and 7DS. Many of these genes are still uncharacterized. Some of those that are characterized within the first 500 Kb include transporters, regulatory proteins, tRNAs, nuclear factors and metabolic genes. When the analysis was extended to a 5 Mb region, we observed a significant reduction of gene density, when compared to the gen density within the distal 500 Kb, in all the subtelomeres analyzed (Table 2).

3.3 | Transposable elements

The presence and distribution of TEs (transposable elements), both retrotransposons and transposons (LTR, LINE, SINE and class II DNA transposable elements), was also analyzed in the distal subtelomeric region of the chromosome arms subject to this study (Supplemental Figure S1). As a relevant differential feature, we found that the distribution pattern of TEs is chromosome-specific, both when comparing different chromosomes within the same genome and when comparing homoeologues 7AS, 7BS and 7DS. A more detailed quantitative analysis of TEs is shown in Table 3. The relative abundance of TEs in the 500 Kb distal subteloremic region varies among chromosomes, with a minimum of 15% in 1AS and a maximum of 53% in 4AS.

TABLE 1 Sequences of wheat (Triticum aestivum) chromosome ends. All 42 chromosome ends are displayed, including both short and long arms of chromosomes of the three A, B and D genomes. Sequences were obtained from the International Wheat Genome Sequencing Consortium (IWGSC RefSeq v1.0). Telomere repeat sequence is shown in bold font. Regions of degenerate telomere repeats are underlined

1AS: **CTAAACCCTAAACCCTAAACCCTAAACCCTAAAC** **CCCTAACCCTAAACCCTAACCAAAAC** **CCTAAACCCTAAACCCCTAAACCCTAAA**CCCT AACCTAAACCTACCTAAACCATACATTG	1BS: ACTGCCAAAACTATTGTTTTTCATCCTGTA GTCCCATTTAGAATTACTAAACGTCCTTTT TTTGGGCCGAGATTTAGAGGAACGCTTTCTC GGGGTTTGTGTGCAAGTATTTCAAACCAT	1DS: CTAGGCTTCTTGGGCGTGTATGGGAAGACAA AAGATACACCTGAGGCACGGGAGGACCTG CAACGTTTGCACGAAAAAGACGGCATGCCTCC GAAGCAGTATGAAGGTCCTGCCAGCTAC
1AL: CTCCGAGGCTGCCATGTACTCCGCTTCACAC GTAGATCCCGCCACGACGCTCTGCTTGCA GCTGCACCAGCTTACTGCTCC	1BL: CCCTCCGACACGGCTTCCACCGTGGAGTTCC AGCCGCTGTGCGTGATGAAGGCGCCTACC GCAGGGTGGGCGAGCACCTCC	1DL: CATCAACAATCACAGTTTGGTCCGATTCTAGC ATGTTTCATGGACTATTACTCAGTTTTG GGGTCCTGGAGGCATTTCCAT
2AS: CAGTTCCTAAACTGCTCCAGTCGGCGCACGTT ATCATTCTTGTTAGTTCTGAGCGAATTC CCTCGGTATGATCATTCTTTCATCAGTTTGTTC CCCGTTTCTTCTCTAGTTTGAGCCTAG 2AL: CGTACTACTTCAGAGTGGTACATAATCCTGTCT TGGAAGTCATGTTGCTAGACAACAATG AACCTACGAGCTATGGAGAAG	2BS: GCATAGCCACGCCCCGAAGGCCACCCCGAAT CCCAGTTGAACAGGAGATCTAGCCCTTTG ACTTTTGCCGGACGGGCTTTGACCAGTG- GTCT TTTCCACCTGGTTGCGACAGGTCTGCCC 2BL: CGACCGAGAGAGTGCCGACCGAGAGACTAGG CGTGCCGACCGAGAGAGCGCCGCCCGCGC CGCTTCCCGGTGCAGGCGACT	2DS: TCCCTTGAAGAGGAAAGGGTGATGCAGCA AAGTAGCGTAAGTATTTCCCTCAGTTTTTGA GAACCAAGGTATCAATCCAGTAGGAGGCCA CACGCAAGTCCCTCGTACCTACACAAACAA 2DL: ACGGCAAAGAGGGCGGACGGCAATATTTGC GCTGTCAGTCCGTTAAGTGGCTAACGGCAG ACCTTTGCCGTCCGCCGCAGA
3AS: AACGGGTTCGCCGAAACCTACCTTGTGAGCA TCGGCGATGTTACGGCTGTTGGCACTGGT CCTCCGTCTGACAAGACCACTGTCCTAGGCG ATGGAGATGCAAATGGTATGTAAGCACGC 3AL: GACCTTTCGGTCTCCGTGTTCCGAGGCCA TGTCTGCACATGCTAGGCTCGTCAAGTTAAC CCTAAGTGTTTTCGCTGTGTA	3BS: GAGTAATAATGCACAAAACGGACCA- GAATCGG CCAAAATTGTGAGTGTTGATGACCGACA CATAAACGCACCACCGGGTTCGTCAATCGT GAAATCACTTTGGGACCCCAACACTGAGTA 3BL: AAAGTGGCCAGCGCCCATGCCACACCCCTC TCTCTCTCCCCTCTGCCTCTCTCTCCCCCC TCGCCTCACCCGTGCCACCGG	3DS: GCATCCCCGTGATCAAATCACAAAGTGTCTG GCCAGACGATGATGGATACCGTAACACCG AGAGGGCCCAGAGATATCTCTCCATCGTCGGA GGAGCAAATCCCAATCTTGAGCTATCAA 3DL: ACATTCATGCCACAATAAAGAAAATTACATTGA AGAATATGCTAAGTAGCATTCCACATC AAAAATTCTGTTTTTATCATT

(Continues)

TABLE 1 (Continued)

4AS: **AACCCTAAACCCTAAACCCTAAACCCTAAC** **CCTAAACCCTAAACCCTAAACCCTAAACCC** **TAAACCCTAAACCCTAAACCCTAAA**CCCAAAA CAGTGTGTTATACCTTACGAAAACGGTC 4AL: GCAGTTTTGGCCGATTCTGGCCCATTTCGTG GACTATTACTTACTGTTTTGGGGTCCCGG AGTGGTTTCCATGATTGATGA	4BS: AAGGAGTTAGTTAATGATGTTTGAGCATGTA CTTTTTTTGAGTGTCTATTTATTTTCGAT TGGTATCTATGGATTGATCACAAGCCGAAAC ATGGTTAGAGCTCTAATATGTCTTGAACT 4BL: ACGGGCTAGAATCGGCCAAAACTGCGAGTGT TGATGACCGACACGTAAGCGCACCTTGGG GTTCAACAACTATGCAAATCG	4DS: CAAAAATAAAAAATCTACCATGATGGATCAA AACGGTTAAAAATGTATCAAACTATCAGA TTTCTATAGGTAATTTGCTATCTGAGGTGCA TATAGATGCATGGACAAACCTCGAAGCTT 4DL: TTGAATCCTCGACACCGTGGTTCATCCGATGA GATCATCGTGGAACATGTGGGAGCCAAC ATGGGTATCCAGATCCCGCTG
5AS: TATATATATATATATATATATATATATATAT ATATGTACAGAATCTGAAAGATGGCAGTC GTTGGCTTCAGCCCCCATGCACATAATGCGG GGGTGCTCGCAGAAGACGCATTTTCACAC 5AL: GACGGTGGCGGTGGCGAGCAATGCATCACCG ACGGATTCCTTCCTTCCATCCCCTCCCCT CTTGGTGCAAATTTCTTCTCT	5BS: AAAAAATTCAAAATAACAAAGTTTTTCTAAG AAAAAATGAATCAGTGGCATTGGTATTAC CCTTCTAGAAGAAAAACAAATAATAATGAGA AAATTACACAAATTTTACAAAGAAATTGC 5BL: GGTTAGGGTTTAGGGTTTTTAGGGTTTT TA **GGGTTTAGGGTTTAGGGTTTTGGGTTTAGG** **GTTTAGGGTTTAGGGTTTAGG**	5DS: CATATCTGCTCAAATCATCTGTGAAGGTGA GAAAATAACGATATCCGCCACGAGCTTCAA TATTCATCGGACCACATACATCTGTATGTAT GATTTCCAACAAATCTTTTGCTCTCTCCA 5DL: TTTACCGATTTGTTTAGAGAGCTAAATGACC GTGAAATTAAAAATCACTACAAAATGAAC TCTGAAAATGTCGAAACTTGG
6AS: TGCAGATATATTAAATAAAAACAGAAACAGAA AAAAAGGGAAAAACTTACCTGTGCGGCC CAGCCAGCTGGCCGGCCCAGCAAGCCAGCCCA GCCCACCACCGCCGTCGTCCTCCTGGCG 6AL: CAACGGACTGCGCCGTCAAAACTGTTGGCGGTT CAGGGGGCGACACGTGGCGAAGATATG CCTACGGCAAAGCCGTCGGCA	6BS: AGGGTCCTTTCACAATTATCTCACGTACCAAT CCATCCATTCTTCGTCCAGTTTAATCCA TCCATCCTTTGTCTAGCTTACTACATATCAT TTCACACATGTATGCAAATTGGTACGATC 6BL: ATCCTCGCTACCGCGACGACGTCCGCCTCGA CTCGCGTTACGCTAAGACAGGGTGGCAGT GGTGTCACGTGCTCGTCTCGG	6DS: GCGGGCACATGTGCTGTAGGACTGATGGTA ATTTTCATAATTGTTCGTGATGGAGTAGTA TCTGAACAATTCCCTCTATGCGGATTGCTC TTCGCGTGTCTACGAGTGTGGCCGGCGGCC 6DL: AAGAATTAGCCAAAACTGCGTGTGTTGATGA CCGACACGTAAATGCACCCCGGGGTTCAT CAATCGCGGAAATCAGCCCGG
7AS: **ACCCTAAACCCTAAACCCTAAACCCTAAACCC** **TAAACCCTAAACCCTAAACCCTAAACCCT** **AAACCCTAAACCCTAAA**TCCTAAACCCTAAC CCTAACCCTAACCCTAACCCTAACCCTAA 7AL: AAGGATTATGAATTCTCTGTTGATCCTGATA TAATTACTTTGGTTGAATCTGATCGTTTC CATGGCTATGAATCTGAAACT	7BS: **AACCCTAAACCCTAAACCGTAAACCCTAAAC** **CCTAAACCCTAAACCCTAAAAACCCTAAA** **CCCTAAACCCTAAACCCTAAA**GCCTAAAAA CCCTAAACCCTAAACCCTAAACCTATTCCG 7BL: AACAACAACAACAACAACAACAACAACAACA ACAACAACAACAACAACAACAACAACAAC AACAGCAACTACAATAACAGC	7DS: **AACCCTAAACCCTAAACCCTAAACCCTAAACC** **CTAACCCTAAACCCTAAACCCTAAACCC** **TAAACCCTAAACCCTAAACCCTAAA**CCCGGG TTCACAAACCATCACTAACAACTTCGGTT 7DL: GTGGACTTGCAACTGAAGCGGAAAAAATCGAG CCTAGTCGTGAGTCAAAGGATGGACTTG TATACTGGGGCAAAAAAAAAA

