New BIOTECHNOLOGY xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

New BIOTECHNOLOGY

journal homepage: www.elsevier.com/locate/nbt

Integrated physical map of bread wheat chromosome arm 7DS to facilitate gene cloning and comparative studies

Zuzana Tulpová[a], Ming-Cheng Luo[b], Helena Toegelová[a], Paul Visendi[c], Satomi Hayashi[d], Petr Vojta[e], Etienne Paux[f], Andrzej Kilian[g], Michaël Abrouk[a], Jan Bartoš[a], Marián Hajdúch[e], Jacqueline Batley[h], David Edwards[h], Jaroslav Doležel[a], Hana Šimková[a,*]

[a] Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, 783 71 Olomouc, Czech Republic
[b] Department of Plant Sciences, University of California, Davis, USA
[c] Natural Resources Institute, University of Greenwich, Chatham Maritime, Kent ME4 4TB, UK
[d] Earth, Environmental and Biological Sciences, Queensland University of Technology, Brisbane, Queensland, Australia
[e] Laboratoty of Experimental Medicine, Institute of Molecular and Translational Medicine, Hněvotínská 5, 779 00 Olomouc, Czech Republic
[f] GDEC, INRA, Université Clermont Auvergne, 5 Chemin de Beaulieu, 63000 Clermont-Ferrand, France
[g] Diversity Arrays Technology Pty Ltd, University of Canberra, Bruce, Australia
[h] School of Biological Sciences and Institute of Agriculture, University of Western Australia, Crawley, WA, Australia

ARTICLE INFO

Keywords:
Triticum aestivum
BAC
BNG map
Aegilops tauschii
Centromere

ABSTRACT

Bread wheat (*Triticum aestivum* L.) is a staple food for a significant part of the world's population. The growing demand on its production can be satisfied by improving yield and resistance to biotic and abiotic stress. Knowledge of the genome sequence would aid in discovering genes and QTLs underlying these traits and provide a basis for genomics-assisted breeding. Physical maps and BAC clones associated with them have been valuable resources from which to generate a reference genome of bread wheat and to assist map-based gene cloning. As a part of a joint effort coordinated by the International Wheat Genome Sequencing Consortium, we have constructed a BAC-based physical map of bread wheat chromosome arm 7DS consisting of 895 contigs and covering 94% of its estimated length. By anchoring BAC contigs to one radiation hybrid map and three high resolution genetic maps, we assigned 73% of the assembly to a distinct genomic position. This map integration, interconnecting a total of 1713 markers with ordered and sequenced BAC clones from a minimal tiling path, provides a tool to speed up gene cloning in wheat. The process of physical map assembly included the integration of the 7DS physical map with a whole-genome physical map of *Aegilops tauschii* and a 7DS Bionano genome map, which together enabled efficient scaffolding of physical-map contigs, even in the non-recombining region of the genetic centromere. Moreover, this approach facilitated a comparison of bread wheat and its ancestor at BAC-contig level and revealed a reconstructed region in the 7DS pericentromere.

Introduction

Bread wheat (*Triticum aestivum* L.) is the staple food for 40% of the world's population. Improvements in yield and tolerance to biotic and abiotic stresses are essential to improve its production to meet the demands of the growing human population. Knowledge of the genome sequence would provide a deep insight into the genome composition and a detailed gene catalogue to facilitate genomics-assisted breeding. Bread wheat is an allohexaploid species (2n = 6x = 42, AABBDD), whose genome arose through spontaneous hybridization between tetraploid durum wheat, *Triticum turgidum* (AABB), and diploid goatgrass, *Aegilops tauschii* (DD), less than 400,000 years ago [1]. Allotetraploid *T.*

Abbreviations: 7DS, short arm of the bread wheat chromosome 7D; BAC, bacterial artificial chromosome; HICF, High Information Content Fingerprinting; FPC, FingerPrinted Contigs; LTC, Linear Topological Contig software; MTP, minimal tiling path; BNG map, Bionano genome map; CS, Chinese Spring; CsxRe genetic map, Chinese Spring x Renan genetic map; RH map, radiation hybrid map; CSS sequence, Chinese Spring survey sequence

* Corresponding author at: Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ−783 71 Olomouc, Czech Republic.

