Published October 2, 2015

ORIGINAL RESEARCH

Physical Mapping of Bread Wheat Chromosome 5A: An Integrated Approach

Delfina Barabaschi,* Federica Magni, Andrea Volante, Agata Gadaleta, Hana Šimková, Simone Scalabrin, Maria Lucia Prazzoli, Paolo Bagnaresi, Katia Lacrima, Vania Michelotti, Francesca Desiderio, Luigi Orrù, Valentina Mazzamurro, Agostino Fricano, AnnaMaria Mastrangelo, Paola Tononi, Nicola Vitulo, Irena Jurman, Zeev Frenkel, Federica Cattonaro, Michele Morgante, Antonio Blanco, Jaroslav Doležel, Massimo Delledonne, Antonio M. Stanca, Luigi Cattivelli, and Giampiero Valè

Abstract

The huge size, redundancy, and highly repetitive nature of the bread wheat [*Triticum aestivum* (L.)] genome, makes it among the most difficult species to be sequenced. To overcome these limitations, a strategy based on the separation of individual chromosomes or chromosome arms and the subsequent production of physical maps was established within the frame of the International Wheat Genome Sequence Consortium (IWGSC). A total of 95,812 bacterial artificial chromosome (BAC) clones of short-arm chromosome 5A (5AS) and long-arm chromosome 5A (5AL) arm-specific BAC libraries were fingerprinted and assembled into contigs by complementary analytical approaches based on the FingerPrinted Contig (FPC) and Linear Topological Contig (LTC) tools. Combined anchoring approaches based on polymerase chain reaction (PCR) marker screening, microarray, and sequence homology searches applied to several genomic tools (i.e., genetic maps, deletion bin map, neighbor maps, BAC end sequences (BESs), genome zipper, and chromosome survey sequences) allowed the development of a high-quality physical map with an anchored physical coverage of 75% for 5AS and 53% for 5AL with high portions (64 and 48%, respectively) of contigs ordered along the chromosome. In the genome of grasses, *Brachypodium* [*Brachypodium distachyon* (L.) Beauv.], rice (*Oryza sativa* L.), and sorghum [*Sorghum bicolor* (L.) Moench] homologs of genes on wheat chromosome 5A were separated into syntenic blocks on different chromosomes as a result of translocations and inversions during evolution. The physical map presented represents an essential resource for fine genetic mapping and map-based cloning of agronomically relevant traits and a reference for the 5A sequencing projects.

Published in The Plant Genome 8
doi: 10.3835/plantgenome2015.03.0011

5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

D. Barabaschi, A. Volante, M.L. Prazzoli, P. Bagnaresi, K. Lacrima, V. Michelotti, F. Desiderio, L. Orrù, A.M. Stanca, L. Cattivelli, and G. Valè, Council for Agricultural Research and Economics (CREA)–Genomics Research Centre, Fiorenzuola d'Arda, Piacenza, I-29017; M.L. Prazzoli, present address: Fondazione E. Mach (IASMA), S. Michele,Trento, I-38010; F. Magni, S. Scalabrin, I. Jurman, F. Cattonaro, and M. Morgante, Institute of Applied Genomics, Udine, I-33100; A. Gadaleta and A. Blanco, Dep. of Soil, Plant and Food Sciences, Section of Genetic and Plant Breeding, Univ. of Bari, Bari, I-70126; H. Šimková and J. Doležel, Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, CZ-77200; V. Mazzamurro and A.M. Stanca, Dep. of Life Sciences, Univ. of Modena and Reggio Emilia, Reggio Emilia, I-42100; A. Fricano, Parco Tecnologico Padano, Lodi, I-26900, present address: Council for Agricultural Research and Economics (CREA)–Unità di ricerca per la maiscoltura, Bergamo, I-24126; A.M. Mastrangelo, Council for Agricultural Research and Economics (CREA)–Cereal Research Centre, Foggia, I-71122; P. Tononi and M. Delledonne, Dep. of Biotechnology, Univ. of Verona, Verona, I-37129; N. Vitulo, CRIBI Biotechnology Center, Univ. of Padova, Padova, I-35121; Z. Frenkel, Institute of Evolution and Dep. of Evolutionary and Environmental Biology, Univ. of Haifa, Haifa, IL-3498838; G. Valè, Council for Agricultural Research and Economics (CREA)–Rice Research Unit, Vercelli, I-13100. Received 9 Mar. 2015. Accepted 21 June 2015. *Corresponding author (delfina.barabaschi@entecra.it).

Abbreviations: 5AL, long arm of chromosome 5A; 5AS, short arm of chromosome 5A; BAC, bacterial artificial chromosome; BES, bacterial artificial chromosome end sequence; COS, conserved ortholog set; CS, *Triticum aestivum* L. 'Chinese Spring'; CS5A, 'Chinese Spring'–*Triticum dicoccoides* disomic substitution 5A; CSS, chromosome survey sequences; DArT, diversity array technology; dDT5A, double ditelosomic line 5A; dNTP, deoxynucleoside triphosphate; DV, *Triticum monococcum* L. 'DV92'; EST, expressed sequence tags; FPC, FingerPrinted Contigs (software); G3, *Triticum monococcum* L. 'G3116'; GZ, genome zipper; IDP, insertion-deletion polymorphisms; ISBP, insertion site-based polymorphism; IWGSC, International Wheat Genome Sequence Consortium; kb, kilobase; L_{50}, length of the smallest contig needed to cover 50% of the assembly; LG, linkage group; Lt, *Triticum turgidum* ssp. *durum* 'Latino'; LTC, Linear Topological Contig (software); Mb, megabase; MG, *Triticum turgidum* ssp. *dicoccum* 'MG5323'; MTP, minimal tiling path; N_{50}, the contig number needed to cover 50% of the assembly; PCR, polymerase chain reaction; POPSEQ, population sequencing; Re, *Triticum aestivum* L. 'Renan'; RFLP, restriction fragment length polymorphism; RIL, recombinant inbred line; RJJM, repeat junction–junction marker; RJM, repeat junction marker; SNP, single nucleotide polymorphisms; SSR, simple-sequence repeat; STS, sequence tag site; TE, transposable element; URGI, Unité de Recherche Génomique Info.

The key for access to the complete gene catalog and regulatory portion of a species is a high-quality reference genome sequence. Such a resource is also fundamental to disclose genomic structural variations among different genotypes and how it relates to phenotypic variation and for providing genome-wide molecular tools to assist crop improvement and to explore genome organization, evolution, speciation, and domestication (Barabaschi et al., 2012). Despite recent advances in sequencing technologies and the evolution of related protocols and equipment (Egan et al., 2012), to achieve high-quality sequences of a complex genome like that of allohexaploid bread wheat (2*n* = 6x = 42, 1C genome size 17 gigabase, and repetitive fraction representing more than 80%) remains a challenge (Feuillet et al., 2011). The large amount of repeated sequences present in the wheat genome makes de novo assembly of nongenic regions extremely difficult. Currently, the gold standard approach for a complete bread wheat reference genome sequence is based on BAC by BAC sequencing, which in turn relies on the availability of robust physical maps as a prerequisite for the correct assembly of the sequences (Mascher and Stein, 2014; Bolger et al., 2014).

An international collaborative program coordinated by the IWGSC (http://www.wheatgenome.org) was established to accelerate the release of the reference genome sequence of wheat (Feuillet and Eversole, 2007). The guidelines of the consortium proved to be successful in terms of number and quality of physical maps produced with the reference cultivar Chinese Spring (CS) for several bread wheat chromosomes (Paux et al., 2008; Lucas et al., 2013; Philippe et al., 2013; Raats et al., 2013; Breen et al., 2013; Poursarebani et al., 2014).

The IWGSC strategy is based on the purification of chromosomes or chromosome arms by flow cytometry (Doležel et al., 2007, 2012) for preparation of specific BAC libraries (Šafář et al., 2010), the fingerprinting analysis of BACs with SNaPshot labeling or whole-genome profiling for contig assembly, definition of a minimal tiling path (MTP) of BAC clones (Luo et al., 2003; Van Oeveren et al., 2011), and anchoring of contigs to genome maps (e.g., genetic maps, gene order predicted by the genome zipper).

The quality of physical maps and MTPs (Ariyadasa and Stein, 2012) was improved through the use of two complementary analytical approaches applied to high-information-content fingerprinting: (i) FPC, generally used for the cereal MTP assemblies published so far (Soderlund et al., 1997, 2000); and (ii) LTC, recently developed to improve physical assembly in complex genomes, enabling longer, better ordered, and more robust contigs compared with FPC (Frenkel et al., 2010).

Chromosome 5A has an estimated size of 827 megabase (Mb) (Šafář et al., 2010), more than twice that of the rice genome, and contains an abundance of loci for agronomically important traits such as vernalization requirement (*TaVRT-1*, Daniluk et al., 2003), cold tolerance (*Cbf-DREB* and *dehydrin*-related genes, Vágújfalvi et al., 2003; Tondelli et al., 2011), domestication traits (e.g., free-threshing *Q* gene, Simons et al., 2006), and resistance to Fusarium head blight (Buerstmayr et al., 2003; Salameh et al., 2011). Several genes encoding soluble proteins influencing kernel texture, like puroindoline and grain softness protein (Tranquilli et al., 2002; Massa and Morris, 2006) with important effects on the rheological properties of flour and bread and pasta-making quality, are also located on 5A and its two homeologous chromosomes 5B and 5D. In addition, 5A and 5B carry the phytoene synthase (*PSY*) gene coding for an important enzyme involved in the carotenoid biosynthetic pathway, which is strongly associated with the yellow pigment content in wheat grain (Cenci et al., 2004).

A first sequencing survey of the 5A chromosome was provided by Vitulo et al. (2011) and lead to an estimation of the number of genes (about 5000), microRNAs, and overall transposable element (TE) content (about 72%). On the basis of synteny analysis with model grass genomes, *Brachypodium*, rice, and sorghum, 5AS is related to chromosomes Bd4, Os12, and Sb8, respectively, while 5AL syntenic regions have been identified in *Brachypodium* chromosomes Bd4 and Bd1, rice chromosomes Os9 and Os3, and sorghum chromosomes Sb1 and Sb2. Supposing long-time conservation of local gene order, a gene order on 5A was predicted based on synteny with the *Brachypodium*, rice, and sorghum genome (392 genes for 5AS and 1480 genes for 5AL), in a tool referred to as the genome zipper (GZ) (Vitulo et al., 2011).

In this work we present anchored physical maps for both arms of bread wheat chromosome 5A. Overall, 75,995 fingerprinted BACs were assembled into contigs independently by FPC and LTC programs and two versions (partially overlapping) of MTPs were established for each chromosome arm. Anchoring of the contigs was performed with three-dimensional pools of MTP clones by: (i) PCR screening of markers from four 5A genetic maps and a 5A consensus genetic map constructed with a neighbor approach and (ii) hybridization of BAC pool DNA onto a high-throughput array containing sequences belonging to the 5A-GZ. Markers used for anchoring were also assigned to chromosome 5A deletion bins to highlight possible discrepancies with the physical map generated.

The physical map was also integrated with the data of about 5000 BESs and the chromosome survey sequencing data from the IWGSC consortium (IWGSC, 2014). Use of different assembly tools and multiple anchoring platforms and resources generated a high-quality integrated physical map for chromosome 5A. This is the most up-to-date resource currently available for map-based cloning of genes of interest from chromosome 5A or for generating markers for these genes for breeding and for progressing the complete sequencing of the chromosome.

Materials and Methods

Purification of Chromosome Arms and Construction of Bacterial Artificial Chromosome Libraries

The short and long arms of chromosome 5A were purified separately by flow cytometric sorting from a double ditelosomic line 5A of wheat cultivar CS ($2n = 40 + 2t5AS + 2t5AL$), in which both arms were stably maintained as telocentric chromosomes (Sears and Sears, 1978). Seeds of the double ditelosomic line 5A were kindly provided by Prof. Bikram S. Gill (Kansas State University, Manhattan, USA). Preparation of samples for flow cytometry and chromosome sorting was done as described by Vrána et al. (2000). The identity and purity of sorted fractions was checked by fluorescence in situ hybridization on chromosomes sorted onto microscope slides with probes for a telomeric repeat and GAA microsatellite (Janda et al., 2006). Two chromosome arm-specific BAC libraries containing 46,080 (5AS: coded as TaaCsp5AShA library) and 90,240 (5AL: coded as TaaCsp5ALhA library) clones were constructed as described by Šimková et al., (2011; Table 1). To estimate average insert size, 200 and 280 BAC clones from 5AS and 5AL, respectively, were randomly selected and analyzed as described in Janda et al. (2006).

The two libraries are stored at the French Plant Genomic Resources Center, Toulouse, France.

Fingerprinting Reaction and Data Analysis

Using the SNaPshot high-information-content fingerprinting procedure (Luo et al., 2003) as described in Paux et al. (2008), 44,740 clones of the 5AS-specific BAC library and 51,072 clones of the 5AL-specific BAC library were analyzed. Briefly, about 100 ng of purified BAC DNA were digested with five restriction enzymes (*Bam*HI-HF, *Eco*RI-HF, *Xba*I, *Xho*I, and *Hae*III, 1 U each, New England Biolabs, NEB.) and the restricted fragments were end labeled with the SNaPshot multiplex labeling kit (Applied Biosystems). Fragment lengths were estimated on an ABI 3730XL DNA capillary sequencer (Applied Biosystems) with GeneScan 500 LIZ size standard (Applied Biosystems). Data were processed with ABI GeneMapper 3.5 (Applied Biosystems), fingerprinting background removal (Scalabrin et al., 2009), and GenoProfiler 2.1 (You et al., 2007) programs to size the fragments and remove background noise and contamination. Raw electropherograms were first analyzed with GeneMapper program and peak areas, peak heights, and band sizes of each BAC fingerprint profile were exported in text format. Spurious peaks (background noise, vector bands, and partial or nonspecific digestions) and bands outside of the range 50 to 500 bp were removed by fingerprinting background removal. This program was also used to discard low-quality fingerprints that may have negatively affected contig assembly and to convert data to a format that was compatible with GenoProfiler, FPC, and LTC programs. GenoProfiler was used to detect cross-contaminating clones arrayed in 384- and 96-well plates and to remove negative controls. At the end of the processes, a total of 36,165 high-quality fingerprints (80.8% of the original BAC clones) for 5AS and 39,830 high-quality fingerprints (77.9% of the original BAC clones) for 5AL were obtained (Table 1).

Table 1. Description of the two 5A bacterial artificial chromosome (BAC) libraries and comparison between FingerPrinted Contigs (FPC) and Linear Topological Contig (LTC) assembly.

	5AS[†]		5AL[†]	
Chromosome arm size (Mb)[‡]	295		532	
BAC library size (clones)	46,080		90,240	
Average insert size (kb)	120		123	
Library clone depth	16.5×		18.3×	
Fingerprinting technique	SNaPshot HICF		SNaPshot HICF	
Fingerprinted clone no.	44,740		51,072	
Useful fingerprint no.	36,165		39,830	
Chromosome arm equivalents[§]	13.2×		8.1×	
Assembly method	**FPC**	**LTC**	**FPC**	**LTC**
Assembly protocol	IWGSC guideline	Frenkel et al. (2010)	IWGSC guideline	Frenkel et al. (2010)
Assembly stringency	1.00×10^{-45}	1.00×10^{-20}	1.00×10^{-45}	1.00×10^{-20}
Contig no.	1308	652	2556	1504
Clones in contig no.	25,084	26,659	27,764	29,610
Singletons no.	11,081	9506	12,066	10,220
Questionable clone no.	468	355	522	642
Minimal tiling path clone no.	4201	5412	6560	8709
Longest contig (kb)[¶]	1297	3391	1027	2303
Estimated chromosome arm coverage (Mb)[¶,††]	342	330	601	678
Estimated chromosome arm coverage (%)[#,††]	116	112	113	127
Contig N_{50}[¶,††]	354	128	823	407
Contig L_{50} (kb)[¶,††]	296	820	251	563
Average contig size (kb)[¶,††]	261	506	235	451
Contigs ≥ 5 clone no.	948	460	1791	1162
Estimated chromosome arm coverage (Mb)[¶,‡‡]	291	297	483	611
Estimated chromosome arm coverage (%)[†,‡‡]	99	101	91	115
Contig N_{50}[¶,‡‡]	286	109	603	350
Contig L_{50}(kb)[¶,‡‡]	418	900	282	601
Average contig size (kb)[¶,‡‡]	307	646	270	526

[†] 5AS, short arm of 5A chromosome; 5AL, long arm of 5A chromosome; HICF, high information content fingerprinting; MTP, minimal tiling path; IWGSC, International Wheat Genome Sequence Consortium.

[‡] Šafář et al. (2010).

[§] Derived by the ratio between average insert size × fingerprint no. × proportion of chromosomes in the sorted fraction and the chromosome arm size.

[¶] Based on a CB unit size of 1.2 kb for SNaPshot HICF method, as defined in the IWGSC guideline.

[#] Based on the expected size of the chromosome arm.

[††] Based on the total number of contigs (≥2 clones).

[‡‡] Based on the total number of contigs (≥5 clones).