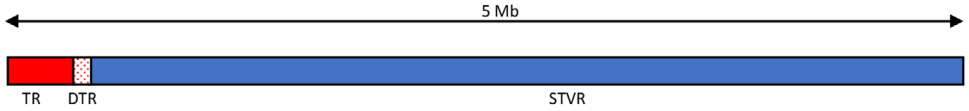

FIGURE 1 Structure of wheat chromosome telomere-subtelomere junction. TR: telomere region including telomeric tandem repeat sequence (TTTAGGG). DTR: subtelomere region including degenerate telomere repeat sequence. STVR: subtelomere variable region including multiple elements (genes, transposons, retrotransposons, satellite repeats, simple repeats and low complexity DNA arranged in chromosome-specific patterns)

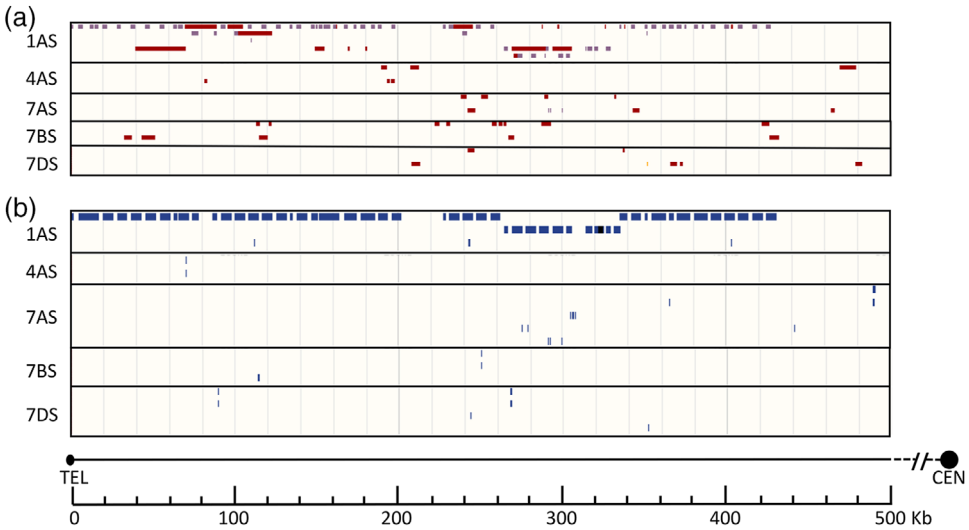

FIGURE 2 Genes. Distal subtelomere sequence (500 Kb) of chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS were analyzed for the presence and distribution of both coding and non-coding genes. Direct and reverse complementary sequences were considered. (a) Results include RNA genes produced by filtering alignments of Rfam (release 12.2) covariance models, genes annotated with high confidence by IWGSC, and RNA genes produced by filtering predictions from tRNAscan-SE v1.23. Protein coding genes are shown in red color and RNA genes in pink color. (b) Results include long noncoding RNA, Rfam Models (LCA) and tRNA Models

LTR-type retroelements (Ty1/*copia*, Ty3/*gypsy* and DIRS1) are the most abundant TEs (more than 76% of the total) in all the distal wheat subtelomeres analyzed. These regions are also rich in class II DNA transposable elements, while LINE and SINE retrotransposon sequences are rare.

3.4 | Repeat DNA sequences

The presence and distribution of repeats along the 500 Kb distal subtelomeric region of wheat chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS, was also investigated (Supplemental Figure S2). A quantitative analysis of these repeats is shown in Table 4. Satellite repeats are not common in the distal region of subtelomeres in the chromosome arms subject to this study, except for 4AS and 7DS (Table 4), where they are present close to the chromosome end (Supplemental Figure S2A). Simple repeats are not frequent either (Table 4) and they are not present in the distal subtelomere close to the telomeric repeats (Supplemental Figure S2B). Again, we found that the distribution patterns of repeat sequences are chromosome-specific, so that

TABLE 2 Genes in distal subtelomeres of wheat chromosomes. The distal subtelomeric region adjacent to telomeric repeats of chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS were analyzed for the presence of genes (IWGSC RefSeq v1.1, /www.wheatgenome.org)

Chr. arm	Nr. Genes (in distal 500 Kb)	Nr. Genes (in distal 5 Mb)	Gene density (in distal 500 Kb)	Gene density (in distal 5 Mb)
1AS	141	217	1 per 3.5 Kb	1 per 23 Kb
4AS	18	81	1 per 27.8 Kb	1 per 62 Kb
7AS	39	114	1 per 12.8 Kb	1 per 44 Kb
7BS	40	84	1 per 12.5 Kb	1 per 60 Kb
7DS	26	104	1 per 19.2 Kb	1 per 48 Kb

TABLE 3 Transposable elements in wheat distal subtelomeres. The 500 Kb distal subtelomeric region adjacent to telomeric repeats of chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS were analyzed for the presence of transposable elements

	Element	Nr.	L(bp)	%
1AS	Retroelements	21	68511	13.70
	SINEs:	0	0	0.00
	LINEs:	0	0	0.00
	LTR elements:	21	68511	13.70
	Ty1/*Copia*	6	13663	2.73
	Gypsy/DIRS1	15	54848	10.97
	DNA transposons	12	9213	1.84
	Total:		77724	15.54
4AS	Retroelements	89	255838	51.17
	SINEs:	0	0	0.00
	LINEs:	0	0	0.00
	LTR elements:	89	255838	51.17
	Ty1/*Copia*	28	78555	15.71
	Gypsy/DIRS1	61	77283	35.46
	DNA transposons	36	12713	2.54
	Total:		268551	53.71
7AS	Retroelements	91	131718	26.34
	SINEs:	1	181	0.04
	LINEs:	13	8377	1.68
	L1/CIN4	13	8377	1.68
	LTR elements:	77	123160	24.63
	Ty1/*Copia*	35	71168	14.23
	Gypsy/DIRS1	42	51992	10.40
	DNA transposons	55	41409	8.28
	Total:		173127	34.63
7BS	Retroelements	90	229306	45.86
	SINEs:	0	0	0.00
	LINEs:	3	517	0.10
	L1/CIN4	3	517	0.10
	LTR elements:	87	228789	45.76
	Ty1/*Copia*	31	124025	24.80
	Gypsy/DIRS1	56	104764	20.95
	DNA transposons	53	30192	6.04
	Total:		259498	51.90
7DS	Retroelements	54	84046	16.81
	SINEs:	0	0	0.00
	LINEs:	4	551	0.11
	L1/CIN4	4	551	0.11
	LTR elements:	50	3495	16.70
	Ty1/*Copia*	25	61654	12.33
	Gypsy/DIRS1	25	21841	4.37
	DNA transposons	55	24849	4.97
	Total:		108895	21.78

TABLE 4 Repeats in wheat distal subtelomeres. The 500 Kb distal subtelomeric region adjacent to telomeric repeats of chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS were analyzed for the presence of repeated elements. Satellite repeats: Multiple copies of the same DNA sequence. The repeated pattern can vary in length from a single base to several thousand bases long. Simple repeats: Duplications of simple sets of DNA bases (1–5 bp) such as A, CA, CGG etc. Low complexity regions: Poly-purine or poly-pyrimidine stretches, or regions of extremely high AT or GC content

	Element	Nr.	L(bp)	%
1AS	Satellites:	0	0	0.00
	Simple repeats:	12	1009	0.20
	Low complexity:	1	54	0.01
	Total:	13	1063	0.21
4AS	Satellites:	15	32940	6.59
	Simple repeats:	75	3701	0.74
	Low complexity:	15	861	0.17
	Total:	105	37502	7.50
7AS	Satellites:	7	4030	0.81
	Simple repeats:	115	6052	1.21
	Low complexity:	26	1661	0.33
	Total:	148	11743	2.35
7BS	Satellites:	1	26	0.01
	Simple repeats:	129	5616	1.12
	Low complexity:	14	893	0.18
	Total:	144	6535	1.31
7DS	Satellites:	2	341	0.07
	Simple repeats:	91	4551	0.91
	Low complexity:	18	1408	0.28
	Total:	111	6300	1.26

different chromosomes of the same genome and even homoeologues (7AS, 7BS, and 7DS) can be clearly differentiated by this feature.

3.5 | Other DNA features

Distal subtelomere sequence (500 Kb) of chromosomes 1A, 4A, 7A, 7B and 7D short arms were analyzed for the GC content and the identification of predicted CpG islands (Figure 3). Emboss CpG plot was used for GC content calculation and CpG island prediction. Both GC content and predicted CpG islands were dramatically different among all chromosomes, particularly relevant is the high density of GC-rich DNA stretches and CpG islands in 1AS and the low density within the 140 Kb most distal region of 7DS. We also scanned the distal subtelomeric regions for the presence of DSB hotspot motif (GVSGRSGNSGRSGVSGRSG). As shown in Table 5, the density of these hotspots is about 1 site per hundreds of Kb. Even though the 500 Kb distal region shows high differences among chromosomes, density is more similar among the chromosomes examined when a larger region was considered (5 Mb). Regarding a second DSB motif (C-rich motif CGSCCGSCGSSGCCGCCGCCS) we found that this motif must be very rare in wheat, because we detected none within the 5 Mb distal subtelomeric region in the chromosomes studied.