E-mail addresses: tulpova@ueb.cas.cz (Z. Tulpová), mcluo@ucdavis.edu (M.-C. Luo), toegelova@ueb.cas.cz (H. Toegelová), P.Muhindira@greenwich.ac.uk (P. Visendi), satomi.hayashi@qut.edu.au (S. Hayashi), petr.vojta@upol.cz (P. Vojta), etienne.paux@inra.fr (E. Paux), a.kilian@diversityarrays.com (A. Kilian), abrouk@ueb.cas.cz (M. Abrouk), bartos@ueb.cas.cz (J. Bartoš), marian.hajduch@upol.cz (M. Hajdúch), Jacqueline.batley@uwa.edu.au (J. Batley), Dave.Edwards@uwa.edu.au (D. Edwards), dolezel@ueb.cas.cz (J. Doležel), simkovah@ueb.cas.cz (H. Šimková).

https://doi.org/10.1016/j.nbt.2018.03.003

1871-6784/

turgidum originated *via* hybridization between diploid *Triticum urartu* (AA) and a diploid species from Sitopsis section (BB). The presence of three homoeologous subgenomes A, B and D and a high content of repetitive sequences (~85%) contribute to the huge genome size (16 Gb/1C) [2], which together impede its mapping and sequencing.

The first coordinated efforts towards obtaining a reference wheat genome date to 2005, when the International Wheat Genome Sequencing Consortium (IWGSC) was established. At that time, a proven strategy to obtain high-quality reference sequences of large genomes was the clone-by-clone approach, i.e. sequencing clones from large-insert DNA libraries ordered in physical maps. This procedure was initially used to produce reference sequences of *Arabidopsis* and rice and more recently also maize and barley [3–5]. To overcome problems due to polyploidy and genome complexity, the IWGSC adopted a chromosome-based strategy, which relied on dissecting the genome by flow-cytometric sorting of chromosomes and/or their arms [7,8], thus significantly reducing sample complexity. The first wheat chromosome-specific BAC library was constructed from chromosome 3 B by Šafář et al. in 2004 [9] and was used to construct a BAC-based physical map of the largest wheat chromosome [10]. Chromosomal BAC libraries were then constructed from all wheat chromosome arms of cv. Chinese Spring [11] (http://olomouc.ueb.cas.cz/dna-libraries/cereals) and laid the basis for constructing chromosomal physical maps for the whole wheat genome [10,12–20] (www.wheatgenome.org). They proved a favorable resource for map-based gene cloning [21] as well as generating a whole-chromosome sequence [22]. The availability of a new assembling algorithm DeNovoMAGIC, which is powerful enough to produce quality assemblies from whole-genome shotgun Illumina data even in large polyploid genomes, such as tetraploid wild emmer wheat [23], led to a change in the IWGSC plan and integration of the whole-genome assembly (WGA) approach as a key component of the sequencing strategy. WGA together with chromosomal maps, available BAC sequences and other resources resulted in a superior bread wheat reference sequence (www.wheatgenome.org), comparable to that obtained for wild emmer wheat.

Even with the progress in genome sequencing and assembly technologies, wheat chromosomal physical maps and the BAC clones they comprise remain a valuable resource, enabling a fast access to and a detailed analysis of a region of interest. This is because even the latest genome assemblies do not completely cover the genome. In the case of emmer wheat, the reference sequence represents 87.5% of the estimated genome size and the missing part is considered a combination of both unresolved repetitive sequences and difficult-to-sequence regions [23]. Moreover a part of sequence scaffolds (4.1% of emmer wheat assembly) remains non-assigned to a specific genome position. A highly complete and accurate sequence of the region of interest is a prerequisite for successful gene cloning. The availability of BAC clones with known genomic context enables focused and affordable re-sequencing of the critical region with more advanced technologies, such as PacBio and Oxford Nanopore, facilitating the identification of the desired gene or gene cluster [24,25].