Automated Bacterial Artificial Chromosome Assembly with FingerPrinted Contigs and Linear Topological Contig and Definition of Four Minimal Tiling Paths

The 5AS and 5AL high-quality fingerprints were analyzed with FPC (Soderlund et al., 1997) and LTC software tools (Frenkel et al., 2010) to independently obtain two contig assemblies. The first version of the MTP was generated by means of FPC. This MTP was used for anchoring of markers to BACs. Later on, the second version of the MTP was generated by using LTC, which became available only after the FPC-based MTP was selected and arranged into three-dimensional pools. The FPC-based assembly was done following the IWGSC guidelines for physical mapping (described in Paux et al., 2008; Appels et al., 2010) with the following parameters: scaling value for band size 30, tolerance 12 (corresponding to real tolerance 0.4 bp) and gel length 54,000 [= (500 − 50) × 4 × 30]. An initial assembly of all clones was performed using the incremental contig building function under stringent conditions applying a Sulston Score probability cut-off of 1×10^{-75}. Contigs containing more than 10% of questionable (Q) clones were broken up by the DQer function with increasing stringency (1×10^{-3} at each step, up to 3 times: 1×10^{-78}, 1×10^{-81}, and 1×10^{-84}). This first process resulted in robust (to start point selection) contigs used as highly reliable cores for the next iterative procedure. Step-wise merging and DQing were performed under relaxing stringency until a cut-off of 1×10^{-45} (1×10^{-5} at each step: 1×10^{-70}, 1×10^{-65}, 1×10^{-60}, 1×10^{-55}, 1×10^{-50}, and finally 1×10^{-45}). The MTP was generated with the FPC-MTP module with the following parameters: minimum FPC overlap on band map 30, maximum overlap 250, FromEnd = 0, minimum shared bands 12, and preference given to large clones. For 5AS, the MTP consisted of 4201 clones, and for 5AL it consisted of 6560 clones (Table 1).

Automatic LTC assembly was run with the following parameters: tolerance 12 (corresponding to 0.4 bp), gel length for data reading 60,000 (= 500 × 4 × 30) and total number of different bands for Sulston score calculation 2250 [= (500 − 50) × 4 × 30/(2 × 12)]. A network of significant clone overlaps was constructed under liberal conditions assuming a Sulston score cut-off of 1×10^{-15}. The questionable overlaps and clones were excluded by assuming basic cut-off of 1×10^{-15} and an additional more-stringent cut-off of 1×10^{-25}. After exclusion from the analysis of questionable clones and overlaps, automated adaptive clustering of clones was performed, applying an initial cut-off of 1×10^{15} for the long arm and of 1×10^{-20} for the short arm and increasing stringency at each cycle (1×10^{-3} at each step, up to 10 times). The MTP was selected with the LTC default settings (double covering of contig ends, significant overlap of consequent MTP clones at cut-off of contig construction in the adaptive clustering).

Contigs with only two to four clones have usually much lower coverage than most of contigs with five clones and more. A low coverage increases a risk of false-positive overlaps that can derive from clone-by-clone DNA contamination or errors in fingerprinting. Ordering of clones within short contigs and estimation of contig length is also problematic (usually overestimated by two to three times). Hence, we considered contigs with less than five clones as less reliable.

The average size of the anchored contigs was calculated based on the estimation of a virtual band length ~1.2 kilobase (kb) (as recommended by internal communication of the IWGSC consortium).

Three-Dimensional Pooling of the Four Minimal Tiling Paths

The clones of the four MTPs, defined either with FPC (5AS-FPC-MTP, 5AL-FPC-MTP) or LTC (5AS-LTC-MTP, 5AL-LTC-MTP), were recovered manually from 5AS and 5AL BAC libraries and arranged into 384-well microtiter plates (Greiner Bio One) containing 80 μL of 2YT medium supplemented with 12.5 μg mL^{-1} chloramphenicol and 7% glycerol. The plates were then incubated at 37°C for 16 h and finally stored at −80°C. To speed up the screening, three-dimensional pools were created (plate, row, and column) for each MTP (Yim et al., 2007). The three-dimensional pools belonging to the FPC-MTPs were made from the DNA extracted from each clone, after being grown individually, while those constituting the LTC-MTPs were set up directly from the individual BAC cultures. The two different pipelines were used to compensate for possible problems resulting from nonoptimal extraction of DNA or lack of growth of a clone.

FingerPrinted Contigs Minimal Tiling Paths

DNA extraction from the 4201 (5AS-FPC-MTP) and 6,560 (5AL-FPC-MTP) clones was performed with the NucleoSpin 96 Flash MN kit (Macherey-Nagel) according to the manufacturer's instructions. The purified DNA samples were then resuspended in 40 μL of FE buffer (5 mM Tris-HCL, pH 8.5), quantified with a NanoDrop ND-1000 spectrophotometer and diluted to 10 ng μL^{-1}. Pooling resulted in 24 column pools, 16 row pools, and 11 plate pools for the 5AS MTP and 24 column pools, 16 row pools, and 18 plate pools for the 5AL MTP. For better handling, the pools (51 for 5AS and 58 for 5AL) were organized in 96-well plates.

Linear Topological Contig Minimal Tiling Paths

To produce the pools for 5AS-LTC-MTP and 5AL-LTC-MTP, BAC clone colonies (5412 for 5AS and 8709 for 5AL) were individually grown in 2YT media (Life Technologies), supplemented with 12.5 μg mL^{-1} chloramphenicol. The plate, column, and row pools (15, 24, and 16 for 5AS and 23, 24, and 16 for 5AL, respectively) were obtained by mixing 10 μL of each bacterial culture then adding 1 mL of fresh medium and growing for 16 h at 37°C. DNA from each pool was obtained employing

phi29 DNA polymerase rolling circle amplification (GenomiPhi V2 DNA Polymerase Kit, GE Healthcare; Dean et al., 2001). Briefly, 1 μL of each BAC pool was denatured in 9 μL sample buffer at 95°C for 5 min and then cooled on ice for 5 min. The denatured samples were mixed with 1 μL GenomiPhi phi29 DNA polymerase and 9 μL reaction buffer and then incubated at 30°C for 2 h; reaction was terminated at 65°C for 10 min. The yield and quality of the amplified DNA were determined with NanoDrop ND-1000 and Qubit Fluorometer (Invitrogen). The obtained pools, in total 55 and 63 for the 5AS and 5AL, respectively, were arranged in 96-well plates.

Bacterial Artificial Chromosome End Sequences and Marker Development

The BAC clones positioned at the ends of contigs identified in the chromosome 5A FPC-MTPs were recovered and BESs (5′ and 3′) were obtained with the Sanger technology. DNA extraction was performed with the NucleoSpin 96 Flash MN kit (Macherey-Nagel) according to the manufacturer's instructions. Sequencing reactions were set up with BigDye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems) and performed with 99 PCR cycles (denaturation at 96°C for 10 sec; annealing at 50°C for 5 s; extension at 60°C for 3 min and 30 s). Reactions were purified by ethanol-EDTA precipitation and sequenced on an ABI 3730XL DNA capillary sequencer (Applied Biosystems).

The total number of raw BESs was 5473 and 6412 for 5AS and 5AL, respectively. Raw BESs were subsequently quality filtered for presence of nondetermined nucleotides, duplications, extreme biases in base composition, and length below 50 bp. This resulted in 3861 and 5194 final high-quality BESs for 5AS and 5AL, respectively. The BESs were screened computationally to develop a panel of TE-based markers (insertion site-based polymorphism [ISBP], see below) and to design simple-sequence repeat (SSR; di- and trinucleotides) markers with the program Imperfect SSR Finder (http://ssr.nwisrl.ars.usda.gov/). The ISBP and SSR markers deriving from BESs were labeled with the prefixes Bgpg and Bldk, respectively.

Genetic Mapping

Plant Material and DNA Extraction

Four wheat segregating populations were employed to develop genetic maps for the 5A chromosome: (i) 383 F_2 lines from a cross between CS and 'Renan' (CS × Re; *T. aestivum* × *T. aestivum*), kindly provided by Catherine Feuillet (Institut National de la Recherche Agronomique, Clermont-Ferrand, France); (ii) 188 F_7 recombinant inbred lines (RILs) from a cross between CS and CS–*T. dicoccoides* disomic substitution 5A (CS × CS5A; *T. aestivum* × *T. turgidum* ssp. *dicoccoides*); (iii) 132 F_8 RILs from a cross between 'DV92' and 'G3116' (DV × G3; *T. monococcum* × *T. monococcum*; the DNA of the RILs and parents was kindly provided by Miroslav Valárik, Institute of Experimental Botany, Olomouc, Czech Republic); and (iv) 122 F_9 RILs coming from the cross 'Latino' × 'MG5323' (Lt × MG; *T. turgidum* ssp. *durum* × *T. turgidum* ssp. *dicoccum*). The DNA from all genotypes was extracted from young leaves following the CTAB-based protocol (Stein et al., 2001).

Development of New Chromosome 5A-Specific Transposable-Element-Based, Simple-Sequence Repeat and Conserved Orthologous Set Markers

Chromosome-5AS-specific sequences, obtained by 454 GS-FLX Titanium sequencing (Vitulo et al., 2011), were searched to design TE junction-based markers. The software Isbp Finder (Paux et al., 2010) was first used to develop an initial panel of ISBP markers then enriched with the RJPrimers software (You et al., 2010) and additional markers like ISBPs, repeat junction markers (RJMs), and repeat junction–junction markers (RJJMs) were developed. Both pieces of software identify unique junctions through a BLASTn search against databases of annotated repeats (such as TREP; Wicker et al., 2002) and then generate primers for the PCR amplification of the genomic fragment spanning the junction with the software Primer3 (http://bioinfo.ut.ee/primer3–0.4.0/). More than 2000 markers developed with this procedure were analyzed on the seven parents of the segregating populations (CS, Re, CS5A, DV, G3, Lt, and MG) to search for polymorphisms (insertion-deletion polymorphisms [IDPs] or single nucleotide polymorphisms [SNPs]) useful for mapping. The ISBPs, RJMs, and RJJMs were labeled with the prefixes gpg, jfio or jjfio, respectively. The SSR markers were developed from the 454 5AS sequences with the software Imperfect SSR Finder. About 300 of the SSRs were analyzed for polymorphisms between the parental lines and were labeled with the prefix ldk.

Conserved ortholog set (COS) markers for chromosome 5A were, in part, gathered from literature (Quraishi et al. [2009]; http://wheat.pw.usda.gov/SNP/new/index.shtml) and, in part, newly developed in this work using genes from the rice orthologous regions to identify bread wheat expressed sequence tags (ESTs).

Genotyping of Molecular Markers

The ISBPs, RJMs, RJJMs, and COSs were amplified with the following protocol. The PCR reactions contained 15 ng template DNA, 1.5 U of Taq polymerase (GoTaq DNA Polymerase, Promega), 1× Green GoTaq reaction buffer, 1.5 mM $MgCl_2$, 400 mM deoxynucleoside triphosphate (dNTP) mix, and 500 nM of each primer in a total reaction volume of 10 μl with the following touch-down profile: 4 min at 95°C, followed by 10 cycles (30 sec 95°C, 30 sec 65°C minus 0.5°C each cycle, 30 sec 72°C), 25 cycles (30 sec 95°C, 30 sec 60°C, 30 sec 72°C), and a final extension of 7 min at 72°C. Amplification products were then resolved on a 2% agarose gel and sequenced on ABI 3130XL Genetic Analyzer (Applied Biosystems) for SNP and small IDP detection.

Different methods were adopted for SNP genotyping of TE-based and COS markers, including Sanger sequencing, enzymatic restriction (cleaved amplified polymorphic

sequences or derived cleaved amplified polymorphic sequences) as well as melting curve analysis (KAPA HRM FAST PCR Kit, Resnova) and the KASPar SNP Genotyping System (http://www.lgcgroup.com/our-science/genomics-solutions/genotyping/#.VNJQzSzLIoE). The genotyping of SSRs and SSR-ESTs (derived from GrainGenes 2.0 database http://wheat.pw.usda.gov/GG2/index.shtml or developed in the present work) was performed with the M13-tailed primer method (Boutin-Ganache et al., 2001). Reactions of 10 μL contained 30 ng of genomic DNA, 1.2 U of Taq polymerase (GoTaq DNA Polymerase, Promega), 1× of Colorless GoTaq reaction buffer, 1.5 mM $MgCl_2$, 400 mM dNTP mix, 100 nM FAM- or HEX-labeled 19-bp M13 primer (5′-CACGACGTTGTAAAACGAC-3′), 40 nM M13-tailed forward primer, and 200 nM of the reverse primer. The amplification protocols were optimized for each SSR, but the touch-down profiles described above for other markers were the most widely used. Dye-labeled PCR products were visualized on a ABI 3130XL after denaturation in a mix of Hi-Di Formamide and ROX standard (Applied Biosystems). The amplification and size of the fragments were verified with the GeneMapper v. 4.0 program (Applied Biosystems). Specificity of the polymorphic markers for chromosome 5A was verified employing two 5A nullisomic lines (described below) as a negative control and CS as positive control. Only polymorphic and 5A specific markers were screened on the lines of the segregating populations.

Construction of Four Genetic Maps

Four genetic maps were developed for chromosome 5A. Markers and lines showing more than 10 and 5% of missing data, respectively, were discarded; for RIL populations, markers showing a genotypic frequency lower than 30% were discarded, while for F_2 populations, markers showing a frequency of homozygotes lower than 15% were not considered. JoinMap4 (Van Ooijen, 2006) was used to estimate linkage distances based on the linear regression mapping algorithm, with the Kosambi mapping function and logarithm of the odds and recombination frequency thresholds of 5 and 0.40, respectively.

Development of a Neighbor Map for Chromosome 5A

A neighbor mapping approach was undertaken to achieve a densely covered map for an efficient anchoring of the physical map according to the method introduced by Cone et al. (2002). The neighbor maps, one for each chromosome arm, were produced integrating the CS × CS5A framework map (Gadaleta et al., 2014) with maps from the following sources: (i) an SSR-based consensus map developed from four maps ('Synthetic' × 'Opata', 'RL4452' × 'AC Domain', 'Wuhan #1' × 'Maringa', and 'Superb' × 'BW278'; Somers et al., 2004); (ii) six maps developed from biparental populations: 'Nanda2419' × 'Wangshuibai' (Xue et al., 2008), CS × Re, DV × G3 (both developed in this work), and a Lt × MG map made using the Illumina Infinium 90K SNP array (Desiderio et al., 2014), and (iii) the 90K_SNP-based consensus map obtained from eight doubled-haploid populations published by Wang et al. (2014).

Marker integration was performed by comparing the marker order of the above-mentioned maps with that of the CS × CS5A map. The common markers sharing the same order were used as a fixed backbone to which additional loci were added.

If M_1 and M_2 are the positions of the two backbone markers surrounding a noncommon marker on a published map, and M_3 and M_4 are the positions of the same two markers on the neighbor map, the noncommon marker was positioned on the neighbor map by means of the following proportion between distances: (unpositioned marker $- M_1$):($M_2 - M_1$) = (unpositioned marker $- M_3$):($M_4 - M_3$). The neighbor map approach assumes the recombination frequency to be homogeneous along the whole inference interval. The longer the interval, the lesser this assumption is supposed to be true. Therefore, only markers inferred from intervals equal or shorter than 10 cM were assumed as reliable positions and were integrated in the final neighbor map.

Deletion Mapping

The physical position of the 5A-specific markers employed in the genetic map construction and of a subset of ISBPs developed from BESs of the 5AS-FPC-MTP was determined with a set of aneuploid lines derived from the hexaploid wheat cultivar CS, consisting of nullitetrasomic, ditelosomic, and deletion lines (Gadaleta et al., 2012). Each chromosome bin was defined by adjacent breakpoints from two separate addition lines, with the name of the bin being comprised of the name of the proximal deletion line followed by the percentage of the arm present in the distal deletion line (Qi et al., 2003). Two nullitetrasomic lines, N5AT5B and N5AT5D (Sears, 1966), were used to assess specific assignment of markers to chromosome 5A, while the ditelosomic line DT5AL (Sears and Sears, 1978) allowed markers to be positioned on the short or long arm. Physical location of markers to 5AS bins was obtained using a set of four deletion lines (C-5AS1-0.40, 5AS3-0.75, 5AS6-0.97, and 5AS9-0.98) dividing the short arm into five bins (Endo and Gill, 1996), while location on 5AL was determined using nine deletion lines (5AL1-0.11, 5AL12-0.35, 5AL9-0.43, 5AL5-0.46, 5AL4-0.55, 5AL8-0.64, 5AL15-0.67, 5AL17-0.78, and 5AL7-0.87).

Anchoring by Polymerase Chain Reaction and Data Deconvolution

Markers genetically mapped in the four populations described above were amplified in the 51 and 58 5AS- and 5AL-FPC-MTP DNA pools, respectively. Polymerase chain reaction amplifications and scoring were as indicated above. Furthermore, a set of 90K_SNPs (43 for 5AL and 30 for 5AS), distributed along the 5A neighbor map at approximately constant intervals, were transformed into PCR-based markers and used to screen the 55 and 63 5AS- and 5AL-LTC-MTP DNA pools, respectively. A script was created to deconvolute the positive plate,

column and row pools data to identify of the BAC clone containing the marker sequence. When multiple possible combinations coming from multiple positive pools were detected, the presence of overlapping clones among contigs in the fingerprinting database was verified. The PCR was repeated directly on the bacterial cultures of the putative positive clones from the original library, to avoid any assignment errors, in the following cases: (i) only a single clone was matched, (ii) no partial overlapping clones were observed in the same contig (FPC or LTC) among the possible combinations arising from the script, or (iii) doubtful results.