3.6 | Crossover and recombination hot spots

Distal subtelomere sequence (500 Kb) of chromosomes 1A, 4A, 7A, 7B and 7D short arms were analyzed for the distribution of predicted hot and cold recombination regions (Figure 4a). iRSpot-EL was used for prediction of hot and cold recombination spots. The distribution of three short sequence motifs (CCGCCGCCG; CTCCCTCC; TTAGTCCCGGTT) associated with hot recombination regions in wheat were also analyzed both within the 500 Kb distal subtelomere and within a larger region of 5 Mb. Sequences associated with hot recombination spots were identified by MAST (MEME Suite 5.0.5). A good correlation between these sequence motifs and hot recombination spots was detected. Clear differences among chromosomes, both regarding localization and extension of these regions, can be observed. A few DSBs hot spots were also localized, and their distribution showed clear differences among

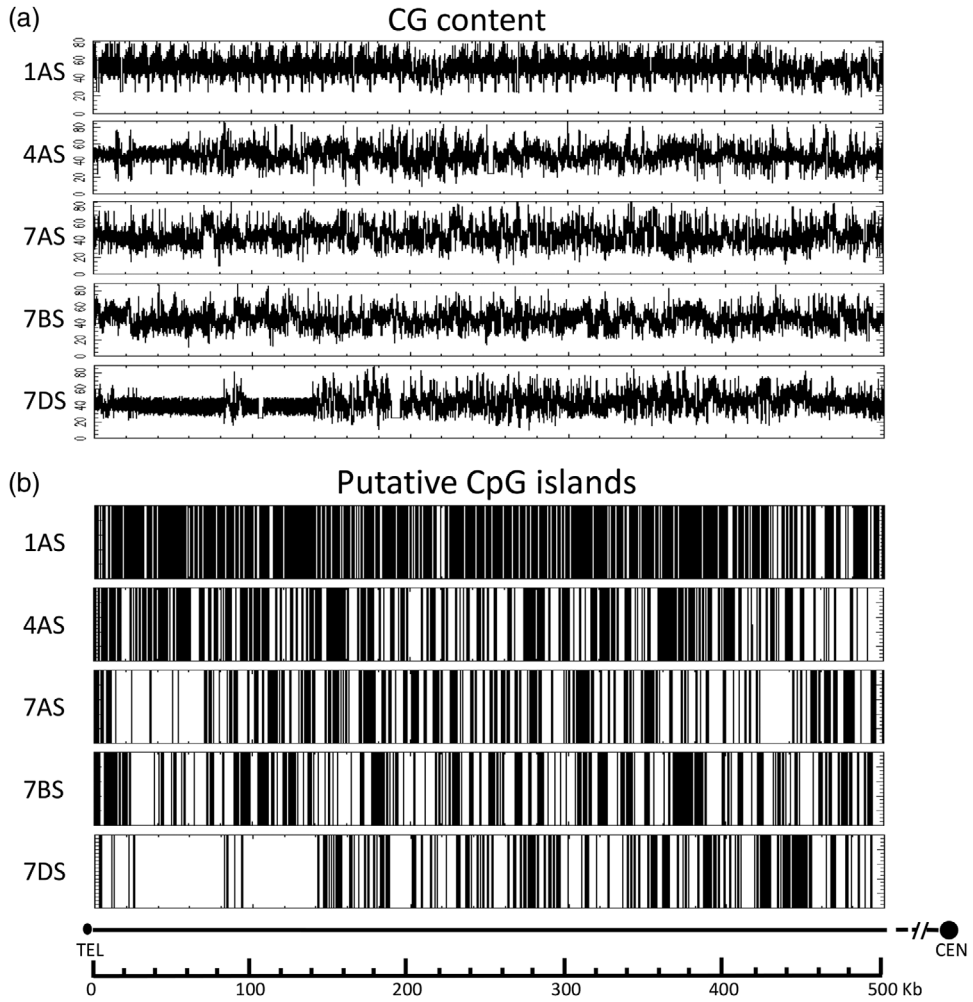

FIGURE 3 GC content and CpG islands. Distal subtelomere sequence (500 Kb) of chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS were analyzed for the GC content and the identification of predicted CpG islands. Emboss CpG plot was used for GC content calculation (panel a) and CpG island prediction (panel b). Predicted CpG islands are marked in black

TABLE 5 Frequency distribution of DSB hotspots. Sequences of chromosome arms (1AS, 4AS, 7AS, 7BS, 7DS) were scanned with CONREPP for the presence of DSB hotspots (GVSGRSGNSGRSGVSGRSG). Scans were performed on 500 Kb and 5 Mb distal subtelomere regions. Figures represent number of sites

	1AS	4AS	7AS	7BS	7DS
500 Kb	0	4	4	2	1
5 Mb	34	45	50	39	40

chromosomes (Figure 4). All DSBs colocalized with hot recombination spots.

3.7 | DNA-binding proteins

Figure 5 shows the distribution of predicted binding sites of relevant DNA-binding proteins, regulatory or structural proteins that could be involved in the architecture of chromosomes and the process of chromosome approaching and interaction during chromosome pairing. The analysis was initially performed on the 500 Kb distal subtelomere and extended to a larger 5 Mb region towards the centromere. We chose these proteins because of their putative role in these processes and because of the availability of identified DNA-binding sites. Distal subtelomere sequences were analyzed for the distribution of predicted DNA-binding sites of putative wheat proteins homologous to human SMC1β meiosis-specific cohesion (Revenkova, Eijpe, Heyting, Gross, & Jessberger, 2001), Ying Yang 1

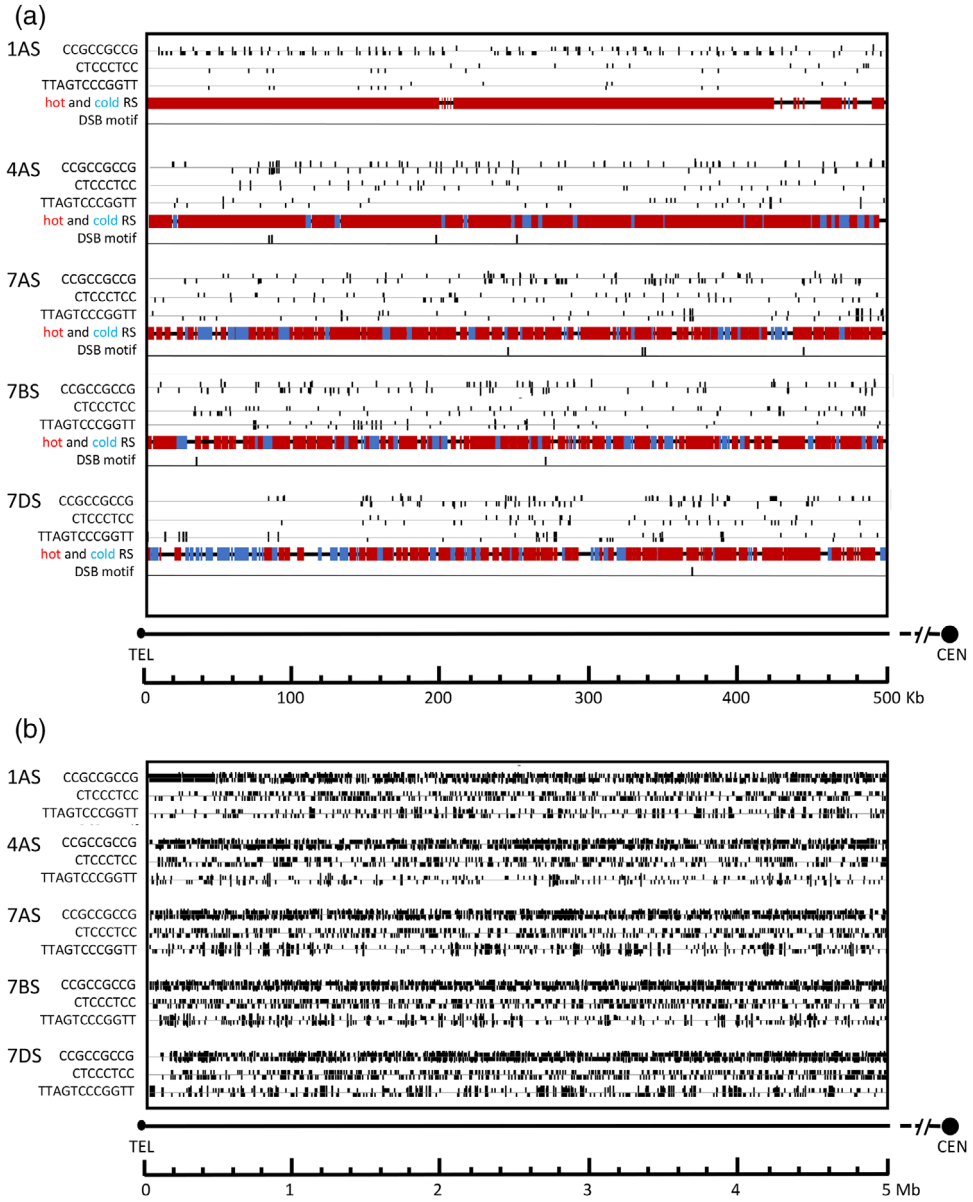

FIGURE 4 Distribution of predicted hot-cold recombination spots. (a) Distal subtelomere sequence (500 Kb) of chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS were analyzed for the distribution of predicted hot and cold recombination regions and sequences associated to them. iRSpot-EL was used for prediction of hot and cold recombination spots. Sequences associated with hot recombination spots were identified by MAST (MEME Suite 5.0.5). Red and blue segments represent hot and cold spots, respectively. (b) An extension of the study on the 5 Mb distal subtelomeric region

protein (Beagan et al., 2017) and high mobility group proteins HMG (Subirana & Messeguer, 2011). Sequences were identified and plotted using MAST (MEME Suite 5.0.5). We found that the distribution of putative binding sites for the proteins investigated was chromosome-specific, with evident differences in density and distribution of these putative binding sites among chromosomes, both among nonhomologous chromosomes of the same genome and among homoeologues, which is in agreement with their putative role as determinants of the specificity of chromosome recognition and pairing. In addition, we found similar distribution patterns of binding sites for SMC1β cohesin and YY1 in every chromosome, with a higher density of sites in regions enriched in genes and retrotransposons (chromosome arms 1AS, 4AS) and a particularly low density in regions poor in genes and retrotransposons (distal subtelomeric region of chromosome arm 7DS). We also found a striking differential distribution of putative binding sites for HMG proteins, with a very low concentration of these putative sites in regions with a high concentration of genes (1AS) or with a low concentration of LTR retrotransposons (distal region