In the current work, we targeted the short arm of bread wheat chromosome 7D (7DS) with the size of 381 Mb [11], which is known to harbor numerous agronomically important genes and QTLs [26–29]. An early version of a physical map constructed from a 7DS-specific BAC library [30], was used to delimit a region of a gene underlying an aphid resistance [29]. BAC clone Illumina sequences generated in the current study supplemented with OxfordNanopore sequences of a selected BAC clone are being used to identify candidate genes for the pest resistance (Tulpová, unpublished). The presented version of the 7DS physical map integrates markers from a radiation hybrid (RH) map and three genetic maps, including one from the D-genome ancestor *Ae. tauschii*. Our approach to physical map assembly included the previously published integration of a 7DS physical map with a whole-genome physical map of *Ae. tauschii* [31]. Co-assembly with *Ae. tauschii*, supported by a recently generated Bionano genome (BNG) map of the 7DS arm [32], facilitated partial ordering of physical map contigs even in the non-recombining region of the genetic centromere, which generally poses the major challenge in whole-genome assemblies. Moreover, this approach enabled comparison of genomes of bread wheat and its ancestor at a BAC-contig level and indicated a megabase-size region in the genetic centromere showing structural variation between *Ae. tauschii* and bread wheat 7DS.

Materials and methods

Physical map assembly

A total of 49,152 clones from the 7DS-specific BAC library TaaCsp7DShA [30] were previously fingerprinted using SNaPShot-based HICF technology [33], the resulting fingerprints were processed as described in Šimková et al. [30] and 39,765 useful fingerprints were used for automatic contig assembly in FPC [34]. The initial assembly was performed by incremental contig building with a cut-off value of 1×10^{-75}, tolerance of 0.4 bp and gel length 3600 followed by six iterations of single-to-end and end-to-end (Match: 1, FromEnd: 50) merging with decreasing cut-off up to the terminal value 1×10^{-45}. The DQer function was used after each merge to break up all contigs, that contained more than 10% questionable (Q) clones (Step: 3). In parallel, fingerprints of clones from the TaaCsp7DShA library were FPC-assembled together with fingerprints of all clones from *Ae. tauschii* BAC libraries previously used to produce the *Ae. tauschii* physical map [31]. This co-assembly resulted in *Ae. tauschii* supercontigs with embedded 7DS clones (Fig. 1), which were associated with contig signatures from the initial 7DS assembly. 7DS contigs showing apparent overlap in the co-assembly were subjected to end-merging in FPC with decreased stringency (Sulston-score value rising up to 1×10^{-15}).

As the next step, the 7DS assembly was validated using LTC software [35]. 3D view in LTC allowed visualization of missassembled or chimeric contigs, which were disjoined in FPC. Subsequently, contigs of two clones, likely originating from low-quality fingerprints or other wheat chromosomes that contaminated the 7DS fraction in the process of BAC library preparation, were removed (kill function). MTP clones were selected from this (pre-final) 7DS assembly using FPC with the following parameters: min FPC overlap = 25, from end = 0, min shared bands = 12. Final contig assembly merging was done upon the availability of MTP sequence data (see below). Deconvoluted sequence contigs of MTP BAC clones or whole BAC pools were compared using BLASTn (http://blast.ncbi.nlm.nih.gov/) with default parameters. Subjects of comparison were i) clones from different contigs sharing a marker, ii) clones proposed as overlapping by *Ae. tauschii* co-assembly, that could not be joined at 1×10^{-15}, and iii) clones, whose overlap was indicated by 7DS BNG map [32]. Physical-map contigs were merged if fulfilling at least three out of four criteria: marker sharing, sequence overlap, overlap in *Ae. tauschii* co-assembly, overlap supported by BNG map.