Anchoring by Microarray Technology and Data Processing

A custom Agilent microarray, specific for chromosome 5A, was designed using 4722 sequences derived from several resources: (i) ESTs located on chromosome 5A recovered from http://wheat.pw.usda.gov/wEST/; (ii) sequences derived from the wheat GZ v.2 produced in the frame of the consortium and kindly provided by Klaus Mayer (http://wheat-urgi.versailles.inra.fr/Seq-Repository/Genes-annotations); (iii) restriction fragment length polymorphism (RFLP), SSR, diversity array technology (DArT) markers and exome-based SNPs from the literature; (iv) SSRs and TE-based junction markers developed in this work for 5AS and already genetically mapped in our maps; and (v) the 5AL and 5AS GZ (GZ_1) sequences produced by Vitulo et al. (2011).

The ESTs and sequences of the 5A GZs were BLASTed against the 5A survey sequence assemblies of CS available at the Unité de Recherche Génomique Info (URGI) database (http://wheat-urgi.versailles.inra.fr/Chr5A; IWGSC 2014) to extend the region at both sides and to include the introns. A tiling approach was employed to design different 60-mer probes for each target sequence with the Agilent eArray software (https://earray.chem.agilent.com/earray). A total of 12,676 probes were selected based on melting temperature, specificity, and base composition score. Sixty negative control probes were designed on *Homo sapiens* genes and replicated in random positions on the microarray. The array design also included positive control probes designed on molecular markers already anchored. One-channel processing of the microarrays was used. In total, 15 arrays (8 × 15K format) were hybridized with 55 plus 63 three-dimensional pools of 5A LTC-MTPs. Briefly, 450 ng of DNA polymerase-amplified DNA of each pool was fragmented for 10 min at 95°C and then labeled with cyanine dye 3 with the Genomic DNA ULS labeling kit according to manufacturer instructions (Agilent Technologies). After a cooling step on ice of 3 min, nonreacted ULS-Cy3 was removed through KREApure columns (Kreatech Biotechnology), and the yield and degree of labeling was evaluated by the Nanodrop ND-1000. Hybridizations and washing were performed as recommended (Agilent Oligonucleotide Array-Based cgh for Genomic DNA Analysis v3.4). Scanning was performed with an Axon GenePix 4400A (MDS, Analytical technologies) and data extracted with GenePix Pro7 program.

For each hybridization, a threshold value was calculated as the mean plus 2× standard deviation of signal intensity values of the negative control probes. Signals above the threshold values in more than three plate pools, three column pools, or three row pools were considered as nonspecific and therefore not considered in subsequent analysis. After filtering of nonspecific probes, a sequence was called as present in a pool when all the relative probes showed a signal higher than the cut-off level. To address the clones, and therefore the anchored contigs, the same script described for anchoring by PCR was employed. Only the coordinates corresponding to overlapping clones were considered as robust data for the script, while for the remaining case, a deeper manual check was done to avoid losing any information. Sequences anchoring only a single positive clone (only three coordinates, one for each pool) were discarded to proceed only with the most robust data.

Update of the 5A Genome Zipper

A new version of the 5A GZ produced by Vitulo et al. (2011; GZ_1) was created also including the BESs of 5AS and 5AL-FPC-MTPs and the sequences of GZ v.2 produced by the IWGSC and employed in the chip construction (GZ_2). The search was performed at the DNA level with the latest version of the *Brachypodium* genome assembly (http://www.plantgdb.org; Bdistachyon_192), the rice genome annotation (rice.plantbiology.msu.edu, version 7.0), and BLASTn similarity E-value cutoff of 1×10^{-10}. The mapping against *Brachipodium* genome identified for some genes multiple matches in different part of the genome. To discriminate between paralogous and orthologous relationships we used the best-reciprocal-hit method.

In Silico Anchoring

Multiple analyses of in silico anchoring were performed during this project. A first experiment was performed with BESs of FPC-MTPs as queries for a BLASTn search where the expressed sequences used for the Agilent chip realization were used as subjects. A similarity E-value cutoff lower than 1×10^{-10} and an identity higher than 95% were applied.

No more than six mismatches were allowed. The contigs in which the BES are included were linked for inference to the BLASTed ESTs.

A second in silico anchoring was performed by BLASTing 90K_SNP markers from the neighbor maps against all the expressed reads that had been anchored (regardless their presence in the GZ). For this search, a BLASTn algorithm with a similarity E-value cutoff lower than 1×10^{-10}, an identity higher than 95%, and no more than three mismatches was applied.

In the third approach, 5A available sequences were analyzed by BLASTn against the IWGSC chromosome survey sequences (CSSs) (http://wheat-urgi.versailles.inra.fr/Seq-Repository; IWGSC 2014). BLASTn searches (E-value >

1×10^{-10}) were performed for (i) all the BESs of FPC-MTPs, a minimum similarity of 600 bp, at least 99% sequence identity, and no more than five mismatches were considered; (ii) all the sequences that were anchored, belonged to GZs, the neighbor map or mapped on deletion bins, considering only hits of more than 200 bp, a sequence identity of at least 95%, and no more than three mismatches.

Construction of Short- and Long-Arm Chromosome 5A Physical Scaffolds

The scaffolding for LTC-based contigs was conducted via end-to-end contig merging and contig–end elongation. This procedure was done with less significant clone overlaps (down to 1×10^{-15}), clone overlaps that were unproven by parallel clones (previously excluded as possible false significant), and clones unproven by parallel clones (previously excluded as possible chimerics). In the case of several possible variants of contig–end elongation or end-to-end contig merging we selected a variant proven by (lower) significant clone overlaps (down to 1×10^{-8}). If more than one variant was proven by parallel low significant overlaps, we selected a variant proven by substantially highly (in three orders of magnitude) significant overlaps. If no variants were proven by parallel overlaps, then we selected a variant based on substantially high (in three orders of magnitude) significant overlaps. In questionable situations (e.g., with variants with similar *p*-value of clone overlaps) contig–end elongation or end-to-end contig merging was not performed. The physical scaffolds generated from this process are expected to be less reliable than the original physical contigs, owing to their less-proven clone overlaps. Additional proofs or hints for editing of these scaffolds can be obtained from partial ordering provided by other maps, that is, genetic maps, deletion bin map, radiation hybrid maps, GZ, etc. (see Results section). To produce scaffolds from LTC contigs, genetic and physical maps were linked using all available markers.

Results and Discussion

Construction of Two Chromosome-5A-Specific Bacterial Artificial Chromosome Libraries

Short and long arms of chromosome 5A were purified from CS by flow-cytometric sorting. The analysis of aqueous suspensions of mitotic chromosomes stained by 4′,6-diamidino-2-phenylindole gave histograms of fluorescence intensity (flow karyotypes) on which peaks of 5AS and 5AL were clearly discriminated from other chromosomes. This permitted their purification by flow sorting and a total of 8.6 million and 4.95 million 5AS and 5AL arms were collected, representing 5 and 5.2 μg of DNA, respectively. Analysis of random samples of sorted chromosome fractions by fluorescence in situ hybridization revealed that 5AS and 5AL arms were sorted with an average purity of 90 and 88%, respectively. The contaminating fraction was assessed to be a random mixture of various wheat chromosomes. Flow-sorted chromosome arms were used to construct chromosome arm-specific BAC libraries TaaCsp5AShA (5AS: 46,080 clones, average insert size of 120 kb, *HindIII* cloning site) and TaaCsp5ALhA (5AL: 90,240 clones, average insert size of 123 kb, *HindIII* cloning site). Considering a size of 295 Mb for 5AS and of 532 Mb for 5AL (Šafář et al., 2010) and the level of contamination due to other chromosomes, the libraries represent 16.5 equivalents of 5AS and 18.3 equivalents of 5AL (Table 1).

Assembly of Short- and Long-Arm Chromosome 5A Physical Map with FingerPrinted Contigs and Linear Topological Contig Tools

A total of 44,740 clones (16.3×) for 5AS and 51,072 clones (10.4×) for 5AL were fingerprinted by high-information-content fingerprinting (Luo et al., 2003). The analyzed genome equivalents are similar—for both arms—to what has already been done for 6A (Poursarebani et al., 2014) and 3B (Paux et al., 2008) wheat chromosomes. This work yielded 36,165 (80.8%) and 39,830 (77.9%) high-quality fingerprints for the short and long arms, respectively, representing 13.2 5AS and 8.1 5AL equivalents calculated according to the average BAC clone length (120 and 123 kb) multiplied by the number of fingerprints and the proportion of chromosome arm in the sorted fraction (0.90 and 0.88) divided by the expected chromosome arm size (295 and 532 Mb; Table 1).

A first automated assembly with the program FingerPrinted Contigs (Soderlund et al., 2000) generated 1308 contigs (342 Mb, MTP of 4201 clones) and 11,081 singletons for 5AS and 2556 contigs (601 Mb, MTP of 6560 clones) and 12,066 singletons for 5AL. The N_{50}, defined as the minimum number of contigs needed to cover at least 50% of the assembly, was 354 for 5AS and 823 for 5AL, while the L_{50}, defined as the size of the smallest contig needed to cover 50% of the assembly, was 296 and 251 kb for short and long arm respectively (Table 1). Considering only contigs containing at least five clones (948 and 1791 for 5AS and 5AL, respectively), the map coverage dropped to 291 Mb for 5AS and 483 Mb for 5AL, which is more similar to the cytogenetically estimated chromosome arm sizes. For contigs with five clones and more (presumably more reliable, see Materials and Methods section), the average contig size was 307 kb for 5AS and 270 kb for 5AL, and consequently, L_{50} values were 418 kb for 5AS and 282 kb for 5AL. The longest contig was made of 126 clones (1297 kb) for 5AS and of 76 clones (1027 kb) for 5AL (Table 1).

The quality of a physical map depends on the accuracy of the fingerprinting data assembly, which is often complicated by fingerprints of limited quality, presence of segments from two different portions of the genome in the same BAC (chimeric clones), false clone overlaps due to repeats and duplications, and contamination of the library by other chromosomes. To improve the physical assembly produced by FPC, LTC was employed, since it had been shown to produce more accurate assemblies even with nonoptimal fingerprints quality in wheat (Lucas et al., 2013; Breen et al., 2013; Poursarebani et al., 2014) and in other crops (Frenkel et al., 2010; Ariyadasa et al., 2014).

Linear Topological Contig, which became available after our initial FPC assembly and MTP selection, was used to reassemble 5A high-quality fingerprints. The assembly statistics obtained for both chromosome arms with FPC and LTC are summarized in Table 1. Linear Topological Contig defined a second version of the 5A physical map and a new MTP (LTC-MTP) for both chromosome arms. The number of contigs for each MTP was reduced, while the number of clones grouped into contigs was increased (Table 1); the same trend was observed in the LTC assembly of 1AS, 1AL, and 1BS (Lucas et al., 2013; Breen et al., 2013, Raats et al., 2013). The LTC version of 5AS physical map includes 26,659 clones (9506 singletons) and is composed of 652 contigs (5AS-LTC-MTP, 5412 clones), the longest of which contains 357 clones and covers 3391 kb. The 5AL map incorporates 29,610 clones (10,220 singletons), grouped into 1504 contigs (5AL-LTC-MTP, 8709 clones), the longest of which contains 156 clones and covers 2303 kb (Table 1). FingerPrinted Contigs and LTC MTPs shared 1414 clones and 2573 clones for 5AS and 5AL, respectively.

The current FPC and LTC assemblies are available at the wheat genome browser of the URGI website for the wheat 5AS and 5AL physical map (https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_pub/).

Comparison of the Two Assembly Tools: FingerPrinted Contigs versus Linear Topological Contig

Compared with FPC assemblies, the LTC assemblies showed an almost doubling of the average size of contigs (506 and 451 kb for 5AS and 5AL, respectively), more than doubled the L_{50} (820 and 563 kb for 5AS and 5AL, respectively), and halving of N_{50} (128 and 407 for 5AS and 5AL respectively).

The length of the 5AS LTC map (330 Mb) was slightly lower compared with the map built with FPC, while the 5AL LTC map was 678 Mb, showing a slight increase compared with the map built with FPC (Table 1). The introduction of additional BAC overlaps allowed by LTC (that uses more liberal cutoffs for clone overlap but which filters out some clones before contig assembly) together with the lower coverage of the long arm, could explain the higher number of small and fragmentary contigs, which FPC is not able to join into longer contigs, thus resulting in an increased sized map. However, if only the contigs made by at least five clones are considered, the arm coverage approaches the estimated physical sizes.

As already observed for chromosome 6A, the coverage of the 5A assembly obtained by LTC, excluding the contigs with less than five clones, was higher than 100% for both arms (101 and 115% for 5AS and 5AL, respectively). This could result from contigs with overlaps not being merged due to the repetitive fraction or could be due to the lower number of fingerprinted BACs, especially for 5AL (8.1 equivalents compared with 13.2 equivalents for 5AS), or to an overestimation of the virtual band length. Another explanation is that the arms overlap a bit, that is, contain the same part of centromere, and this depends on the position of the break during their formation. The performance of LTC vs. FPC was outlined in Fig. 1 binning the number of clones included in each contig (Fig. 1a) and relative physical size (Fig. 1b). Linear Topological Contig provided an improved clustering contiguity than FPC, decreasing the number of contigs in all bins with less than 25 and 50 clones for 5AS and 5AL, respectively. The largest LTC contigs include a higher number of clones and exhibit a longer physical size. Improvement of the LTC assembly allowed about 127 Mb of the 5AS assembly (half of the chromosome arm) to be covered by contigs longer than 1 Mb and more than 100 Mb of 5AL (about 20% of the chromosome arm) to be covered by contigs of >1 Mb (Fig. 1b).

A high proportion of FPC contigs were fully incorporated into the contigs obtained with LTC (approximately 80% for 5AS and 85% for 5AL). These data support both the robustness of contigs obtained by high stringency FPC assembly and the reproducibility of the less stringent LTC approach. A small portion of the FPC contigs was broken into blocks with LTC because of the presence of internal clones incorrectly assigned to a contig (probably chimeric clones and responsible for false assemblies, accounting for about 5% of both 5AS and 5AL).

Effectiveness of LTC in merging multiple FPC contigs into one LTC contig is highlighted in a network graphical representation of two contigs of 5AS and 5AL LTC assembly (Fig. 2) where each LTC contig incorporates five FPC contigs. In both examples, clones excluded from FPC assembly and thus marked as singletons were included in the new assembly, as already shown for chromosome 1AL (Breen et al., 2013). This partially explains the higher number of clones included in the LTC map compared with that of FPC (Table 1).

Since overall LTC outperformed FPC (longer contigs, bigger coverage with a higher L_{50}, and a reduced N_{50}) all the anchoring results and the production of the physical map were presented in context of the LTC assembly.

Bacterial Artificial Chromosome End Sequencing to enrich the Physical Map

Bacterial artificial chromosome end sequencing undertaken before the LTC assembly was performed on 2487 and 2835 clones recovered from 5AS-FPC-MTP and 5AL-FPC-MTP, respectively. For 5AS, 3861 high-quality sequences (about 78%) were obtained representing a cumulative length of 2.8 Mb, with an average size of 725 bp and a guanine–cytosine content of 44.3%. For 5AL, 5194 high-quality sequences (about 92%) were produced corresponding to a cumulative length of 3.4 Mb, with an average size was 654 bp and the guanine–cytosine content of 43.8%. The bias toward AT content, associated with repetitive sequences, is in agreement with that reported for other wheat chromosomes and chromosome arms, for example, 3AS, 4A, and 1AL (Sehgal et al., 2012; Hernandez et al., 2012; Lucas et al., 2012).