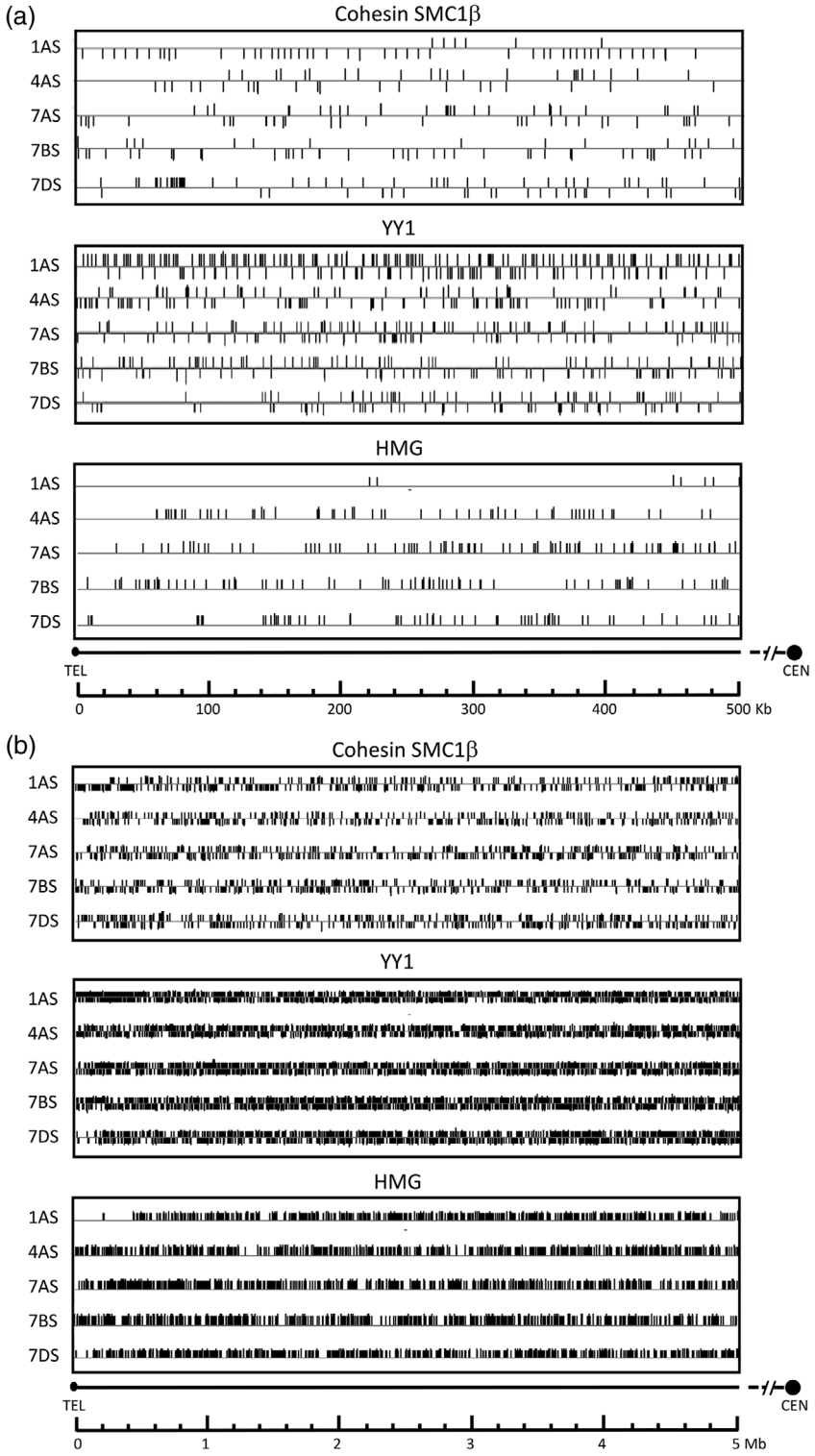

FIGURE 5 Distribution of predicted binding sites of relevant DNA-binding proteins. (a) Distal subtelomere sequence (500 Kb) of chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS were analyzed for the distribution of predicted DNA-binding sites of putative wheat proteins homologous to SMC1β, YY1 and HMG proteins. Sequences were identified and plotted using MAST (MEME Suite 5.0.5). (b) A similar study on the 5 Mb distal subtelomeric region

of 7DS). We also observed that, despite the highly different density of sites among chromosomes when comparing the 500 Kb distal region, a more similar density among chromosomes was revealed when a broader region of 5 Mb was considered. The average density of sites is around 1 per 50 Kb, like the figures found in most eukaryotic organisms.

4 | DISCUSSION

The analysis of available wheat sequences at IWGSC has revealed that most chromosome ends are not fully sequenced or properly assembled yet, since the current assemblies do not contain the plant terminal telomeric repeat (5′-TTTAGGG-3′, or 3′-AAATCCC-5′ on the

complementary strand) (IWGSC, 2018). The telomeric conserved repeat from *A. thaliana* was detected only in fourteen out of 24 rice chromosome ends when the full rice genome was sequenced (Mizuno et al., 2006; Mizuno et al., 2008b). Previous published studies of wheat chromosomes at whole scale ignored this fact, and a comparison of all chromosomes, including chromosome ends, was done (IWGSC, 2018; Wicker et al., 2018). We strongly consider that, to be sure that actual distal subtelomeric regions are being accurately compared, the study should be done only among chromosome ends whose sequence includes a telomeric repeat. Fortunately, we have identified the short arm ends of the chromosomes from homoeology group 7 (7AS, 7BS, 7DS) plus the short arm ends of two more chromosomes from the A subgenome (1AS and 4AS), to perform our analysis.

The region adjacent to the telomere contained a short stretch (10–30 bp) of variable degenerate telomere repeats, a feature that was found in all organisms with complex genomes previously studied. Numerous copies of the telomeric repeat and its derivatives were detected in the subtelomeres of human, *Arabidopsis* and wheat chromosomes (Ambrosini, Paul, Hu, & Riethman, 2007; Uchida, Matsunaga, Sugiyama, & Kawano, 2002). In organisms previously studied, the subtelomeric region is defined as the extension between the telomere and the most distal chromosome-specific sequence (Kuo et al., 2006; Louis, 2014; Mefford et al., 2002; Pryde, Gorham, & Louis, 1997). A common feature supposed to be shared by plants is a stretch of tens of kilobases of highly rearranged and repetitive DNA before the first active gene at the distal part of subtelomeres (Alkhimova et al., 2004; Ohmido, Kijima, Ashikawa, de Jong, & Fukui, 2001; Pearce, Pich, Harrison, Flavell, & Heslop-Harrison, 1996; Röder et al., 1993; Vershinin, Schwarzacher, & Heslop-Harrison, 1995; Wu & Tanksley, 1993). However, the model plant *A. thaliana*, with its small genome, shows a different picture. *Arabidopsis* subtelomeric regions are small (less than 5 Kb) and relatively simple. They include a short stretch of BAAAA repeats (where B = C, T, or G) and a 32-bp tract composed almost entirely of G (Kuo et al., 2006). In wheat, we have not identified a consistent sequence pattern among the different chromosomes that could be indicative of a clear border between the proper telomeric sequence and the subtelomere, nor between the distal subtelomere region and the rest of the chromosome. Within the 3 Kb ends, we did not find a feature that would be shared by all five chromosome arms analyzed either. We found none of the short stretches of DNA described in *Arabidopsis*. Instead, we have seen a high diversity of features, including coding genes, retrotransposons, transposons, tandem repeats and low complexity DNA, right adjacent to the telomere repeats and degenerate telomere repeats, among the five chromosome ends analyzed as it was observed at a whole genome scale (Alaux et al., 2018; IWGSC, 2018). We decided to start our analysis focusing on the most distal 500 Kb of chromosome ends, as done in a model cereal species like rice (Mizuno et al., 2006) and other organisms like humans and *Arabidopsis* (Kotani et al., 1999; Macina et al., 1994), but we later expanded the analysis to a larger chromosome region (5 Mb).

Previous evidences have suggested the importance of subtelomeres in homologue-specific chromosome recognition and pairing (Calderón et al., 2014). When subtelomeres were absent, homologous chromosomes failed to recognize and associate in pairs, suggesting an important function of subtelomeres in these processes (Calderón et al., 2014). In rye, heterochromatin blocks at the subtelomeres have also been visualized clustering at the beginning of meiosis, suggesting a putative contribution of these regions in chromosome associations (Mikhailova et al., 2001). In this work, we have focused on different features within the subtelomeric region of wheat chromosomes that might contribute to specificity: genes, transposable elements, repeat sequences, GC content and CpG islands, distribution of binding sites of proteins putatively involved in chromosome pairing and recombination, as well as predicted hot/cold recombination spots. All the features analyzed within the subtelomeric region revealed a chromosome-specific pattern among homoeologous chromosomes, which might contribute to the specificity between homologous chromosomes at the onset of meiosis in a polyploid like wheat.

Regarding the localization of gene sequences within the 500 Kb distal subtelomeric regions of chromosomes 1AS, 4AS, 7AS, 7BS and 7DS, a total of 264 genes were predicted in these five regions altogether (IWGSC, RefSeq v1.1, https://www.wheatgenome.org). Considering an overall length of 2,500 Kb for the five subtelomeric regions analyzed, it gives an average density of 1 gene per 9.5 Kb. These figures are like those found in rice, where a total of 598 putative genes were found in 500 Kb regions inward in seven chromosome ends (Fan et al., 2008; Mizuno et al., 2006). In a total of 3,500 Kb for seven chromosomes, it gives an average gene density of one gene per 5.9 Kb, though gene density varies among the 5 wheat chromosome ends studied. Concerning the nature of the genes localized in these regions, it is worth mentioning the fact that chromosome arm 1AS contains an array of ribosomal RNA genes that starts right after the telomere repeats. The subtelomeres of *Arabidopsis* chromosomes 2 and 4 short arms contain tandemly arranged rRNA genes, NOR2 and NOR4. As in wheat, NOR4 is located just close to the telomeric repeat, and repetitive DNA is absent at the junction between the telomere and rDNA (Copenhaver & Pikaard, 1996). Chromosome arm 1AS subtelomere also harbors another set

of 31 genes that remain uncharacterized. With a total of 141 genes, chromosome 1A distal subtelomere shows the highest gene density among the 5 chromosome arms studied. Gene density is different among chromosomes but about the same order of magnitude: 1 gene per tens of Kb. Remarkably, gene density is reduced similarly in all chromosomes when a larger region (5 Mb) was analyzed, in agreement with other studied performed at a larger scale (Alaux et al., 2018; IWGSC, 2018). Another relevant feature is the fact that all chromosomes have a different and specific pattern of gene distribution. Among homoeologous chromosome arms 7AS, 7BS and 7DS, great differences have been found too. Many of these genes are still uncharacterized. The presence of transporters, regulatory proteins, tRNAs, nuclear factors and metabolic genes agrees with the idea that subtelomeres seem to harbor genes that are subject to a high evolutionary rate, probably because of their relevance for adaptation and interaction with the environment (Brown, Murray, & Verstrepen, 2010).