MTP sequencing and BNG map alignments

A total of 4608 MTP BAC clones were pair-end sequenced using a pooling strategy in which 96 pools, each consisting of four non-overlapping clones, were indexed and sequenced on a single lane of the Illumina HiSeq2000 platform [32]. DNA sequences were de-multiplexed and assembled using SASSY as described in detail by Visendi et al. [36]. DNA sequence deconvolution was supported by BAC-end sequences (BES), obtained by Sanger sequencing of all MTP clones from both ends, and by utilizing overlaps between BAC clones in physical map contigs. Mate-pair data was obtained by Illumina sequencing MTP-plate pools (384 clones per pool), and mate-pair reads were then applied to build scaffolds by SSPACE as described in Visendi et al. [36]. Alignments of sequence scaffolds to the 7DS BNG map were performed in IrysView 2.1.1. *Cmaps* were generated from *fasta* files of individual

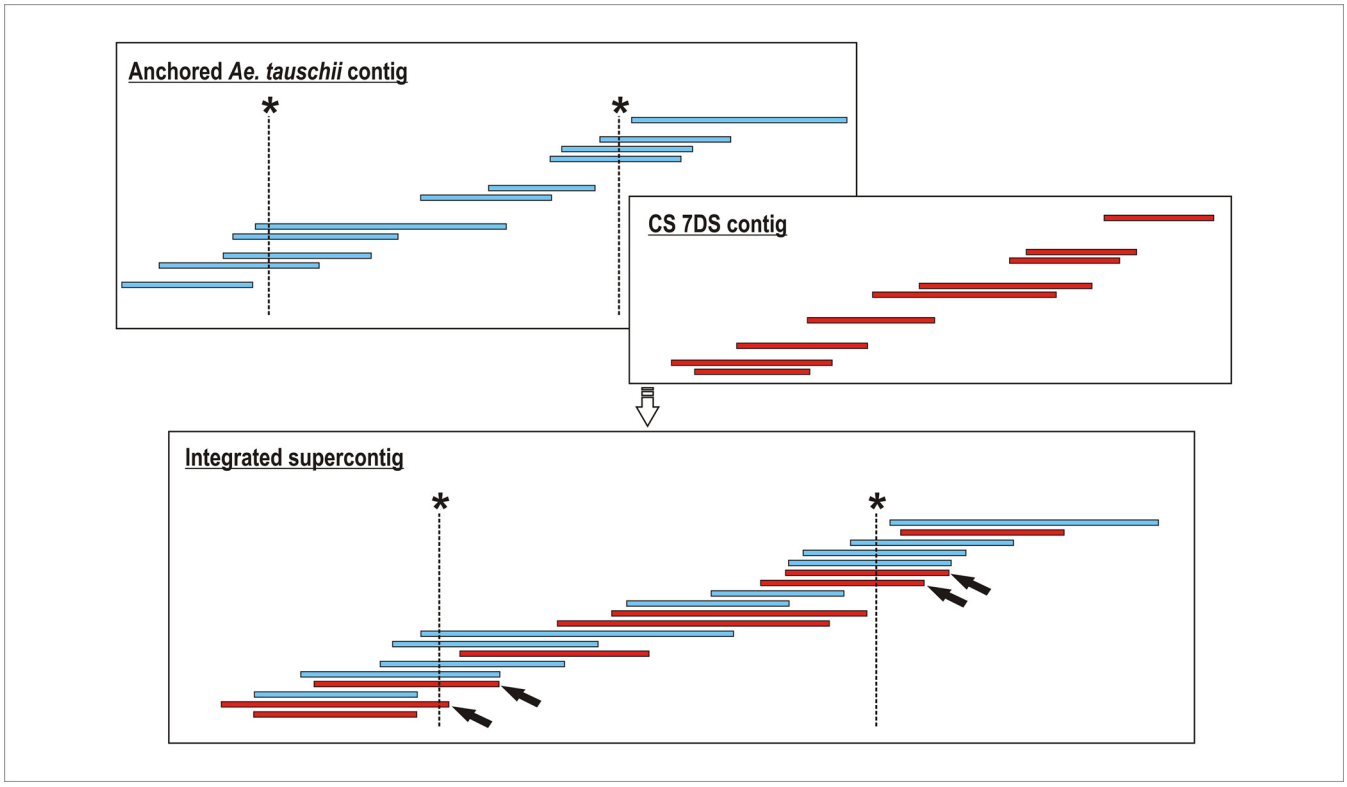

Fig. 1. Contig anchoring based on integration of 7DS and *Ae. tauschii* physical maps. A supercontig resulting from a co-assembly of BAC clones from *Ae. tauschii* (blue bars) and CS 7DS (red bars) comprises *Ae. tauschii* markers (asterisks) anchored to *Ae. tauschii* clones that can be tentatively assigned to 7DS clones (arrows) based on their overlap in FPC visualization. The assignment was validated by BLASTn of marker sequences to 7DS MTP BAC sequences.

MTP BAC clones or BAC pools. Query-to-anchor comparison was performed with default parameters and variable *P*-value threshold ranging from $1e^{-6}$ to $1e^{-10}$.

Physical map anchoring and inter-map comparison

The physical map of 7DS was anchored by combining three approaches. First, the TaaCsp7DShA library was screened with 35 publically available markers (Suppl. Table A) from GrainGenes database (https://wheat.pw.usda.gov/GG2/) using PCR on three-dimensional BAC pools following the procedure described in Šimková et al. [30]. The second approach utilized the 7DS-*Ae. tauschii* map integration described above. *Ae. tauschii* contigs were previously anchored to a linkage map comprising 1159 7D-specific SNP markers [31], part of which could be assigned to particular 7DS BAC clones based on *Ae. tauschii*-7DS clone overlap, as visualized by FPC (Fig. 1). The third approach – *in silico* anchoring – was used to anchor physical map contigs to *Ae. tauschii* linkage map, Chinese Spring x Renan F_6 RIL map (2413 7D SNP markers [37]), Chinese Spring consensus map v.3 (2323 7D DArTseq markers, Suppl. Table B) and Chinese Spring radiation hybrid map (485 7DS SNP markers from iSelect 90 K SNP array, [38,39]). Sequences of all markers were analyzed for homology with 7DS MTP sequences using BLASTn. Only unique hits with > 98% homology along the whole marker length were considered. For markers from the CsxRe map, CSS sequence contigs [40] comprising the markers were applied. To anchor *Ae. tauschii* markers, overlap with marker-bearing *Ae. tauschii* clones in the FPC visualization was required.

To perform inter-map comparison, anchored markers from individual maps were divided into groups of syntenic markers, which were compiled on the basis of co-localization in particular 7DS contigs. Information on genetic position of each marker and its syntenic relationship (Suppl. File A) were processed by Strudel software [41].

Results

Physical map assembly

A total of 39,765 high-quality fingerprints previously generated by SNaPShot HICF technology from a 7DS-specific BAC library [30] were automatically assembled into contigs using FPC, which resulted in assembly of 29,850 clones ordered in 1767 physical map contigs with average size of 264 kb, totaling approximately 468 Mb. This corresponded to 123% of the estimated 7DS chromosome arm length. The largest contig had a size of approximately 2382 kb. The assembly was manually end-merged based on the integration with the *Ae. tauschii* physical map and verified by 3D visualization in LTC. This reduced the number of contigs from 1767 to 1481 and increased their average size to 300 kb, totaling 445 Mb assembly length (117% of 7DS arm length). In the next step, contigs containing two clones were killed, which resulted in 931 physical map contigs comprising 28,339 clones with average contig size of 388 kb, N50 of 528 kb and L50 of 205. This pre-final assembly had estimated assembly length of 362 Mb, which corresponded to 95% of the 7DS arm length (Table 1).