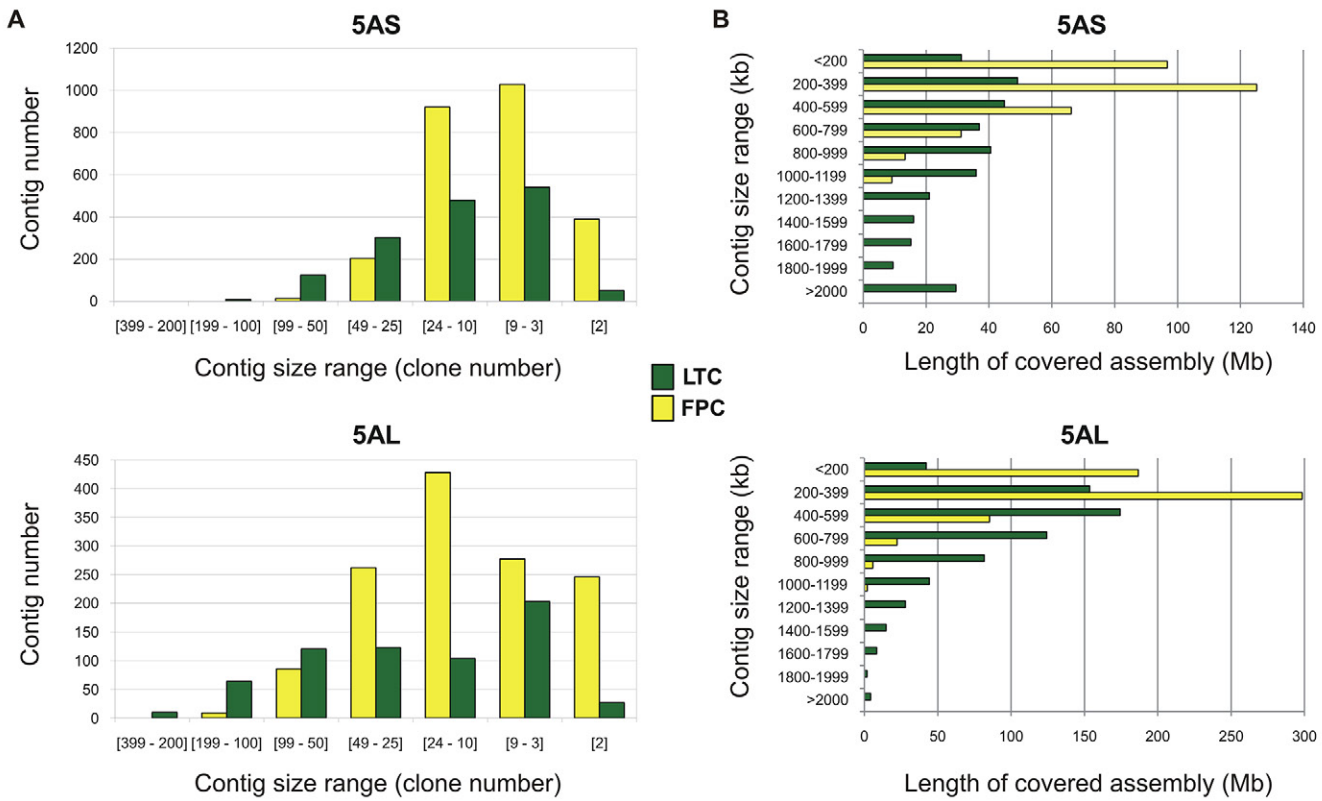

Figure 1. FingerPrinted Contigs (FPC) vs. Linear Topological Contig (LTC): Comparison of assembled contig lengths. (A) The number of contig per contig size range. (B) Contigs were grouped by estimated length (based on an average consensus band length of 1.2 kilobase [kb]) into ranges of 200 kb and the total physical portion of each range was calculated and plotted. FPC contigs are marked in yellow and LTC contigs are marked in green.

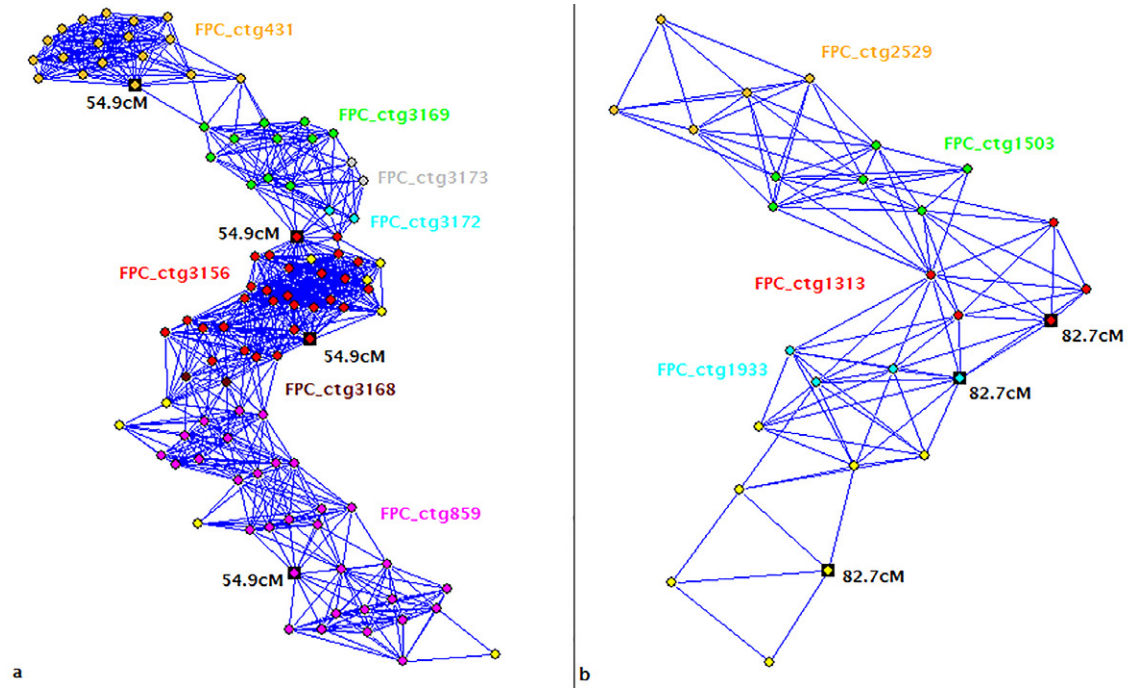

Figure 2. Graphical representation of Linear Topological Contig (LTC) and corresponding FingerPrinted Contigs (FPC). Bacterial artificial chromosomes (BACs) are represented as vertices and overlaps between them as connecting lines. The BACs from the same FPC contig share the same color. Singletons are colored in yellow. The BACs genetically anchored are labeled with corresponding genetic distance. (A) The LTC contig 467 of 5AS includes 103 BACs and seven FPCs. Four markers, genetically anchored in the same region, cover the cluster at different locations. (B) The LTC contig 871 of 5AL includes 27 BACs and four FPCs . Seven singletons of FPC are included in the LTC. One of the three markers covering the LTC 871 genetically anchors an FPC singleton.

Bacterial artificial chromosome end sequences proved to be an important source of new markers in bread wheat, that is, SSRs and inserted TE-junction-based markers, as experienced previously (Šafář et al., 2010; Lucas et al., 2012; Sehgal et al., 2012; Periyannan et al., 2014). Once these markers are genetically mapped or located within a virtual gene order, they greatly facilitate anchoring of contigs to the physical map. Bacterial artificial chromosome end sequences were processed with the RJPrimers program (You et al., 2010) to identify TE junctions; overall, almost 4000 ISBP markers were identified: 1814 for 5AS and 2117 for 5AL. In 5AS, 896 BESs carried at least one ISBP (corresponding to 791 BAC clones or 31.8% of the sequenced BAC), with a maximum of seven possible ISBPs in the same sequence, while for 5AL, 1068 BESs had at least one ISBP (961 BACs or 33.9% of the sequenced BAC), with a maximum of seven ISBPs per sequence, as well. Moreover, 124 SSRs were identified with the Imperfect SSR Finder program, 72 for 5AS and 52 for 5AL, including di-, tri-, and tetranucleotide repetitions (5AS: 34 di-, 37 tri-, 1 tetra-; 5AL: 32 di-, 20 tri-). The primer sequences of newly developed markers, ISBPs (named Bgpg) and SSRs (named Bldk), were reported in Supplemental Table S1.

A total of 2086 BESs (around 25% of the sequences produced) had at least one potentially useful molecular marker. The large number of ISBP markers generated from BESs (3931) represents a very promising tool to further refine the physical map herein presented, by allowing the gaps to be filled by additional (as yet un-anchored) contigs.

Insertion site-based polymorphism markers have already been developed from BESs to anchor gene poor chromosome regions (Lucas et al., 2012); apart from this purpose, the ISBPs derived from minimum tiling paths are supposed to be evenly located along the chromosome arms, so they may be also of great value to assist the map-based cloning of agronomically important genes. Also the SSRs developed from BES have been very useful in fine-mapping projects in common wheat (Diéguez et al., 2014). The BESs were employed in the present work (i) to develop a new version of 5A GZ, anchoring contigs (in which the BESs are included) to the *Brachypodium* reference zipper and (ii) for anchoring contigs (developed with FPC and LTC assemblies) to the Illumina sequence contigs obtained from the IWGSC CSS, as detailed below.

Development of Four Genetic Maps of Chromosome 5A

New genetic maps were developed for the 5A chromosome to anchor the identified BAC contigs. To increase the likelihood of identifying polymorphisms for linkage mapping, four segregating populations involving different wheat species (*T. aestivum*, *T. turgidum*, *T. monococcum*) and progenitors (*T. turgidum* ssp. *dicoccoides*, *T. turgidum* ssp. *dicoccum*), were employed to compensate for the low level of polymorphism in the cultivated gene pool. Beside CS × RE, which is the reference population of the IWGS consortium, another three RIL populations were used: CS × CS5A, Lt × MG, and DV × G3, described in Materials and Methods section.

Since it is well known that different types of molecular markers preferably map on different areas of the genome (Paux et al., 2008; Kumar et al., 2012), several approaches for marker development were implemented. One hundred forty-six SSRs and SSR-ESTs were sourced from the public GrainGenes database (http://wheat.pw.usda.gov/GG2/index.shtml); 81 COS markers were obtained from the literature (Quraishi et al., 2009) and from an on-line database (http://wheat.pw.usda.gov/SNP/new/index.shtml) or were developed by the procedure of Quraishi et al. (2009). Furthermore, new molecular markers were developed from the 5AS sequences of CS produced by 454 sequencing (Vitulo et al., 2011). Markers developed on TE junctions represent the most frequent marker source in genomes composed mostly by repetitive sequences (Paux et al., 2010; You et al., 2010), with a density of one TE junction every 3.8 kb, a genome-specificity rate of 70%, and an estimated total number of about 3 million ISBPs in the hexaploid wheat genome.

More than 25,000 high-confidence, TE-based markers were automatically generated with IsbpFinder on 454 reads (about 23,000 ISBPs) and with RJPrimers on 454 contigs (about 2000 ISBPs and 400 RJMs-RJJMs). A total of 1300 plus 834 nonredundant ISBPs obtained from 454 reads and contigs, respectively, and 48 nonredundant RJMs-RJJMs (Supplemental Table S2) were randomly selected for screening on the parents of the segregating populations. Furthermore, about 8000 new SSRs were produced with Imperfect SSR Finder starting from 454 contigs, and 290 microsatellites were randomly chosen (Supplemental Table S2) for polymorphisms search.

A total of 2699 primer pairs (SSRs, SSR-ESTs, and IDPs or SNPs for TE-based and COS markers) were first tested for 5A specificity through their amplification on genomic DNA of CS aneuploid lines N5AT5B and N5AT5D (nullisomic for 5A and tetrasomic for 5B or 5D). Only markers amplifying in CS and absent in the nullitetrasomic lines were considered specific for 5A and were used in a polymorphism search in the seven parents of the four segregating populations (CS, Re, CS5A, DV, G3, Lt, and MG). Overall, 299 (11%) of the tested markers were both specific for chromosome 5A and polymorphic in at least one population (126 SSR or SSR-ESTs, 23 COSs, and 150 TE-based). A total of 153 unique loci were successfully mapped. A further localization along the 5AS and 5AL chromosome arms was also determined for 240 markers by a deletion-mapping approach performed on four and nine deletion lines for 5AS and 5AL, respectively, representing 14 deletion bins (Supplemental Table S3).

The limited level of specificity of PCR primers for chromosome 5A, already observed for other wheat chromosomes (Sehgal et al., 2012), can be ascribed to possible duplications between wheat chromosomes (Zhang, 2003; Choulet et al., 2014) or to amplification from homeologous chromosomes due to their possible presence in the flow-sorted fraction (the 5A chromosome arm-based libraries showed a contamination of about 10%).

The mapping population that proved to be the most informative, in terms of polymorphisms, was CS × CS5A, while CS × Re was the least informative (Supplemental Fig. S1). Considering the different marker types, SSR and TE-based markers showed similar effectiveness in detecting polymorphisms in the CS × CS5A population, while for CS × Re and DV × G3, SSRs were more polymorphic than TE-based markers. The population Lt × MG was not included in this comparison because it was introduced later in the study and many molecular markers, more than 35%, were not tested in this population (see ND in Supplemental Table S3). The population CS × CS5A represents a very interesting genetic resource in the frame of this project since it segregates only for the chromosome 5A and shows an outstanding level of polymorphisms, due to the donor of 5A in CS5A being an accession of *T. turgidum* ssp. *dicoccoides*, a wild progenitor retaining a high level of genetic diversity (Haudry et al., 2007) and consequently significantly increased the likelihood of identifying useful polymorphisms.

Linkage analysis generated four genetic maps, each divided in two or more linkage groups (LGs): (i) CS × Re: two LGs, covering a total of 159.3 cM with an average of one marker every 2.8 cM; (ii) CS × CS5A: two LGs, 138.54 cM long, one marker every 2 cM on average; (iii) DV × G3: five LGs, 187 cM in total, one marker every 2.6 cM on average; and (iv) Lt × MG: three LGs, 116.7 cM, one marker every 1.9 cM on average (Supplemental Table S3). The maps from CS × CS5A and Lt × MG had the highest marker density, while the DV × G3 map was fragmented into five LGs and had the lowest marker density.

Construction of a High-Density Neighbor Map for Chromosome 5A

To increase the integration of genetic and physical data, a neighbor map for each chromosome arm was developed following procedures already adopted for wheat and other cereals (Mayer et al., 2011; Paux et al., 2008; Philippe et al., 2012). A high-density genetic map of CS × CS5A_90K recently released (Gadaleta et al., 2014) based on EST-derived 90K_SNPs (Infinium iSelect 90K BeadChip array) and SSRs was employed as a framework on which markers from additional genetic maps were individually integrated. In the CS × CS5A map population, 5A loci segregate in a CS background, therefore this population is likely to be the best for anchoring and ordering MTP contigs made of CS BAC clones. In addition to CS × CS5A, nine different maps contributed markers for the construction of the neighbor map. The Lt × MG 90K map (Desiderio et al., 2014) and the consensus map from Wang et al. (2014) provided the most important contribution in terms of number of markers. Indeed, approximately 55 and 66% of markers located on these two maps, respectively, were successfully positioned in the neighbor map framework (Supplemental Table S4). Moreover the wealth of markers in both maps (both based on 90K_SNPs) allowed a quite reliable inference based on relatively short intervals, usually <10 cM. This latter interval corresponds to the maximum distance within which, even choosing different backbone markers, the order of inferred markers does not change. The CS × CS5A 90K map (Gadaleta et al., 2014) was integrated with the CS × CS5A map developed in this work with SSRs, ESTs, COSs, and TE-based markers to define a robust scaffold of common markers. Because of the lack of a sufficiently conserved subset of common markers, only one marker was integrated from the DV × G3 map into the neighbor map. This result can be explained considering that both parents are *T. monococcum* lines, therefore, phylogenetically less related to *T. aestivum* with respect to the other genotypes or species used. Nevertheless, the *T. monococcum* map contains a number of TE-based and noncommon markers, which were anchored on the physical map and could be useful for further studies.

Table 2. Statistics of the entire 5A neighbor map.

	Marker no. per LG† including cosegregants	Marker no. per LG excluding cosegregants	Size	Density‡	Longest distance§
			cM		cM
LG1	1168	184	83.5	0.45	13.1
LG2	1169	303	117.8	0.38	10.3
LG3	244	101	47.4	0.47	6.0
Total map	2581	588	248.7	0.42	10.1

† LG, linkage group.

‡ Ratio between size of LGs and number of unique markers.

§ Between two consecutive markers.

Overall, the 5A neighbor map consisted of 2581 markers (including 2451 EST-based SNPs, 87 SSRs, 16 sequence tag sites (STSs), and 27 TE-based markers; Table 2; Supplemental Table S4) with an average point density of 0.42 cM, which is comparable to the density achieved for the 3B neighbor map (Paux et al., 2008).

Among the three LGs obtained, LG2 showed the highest marker density (expressed as ratio between the total length in cM and the number of markers) calculated removing all cosegregating loci. Even considering that a neighbor map provides a less accurate position of a marker (Cone et al., 2002), the availability of a high number of markers and even of a high number of cosegregating markers is however advantageous as it increases the probability of anchoring physical contigs to chromosomal regions where recombination is suppressed, for example in centromeric and pericentromeric regions (Philippe et al., 2012). Bin locations were available for 165 and 331 markers of the 5AS and 5AL neighbor maps, respectively (Supplemental Table S4).