The analysis of TEs revealed a relevant differential feature: the distribution pattern of TEs is chromosome-specific, both when comparing different chromosomes within the same genome and when comparing homoeologues 7AS, 7BS and 7DS. The relative abundance of TEs varies among chromosome arms, with a minimum of 15% in 1AS distal subtelomere and a maximum of 53% in 4AS. LTR-type retroelements (Ty1/*Copia*, Ty3/*Gypsy* and DIRS1) are the most abundant TEs (more than 76% of the total) in all the distal wheat subtelomeres analyzed, as it happens in maize (Lamb et al., 2007; SanMiguel et al., 1996). In fact, LTR-type retroelements are the most abundant TEs in plants, including wheat (IWGSC, 2018; Vicient & Casacuberta, 2017). In a genome scale, TEs represent more than 80% of the wheat genome, including 70% retroelements and 13% DNA transposons (Clavijo et al., 2017; IWGSC, 2018; Li, Zhang, Fellers, Friebe, & Gill, 2004). Among the retroelements, *Gypsy* and *Copia* LTR retroelements are predominant in wheat, while CACTA DNA elements are the most abundant DNA transposons (Clavijo et al., 2017; IWGSC, 2018). A few plant genes result from the activity of TEs, which have been involved in the evolution of promoters, exons and introns (Zhao, Ferguson, & Jiang, 2016). It is worthy to mention that TEs can contribute to genome organization. Thus, chromosomal reorganizations or deletions can possibly be obtained by recombination between two TEs as described in maize, where the Ac elements are responsible for chromosome rearrangements and deletions (Weil & Wessler, 1993). Subtelomeric regions, together with pericentromeric regions, are usually rich in TEs, which seems to be the consequence of diverse mechanisms. In the heterochromatic regions, TEs are removed by intra- or inter-element recombination at a much lower rate as these regions display a lower recombination rate (Zamudio et al., 2015). This would explain the aggregation of TES within the heterochomatin observed near telomeres. Finally, a negative correlation among TEs density and recombination rate has been found (Daron et al., 2014; Tian et al., 2009). However, the pattern of these TEs on each subtelomeric chromosome, especially between homoeologous chromosomes, might contribute to specificity of initial chromosome interactions between homologous chromosomes at the onset of meiosis.

The number and complexity of repetitive sequences varies among species and they are more abundant in species with larger genomes (Kidwell, 2002). Diverse families of repeats were identified in the subtelomeres of plants including *Arabidopsis*, tobacco, barley, wheat and potato (Heacock et al., 2004; Mao et al., 1997; Koukalova, Reich, Matyasek, Kuhrova, & Bezdek, 1989; Belostotsky & Ananiev, 1990; Torres et al., 2011). Repetitive sequences account for more than 90% of wheat genome (Li et al., 2004). These repeated sequences include satellites, simple repeats (1–5 bp long) and low complexity repeats (poly-purine or poly-pyrimidine stretches, or regions of extremely high AT or GC content). Our analysis in wheat revealed the presence and distribution of repeats along the 500 Kb distal subtelomeric region of chromosomes 1AS, 4AS, 7AS, 7BS and 7DS. Again, we found that the distribution patterns of repeat sequences are chromosome-specific, so homoeologues (7AS, 7BS, and 7DS) can be clearly differentiated by this feature at the subtelomeric region. Repeated sequences can be visualized as heterochromatic regions in cereal chromosomes. In maize, tandem repeats are very common across the chromosomes, but mostly localized to knob regions, while they are less abundant in the subtelomeric regions (Ananiev, Phillips, & Rines, 1998; Lamb et al., 2007). Close related species like rye, barley and wheat, display very different patterns. Rye chromosome ends have large heterochromatin blocks as a specific feature (Vershinin & Evtushenko, 2014). Wheat and barley chromosomes show a complex pattern of heterochromatin distribution, but their subtelomeres lack large blocks of heterochromatin (Gill & Kimber, 1974; Linde-Laursen, 1978).

Taken together, genes, transposable elements and tandem repeat sequences of various types draw an overall complex and dynamic structure of distal subtelomeres in wheat that seems to be chromosome specific and might contribute to chromosome interactions specificity at the onset of meiosis. Other plants, even rice with its small genome, share this complex and dynamic structure of subtelomeres (Vershinin & Evtushenko, 2014). However, none of these elements is exclusive of subtelomeres, suggesting that plant subtelomeres do not differ from the rest of the chromosome in the composition of these structural elements but in the distribution pattern, which can contribute

to chromosome specificity. In fact, and as reminded previously, subtelomeres are involved in chromosome initial interactions and pairing, safeguard distal genes against the active loss/gain processes within the terminal regions, or stabilize the chromosomal ends when canonical telomeric repeats are absent (Garrido-Ramos, 2015; Henderson, 1995; Jain, Hebden, Nakamura, Miller, & Cooper, 2010; Kipling, 1995; Louis & Vershinin, 2005). Thus, it seems evident that all the DNA characteristics analyzed in this work and their distribution pattern contribute to the high specificity of the distal chromosome region, but other DNA sequence features can also be considered.

Simple features like GC versus AT content could have a relevant role in determining the functionality of subtelomeres. Both GC content and predicted CpG islands were dramatically different among all the distal subtelomere sequence (500 Kb) of wheat chromosomes 1A, 4A, 7A, 7B and 7D short arms, supporting the high polymorphism of these subtelomeric regions. Particularly relevant is the high density of GC-rich DNA stretches and CpG islands in 1AS and the low density within the 100 Kb most distal region of chromosome 7DS. GC content and CpG islands were found to be closely correlated with the presence of genes in animals and plants (Ashikawa, 2001; Larsen, Gundersen, Lopez, & Prydz, 1992; Mouchiroud et al., 1991). In our work, this correlation seems evident (Figures 2 and 3), particularly when comparing the subtelomeric regions of 1AS and 7DS.

Yet, most importantly, GC content is also correlated with recombination and crossover, two key processes, occurring mainly at the subtelomeric regions, which are greatly influenced by the correct chromosome association earlier in meiosis. In fact, recombination and GC content seem to be reciprocally interrelated (Sundararajan et al., 2016), and this has been reported in Triticeae (Escobar et al., 2010; Haudry et al., 2008; Muyle, Serres-Giardi, Ressayre, Escobar, & Glemin, 2011). Interesting enough, a study on *Brachypodium*, maize and rice revealed a substantial connection between high GC content, local recombination and crossover rate (Serres-Giardi, Belkhir, David, & Glemin, 2012). Crossovers arise specifically in meiotic prophase I, once some DSBs (double strand breaks) are established and resolved after homologous chromosomes interact and pair (Zickler & Kleckner, 2015). A correlation between DSBs and crossovers with high GC content has been described in many organisms, including plants (Sundararajan et al., 2016) except for *A. thaliana* (Choi et al., 2018). Multiple evidences support that DSBs are necessary for recombination, though most DSBs will not yield crossovers. In maize, a couple of GC-rich motifs (GVSGRSGNSGRSGVSGRSG and CGSCCGSCGSSGCCGCCGCCS) associated to cross-overs were identified (He et al., 2017). Analogous C-rich motifs have been associated with crossover in *Arabidopsis* (Choi et al., 2013; Wijnker et al., 2013; Shilo, Melamed-Bessudo, Dorone, Barkai, & Levy, 2015), although these DSB-associated motifs are not in GC-rich but in AT-rich regions (Choi et al., 2018). After all these findings, DSBs and crossovers seem to be determined by the presence of specific sequence motifs and not by GC content. These specific sequences could be related to a more relaxed chromatin that determines SPO11 accessibility resulting in high levels of DSBs (Choi et al., 2018). We scanned the distal subtelomeric regions of wheat chromosome arms 1AS, 4AS, 7AS, 7BS and 7DS for the presence of this DSB hotspot motif (GVSGRSGNSGRSGVSGRSG). The density of these hotspots is about 1 site per hundreds of Kb, and the 500 Kb distal region shows high differences among chromosomes, contributing to specificity at the distal chromosome regions, taking also into account that the density of these hotspots is more similar among the chromosomes examined when a larger region was considered (5 Mb).

We analyzed the distribution of short sequence motifs associated with hot recombination regions in wheat described by Darrier et al. (2017). Sequences associated with hot recombination spots were identified, founding a good correlation between these sequence motifs and hot recombination spots. Clear differences among chromosomes, regarding both localization and extension of these regions, were observed. A few DSBs hot spots were also localized, revealing strong differences of distribution among chromosomes, and all DSBs colocalized with hot recombination spots. Interestingly, an apparent positive correlation was found between frequency of DSB hot spots and transposable elements, but not with genes (particularly in chromosome arms 1AS and 4AS; Figure 2, Figure 4; Supplemental Figure S1). The clear differences of recombination-associated DNA sequence patterns among homoeologues support the idea that the determinants of complex cytological processes like crossover and chromosome pairing are related to the very sequence of DNA, probably influenced by DNA methylation status, contributing to specificity of the terminal region of the chromosomes.