The pre-final assembly was subject to MTP selection using FPC software. This resulted in an MTP of 4608 clones, which were sequenced. This provided an assembly totaling 9063 scaffolds with N50 of 117 kb, average scaffold size 63 kb and 2.5% Ns. On average, we obtained 1.9 scaffolds per BAC clone. The availability of MTP sequences enabled identification of new contig overlaps based on homology revealed by BLASTn search. In this way, 61 cases of potential overlaps were identified on edges of 121 contigs, out of which 71 could be end-merged under a lower-stringency condition in FPC. The remaining 50 contigs did not meet at least two out of three requirements (shared markers, fingerprint overlap with cutoff value $\leq 1 \times 10^{-15}$, clone overlap supported by BNG map) and were not merged. BNG map alignment revealed several misassemblies and proposed one contig

Table 1
Development of 7DS physical map assembly.

Assembly	Initial ($1e^{-45}$)	After integration with *Ae. tauschii* and LTC validation	After killing 2-clone contigs	After manual contig merging
No. contigs	1767	1481	931	895
Average contig size	264 kb	300 kb	388 kb	401 kb
Total assembly length (% arm length)	468 Mb (123%)	445 Mb (117%)	362 Mb (95%)	359 Mb (94%)
Contig N50	308 kb	424 kb	528 kb	555 kb
Contig L50	418	294	205	197

split. The final physical map assembly consisted of 895 contigs with average size of 401 kb, N50 of 555 kb and L50 of 197, representing approximately 359 Mb (94%) of the estimated chromosome-arm length (Table 1).

Physical map anchoring and inter-map comparison

Several approaches were combined to order 7DS physical-map contigs on the chromosome. First, 3-dimensional (3D) BAC pools were screened using PCR with 35 genetically mapped markers from GrainGenes database. Of these (Suppl. Table A), 21 were unambiguously assigned a distinct genetic-map position. The remainder were not anchored because they provided PCR products from two or more loci on 7DS. The second approach made use of integration of the 7DS physical map with the map of *Ae. tauschii*, which had been anchored to a high-density genetic SNP map [31]. The *Ae. tauschii* map of chromosome 7D spans 228.104 cM in total and comprises 1159 markers mapped to 488 distinct positions. The integration of 7DS and *Ae. tauschii* BAC clones in one physical map made it possible to delimit the short-arm part of the genetic map, which spans from 0 to 114.563 cM, and to identify 635 SNP markers in 7DS, which were mapped to 285 distinct positions. Out of the 635 markers, 582 could be tentatively assigned to 250 contigs of the 7DS physical map based on the graphical view in FPC (Fig. 1). Verification of this anchoring approach involved PCR screening of the 7DS-specific BAC library with bread wheat STS markers derived from their *Ae. tauschii* orthologs [29] and *in silico* anchoring based on MTP sequence data. The expected positions were confirmed for all STS markers and 460 out of 582 *Ae. tauschii* SNP markers. A failure to confirm additional *Ae. tauschii* markers was mainly due to an inability to deconvolute BAC sequence data or to find a significant sequence match for those markers. *In silico* anchoring positioned 621 SNP markers from the CsxRe genetic map, 147 SNP markers from the CS RH map and 464 DArTseq markers from the CS consensus map into contigs of the 7DS physical map. These efforts enabled anchoring 460 contigs of the 7DS physical map with 1713 markers (Table 2). The anchoring enabled delimiting the 7DS-specific part in the applied maps as follows: 0–117.597 cM (out of total 226.128 cM for 7D), 0–190.086 cM (out of total 371.634 cM) and 0–516.3 cR (out of total 758.7 cR) for CsxRe, CS consensus and CS RH maps, respectively.