An Upgraded Version of the 5A Genome Zipper

A 5A GZ, GZ_1, was previously created by low-pass survey sequencing coupled with the identification of genomic regions of conserved homologies across *Brachypodium*, rice and sorghum (Vitulo et al., 2011). This GZ was upgraded (GZ_2) integrating the sequences derived from the wheat GZ v.2, realized by the IWGSC consortium (http://wheat-urgi.versailles.inra.fr/Seq-Repository/

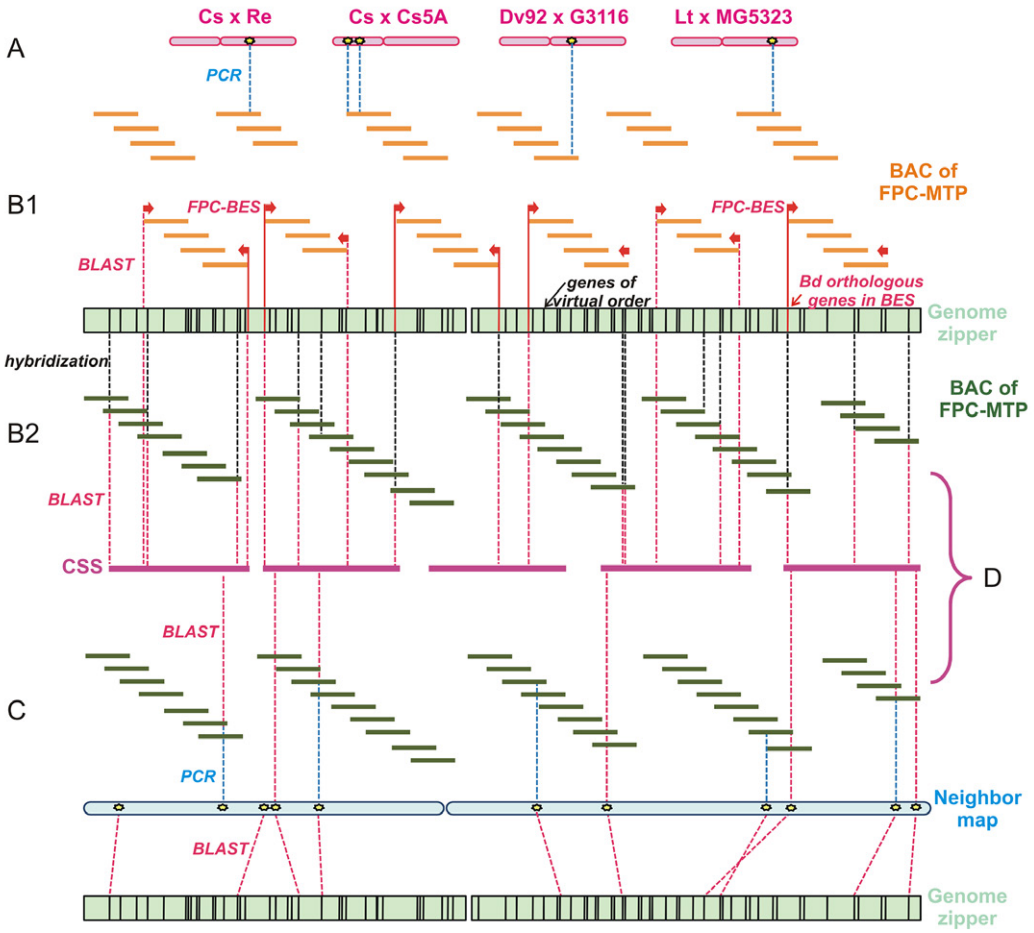

Figure 3. Graphical representation of different anchoring approaches employed for the realization of integrated 5A physical map. (A) Molecular markers genetically mapped to four genetic maps were anchored to FingerPrinted Contigs (FPC)–minimal tiling path (MTP) physical contigs by polymerase chain reaction (PCR) screening. (B1) The bacterial artificial chromosome end sequences obtained from FPC-MTP and containing *Brachypodium* orthologs genes were integrated into the version of 5A genome zipper released by Vitulo et al. (2011; GZ_1). (B2) The Linear Topological Contig (LTC)–MTP physical contigs were anchored to the new version of genome zipper (GZ_2) through hybridization with the Agilent microarray. (C) The neighbor map was anchored to LTC-MTP physical contigs by PCR screening and linked to genome zippers through BLAST search. (D) The anchored markers and the anchored genes, present onto the genome zippers or mapped along the neighbor map were integrated by BLAST search with the Illumina International Wheat Genome Sequence Consortium (IWGSC) contigs.

Genes-annotations) and the available BESs. At first, the sequences of GZ_1 released by Vitulo et al. (2011) were mapped against the latest version of the *Brachypodium* genome (Bdistachyon_1.2; available at http://www.plantgdb.org). The additional sequences available from wheat GZ v.2 were subsequently searched against the *Brachypodium* transcriptome by the same procedure and parameters employed to update GZ_1. Finally a total of 181 BESs (67 for 5AS and 114 for 5AL) were identified as showing homologies with *Brachypodium* genes, most of them already belonging to GZ_1 (Supplemental Table S5). The final GZ, GZ_2, was sorted according to the *Brachypodium* gene order and considering only the two orthologous chromosomes 1 and 4. The new virtual order of syntenic sequences integrated 462 and 1461 sequences for the 5A short and long arms respectively (Supplemental Table S5). With respect to the previous version, the GZ_2 of 5AS was implemented with 23 BESs and 98 ESTs of GZ v.2, while 60 BESs plus 72 ESTs of GZ v.2 were added to the 5AL GZ_2.

Anchoring of the FingerPrinted Contigs Minimal Tiling Path through Polymerase Chain Reaction

The four genetic maps developed specifically for chromosome 5A during this work were employed for a first anchoring of the physical map derived from FPC assembly (Fig. 3a). All 5A-specific and polymorphic markers were tested by PCR screening three-dimensional pools of BAC clones from the 5AS- and 5AL-FPC-MTP. With this procedure, 178 markers (87 TE-based markers, 70 SSRs or SSR-ESTs, 10 ESTs, 9 COSs, and 2 STSs) were assigned to individual BACs and anchored to 168 5A FPC physical contigs (Supplemental Table S6). Specifically, 122 5AS unique contigs were anchored, corresponding to 47 Mb, while 17.7 Mb of 5AL were anchored covered by 46 unique contigs (evaluations based on the estimated length of the anchored contigs) (Table 3; Supplemental Table S6; Fig. 3a). The greater coverage of 5AS compared with 5AL can be attributed to the greater number of molecular markers mapped on the short arm (130 vs. 48 on the long arm).

Table 3. Anchoring results of FingerPrinted Contigs (FPC)–minimal tiling path (MTP) and the corresponding estimation calculated in Linear Topological Contig (LTC)–MTP.

FPC Assembly†	5AS	5AL
Anchored markers	130	48
EST and COS	9	10
SSR and SSR-EST	33	37
STS	1	1
TE-based	87	–
FPC unique contigs	122	46
Total length anchored (Mb)	47	17.7
Corresponding total length anchored (Mb) by LTC	105.4	29.1
Total Mb ordered on genetic maps	70.0	15.5

† EST, expressed sequence tag; COS, conserved orthologous set; SSR, simple sequence repeat; STS, sequence tag site; TE, transposable element.

When the new assembly produced by LTC became available, the anchoring data obtained with the FPC assembly were integrated into the LTC contigs through an iterative process conducted for each chromosome arm. By this procedure, it was possible to create a bridge between the FPC and LTC contigs, without losing any information. This integration allowed an increase of the anchored portion so that it then covered 105.4 Mb for 5AS (70.0 Mb ordered in genetic maps) and 29.1 Mb for 5AL (15.5 Mb ordered in genetic maps) (Table 3).

For only nine markers detecting FPC contigs, no link was identified to LTC contigs; these markers were therefore directly tested by PCR on three-dimensional LTC-MTP pools and assigning all but one marker (wmc727) to LTC contigs (Supplemental Table S6). The FPC contig linked by wmc727 (ctg3461) was detected by LTC as a questionable overlap and hence split into two pseudo LTC contigs, ctg_7409 and ctg_7304, indeed singletons.

High-Throughput Anchoring of the Linear Topological Contig Minimal Tiling Path through a 5A-Specific Array

A 15K chromosome-5A-specific Agilent microarray was built employing sequences from 5A ESTs as well as molecular markers, including some already screened on the FPC-MTP and LTC-MTP via PCR as internal controls to ensure a proper allocation on MTP clones through hybridization. In detail, the chip consisted of 4722 sequences obtained from several resources: (i) 574 ESTs located on chromosome 5A recovered from http://wheat.pw.usda.gov/wEST/; (ii) 1414 sequences of genes from the 5AL GZ_1 (Vitulo et al., 2011) and 962 5AL genes obtained by BLAST search against the URGI survey sequences; (iii) 373 sequences of genes from the 5AS GZ_1 (Vitulo et al., 2011) and 288 5AS genes obtained from BLAST search against the URGI survey sequences; (iv) 918 genes from the 5A GZ v.2 produced by the IWGS consortium (http://wheat-urgi.versailles.inra.fr/Seq-Repository/Genes-annotations); (v) 23 RFLPs and 64 SSRs selected from GrainGenes 2.0; (vi) 23 exome-based SNPs (Allen et al., 2012); (vii) six SSRs and 75 TE-based junction markers developed for 5AS in this work; and (viii) two DArT markers.

The 5A-specific Agilent array was then hybridized with the 118 three-dimensional pools representing the two LTC-MTPs (Fig. 3b.2), comprising 55 pools from the short arm and 63 from the long arm, and the hybridization signals were quantified and normalized following the procedure described in materials and method. Data deconvolution, performed through an ad hoc script written in R, allowed the assignment of 138 different sequences (corresponding to 164 probes) to 89 individual 5AS contigs corresponding to 76.2 Mb of the 5AS-LTC-MTP and 443 different sequences (corresponding to 570 probes) to 283 individual 5AL contigs for a total of 169.1 Mb (Supplemental Table S7). Five unique sequences showed hits with contigs on both the 5AS and 5AL map (Supplemental Table S7), which may be due to sequence duplications between the two chromosome arms.

As previously experienced (Liu et al., 2011; Rustenholz et al., 2010), screening based on the microarray gave a significant boost to the anchoring; accurate gene-to-BAC addresses were obtained in little time and at relatively small cost.

Since very stringent parameters were adopted for declaring positive hybridization results, we are confident that the array hybridization approach produced very robust and reliable anchoring data. The robustness of the obtained data was also confirmed by six positive controls included in the array for which a contig was already identified by PCR to FPC-MTP and LTC-MTP pools. These controls revealed a perfect correspondence between the contigs identified by PCR and those identified by array-based screening. The reliability of the array was also shown by the fact that probes of the same sequence anchored the same contigs, for example, a sequence from GZ_1 and the same sequence BLASTed against the URGI assembly. Frequently, multiple expressed sequences anchored to a clone or group of partially overlapped clones, so limiting the number of unique anchored contigs. Nevertheless, this finding is in agreement with the knowledge that genes are frequently clustered in the wheat genome (Rustenholz et al., 2010; Gottlieb et al., 2013).

In only one case did probes designed to two separated regions of the same sequence anchor two distinct contigs (ctg_11831 and ctg_832). This could reflect failure of contigs to merge in the LTC assembly due to the high stringency applied or low fingerprint quality.

Hybridization data are accessible at NCBI GEO database, accession GSE66996 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE66996).

Anchoring of the 5A Genome Zipper and Neighbor Map to the Physical Map

Various strategies were explored to anchor the physical contigs using the new GZ_2 sequences. Most of the GZ_2 sequences were linked to the physical map through hybridization to the Agilent array (few ESTs were anchored to FPC or LTC assemblies by PCR analysis), while BESs were anchored to FPC or LTC maps by using the anchor data of the contigs from which the BESs were

obtained. Bacterial artificial chromosome end sequences were also used in a BLAST search against all the expressed sequences loaded on the Agilent array (seven of them were also included in the GZ_2), thus allowing an in silico anchoring of 11 BESs (Supplemental Table S8; Fig. 3b.1).

Overall, 277 contigs were identified as containing sequences with homologies to genes present in the GZ_2. A total of 57 contigs were mapped along the 5AS GZ, corresponding to 52.5 Mb anchored, while 219 contigs were anchored to the 5AL GZ, accounting for a length of 132.3 Mb.

With regard to the neighbor maps, the already available anchoring information was retrieved for markers positioned on them (Supplemental Table S8). Additional contigs were anchored to the neighbor maps through PCR screening of LTC-MTP three-dimensional pools with primers for EST-based SNP markers (10 for 5AS and 14 for 5AL), distributed along the 5AS and 5AL neighbor maps at approximately constant intervals (Fig. 3c). Furthermore, since we aimed at including in the neighbor map the anchoring data obtained with the new GZ version (GZ_2), a second level of in silico anchoring was performed. A BLASTn search of the EST-based SNP markers, positioned on the neighbor map, was performed against the ESTs of the GZ_2 already anchored to the LTC assembly (Fig. 3c) for 5AS and 5AL separately (Supplemental Table S8, last sheet). With this approach, the 5A neighbor maps were enriched of 37 (5AS) and 46 (5AL) markers anchored to LTC contigs.

The 5AS neighbor map (corresponding to LG1) anchored 74 different contigs (two of them linked by the same marker) for a total of 71 Mb. On average, a density of a contig every 0.9 cM was obtained, and, in the pericentromeric region, 13 different contigs were linked to cosegregant markers.

For the 5AL neighbor map (LG2 and LG3), 69 different anchored contigs were identified (two of them linked by the same marker) for a total of 41.8 Mb anchored. The contig density in 5AL was a bit lower, with a contig every 1.9 cM (considering LG2 and LG3 as a single linkage group and summarizing all the contigs). Although the number of contigs positioned was almost the same for the 5AS and 5AL neighbor maps, the estimated anchored length was higher for the short arm because of the different average contig sizes (larger for short arm).

Anchoring to the Deletion Bin Map

Four deletion lines for 5AS and nine deletion lines for 5AL (lacking different segments of 5A chromosome) were employed to assign contigs to deletion bins. Two hundred forty genomic markers used in genetic mapping and 14 TE-based markers developed on BESs of the 5AL-FPC-MTP were successfully assigned to deletion bins. Further, bin localizations were obtained for 90K_SNPs mapped onto the neighbor map (Gadaleta et al., 2014), for ESTs (Lazo et al., 2004; Qi et al., 2004), and for gene sequences from GZ_1 (Vitulo et al., 2011).

Since some deletion bins partially overlap (e.g., 5AL12, 0.35–0.57 and 5AL5, 0.46–0.55), for simplicity, we defined three bin intervals on 5AS (5AS_0–0.40, 5AS_0.40–0.75, and 5AS_0.75–1.00) and on 5AL (5AL_0–0.57, 5AL_0.57–0.78, and 5AS_0.78–1.00), representing respectively the centromeric, central and telomeric bins of each arm (Supplemental Table S9). Combining this physical information with the contig data, we estimated the size of the anchored portion along these six regions of the chromosome (Supplemental Table S9; Supplemental Fig. S2a).

The whole coverage, based on the estimated chromosome arm length, was 109.2 Mb for 5AS (37%) and of 85.9 Mb for 5AL (16.1%). On 5AS, we anchored 35.1 Mb in the telomeric region, 37 Mb in the central part, and 46.2 Mb in the centromeric region, while on 5AL we anchored 19.7 Mb in the telomeric region, 17.6 Mb in the central interval, and 50.5 Mb in the centromeric region. These data were normalized by the percentage of the chromosome arm covered by each bin (e.g., 5AS_0–0.40 represents approximately 40% of 5AS). A homogeneous density of the physical coverage within each chromosome arm was evident: 0.9, 0.8, and 0.9 Mb per percentile, respectively, for centromeric, central, and telomeric bins of 5AL (Supplemental Fig. S2b) and 1.2, 1.1, and 1.4 Mb per percentile for the corresponding bin intervals of 5AS (Supplemental Fig. S2b).

Anchoring of the 5A Physical Contigs to the Illumina Chromosome Survey Sequences

The 5A physical map was also integrated with the Illumina sequence contigs produced by the CSS of the IWGSC (IWGSC, 2014). BLASTn searches against the CSSs were conducted for a total of about 9000 BESs obtained from BACs in the FPC-MTPs (Fig. 3d) with different level of stringency depending on the repetitive or expressed nature of the query (as detailed in Materials and Methods section). For a total of 218 BESs (160 for 5AS and 58 for 5AL), the corresponding sequence in the Illumina contigs was identified, allowing anchoring to the corresponding LTC contigs. The procedure allowed anchoring of 218 sequence contigs to 188 different LTC contigs: 131 for the 5AS and 57 for the 5AL (Supplemental Table S10). Of these, 77 and 37 contigs belonging to the 5AS and 5AL LTC assembly, respectively, were not previously anchored to any sequence or marker. This in silico approach increased the percentage of the anchored portion of chromosome 5A, linking an additional 49.8 Mb for 5AS and 25.7 Mb for 5AL. Moreover, all gene sequences and markers for which the sequences were available, belonging to the genome zippers, neighbor map, or mapped into the deletion bins and already anchored, were also BLASTed against the CSSs (Fig. 3d; Supplemental Table S10). These alignment data, together with the previous ones obtained with BESs, were joined to create a physical map in silico anchored to the chromosome survey sequences. Also, the genetic position obtained by the IWGS consortium with the population sequencing (POPSEQ) approach (Mayer et al., 2014) was reported,

when available, to add a further level of contig ordering along the chromosome (Supplemental Table S10). Overall, 260 Illumina 5A sequence contigs (only two of them belonging to 5AL) (corresponding to 2.7 Mb), 149 of which contained genes, were linked to 183 different 5AS contigs (corresponding to a physical length of 151.2 Mb), while 337 5A CSS contigs (only one deriving from 5AS) (corresponding to 2.7 Mb), 329 of which containing genes, were anchored to 253 different 5AL contigs (corresponding to a physical length of 159 Mb). With this approach, it was possible not only to increase the percentage of anchoring but also to add a further genomic tool to order the physical contigs thanks to the position of the linked CSSs on the POPSEQ map (see later on). This method has been already successfully employed in barley (*Hordeum vulgare* L.) for genetically anchoring and ordering de novo sequence assemblies (Mascher et al., 2013).