DNA sequence determines the architectural features of chromosomes both at a local and at a global scale. The subtelomeric regions of the wheat chromosome arms studied here showed a high polymorphism for all the DNA features analyzed that might contribute to specificity of initial chromosome interactions in a polyploid like wheat. However, initial physical interactions of chromosomes to assure a proper pairing between homologues might also need the contribution of DNA-binding proteins or protein complexes through which chromosomes interact (Ding et al., 2016). Thus, we also analyzed the distribution of predicted binding sites of relevant DNA-binding proteins, regulatory

or structural proteins that could be involved in the architecture of chromosomes. Proteins were selected based on their putative role in initial meiotic events described previously in the bibliography, and the availability of identified DNA-binding sites. Distal subtelomeric sequences were analyzed for the distribution of predicted DNA-binding sites of putative wheat proteins homologous to human SMC1β meiosis-specific cohesin (Revenkova et al., 2001), Ying Yang 1 protein (Beagan et al., 2017) and HMG proteins (Subirana & Messeguer, 2011). Besides their role in sister chromatid cohesion, cohesins are very important for various meiosis-specific events including chromosomal axis formation, synaptonemal complex formation and reciprocal recombination (Ding et al., 2016; Ishiguro, Kim, Fujiyama-Nakamura, Shigeaki Kato, & Watanabe, 2011). Ying Yang 1 is an architectural protein with a critical role in connecting higher-order chromatin folding in animals (Beagan et al., 2017) and in *Arabidopsis* (Wu et al., 2012). HMG proteins have been suggested to be involved in initial interactions between homologues, prior to proper pairing, through their potential interaction with AT-rich sites (Subirana & Messeguer, 2011). Our results revealed that the distribution of putative binding sites for the proteins investigated was chromosome-specific, with evident differences in density and distribution of these putative binding sites among homoeologues, in agreement with their putative role in specific homologous chromosome pairing. We found a striking differential distribution of putative binding sites for HMG proteins, with a very low concentration of these putative sites in regions with a high concentration of genes (1AS) or with a low concentration of LTR retrotransposons (distal region of 7DS). A further analysis of the frequency of HMG binding sites was performed and we observed that, despite the highly different density of sites among chromosomes when comparing the 500 Kb distal region, a more similar density among chromosomes was revealed when a broader region of 5 Mb was considered. These data agree with a putative function of HMG proteins in contributing a molecular structure that might support initial chromosome specific interactions for correct chromosome pairing between homologues (Subirana & Messeguer, 2011). In contrast, similar distribution patterns of binding sites for SMC1β cohesin and YY1 in every chromosome were found, with a higher density of sites in regions enriched in genes and retrotransposons (1AS and 4AS) and a particularly low density in regions poor in genes and retrotransposons (distal subtelomeric region of 7DS).

Chromatin remodeling and chromosome movements allow homologues to find each other (Scherthan, 2001; Naranjo, 2014; Prieto, Shaw, & Moore, 2004). In most organisms, and particularly in plants, although chromosomes also interact at centromeres at the onset of meiosis, homologous chromosomes initiate a physical interaction at the bouquet stage (leptotene) of meiosis, while being all telomeres attached to the inner part of the nuclear envelope. Therefore, subtelomeric regions of chromosomes might benefit of this telomere cluster as they are forced to occupy a restricted space where the interaction and temporal stabilization of unstable chromosome interactions take place. The stabilization of homologues pairing seems to be, in some species, dependent of the later formation of chromosome axis and the establishment of the synaptonemal complex, accompanied by DSB creation and repair of these breaks by recombination and crossover (Barzel & Kupier, 2018; Zickler & Kleckner, 2015). However, there are multiple evidences that suggest that chromosome pairing and crossover are, at least, not completely interdependent (Calderón et al., 2018; Jordan, 2006; Zickler & Kleckner, 2015). This means that there is a feature of the genomic architecture that might facilitate chromosome recognition and pairing independently of recombination and DNA damage. We suggest that HMG proteins could play a major role in homologous chromosome pairing through their putative interaction at AT-rich sites accessible in the protruding DNA loops (Subirana & Messeguer, 2011) and should be the targets in future works. Our hypothesis is coherent with a mechanism in which a set of pairing proteins interact via protruding loops with these AT-rich DNA regions, so stabilizing the associations between homologous chromosomes. Though some other genes like HOP1, REC8 and RED1 have been proposed to play crucial roles in chromosome associations, the initial interactions between chromosomes to recognize a homologue to pair and the molecular factors involved remain elusive (Coutou, Goodyer, & Zetka, 2004; Ding et al., 2016; Jordan, 2006). A recent work, supporting our hypothesis, suggested active telomere forces, which are engaged in a tug-of-war against zippering (Marshall & Fung, 2019). At the onset of meiosis, when telomeres are anchored to the nuclear envelope and shaking, homologous chromosome regions of high affinity compete for zippering with homoeologous regions of lower affinity. When the affinity between the terminal regions of chromosomes is strong enough to oppose shaking, zippering of only true homologs is allowed. This idea is in agreement with the observations that sequence specificity, particularly in regions like subtelomeres where DNA sequences are subjected to rapid change, is fundamental for the pairing process.

Our study shows that all the sequence features analyzed within the subtelomeric region of the five wheat chromosome arms studied have a chromosome-specific pattern, with major differences even among homoeologous chromosomes. These differences might contribute to the specificity of homologous recognition and pairing at the onset of

meiosis in a polyploid like wheat. Nevertheless, the chromosome regions that initiate homologous chromosome interactions remain to be identified. This work contributes to identify the targets that should be studied to go deeper into the knowledge of the initial chromosome interactions during meiosis in wheat.

ACKNOWLEDGMENTS

This research was supported by grant AGL2015-64833R from Spanish Ministerio de Economía y Competitividad (MINECO) and The European Regional Development Fund (FEDER) from the European Union. Authors deeply appreciate the comments from Prof. Tomás Naranjo during the preparation of the manuscript and the observations of independent reviewers during the revision of the manuscript.

AUTHORS CONTRIBUTION

M.A. and P.P. conceived and developed the work and wrote the manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Pilar Prieto https://orcid.org/0000-0002-8160-808X

REFERENCES

Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., … Quesneville, H. (2018). International wheat genome sequencing consortium: Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology*, *19*, 111. https://doi.org/10.1186/s13059-018-1491-4

Alkhimova, O. G., Mazurok, N. A., Potapova, T. A., Zakian, S. M., Heslop-Harrison, J. S., & Vershinin, A. V. (2004). Diverse patterns of the tandem repeats organization in rye chromosomes. *Chromosoma*, *113*, 42–52. https://doi.org/10.1007/s00412-004-0294-4

Ambrosini, A., Paul, S., Hu, S., & Riethman, H. (2007). Human subtelomeric duplicon structure and organization. *Genome Biology*, *8*, R151.

Ananiev, E. V., Phillips, R. L., & Rines, H. W. (1998). A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: Are chromosome knobs megatransposons? *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 10785–10790.

Anderson, J. A., Song, Y. S., & Langley, C. H. (2008). Molecular population genetics of Drosophila subtelomeric DNA. *Genetics*, *178*, 477–487.

Ashikawa, I. (2001). Gene-associated CpG islands in plants as revealed by analyses of genomic sequences. *Plant Journal*, *26*(6), 617–625.

Bailey, T. L., & Gribskov, M. (1998). Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics*, *14*(1), 48–54.

Barzel, A., & Kupier, M. (2018). Finding a match: How do homologous sequences get together for recombination? *Nature Reviews Genetics*, *9*, 27–37.

Bass, H. W., Riera-Lizarazu, O., Ananiev, E. V., Bordoli, S. J., Rines, H. W., Phillips, R. L., … Cande, W. Z. (2000). Evidence for the coincident initiation of homolog pairing and synapsis during the telomere-clustering (bouquet) stage of meiotic prophase. *Journal of Cell Science*, *113*(6), 1033–1042.

Beagan, J. A., Duong, M. T., Titus, K. R., Zhou, L., Cao, Z., Ma, J., & Phillips-Cremins, J. E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Research*, *27*(7), 1139–1152. https://doi.org/10.1101/gr.215160.116

Belostotsky, D. A., & Ananiev, E. V. (1990). Characterization of relic DNA from barley genome. *Theoretical and Applied Genetics*, *80*, 374–380.

Blackburn, E. H. (2005). Telomerase and cancer - Kirk A. Landon - AACR prize for basic cancer research lecture. *Molecular Cancer Research*, *3*(9), 477–482. https://doi.org/10.1158/1541-7786.MCR-05-0147

Blokhina, Y. P., Nguyen, A. D., Draper, B. W., & Burgess, S. M. (2019). The telomere bouquet is a hub where meiotic double-strand breaks, synapsis, and stable homolog juxtaposition are coordinated in the zebrafish, *Danio rerio*. *Plos Genetics*. *15*(1), e1007730.

Brown, C. A., Murray, A. W., & Verstrepen, K. J. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Current Biology*, *20*, 895–903.

Calderón, M. C., Rey, M. D., Cabrera, A., & Prieto, P. (2014). The subtelomeric region is important for chromosome recognition and pairing during meiosis. *Scientific Reports*, *4*, 6488. https://doi.org/10.1038/srep06488

Calderón, M. C., Rey, M. D., Martín, A., & Prieto, P. (2018). Homoeologous chromosomes from two *Hordeum* species can recognize and associate during meiosis in wheat in the presence of the *Ph1* locus. *Frontiers in Plant Science*, *9*, 585. https://doi.org/10.3389/fpls.2018.00585

Choi, K., Zhao, X., Kelly, K. A., Venn, O., Higgins, J. D., Yelina, N. E., & Henderson, I. R. (2013). Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nature Genetics*, *45*, 1327–1336.

Choi, K., Zhao, X., Lambing, C., Underwood, C. J., Hardcastle, T. J., Serra, H., & Henderson, I. R. (2018). Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. *Genome Research*, *28*, 1–15. https://doi.org/10.1101/gr.225599.117

Clavijo, B. L., Venturini, L., Schudoma, C., Garcia, G., Kaithakottil, G., Wright, J., & Clark, Matthew D. (2017). An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Research*, *27*(5), 885–896. https://doi.org/10.1101/gr.217117.116

Copenhaver, G. P., & Pikaard, C. S. (1996). RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *The Plant Journal*, *9*, 259–272.

Corcoran, L. M., Thompson, J. K., Walliker, D., & Kemp, D. J. (1988). Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. *Cell*, *53*, 807–813.

Coutou, F., Goodyer, W., & Zetka, M. (2004). Finding and keeping your partner during meiosis. *Cell Cycle*, *3*(8), 1014–1016.

Daron, J., Glover, N., Pingault, L., Theil, S., Jamilloux, V., Paux, E., & Choulet, F. (2014). Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biology*, *15*, 546. https://doi.org/10.1186/s13059-014-0546-4

Darrier, B., Rimbert, H., Balfourier, F., Pingault, L., Josselin, A. A., Servin, B., & Sourdille, P. (2017). High-resolution mapping of crossover events in the hexaploid wheat genome suggests a universal recombination mechanism. *Genetics*, *206*, 1373–1388.

Ding, D. Q., Matsuda, A., Okamasa, K., Nagahama, Y., Haraguchi, T., & Hiraoka, Y. (2016). Meiotic cohesin-based chromosome structure is essential for homologous chromosome pairing in *Schizosaccharomyces pombe*. *Chromosoma*, *125*, 205–214.