Scaffolding of Linear Topological Contig-Based Physical Contigs and Control of Scaffold Quality by Anchoring

A scaffolding procedure was performed for LTC-based contigs (see Materials and Methods section). Resulting physical scaffolds are less well supported than contigs, but it substantially simplified anchoring and orientation of highly reliable contigs. The ability of scaffolding depends on the coverage, information content of fingerprints, the size of BAC library, the proportion of repetitive sequence, and level of BAC-by-BAC DNA contamination. In the resulting physical map (after the scaffolding), we referred to all contigs as scaffolds (even ones obtained from single contig without any elongation or merging). Scaffolds (contigs) with less than five clones are considered as rather questionable (see Materials and Methods section). The results of scaffolding are as follows (for scaffolds with five clones and more): for 5AS, 324 scaffolds in total containing 26,311 clones and covering 293 Mb, where N_{50} = 65 scaffolds and L_{50} = 1.4 Mb and for 5AL, 742 scaffolds in total containing 29,501 clones and covering 580 Mb, where N_{50} = 187 scaffolds and L_{50} = 0.98 Mb.

The available orders relative to markers (based on expressed and not sequences), including genetic (genetic maps), inferred (neighbor map), virtual (GZ), or physical (deletion bins and physical contigs), contributed to the final phase of the physical map assembly, that is, testing the quality of the physical contigs and scaffolds. In an ideal case, (partial) ordering of markers along the chromosomes should be collinear to (partial) ordering of markers within physical contigs or scaffolds. Inconsistence of orders can point to errors in the map or in the physical contig or scaffold. To highlight inconsistencies in the physical map, we needed at least two independent maps where parts of a contig were consistently anchored to distant regions with these anchors comprising several markers from different clones of the same contig. We found several contradictions between the physical map and virtual marker order provided by GZ, but in these cases the physical map was consistent with other maps (e.g., genetic). These contradictions can be explained by inversions, and translocation occurred in evolution of wheat and *Brachipodium* evolutionary lineage since their split from a common ancestor 32 to 39 million yr ago (IBI, 2010). Few contradictions to genetic maps were caused by single markers; therefore, this type of evidence was not enough to split the contigs. These contradictions may have been the result of a duplication of the marker in the chromosome, an error in the genetic map, an error in anchoring of markers to clones, or chimeric clones (not all of them can be detected based on fingerprinting information only). For more details see Supplemental Table S11 and S12.

Table 4. Anchoring results based on Linear Topological Contig (LTC) and physical scaffolding data per each genomic tool applied. Only contigs with a clone number ≥3 and scaffolds with a clone number ≥5 were included.†

	LTC assembly			Physical scaffolding	
	Marker no.	LTC contigs no.	Anchored size	Scaffold no.	Anchored size
			mb		mb
5AS					
Total anchoring	482	241	197.27	180	220.44
Ordered	379	181	159.31	133	188.05
GZ_3†	313	135	120.7	91	134.98
IWGSC GZ v.2, GZ v.5	90	61	51.48	57	82.92
GZ_2	73	57	52.49	52	82.63
Neighbor map	91	72	70.73	63	96.7
Deletion bin map	161	115	108.79	94	142.73
Genetic maps	72	67	70.01	58	86.69
CSSs‡	273	180	150.56	147	191.88
5AL					
Total anchoring	646	384	233.18	264	280.83
Ordered	606	353	212.05	238	252.47
GZ_3	543	301	183.93	237	251.37
IWGSC GZ v.2,GZ v.5	102	85	47.09	57	65.48
GZ_2	308	213	131.53	182	201.29
Neighbor map	88	69	41.8	64	74.38
Deletion bin map	161	137	88.56	104	122.53
Genetic maps	20	21	15.53	19	22.44
CSSs	358	250	158.64	196	218.7

† GZ, genome zipper.

‡ CSS, chromosome survey sequence.

The scaffolding procedure led to an increase in the anchored portion of the final physical map and comprised 180 physical scaffolds for 5AS, representing about 221 Mb (75%) of the chromosome arm, and 264 physical scaffolds for 5AL, representing about 281 Mb (53%) of the chromosome arm (the estimation is based on the predicted chromosome arm size of 295 Mb for 5AS and 532 Mb for 5AL; Table 4; Fig. 4). This evaluation was done after removing the scaffolds with less than five clones. Among all the genetic and genomic tools developed and employed during this work for anchoring, the major contribution in sorting the scaffolds of the short arm derived from the deletion bin map, while the GZ was the most useful tool for the long

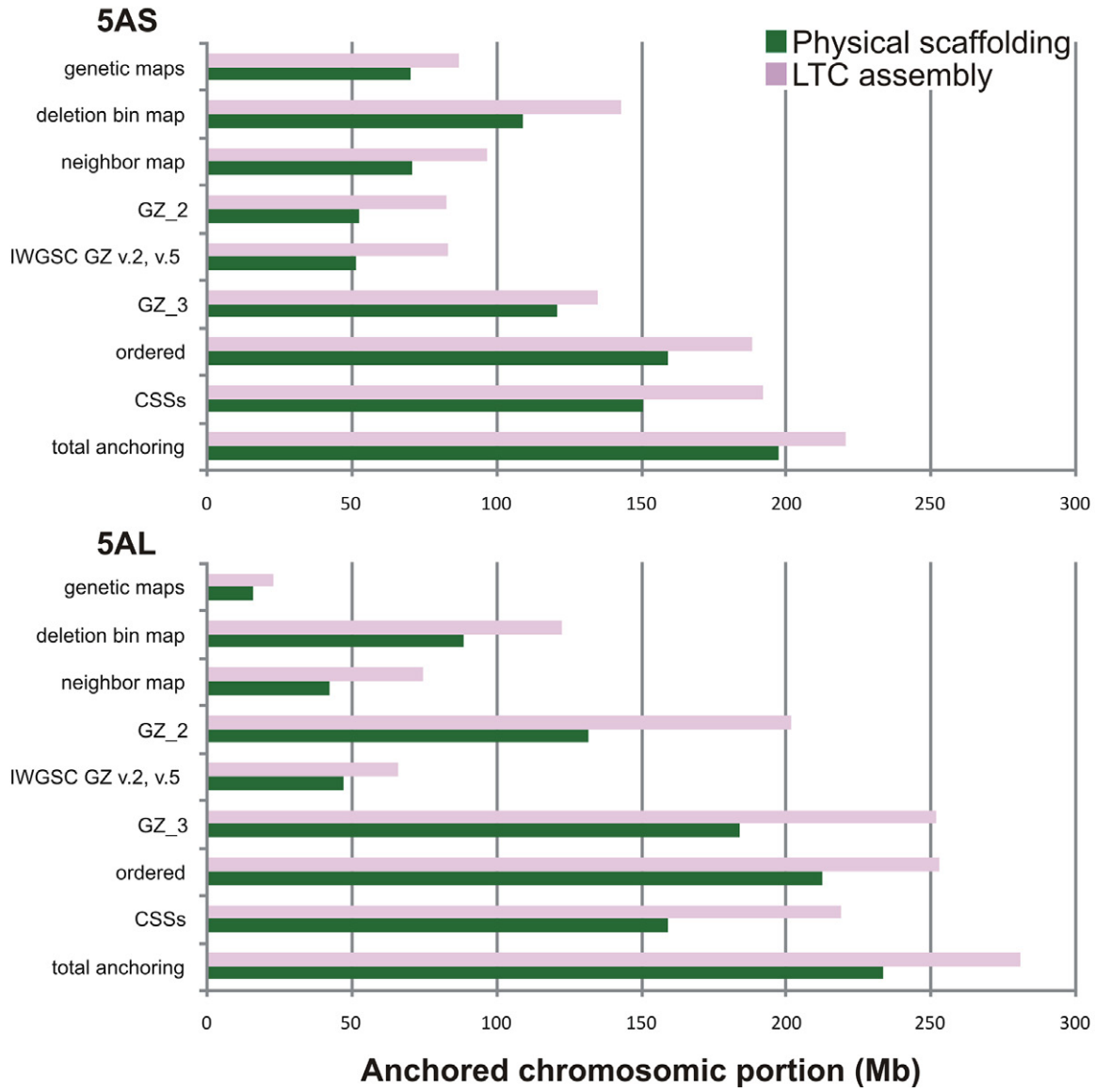

Figure 4. Anchored chromosome portion based on Linear Topological Contig (LTC) assembly and physical scaffolding by means of different approaches. Only contigs with a clone number ≥3 and scaffolds with a clone number ≥6 were included.

arm (Table 4). The percentages reported in Table 4 for each genomic tool employed in anchoring are slightly higher from those already listed previously in each paragraph, since they derive from the integration between different resources (e.g., some contigs linked to GZ_2 genes were included into the neighbor map after in silico analysis).

Comparative Genomics for Creation of an Integrated and Ordered Physical Map for Chromosome 5A

To achieve the final integrated and ordered physical map for the entire chromosome 5A, all the anchoring information obtained by PCR, hybridization, or BLASTn searches were combined. The neighbor map was chosen as the starting point to include all markers and genes linked to physical contigs, since it was considered as the most powerful and reliable resource available, being based on a CS genetic map and having a very high density of markers, most of which were represented by EST-based SNPs. Syntenic relationships of the mapped 90K_SNPs with the three reference genomes (rice, *Brachypodium*, and sorghum) were acquired from literature (Wang et al., 2014).

All the markers inferred into the neighbor map and having an homology with genes located on *Brachypodium* chromosomes different to 1 and 4 were subsequently removed from the neighbor map. The anchored markers, BESs, and genes not previously included into the neighbor map were then added one by one by using the relative physical coordinates of the closest orthologous gene (identified through coding sequence analysis) acquired from the available GZs or obtained from the Wheat Zapper program (Alnemer et al., 2013). Marker positions, as deduced from the syntenic relationships of the orthologous gene, were verified for agreement with the neighboring contigs, the corresponding Illumina contigs, and the position on deletion bins (when available). When no syntenic relationship

Table 5. Major colinearity breaks between 5A chromosome (Chr.) of wheat and *Brachypodium*, rice, and sorghum.

Chr. arm	Syntenic block	Bd† gene up	Bd gene down	Os‡ gene up	Os gene down	Sb§ gene up	Sb gene down	Anchored marker no.	Anchored gene¶ no.
5AS	1	Bradi4g00250	Bradi4g08140	LOC_Os12g44340	LOC_Os12g06640	Sb08g023310	Sb08g003990	144	170
5AL	2	Bradi4g08330	Bradi4g44960	LOC_Os09g02530	LOC_Os09g39640	Sb02g013138	Sb02g032810	38	269
5AL	2b					Sb02g011240	Sb02g012390		
5AL	3	Bradi1g10930	Bradi1g63880	LOC_Os03g49600	LOC_Os03g20580	Sb01g010830	Sb01g036840	2	29
5AL	4	Bradi1g09200	Bradi1g02160	LOC_Os03g52980	LOC_Os03g62500	Sb01g008600	Sb01g002920	13	107
5AL	5	Bradi1g66650	Bradi1g78650	LOC_Os03g11874	LOC_Os03g01100	Sb01g039615	Sb01g050650	15	69

† Bd, *Brachypodium distachyon* (L.) Beauv.

‡ Os, *Oryza sativa* L.

§ Sb, *Sorghum bicolor* (L.) Moench.

¶ 90K_SNPs included.

was available, the link to physical contigs was used to position the markers. The consistency between the LTC physical map and the virtual gene order was evaluated and verified also with the end-to-end merging of LTC contigs into scaffolds, which helped to clarify some uncertainties (e.g., in 5AL, ctg_11831 and ctg_832 merged in scf_162) and confirmed the order of different LTC contigs, linking them to unique scaffolds (e.g., in 5AS, ctg_9966, ctg_379, ctg_10143, ctg_262 merged in scf_19; Supplemental Table S12).

The integrated physical map shown few discrepancies with deletion bin map (reported in detailed and colored manner in Supplemental Table S12).

The integration of the high density neighbor map with the collinear gene order of orthologous *Brachypodium*, rice, and sorghum genes generated a further new virtual gene order (GZ_3) based on a physical map for 5A, including 91 scaffolds for 5AS (135 Mb) and almost all the 5AL anchored scaffolds (237, 251 Mb) (Table 4; Supplemental Table S12) with 170 and 474 anchored genes along the short and long arm, respectively (Table 5).

Besides highlighting syntenic regions in the three reference genomes, the integrated physical map allowed a better definition of the conservation and breakages of syntenic relationships and inversions with respect to the genetic order (Fig. 5). We observed that 5AS was syntenic with a region on rice chromosome 12, encompassing LOC_Os12g44340 and LOC_Os12g06640 (block 1), while chromosome 5AL was syntenic with rice chromosome 9 between LOC_Os09g02530 and LOC_Os09g39980 (block 2) (Table 5; Supplemental Table S12; Fig. 5). A region of 5AS syntenic with the sorghum genome was identified on sorghum chromosome 8 between Sb08g023310 and Sb08g03990 (block 1), while for 5AL, syntenic regions were identified on chromosome 2 between Sb02g013138 and Sb02g032810 (block 2) and between Sb02g011240 and Sb02g012390 (block 2b) (Table 5; Supplemental Table S12).

In *Brachypodium*, 5AS showed syntenic relationships to *Brachypodium* chromosome 4 between loci Bradi4g00250 and Bradi4g08140 (block 1), while a region of the 5AL was syntenic with *Brachypodium* chromosome 4 defined by loci Bradi4g08330 and Bradi4g44960 (block 2).

From the middle of LG2, the long arm is perfectly syntenic to chromosome 1 of *Brachypodium* (from Bradi1g10930 to Bradi1g78650; blocks 3 and 5) and to chromosome 1 of sorghum (from Sb01g10830 to Sb01g050650; blocks 3 and 5). Inserted between the blocks 3 and 5 there is the block 4, which corresponds to the most proximal portion of long arms of the two syntenic chromosomes 1 of *Brachypodium* and sorghum. The order of the loci of this last block is inverted in 5AL relative to *Brachypodium* (from Bradi1g09200 to Bradi1g02160) and sorghum (from Sb01g008600 to Sb01g002920) (Table 5; Fig. 5). With respect to rice, 5AL block 4 is related to the most distal region of chromosome 3, between LOC_Os03g52980 and LOC_Os03g62500, without the inversion observed in the comparisons with the other two genomes. The position of block 4 (in wheat map from 70.4 to 117.8 cM) and its orientation with respect to the *Brachypodium*, rice, and sorghum genomes is supported also from positions of the anchored genes or markers on the POPSEQ map (Supplemental Table S12). Indeed, an almost perfect correspondence was observed between the map position of CSSs in the POPSEQ and in the neighbor map. With the support of the integrated physical map the inversion of this syntenic region in wheat with respect to the *Brachypodium* genome, already previously hypothesized (Vitulo et al., 2011), was better characterized in term of the 5AL distal regions. The same inversion was observed respect to sorghum.

The integrated map (GZ_3) also highlighted scaffolds identified by unlinked markers (i.e., scf_205 identified by markers at 12.7 and 117.8 cM identified by markers at 17.3 cM) and two scaffolds present even on two different syntenic blocks (i.e., scf_20, blocks 2 and 5; scf_112, blocks 2 and 3; Supplemental Table S12). Such results could be explained by the presence of a duplication of related sequences. Furthermore, six other putative duplications of genes were highlighted between the two chromosome arms (Supplemental Table S12), as already observed in a previous study performed on deletion mapping 5A (Gadaleta et al., 2014). Nonetheless, further analyses are needed to exclude the possibility that the presence of the marker in these two different contigs or scaffolds derives from contamination, not detected by the assembly.