Emden, T. S., Forn, M., Forné, I., Sarkadi, Z., Capella, M., Martín Caballero, L., & Braun, S. (2019). Shelterin and subtelomeric DNA sequences control nucleosome maintenance and genome stability. *EMBO Reports*, *20*, e47181. https://doi.org/10.15252/embr.201847181

Escobar, J. S., Cenci, A., Bolognini, J., Haudry, A., Laurent, S., David, J., & Glémin, S. (2010). An integrative test of the dead-end hypothesis of selfing evolution in *Triticeae* (Poaceae). *Evolution; International Journal of Organic Evolution*, *64*, 2855–2872. https://doi.org/10.1111/j.1558-5646.2010.01045.x

Evtushenko, E. V., Levitsky, V. G., Elisafenko, E. A., Gunbin, K. V., Belousov, A. I., Šafář, J., & Vershinin, A. V. (2016). The expansion of heterochromatin blocks in rye reflects the co-amplification of tandem repeats and adjacent transposable elements. *BMC Genomics [Electronic Resource]*, *17*, 337.

Fajkus, J., Kovarík, A., Královics, R., & Bezděk, M. (1995). Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. *Molecular and General Genetics*, *247*(5), 633–638.

Fan, C., Zhang, Y., Yu, Y., Rounsley, S., Long, M., & Wing, R. A. (2008). The subtelomere of *Oryza sativa* chromosome 3 short arm as a hot bed of new gene origination in rice. *Molecular Plant*, *1*, 839–850.

Feuerbach, F., Galy, V., Trelles-Sticken, E., Fromont-Racine, M., Jacquier, A., Gilson, E., … Nehrbass, U. (2002). Nuclear architecture and spatial positioning help establish transcriptional states of telomeres in yeast. *Nature Cell Biology*, *4*, 2014–2221. https://doi.org/10.1038/ncb756

Ganal, M. W., Lapitan, N. L. V., & Tanksley, S. D. (1991). Macrostructure of the tomato telomeres. *The Plant Cell*, *3*, 87–94.

Garrido-Ramos, M. A. (2015). Satellite DNA in Plants: More than just rubbish. *Cytogenetic Genome Research*, *146*, 153–170.

Gaut, B. S., Wright, S. I., Rizzon, C., Dvorak, J., & Anderson, L.K. (2007). Recombination: An underappreciated factor in the evolution of plant genomes. *Nature Reviews Genetics*, *8*, 77–84.

Gill, B. S., & Kimber, G. (1974). Giemsa C-banding and the evolution of wheat. *Proceedings of the National Academy of Sciences of the United States of America*, *71*, 4086–4090.

Glover, L., Alsford, S., & Horn, D. (2013). DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. *Plos Pathogens*, *9*, e1003260.

Griffiths, S., Sharp, R., Foote, T. N., Bertin, I., Wanous, M., Reader, S., … Moore, G. (2006). Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. *Nature*, *439*, 749–752.

Haudry, A., Cenci, A., Guilhaumon, C., Paux, E., Poirier, S., Santoni, S., & Glémin, S. (2008). Mating system and recombination affect molecular evolution in four *Triticeae* species. *Genetical Research*, *90*, 97–109. https://doi.org/10.1017/S0016672307009032

He, Y., Wang, M., Dukowic-Schulze, S., Zhou, A., Tiang, C. L., Shilo, S., & Pawlowski, Wojciech P. (2017). Genomic features shaping the landscape of meiotic double-strand break hotspots in maize. *Proceedings of the National Academy of Sciences*, *114*, 12231–12236. https://doi.org/10.1073/pnas.1713225114

Heacock, M., Spangler, E., Riha, K., Puizina, J., & Shippen, D. E. (2004). Molecular analysis of telomere fusions in Arabidopsis: Multiple pathways for chromosome end-joining. *Embo Journal*, *23*, 2304–2313.

Henderson, E. (1995). Telomere DNA structure. In E. H. Blackburn & C. W. Greider, (Eds.) *Telomeres* (pp. 11–34). Cold Spring Harbor Laboratory Press.

Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. *Nature Reviews Molecular Cell Biology*, *7*, 311–322.

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

International Wheat Genome Sequencing Consortium (IWGSC). (2018). The transcriptional landscape of polyploidy wheat. *Science*, *361*(6403), eaar6089. https://doi.org/10.1126/science.aar6089

Ishiguro, K. I., Kim, J., Fujiyama-Nakamura, S., Shigeaki Kato, S., & Watanabe, Y. (2011). A new meiosis-specific cohesin complex implicated in the cohesin code for homologous pairing. *Embo Reports*, *12*(3), 267–275.

Jain, D., Hebden, A. K., Nakamura, T. M., Miller, K. M., & Cooper, J. P. (2010). HAATI survivors replace canonical telomeres with blocks of generic heterochromatin. *Nature*, *467*, 223–227.

Jordan, P. (2006). Initiation of homologous chromosome pairing during meiosis. *Biochemical Society transactions*, *34*, 545–549.

Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, *115*, 49–63.

Kipling, D. (1995). *The Telomere*. Oxford: Oxford University Press.

Kleckner, N., Zhang, L., Weiner, B., & Zickler, D. (2012). Meiotic chromosome dynamics. In K. Rippe (Ed.), *Genome organization and function in the cell nucleus*. WILEY-VCH Verlag GmbH & Co. KGaA.

Kotani, H., Hosouchi, T., & Tsuruoka, H. (1999). Structural analysis and complete physical map of *Arabidopsis thaliana* chromosome 5 including centromeric and telomeric regions. *DNA Research*, *6*, 381–386.

Koukalova, B., Reich, J., Matyasek, R., Kuhrova, V., & Bezdek, M. (1989). A BamHI family of highly repeated DNA sequences of *Nicotiana tabacum*. *Theorhetical Applied Genetics*, *78*, 77–80.

Kraemer, S. M., Kyes, S. A., Aggarwal, G., Springer, A. L., Nelson, S. O., Christodoulou, Z., … Smith, J. D. (2007). Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: Comparisons of geographically diverse isolates. *BMC Genomics [Electronic Resource]*, *8*, 45.

Kuo, F., Olsen, K. M., & Richards, E. J. (2006). Natural variation in a subtelomeric region of *Arabidopsis*: Implications for the genomic dynamics of a chromosome end. *Genetics*, *173*(1), 401–417.

Lamb, J. C., Meyer, J. M., Corcoran, B., Kato, A., Han, F., & Birchler, J. A. (2007). Distinct chromosomal distribution of highly repetitive sequences in maize. *Chromosome Research*, *15*, 33–49.

Larsen, F., Gundersen, G., Lopez, R., & Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics*, *13*(4), 1095–1107.

Li, W., Zhang, P., Fellers, J. P., Friebe, B., & Gill, B. S. (2004). Sequence composition, organization, and evolution of the core Triticeae genome. *Plant Journal*, *40*(4), 500–511. https://doi.org/10.1111/j.1365-313X.2004.02228.x

Linardopoulou, E. V., Williams, E. M., Fan, Y. X., Friedman, C., Young, J. M., & Trask, B. J. (2005). Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature*, *437*, 94100.

Linde-Laursen, I. (1978). Giemsa C-banding of barley chromosomes. I. Banding pattern polymorphism. *Hereditas*, *88*, 55–64.

Liu, B., Wang, S., Long, R., & Chou, K. C. (2016). iRSpot-EL: Identify recombination spots with an ensemble learning approach. *Bioinformatics*, *33*(1), 35–41.

Louis, E. J. (2014). Introduction. In E. J. Louis & M. M. Becker (Eds.). *Subtelomeres* (pp. 1–12). Springer.

Louis, E. J., Naumova, E. S., Lee, A., Naumov, G., & Haber, J. E. (1994). The chromosome end in yeast: Its mosaic nature and influence on recombinational dynamics. *Genetics*, *136*, 789–802.

Louis, E. J., & Vershinin, A. V. (2005). Chromosome ends: Different sequences may provide conserved functions. *Bioessays*, *27*, 685–697.

Macina, R. A., Morii, K., Hu, X. L., Negorev, D. G., Spais, C., Ruthig, L. A., & Riethman, H. C. (1995). Molecular cloning and RARE cleavage mapping of human 2p, 6q, 8q, 12q, and 18q telomeres. *Genome Research*, *5*(3), 225–232.

Macina, R. A., Negorev, D. G., Spais, C., Ruthig, L. A., Hu, X. L., & Riethman, H. C. (1994). Sequence organization of the human chromosome 2q telomere. *Human Molecular Genetics*, *3*(10), 1847–1853.

Mao, L., Devos, K. M., Zhu, L., & Gale, M. D. (1997). Cloning and genetic mapping of wheat telomere-associated sequences. *Molecular and General Genetics*, *254*, 584–591.

Marshall, W. F., & Fung, J. C. (2019). Modeling meiotic chromosome pairing: A tug of war between telomere forces and a pairing-based Brownian ratchet leads to increased pairing fidelity. *Physical Biology*, *16*, 046005. https://doi.org/10.1088/1478-3975/ab15a7

Martín, A. C., Rey, M. D., Shaw, P., & Moore, G. (2017). Dual effect of the wheat *Ph1* locus on chromosome synapsis and crossover. *Chromosoma*, *126*(6), 669–680.

Mefford, H. C., & Trask, B. J. (2002). The complex structure and dynamic evolution of human subtelomeres. *Nature Reviews Genetics*, *3*, 91–102.

Mikhailova, E. I., Sosnikhina, S. P., Kirillova, G. A., Tikholiz, O. A., Smirnov, V. G., Jones, R. N., & Jenkins, G. (2001). Nuclear dispositions of subtelomeric and pericentromeric chromosomal domains during meiosis in asynaptic mutants of rye (*Secale cereale* L.) *Journal of Cell. Sci*, *114*(10), 1875–1882.

Mizuno, H., Wu, J., Kanamori, H., Fujisawa, M., Namiki, N., Saji, S., … Matsumoto, T. (2006). Sequencing and characterization of telomere and subtelomere regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S. *Plant Journal*, *46*(2), 206–217.

Mizuno, H., Wu, J., Katayose, Y., Kanamori, H., Sasaki, T., & Matsumoto, T. (2008a). Characterization of chromosome ends on the basis of the structure of TrsA subtelomeric repeats in rice (*Oryza sativa* L.) *Molecular Genetics and Genomics*, *280*, 19–24. https://doi.org/10.1007/s00438-008-0341-6

Mizuno, H., Wu, J., Katayose, Y., Kanamori, H., Sasaki, T., & Matsumoto, T. (2008b). Chromosome-specific distribution of nucleotide substitutions in telomeric repeats of rice (*Oryza sativa* L.). *Molecular Biology and Evolution*, *25*, 62–68.

Mizuno, H., Wu, J., & Matsumoto, T. (2014). Characterization of chromosomal ends on the basis of chromosome specific telomere variants and subtelomeric repeats in rice (*Oryza sativa* L.). In E. J. Louis & M. M. Becker (Eds.), *Subtelomeres* (pp. 187–194). Springer.