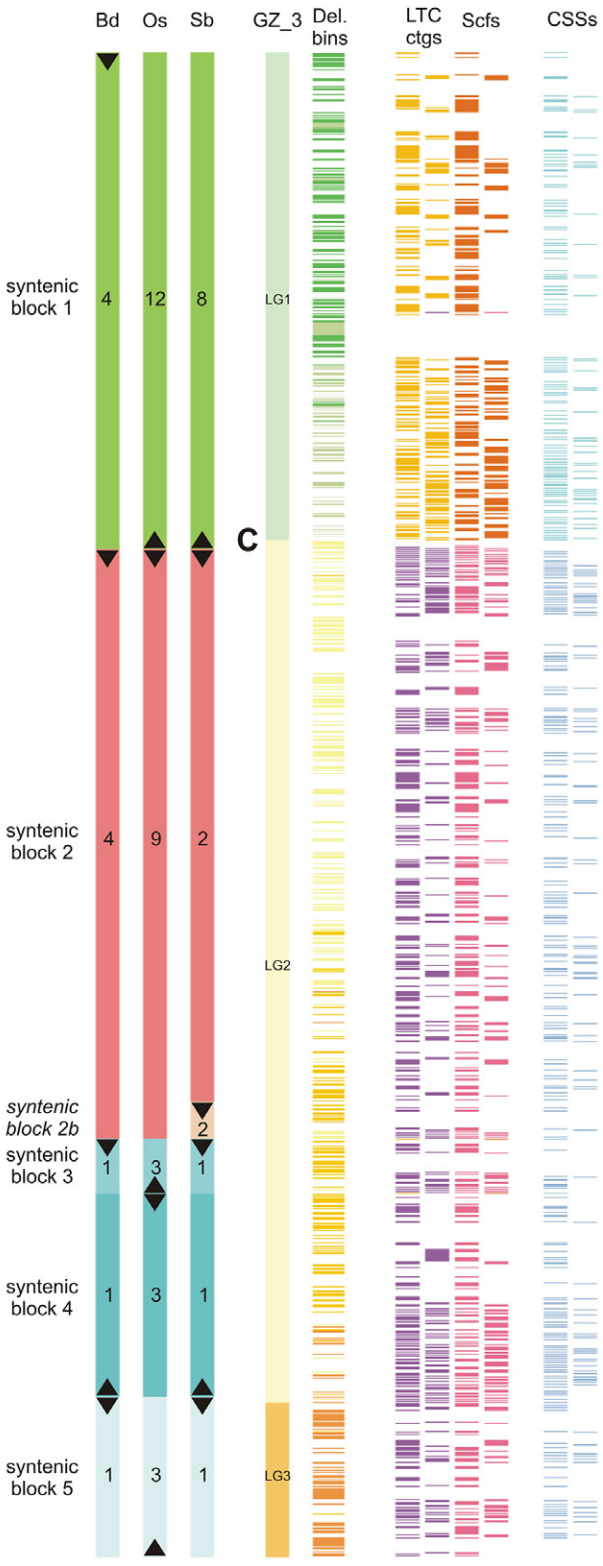

Figure 5. Schematic representation of the syntenic relationships between wheat chromosome 5A (GZ_3: integrated physical map) and the orthologous chromosomes in *Brachypodium distachyon* (Bd), rice (Os), and sorghum (Sb). The number and the black arrow reported in each syntenic block identify the orthologous chromosomes and their orientation respect to wheat (▽ same orientation; Δ opposite orientation). The GZ_3 is made of three linkage groups (LGs) and C identifies the centromere. For each marker present in the GZ_3, the bin interval (Del. bins), the Linear Topological Contig (LTC ctgs), the scaffold (Scfs), and the chromosome survey sequence (CSSs) are reported, when available, as colored rectangles with a thickness corresponding to the estimated length scaled according to the neighbor map. For deletion bins each color corresponds to one of the six bin intervals defined. Also, the LTC, scaffolds, and CSSs are differently colored between the short and long arm, and are reported on two columns to highlight the sequential placement along the GZ_3 without any overlapping.

Conclusions

A variety of assembly strategies, genetic and genomic tools, and complementary anchoring approaches was used to produce an anchored physical map of the wheat chromosome 5A, which comprised 180 physical scaffolds for 5AS (75% of arm length) and 264 scaffolds for 5AL (53% of the arm length). The anchoring level for 5AS (only slightly lower than that obtained for 1BS and 6A) is to be considered satisfactory, while the coverage of 5AL (more similar to the first coverage of 3B) still needs to be improved. The limited number of clones analyzed by fingerprinting could have affected the anchoring procedure. Overall, a placement was assigned to 133 and 238 scaffolds, including 196 and 521 genes or ESTs, for the short and long arm, respectively. The fraction of the 5A physical map, both anchored and oriented, corresponds to 188 Mb (64% of the short arm) and 252 Mb (48% of the long arm) (Table 4; Fig. 4).

The high-density neighbor map developed during this project was revealed to be an efficient tool to integrate all anchoring results of BESs, Illumina sequences, and genes from the GZs. Nevertheless, the pericentromeric region has many cosegregating markers linking different physical contigs. Such a low-recombination region can only be resolved by the alternate recombination-independent mapping method based on radiation-hybrid lines (Faraut et al., 2009). Indeed, radiation-hybrid mapping has been a successful mapping approach also in plant species like wheat where it was used to build high-resolution maps for single chromosomes (3B: Paux et al., 2008; 1D: Kalavacharla et al., 2006; 5A: Zhou et al., 2012).

The development of a physical map for chromosome 5A integrated with a high-resolution genetic and functional map and the establishment of an order among the MTP contigs will help the genetic mapping of relevant agronomic traits and will assist focused sequencing projects. Moreover, a rich platform of 5A genomic resources, including about 9000 high-quality BESs with 3931 ISBP markers and 124 SSRs already developed and 25,000 high-confidence TE-based markers generated from 5AS 454 reads, was produced and is now available for the research community. An enrichment of the present physical map was achieved also by integrating and anchoring 5.4 Mb of CSSs linked to 218 contigs; among these, 49 contain genes orthologous to fully sequenced related genomes (Supplemental Table S11). These sequences could be used to develop new molecular genetic markers.

Moreover, the contigs already genetically anchored and linked to the CSSs provide a useful connection between the 5A genetic maps (and related agronomical relevant loci) with the CSSs (Supplemental Table S11, S12).

The 5A physical map, along with all other single-chromosome (or chromosome arm) wheat physical maps produced so far, provides an invaluable resource for the release of a reliable high-quality genome sequence of bread wheat, as a background for the assessment of genes and other functional elements and as a fundamental step to gain insight into the evolutionary aspects of bread wheat.

Supplemental Information Available

Supplemental material is included with this article.

Supplemental Figure S1. The relative contribution of each segregating population, with the number of total (a) and each type of (b) polymorphic and specific markers. Only populations with less than 5% of missing data are reported.

Supplemental Figure S2. Contig anchoring in deletion bin map.

a. The cumulative length in Mb of anchored contigs in six bins, three for each chromosome arm.

b. The coverage density within each chromosome bin, calculated by normalizing the coverage size on the actual percentage of the chromosome bin interval.

Supplemental Table S1. Primer pairs for potential ISBPs and SSRs derived from BES of 5AS and 5AL-FPC-MTP.

The TE type and the TE source for ISBP markers, and the repetitive motif for SSRs are reported, together with the corresponding BAC clone and the estimated amplicon size.

Supplemental Table S2. ISBP high-quality markers derived from 454 reads (ISBP Finder output), 454 contigs (RJPrimer output); RJM and RJJM markers derived from 454 contigs (RJPrimer output); SSR markers derived from 454 contigs (Imperfect SSR Finder output).

The three outputs produced by the two programs for TE search, report the junction position and junction type. The output of Imperfect SSR Finder reports also the repetitive motif. The primer sequences and the amplicon size are listed for each markers.

Supplemental Table S3. Specific and polymorphic markers for 5AS and 5AL and genetic maps for 4 different segregating populations.

The chromosome arm location, the marker class, the bin position and the reference are reported for each marker, as well as the polymorphism and mapping status in every population. Also the primer sequences are listed in the table.

Supplemental Table S4. Neighbor maps of chromosome arms 5AS and 5AL.

In the first sheet the 5A maps used for the development of the neighbor maps and their relative contribution are summarized. The population type, the length of the map in cM, the number of linkage groups and the corresponding reference are reported for each of the maps employed for the realization of 5A neighbor maps. Moreover the number of markers placed on 5A chromosome of original maps, and out of them, the number of markers included in the neighbor maps (expressed also in percentage) are shown in the last columns.

In the other two sheets the neighbor maps are reported for both arms: for each marker, the type, the LG,

the cumulative distance in cM and the bin location are shown. Depending on the original population, each marker is highlighted with a different color, as indicated in the legend.

Supplemental Table S5. Updated version of 5AS and 5AL genome zipper (GZ_2).

The new version of the genome zipper reports for all reads the corresponding *Brachypodium* gene deriving from the orthologous chromosomes 1 and 4. Corresponding GrainGene EST and bin location are also listed. All the BESs with homologies with the *Brachypodium* genes (included or not in the genome zipper) are listed in the last sheet.

Supplemental Table S6. Results of anchoring performed by PCR of FPC contigs (ctgs) and relative correspondence with LTC ctgs, achieved using markers common between them.

Markers used for anchoring derive from the literature (in black), from 5AS specific sequencing (in red) or from ongoing collaborations (in blue). Most of them are genetically mapped in at least one of the four segregating populations used in this work. For all markers, chromosome arm location, marker type, amplicon size, PCR conditions and the corresponding deletion bin, whenever determined, are also indicated. The physical map data deriving from FPC assembly report the name of contigs directly anchored by each marker, the number of clones belonging to these ctgs and the size of each single ctg. In the second sheet the corresponding LTC contigs (ctg), when present, achieved thanks to common markers between FPC-MTP and LTC-MTP are provided (inferred anchored ctgs), with their size in bp. Some contigs, indeed, have been detected directly (direct anchored ctgs).

Supplemental Table S7. Results of anchoring by hybridization on Agilent chip of LTC contigs.

For each marker or gene that was anchored via chip, the chromosome arm location, the contig id, the number of included clones, the contig length and the probe origin are reported. Possible duplications between 5AS and 5AL are highlighted in bold and their location, when available, is reported.

Supplemental Table S8. Anchoring of genome zipper and neighbor map of 5A chromosome, and the first level of *in silico* anchoring: BES vs. ESTs used for Agilent chip.

In the first two sheets the corresponding contig id and anchoring strategy related to the genome zipper of short and long arm of 5A are reported. For each contig, the number of clones included and the size in kb (considering a virtual Band = 1.2kb) are shown.

In the third and fourth sheets the anchoring results of neighbor maps are reported for both arms divided by anchoring strategy and on the whole (total contig ID). Beside each contig the number of included clones and the estimated physical size are listed. In the last two sheets the results of *in silico* anchoring are listed: neighbor map versus genome zipper and BESs versus the sequences of the chip (for this latter Blast, only hits having relative contig information are shown).

Supplemental Table S9. Contig anchoring in deletion bin map.

All the markers or gene sequences mapped onto deletion bins and anchored to physical contigs are reported. A deletion bin interval (see the text for details) was also defined for each ID.

Supplemental Table S10. Link between physical map and chromosome survey sequences.

In the first sheet the BLASTn results between BESs and the Illumina chromosome survey sequences are reported, together with chromosome arm location and the corresponding anchored contig. In the following two sheets the BLASTn output for all the other already anchored sequences located on GZ or mapped on the neighbor map or deletion bin map (separated per chromosome arm) are listed, again the corresponding contig with its size are reported beside the linked Illumina contig. The last sheet reports all the Illumina sequence contigs that have been anchored and linked to the 5A physical map, with the size in bp and the corresponding genetic position determined using the wheat CSS POPSEQ.

Supplemental Table S11. The complete list of all markers and genes assigned to 5A LTC physical map.

The table lists the chromosome arm, the marker or gene name, the marker type, the anchoring strategy, the ID of the BAC clones, the name of the physical contig to which the marker is assigned, the number of clones included and the size in bp, the scaffold (or super contig) ID, the number of included clones, the relative size in bp, the wheat Illumina contigs matching the marker and the relative size in bp, the ID of the orthologous genes in *Brachypodium*, rice, sorghum, the name of the deletion bin, the location on the genetic tools developed or employed in this work: the genome zipper produced in the present project, the genome zipper v.2 and v.5 realized in the frame of the consortium (Mayer et al., 2014), the genetic maps, the neighbor map, and finally the placement on the integrated and ordered physical map.

In blue the synteny information derived from Wheat Zapper, in green those gathered from Wang et al. (2014), in violet the orthologous genes reported in the present 5A genome zipper.

Supplemental Table S12. Integrated and ordered physical map for 5A chromosome (or GZ_3) based on syntenic integration and putative duplications between the two arms or along 5AL.

The framework was provided by CSxCS5A neighbor map. Non orthologous genes are written in gray while genes highlighted in pink are supposed to be duplicated. Different colors define different deletion

bins and syntenic blocks.

In the second sheet the putative duplications between 5AS and 5AL and along 5AL are listed, with the corresponding LTC contig and scaffold derived from scaffolding and the relative position on integrated and ordered physical map (LG and cM).

Author's Contributions

AMS conceived the study and acquired the funding; GV, LC supervised the entire project; DB, GV, LC drafted the manuscript; HŠ, JD purified wheat chromosome 5A arms and produced the 5AS and 5AL BAC libraries; FC, FM, SS, MM produced the high-information-content fingerprints and the automated assembly of the 5AS and 5AL; JU performed the BAC ends sequencing; DB, MLP, AV developed SSRs and ISBPs from BES; DB, KL, LO, AMM developed TE-based and SSR markers based on 454 sequencing of 5AS; DB, KL, LO, AMM, VM, AF, VM, FD made the genotyping of the SSR, COS, and TE-based markers on four segregant populations; DB, AV produced the four genetic maps; AV developed the 5A neighbor maps; AG, AB performed the deletion bin mapping and kindly provided the CS × CS5A 90K map; FD kindly provided the Lt × MG 90K map; DB, MLP handled the 5A BAC libraries and produced the three dimensional MTP pools for both arms; DB, AV, MLP performed the anchoring of FPC-MTP by PCR; NV produced a new version of the Genome Zipper of wheat chromosome 5A; PT, PB, MD developed the Agilent chip; PB developed the script for anchoring; DB, PT produced hybridization on Agilent microarray and analyzed the hybridization results; DB, AV performed the anchoring of LTC-MTP by PCR and chip; DB, PB performed in silico anchoring; DB performed the comparative genomics analysis; ZF performed the physical scaffolding. All authors read and approved the manuscript.

Acknowledgments

The authors thank Dr. Marie Kubaláková, Dr. Jarmila Číhalíková, Zdeňka Dubská, and Romana Šperková for the preparation of chromosome samples, flow cytometric sorting, and determination of purities in sorted chromosome fractions; the International Wheat Genome Sequencing Consortium initiative for wheat chromosome survey sequencing; and Klaus Mayer (MIPs) for generating the Genome Zipper and making it available for analysis to consortium partners. The authors also wish to thank Moreno Colaiacovo for assistance in deconvolution of anchoring results, Enrico Francia for critical review of the manuscript, Bikram Gill for providing seeds of double ditelosomic line 5A (dDt5A) of wheat cultivar Chinese Spring, and Catherine Feuillet and Miroslav Valarik for providing the seeds of the CS × Re and DV × G3 populations, respectively. This work was supported by the Italian Ministry of Agriculture, Project "Mappa fisica del cromosoma 5A dei frumenti" (D.M. 7398/7303/08), by the Czech Science Foundation (award no. P501/12/G090), and by the Ministry of Education, Youth and Sports of the Czech Republic (grant LO1204 from the National Program of Sustainability I).

References

Allen, A.M., L.A. Gary Barker, P. Wilkinson, A. Burridge, M. Winfield, C. Uauy, et al. 2012. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). Plant Biotechnol. J. 11:279–295. doi:10.1111/pbi.12009

Alnemer, L.M., R.I. Seetan, F.M. Bassi, C. Chitraranjan, A. Helsene, P. Loree, et al. 2013. Wheat Zapper: A flexible online tool for colinearity studies in grass genomes. Funct. Integr. Genomics 13:11–17. doi:10.1007/s10142-013-0317-4

Appels, R., E. Akhunov, M. Alaux, M. Caccamo, F. Cattonaro, J. Dolezel, et al. 2010. IWGSC: Physical mapping standard protocols workshop. AWN 56:5–9.

Ariyadasa, R., M. Mascher, T. Nussbaumer, D. Schulte, Z. Frenkel, N. Poursarebani, et al. 2014. A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. Plant Physiol. 164:412–423. doi:10.1104/pp.113.228213

Ariyadasa, R., and N. Stein. 2012. Advances in BAC-based physical mapping and map integration strategies in plants. J. Biomed. Biotechnol. 2012:184854. doi:10.1155/2012/184854

Barabaschi, D., D. Guerra, K. Lacrima, P. Laino, V. Michelotti, S. Urso, et al. 2012. Emerging knowledge from genome sequencing of crop species. Mol. Biotechnol. 50:250–266. doi:10.1007/s12033-011-9443-1

Bolger, M.E., B. Weisshaar, U. Scholz, N. Stein, B. Usadel, and K.F.X. Mayer. 2014. Plant genome sequencing— Applications for crop improvement. Curr. Opin. Biotechnol. 26:31–37. doi:10.1016/j.copbio.2013.08.019

Boutin-Ganache, I., M. Raposo, M. Raymond, and C.F. Deschepper. 2001. M13-tailed primers improve the readability and usability of microsatellite analyzes performed with two different allele-sizing methods. Biotechniques 31:24–28.

Breen, J., T. Wicker, M. Shatalina, Z. Frenkel, I. Bertin, R. Philippe, et al. 2013. A physical map of the short arm of wheat chromosome 1A. PLoS ONE 8:e80272. doi:10.1371/journal.pone.0080272

Buerstmayr, H., B. Steiner, L. Hartl, M. Griesser, N. Angerer, D. Lengauer, et al. 2003. Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. II. Resistance to fungal penetration and spread. Theor. Appl. Genet. 107:503–508. doi:10.1007/s00122-003-1272-6

Cenci, A., S. Somma, N. Chantret, J. Dubcovsky, and A. Blanco. 2004. PCR identification of durum wheat BAC clones containing genes for carotenoid biosynthetic pathway and their chromosome localization. Genome 47:911–917. doi:10.1139/g04-033

Choulet, F., A. Alberti, S. Theil, N. Glover, V. Barbe, J. Daron, et al. 2014. Structural and functional partitioning of bread wheat chromosome 3B. Science 345:6194. doi:10.1126/science.1249721

Cone, K.C., M.D. McMullen, I. Vroh Bi, G.L. Davis, Y.S. Yim, J.M. Gardiner, et al. 2002. Genetic, physical, and informatics resources for maize. On the road to an integrated map. Plant Physiol. 130:1598–1605. doi:10.1104/pp.012245

Daniluk, J., N.A. Kane, G. Breton, A.E. Limin, D.B. Fowler, and F. Sarhan. 2003. TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiol. 132:1849–1860. doi:10.1104/pp.103.023523

Dean, F.B., J.R. Nelson, T.L. Giesler, and R.S. Lasken. 2001. Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. 11:1095–1099. doi:10.1101/gr.180501

Desiderio, F., D. Guerra, D. Rubiales, L. Piarulli, M. Pasquini, A.M. Mastrangelo, et al. 2014. Identification and mapping of quantitative trait loci for leaf rust resistance derived from a tetraplod wheat *T.* dicoccum accession. Mol. Breed. 34:1659–1675. doi:10.1007/s11032-014-0186-0

Diéguez, M.J., M.F. Pergolesi, S.M. Velasquez, L. Ingala, M. López, M. Darino, et al. 2014. Fine mapping of LrSV2, a race-specific adult plant leaf rust resistance gene on wheat chromosome 3BS. Theor. Appl. Genet. 127:1133–1141. doi:10.1007/s00122-014-2285-z

Doležel, J., M. Kubaláková, E. Paux, J. Bartoš, and C. Feuillet. 2007. Chromosome-based genomics in the cereals. Chromosome Res. 15:51–66. doi:10.1007/s10577-006-1106-x

Doležel, J., J. Vrána, J. Šafář, J. Bartoš, M. Kubaláková, and H. Šimková. 2012. Chromosomes in the flow to simplify genome analysis. Funct. Integr. Genomics 12:397–416. doi:10.1007/s10142-012-0293-0

Egan, A.N., J. Schlueter, and D.M. Spooner. 2012. Applications of next-generation sequencing in plant biology. Am. J. Bot. 99:175–185. doi:10.3732/ajb.1200020

Endo, T.R., and B.S. Gill. 1996. The deletion stocks of common wheat. J. Hered. 87:295–307. doi:10.1093/oxfordjournals.jhered.a023003

Faraut, T., S. de Givry, C. Hitte, Y. Lahbib-Mansais, M. Morisson, D. Milan, et al. 2009. Contribution of radiation hybrids to genome mapping in domestic animals. Cytogenet. Genome Res. 126:21–33. doi:10.1159/000245904

Feuillet, C., and K. Eversole. 2007. Physical mapping of the wheat genome: A coordinated effort to lay the foundation for genome sequencing and develop tools for breeders. Isr. J. Plant Sci. 55:307–313. doi:10.1560/IJPS.55.3-4.307

Feuillet, C., J.E. Leach, J. Rogers, P.S. Schnable, and K. Eversole. 2011. Crop genome sequencing: Lessons and rationales. Trends Plant Sci. 16:77–88. doi:10.1016/j.tplants.2010.10.005

Frenkel, Z., E. Paux, D. Mester, C. Feuillet, and A. Korol. 2010. LTC: A novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. BMC Bioinf. 11:584. doi:10.1186/1471-2105-11-584

Gadaleta, A., A. Giancaspro, S.L. Giove, S. Zacheo, O. Incerti, R. Simeone, et al. 2012. Development of a deletion and genetic linkage map for the 5A and 5B chromosomes of wheat (*Triticum aestivum*). Genome 55:417–427. doi:10.1139/g2012-028

Gadaleta, A., D. Nigro, S.L. Giove, O. Incerti, R. Simeone, L. Piarulli, et al. 2014. A new genetic and deletion map of wheat chromosome 5A to detect candidate genes for quantitative traits. Mol. Breed. 34:1599–1611. doi:10.1007/s11032-014-0185-1

Gottlieb, A., H.G. Muller, A.N. Massa, H. Wanjugi, K.R. Deal, F.M. You, et al. 2013. Insular organization of gene space in grass genomes. PLoS ONE 8:e54101. doi:10.1371/journal.pone.0054101

Haudry, A., A. Cenci, C. Ravel, T. Bataillon, D. Brunel, C. Poncet, et al. 2007. Grinding up wheat: A massive loss of nucleotide diversity since domestication. Mol. Biol. Evol. 24:1506–1517. doi:10.1093/molbev/msm077

Hernandez, P., M. Martis, G. Dorado, M. Pfeifer, S. Galvez, S. Schaaf, et al. 2012. Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. Plant J. 69:377–386. doi:10.1111/j.1365-313X.2011.04808.x

International Brachypodium Initiative (IBI). 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768. doi:10.1038/nature08747

International Wheat Genome Sequencing Consortium (IWGSC). 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788. doi:10.1126/science.1251788

Janda, J., J. Šafář, M. Kubaláková, J. Bartoš, P. Kovářová, P. Suchánková, et al. 2006. Advanced resources for plant genomics: BAC library specific for the short arm of wheat chromosome 1B. Plant J. 47:977–986. doi:10.1111/j.1365-313X.2006.02840.x

Kalavacharla, V., K. Hossain, Y. Gu, O. Riera-Lizarazu, M.I. Vales, S. Bhamidimarri, et al. 2006. High-resolution radiation hybrid map of wheat chromosome 1D. Genetics 173:1089–1099. doi:10.1534/genetics.106.056481

Kumar, A., K. Simons, M.J. Iqbal, M. Michalak de Jiménez, F.M. Bassi, F. Ghavami, et al. 2012. Physical mapping resources for large plant genomes: Radiation hybrids for wheat D-genome progenitor *Aegilops tauschii*. BMC Genomics 13:597. doi:10.1186/1471-2164-13-597

Lazo, G.R., S. Chao, D.D. Hummel, H. Edwards, C.C. Crossman, N. Lui, et al. 2004. Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L.): EST generation, UniGene analysis, probe selection and bioinformatics for a 16,000-locus bin-delineated map. Genetics 168:585–593. doi:10.1534/genetics.104.034777

Liu, H., J. McNicol, M. Bayer, J.A. Morris, L. Cardle, D.F. Marshall, et al. 2011. Highly parallel gene-to-BAC addressing using microarrays. Biotechniques 50:165–174. doi:10.2144/000113627

Lucas, S.J., B.A. Akpinar, M. Kantar, Z. Weinstein, F. Aydinoğlu, H. Šimková, et al. 2013. Physical mapping integrated with syntenic analysis to characterize the gene space of the long arm of wheat chromosome 1A. PLoS ONE 8:e59542. doi:10.1371/journal.pone.0059542

Lucas, S.J., H. Šimková, J. Šafář, I. Jurman, F. Cattonaro, S. Vautrin, et al. 2012. Functional features of a single chromosome arm in wheat (1AL) determined from its structure. Funct. Integr. Genomics 12:173–182. doi:10.1007/s10142-011-0250-3

Luo, M.C., C. Thomas, F.M. You, J. Hsiao, S. Ouyang, C.R. Buell, et al. 2003. High throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378–389. doi:10.1016/S0888-7543(03)00128-9

Mascher, M., G.J. Muehlbauer, D.S. Rokhsar, J. Chapman, J. Schmutz, K. Barry, et al. 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J. 76:718–727. doi:10.1111/tpj.12319

Mascher, M., and N. Stein. 2014. Genetic anchoring of whole-genome shotgun assemblies. Front. Genet. 5:208. doi:10.3389/fgene.2014.00208

Massa, A., and C.G. Morris. 2006. Molecular evolution of the puroindoline-a, puroindoline-b, and grain softness protein-1 genes in the tribe *Triticeae*. J. Mol. Evol. 63:526–536. doi:10.1007/s00239-005-0292-z

Mayer, K.F., M. Martis, P.E. Hedley, H. Simková, H. Liu, J.A. Morris, et al. 2011. Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263. doi:10.1105/tpc.110.082537

Mayer, K.F., J. Rogers, J. Doležel, C. Pozniak, K. Eversole, C. Feuillet, et al. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788. doi:10.1126/science.1251788

Paux, E., S. Faure, F. Choulet, D. Roger, V. Gauthier, J.P. Martinant, et al. 2010. Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol. J. 8:196–210. doi:10.1111/j.1467-7652.2009.00477.x

Paux, E., P. Sourdille, J. Salse, C. Saintenac, F. Choulet, P. Leroy, et al. 2008. A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101–104. doi:10.1126/science.1161847

Periyannan, S., U. Bansal, H. Bariana, K. Deal, M.C. Luo, J. Dvorak, et al. 2014. Identification of a robust molecular marker for the detection of the stem rust resistance gene *Sr45* in common wheat. Theor. Appl. Genet. 127:947–955. doi:10.1007/s00122-014-2270-6

Philippe, R., F. Choulet, E. Paux, J. van Oeveren, J. Tang, A.H. Wittenberg, et al. 2012. Whole genome profiling provides a robust framework for physical mapping and sequencing in the highly complex and repetitive wheat genome. BMC Genomics 13:47. doi:10.1186/1471-2164-13-47

Philippe, R., E. Paux, I. Bertin, P. Sourdille, F. Choulet, C. Laugier, et al. 2013. A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. Genome Biol. 14:R64. doi:10.1186/gb-2013-14-6-r64

Poursarebani, N., T. Nussbaumer, H. Šimková, J. Safář, H. Witsenboer, J. van Oeveren, et al. 2014. Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A. Plant J. 79:334–347. doi:10.1111/tpj.12550

Qi, L.L., B. Echalier, S. Chao, G.R. Lazo, G.E. Butler, O.D. Anderson, et al. 2004. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 168:701–712. doi:10.1534/genetics.104.034868

Qi, L., B. Echalier, B. Friebe, and B.S. Gill. 2003. Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct. Integr. Genomics 3:39–55.

Quraishi, U.M., M. Abrouk, S. Bolot, C. Pont, M. Throude, N. Guilhot, et al. 2009. Genomics in cereals: From genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. Funct. Integr. Genomics 9:473–484. doi:10.1007/s10142-009-0129-8

Raats, D., Z. Frenkel, T. Krugman, I. Dodek, H. Sela, H. Šimková, et al. 2013. The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. Genome Biol. 14:R138. doi:10.1186/gb-2013-14-12-r138

Rustenholz, C., P.E. Hedley, J. Morris, F. Choulet, C. Feuillet, and E. Paux. 2010. Specific patterns of gene space organization revealed in wheat by using the combination of barley and wheat genomic resources. BMC Genomics 11:714. doi:10.1186/1471-2164-11-714

Šafář, J., H. Šimková, M. Kubaláková, J. Číhalíková, P. Suchánková, J. Bartos, et al. 2010. Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenet. Genome Res. 129:211–223. doi:10.1159/000313072

Salameh, A., M. Buerstmayr, B. Steiner, A. Neumayer, M. Lemmens, and H. Buerstmay. 2011. Effects of introgression of two QTL for fusarium head blight resistance from Asian spring wheat by marker-assisted backcrossing into European winter wheat on fusarium head blight resistance, yield and quality traits. Mol. Breed. 28:485–494. doi:10.1007/s11032-010-9498-x

Scalabrin, S., M. Morgante, and A. Policriti. 2009. Automated fingerprint background removal: FPB. BMC Bioinf. 10:127. doi:10.1186/1471-2105-10-127

Sears, E.R. 1966. Nullisomic-tetrasomic combinations in hexaploid wheat. In: R. Riley and K.R. Lewis, editors, Chromosome manipulations and plant genetics. Oliver and Boyd, Edinburgh. p. 20–45.

Sears, E.R., and L.M.S. Sears. 1978. The telocentric chromosomes of common wheat. In: S. Ramanujam, editor, Proceedings of the 5th International Wheat Genetics Symposium. Indian Society of Genetics and Plant Breeding, New Delhi. p. 389–407.

Sehgal, S.K., W. Li, P.D. Rabinowicz, A. Chan, H. Šimkova, J. Doležel, et al. 2012. Chromosome arm-specific BAC end sequences permit comparative analysis of homoeologous chromosomes and genomes of polyploid wheat. BMC Plant Biol. 12:64. doi:10.1186/1471-2229-12-64

Šimková, H., J. Šafář, M. Kubaláková, P. Suchánková, J. Číhalíková, H. Robert-Quatre, et al. 2011. BAC Libraries from Wheat Chromosome 7D: Efficient Tool for Positional Cloning of Aphid Resistance Genes. J. Biomed. Biotechnol. 2011:302543. doi:10.1155/2011/302543

Simons, K.J., J.P. Fellers, H.N. Trick, Z. Zhang, Y.S. Tai, B.S. Gill, et al. 2006. Molecular characterization of the major wheat domestication gene Q. Genetics 172:547–555. doi:10.1534/genetics.105.044727

Soderlund, C., S. Humphray, A. Dunham, and L. French. 2000. Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. 10:1772–1787. doi:10.1101/gr.GR-1375R

Soderlund, C., I. Longden, and R. Mott. 1997. FPC: A system for building contigs from restriction fingerprinted clones. CABIOS Comput. Appl. Biosci. 13:523–535.

Somers, D.J., P. Isaac, and K. Edwards. 2004. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 109:1105–1114. doi:10.1007/s00122-004-1740-7

Stein, N., G. Herren, and B. Keller. 2001. A new DNA extraction method for high-throughput marker analysis in a large-genome species such as Triticum aestivum. Plant Breed. 120:354–356. doi:10.1046/j.1439-0523.2001.00615.x

Tondelli, A., E. Francia, D. Barabaschi, M. Pasquariello, N. Pecchioni, et al. 2011. Inside the CBF locus in Poaceae. Plant Sci. 180:39–45. doi:10.1016/j.plantsci.2010.08.012

Tranquilli, G., J. Heaton, O. Chicaiza, and J. Dubcovsky. 2002. Substitutions and deletions of genes related to grain hardness in wheat and their effect on grain texture. Crop Sci. 42:1812–1817. doi:10.2135/cropsci2002.1812

Vágújfalvi, A., G. Galiba, L. Cattivelli, J. Dubcovsky, et al. 2003. The cold regulated transcriptional activator Cbf3 is linked to the frost-tolerance gene Fr-A2 on wheat chromosome 5A. Mol. Genet. Genomics 269:60–67.

Van Oeveren, J., M. de Ruiter, T. Jesse, H. van der Poel, J. Tang, F. Yalcin, et al. 2011. Sequence-based physical mapping of complex genomes by whole genome profiling. Genome Res. 21:618–625. doi:10.1101/gr.112094.110

Van Ooijen, J.W. 2006. Joinmap 4: Software for the calculation of genetic linkage maps in experimental populations. Kyazma, B.V., Wageningen, the Netherlands.

Vitulo, N., A. Albiero, C. Forcato, C. Campagna, F. Dal Pero, P. Bagnaresi, et al. 2011. First survey of the wheat chromosome 5A composition through a next generation sequencing approach. PLoS ONE 6:e26421. doi:10.1371/journal.pone.0026421

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

Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, et al. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol. J. 12:787–796. doi:10.1111/pbi.12183

Wicker, T., D.E. Matthews, and B. Keller. 2002. TREP: A database for *Triticeae* repetitive elements. Trends Plant Sci. 7:561–562. doi:10.1016/S1360-1385(02)02372-5

Xue, S., Z. Zhang, F. Lin, Z. Kong, Y. Cao, and C. Li. 2008. A high-density intervarietal map of the wheat genome enriched with markers derived from expressed sequence tags. Theor. Appl. Genet. 117:181–189. doi:10.1007/s00122-008-0764-9

Yim, Y.S., P. Moak, H. Sanchez-Villeda, T.A. Musket, P. Close, P.E. Klein, et al. 2007. A BAC pooling strategy combined with PCR-based screenings in a large, highly repetitive genome enables integration of the maize genetic and physical maps. BMC Genomics 8:47. doi:10.1186/1471-2164-8-47

You, F.M., M.C. Luo, Y.Q. Gu, G.R. Lazo, K. Deal, J. Dvorak, et al. 2007. GenoProfiler: Batch processing of high-throughput capillary fingerprinting data. Bioinformatics 23:240–242. doi:10.1093/bioinformatics/btl494

You, F.M., H. Wanjugi, N. Huo, G.R. Lazo, M.-C. Luo, O.D. Anderson et al. 2010. RJPrimers: Unique transposable element insertion junction discovery and PCR primer design for marker development. Nucleic Acids Res. 38:W313–W320. doi:10.1093/nar/gkq425

Zhang, J. 2003. Evolution by gene duplication: An update. Trends Ecol. Evol. 18:292–298. doi:10.1016/S0169-5347(03)00033-8

Zhou, C., W. Dong, L. Han, L.J. Wei, L. Jia, Y. Tan, et al. 2012. Construction of whole genome radiation hybrid panels and map of chromosome 5A of wheat using asymmetric somatic hybridization. PLoS ONE 7:E40214. doi:10.1371/journal.pone.0040214