Mouchiroud, D., D'onofrio, G., Aissani, B., Macaya, G., Gautier, C., & Bernardi, G. (1991). The distribution of genes in the human genome. *Gene*, *100*, 181–187.

Muyle, A., Serres-Giardi, L., Ressayre, A., Escobar, J., & Glemin, S. (2011). GC-biased gene conversion and selection affect GC content in the *Oryza* genus (rice). *Molecular Biology and Evolution*, *28*, 2695–2706. https://doi.org/10.1093/molbev/msr104

Naranjo, T. (2014). Dynamics of rye telomeres in a wheat background during early meiosis. *Cytogenetic and Genome Research*, *143*, 60–68.

Naranjo, T. (2015). Forcing the shift of the crossover site to proximal regions in wheat chromosomes. *Theoretical and Applied Genetics*, *128*, 1855–1863. https://doi.org/10.1007/s00122-015-2552-7

Nasmyth, K., & Haering, C. H. (2005). The structure and function of SMC and kleisin complexes. *Annual Review of Biochemistry*, *74*, 595–648.

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

Ohmido, N., Kijima, K., Ashikawa, I., de Jong, J. H., & Fukui, K. (2001). Visualization of the terminal structure of rice chromosomes 6 and 12 with multicolor FISH to chromosomes and extended DNA fibers. *Plant Molecular Biology*, *47*, 413–421. https://doi.org/10.1023/a:1011632111845

Ong, C. T., & Corces, V. G. (2014). CTCF: An architectural protein bridging genome topology and function. *Nature*, *15*, 234–246.

Page, S. L., & Hawley, R. S. (2003). Chromosome choreography: The meiotic ballet. *Science*, *301*(5634), 785–789.

Pearce, S. R., Pich, U., Harrison, G., Flavell, A. J., Heslop-Harrison, J. S., … Kumar, A. (1996). The Ty1-copia group retrotransposons of Allium cepa are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. *Chromosome Research*, *4*, 357–364. https://doi.org/10.1007/BF02257271

Prieto, P., Martin, A., & Cabrera, A. (2004). Chromosomal distribution of telomeric and telomeric-associated sequences in *Hordeum chilense* by *in situ* hybridization. *Hereditas*, *141*, 122–127.

Prieto, P., Shaw, P., & Moore, G. (2004). Homologue recognition during meiosis is associated with a change in chromatin conformation. *Nature Cell Biology*, *6*, 906–908.

Pryde, F. E., Gorham, H. C., & Louis, E. J. (1997). Chromosome ends: All the same under their caps. *Current Opinion In Genetics & Development*, *7*, 822–828.

Rehmeyer, C., Li, W., Kusaba, M., Kim, Y. S., Brown, D., Staben, C., … Farman, M. (2006). Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. *Nucleic Acids Research*, *34*, 4685–4701.

Revenkova, E., Eijpe, M., Heyting, C., Gross, B., & Jessberger, R. (2001). Novel meiosis-specific isoform of mammalian SMC1. *Molecular and Cellular Biology*, *21*, 6984–6998.

Rey, M. D., Martín, A. M., Higgins, J., Swarbreck, D., Uauy, C., Shaw, P., & Moore, G. (2017). Exploiting the ZIP4 homologue within the wheat *Ph1* locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. *Molecular Breed*, *37*(8), 95.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European molecular biology open software suite. *Trends in Genetics*, *16*(6), 276–277.

Rietchman, H., Ambrosini, A., & Paul, S. (2005). Human subtelomere structure and variation. *Chromosome Research*, *13*, 505–515.

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

Riley, R., & Chapman, V. (1964). The effect of the deficiency of the long arm of chromosome 5B on meiotic pairing in *Triticum aestivum*. *Wheat Info. Serv*, *17*, 12–15.

Riley, R., & Kempana, C. (1963). The homoeologous nature of the non-homologous meiotic pairing in *Triticum aestivum* deficient for chromosome V (5B). *Heredity*, *18*, 287–306.

Röder, M. S., Lapitan, N. L., Sorrells, M. E., & Tanksley, S. D. (1993). Genetic and physical mapping of barley telomeres. *Molecular and General Genetics*, *238*(1–2), 294–303.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., … Krumm, A. (2008). CTCF physically links cohesion to chromatin. *Proceedings of the National Academy of Sciences*, *105*, 8309–8314.

Rudd, M. K., Friedman, C., Parghi, S. S., Linardopoulou, E. V., Hsu, L., & Trask, B. J. (2007). Elevated rates of sister 772 chromatid exchange at chromosome ends. *Plos Genetics*, *3*(2), 319–323.

Said, M., Recio, R., & Cabrera, A. (2012). Development and characterization of structural changes in chromosome 3Hch from *Hordeum chilense* in common wheat and their use in physical mapping. *Euphytica*, *188*, 429–440.

Salina, E. A., Sergeeva, E. M., Adonina, I. G., Shcherban, A. B., Afonnikov, D. A., Belcram, H., … Chalhoub, B. (2009). Isolation and sequence analysis of the wheat B genome subtelomeric DNA. *BMC Genomics [Electronic Resource]*, *10*, 414. https://doi.org/10.1186/1471-2164-10-414

SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., MelakeBerhan, A., … Bennetzen, J. L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science*, *274*, 765–768.

Scherthan, H. (2001). A bouquet makes ends meet. *Nature Reviews Molecular Cell Biology*, *2*, 621–627.

Sears, E. R., & Okamoto, M. (1958). Intergenomic chromosome relationship in hexaploid wheat. *Proceedings of the 10th International Congress of Genetics*, Toronto, ON, Canada, 258–259.

Serres-Giardi, L., Belkhir, K., David, J., & Glemin, S. (2012). Patterns and evolution of nucleotide landscapes in seed plants. *Plant Cell*, *24*, 1379–1397. https://doi.org/10.1105/tpc.111.093674

Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., & Levy, A. A. (2015). DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in Arabidopsis. *Plant Cell*, *27*, 2427–2436.

Subirana, J. A., & Messeguer, X. (2011). The distribution of alternating AT sequences in eukaryotic genomes suggests a role in homologous chromosome recognition in meiosis. *Journal of Theoretical Biology*, *283*, 28–34.

Subirana, J. A., & Messeguer, X. (2010). The most frequent short sequences in non-coding DNA. *Nucleic Acids Research*, *38*(4), 1172–1181.

Sundararajan, A., Dukowic-Schulze, S., Kwicklis, M., Engstrom, K., Garcia, N., Oviedo, O. J., … Mudge, J. (2016). Gene evolutionary trajectories and GC patterns driven by recombination in *Zea mays*. *Frontiers in Plant Science*, *7*, 1433. https://doi.org/10.3389/fpls.2016.01433

Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J. L., Jackson, S. A., … Ma, J. (2009). Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Research*, *19*, 2221–2230. https://doi.org/10.1101/gr.083899.108

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

Uchida, W., Matsunaga, S., Sugiyama, R., & Kawano, S. (2002). Interstitial telomere-like repeats in the *Arabidopsis thaliana* genome. *Genes and Genetic Systems*, *77*, 63–67.

Valenzuela, N. T., Perera, E., & Naranjo, T. (2013). Identifying crossover-rich regions and their effect on meiotic homologous interactions by partitioning chromosome arms of wheat and rye. *Chromosome Research*, *21*, 433–445. https://doi.org/10.1007/s10577-013-9372-x

Vershinin, A. V., & Evtushenko, E. V. (2014). What is the specificity of plant subtelomeres? In E. J. Louis & M. M. Becker (Eds.), *Subtelomeres* (pp. 195–211). Springer.

Vershinin, A. V., Schwarzacher, T., & Heslop-Harrison, J. S. (1995). The large-scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes. *Plant Cell*, *7*, 1823–1833. https://doi.org/10.1105/tpc.7.11.1823

Vicient, C. M., & Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Annals of Botany*, *120*, 195–207. https://doi.org/10.1093/aob/mcx078

Weil, C. F., & Wessler, S. R. (1993). Molecular evidence that chromosome breakage by Ds elements is caused by aberrant transposition. *The Plant Cell*, *5*, 515–522.

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., … Peters, J. -. M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, *451*, 796–801.

Wicker, T., Gundlach, H., Spannag, M., Uauy, C., Borrill, P., Ramírez-González, R. H., … Choulet, F. (2018). Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biology*, *19*, 103. https://doi.org/10.1186/s13059-018-1479-0

Wijnker, E., James, G. V., Ding, J., Becker, F., Klasen, J. R., Rawat, V., … Schneeberger, K. (2013). The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife*, *2*, e01426. https://doi.org/10.7554/eLife.01426

Wu, X., Cheng, Y., Li, T., Wang, Z., & Liu, J. Y. (2012). *In vitro* identification of DNA-binding motif for the new zinc finger protein AtYY1. *Acta Biochimica Et Biophysica Sinica*, *44*(6), 483–489. https://doi.org/10.1093/abbs/gms020

Wu, K. S., & Tanksley, S. D. (1993). Genetic and physical mapping of telomeres and macrosatellites of rice. *Plant Molecular Biology*, *22*, 861–872. https://doi.org/10.1007/BF00027371

Zamudio, N., Barau, J., Teissandier, A., Walter, M., Borsos, M., Servant, N., … Bourc'his, D. (2015). DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes and Development*, *29*, 1256–1270.

Zhao, D., Ferguson, A. A., & Jiang, N. (2016). What makes up plant genomes: The vanishing line between transposable elements and genes. *Biochimica et Biophysica Acta*, *1859*, 366–380.

Zhu, Z., & Wang, X. Roles of cohesin in chromosome architecture and gene expression. *Seminars in Cell & developmental biology*, *90*, 187–193.

Zickler, D (2006). From early homologue recognition to synaptonemal complex formation. *Chromosoma*, *115*, 158–174.

Zickler, D., & Kleckner, N. (2015). Recombination, pairing, and synapsis of homologs during meiosis. *Cold Spring Harbor perspectives in biology*, *7*(6), a016626. https://doi.org/10.1101/cshperspect.a016626

Zickler, D., & Kleckner, N. (1998). The leptotene-zygotene transition of meiosis. *Annual Review of Genetics*, *32*, 619–697.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Aguilar M, Prieto P. Sequence analysis of wheat subtelomeres reveals a high polymorphism among homoeologous chromosomes. *Plant Genome*. 2020;e20065. https://doi.org/10.1002/tpg2.20065