In a previous study, the 7DS BNG map was demonstrated to be a useful tool for validating physical map assembly, scaffolding physical-map contigs, and integrating multiple maps [32]. Furthermore, the BNG map joined here an additional 45 marker-free 7DS contigs to the marker-anchored ones, thus increasing the number of anchored contigs to a total of 506 (Table 2). Combining all approaches, we were able to anchor 263.6 Mb of the 7DS physical map, corresponding to 73.4% of the assembly length.

Comparison of the three genetic maps integrated through the 7DS physical map showed a high collinearity along the 7DS arm (Fig. 2**A**) except for the subtelomeric region, where the CS consensus map showed a collinearity disruption with respect to *Ae. tauschii* and CsxRe maps (Fig. 2**B**). This disruption is roughly delimited by map interval 0–40.87 cM in the CS consensus map, corresponding to 0–31.66 cM and 0–12.994 cM in *Ae. tauschii* and CsxRe maps, respectively (Fig. 2B). The RH map showed overall collinearity but did not have sufficient resolution in the terminal region to resolve the inversion.

Table 2
BAC contig anchoring summary.

Anchoring method		No. anchored markers	No. anchored contigs
Pool screening[a]		21	21
Integration-based prediction	- SNP – *Ae. tauschii* map[b]	582	274
In silico anchoring	- SNP – *Ae. tauschii* map[b]	460	251
	- SNP – CsxRe map[c]	621	254
	- DArTseq – CS consensus map[d]	464	249
	- SNP – CS RH map[e]	147	98
BNG map[f]		NA	45
Overall		1713*	506

Map references.
* only sequence-validated (*in silico* anchored) *Ae. tauschii* markers considered.
[a] GrainGenes (Suppl. Table A).
[b] [31].
[c] [37].
[d] Suppl. Table B.
[e] [38].
[f] [32].

Ordering the centromeric region

Ordering physical map and sequence contigs in pericentromeric regions poses a major challenge in genomic projects due to suppressed recombination. Here we applied an integrated multiple-map approach to order physical map contigs as well as genetic markers around the genetic centromere of the 7D chromosome. It was delimited by a cluster of 88 markers co-segregating at 114.563 cM in the genetic map of *Ae. tauschii*. 85 markers from the cluster were positioned in 35 contigs from *Ae. tauschii* physical-map spanning over 77,090.5 kb in *Ae. tauschii* 7D. Using the physical map co-assembly, we identified 95 7DS contigs embedded in *Ae. tauschii* supercontigs associated with the centromeric marker cluster (Suppl. Table C). An additional five 7DS contigs were assigned to the cluster through BNG map-based scaffolding. A total of 100 7DS physical-map contigs spanned over 59,581 kb in Chinese Spring 7DS, corresponding to 16.6% of the total 7DS assembly length. No 7DS contigs were anchored beyond the 114.563 cM position, confirming affiliation of the cluster to the centromere. Seven out of 35 centromeric *Ae. tauschii* contigs did not comprise 7DS BAC clones and were proposed to belong to the 7DL chromosome arm (Suppl. Table C). The 7DL location was confirmed for all of them by finding matches between markers comprised in these contigs and the 7DL survey sequence [40]. Thus, a total of 15 *Ae. tauschii* markers from the centromeric cluster were assigned to the 7D long arm. The BNG map-based scaffolding also proposed joining an additional three marker-free *Ae. tauschii* contigs to the genetic centromere (Suppl. Table D), expanding it to 82,522.5 kb (including the 7DL part).

Applying all resources, we attempted to resolve the order and orientation of physical map contigs in this region. Besides the unambiguously assigned *Ae. tauschii* markers from the centromeric cluster, the 7DS centromeric contigs comprised 18 DArTseq markers mapping

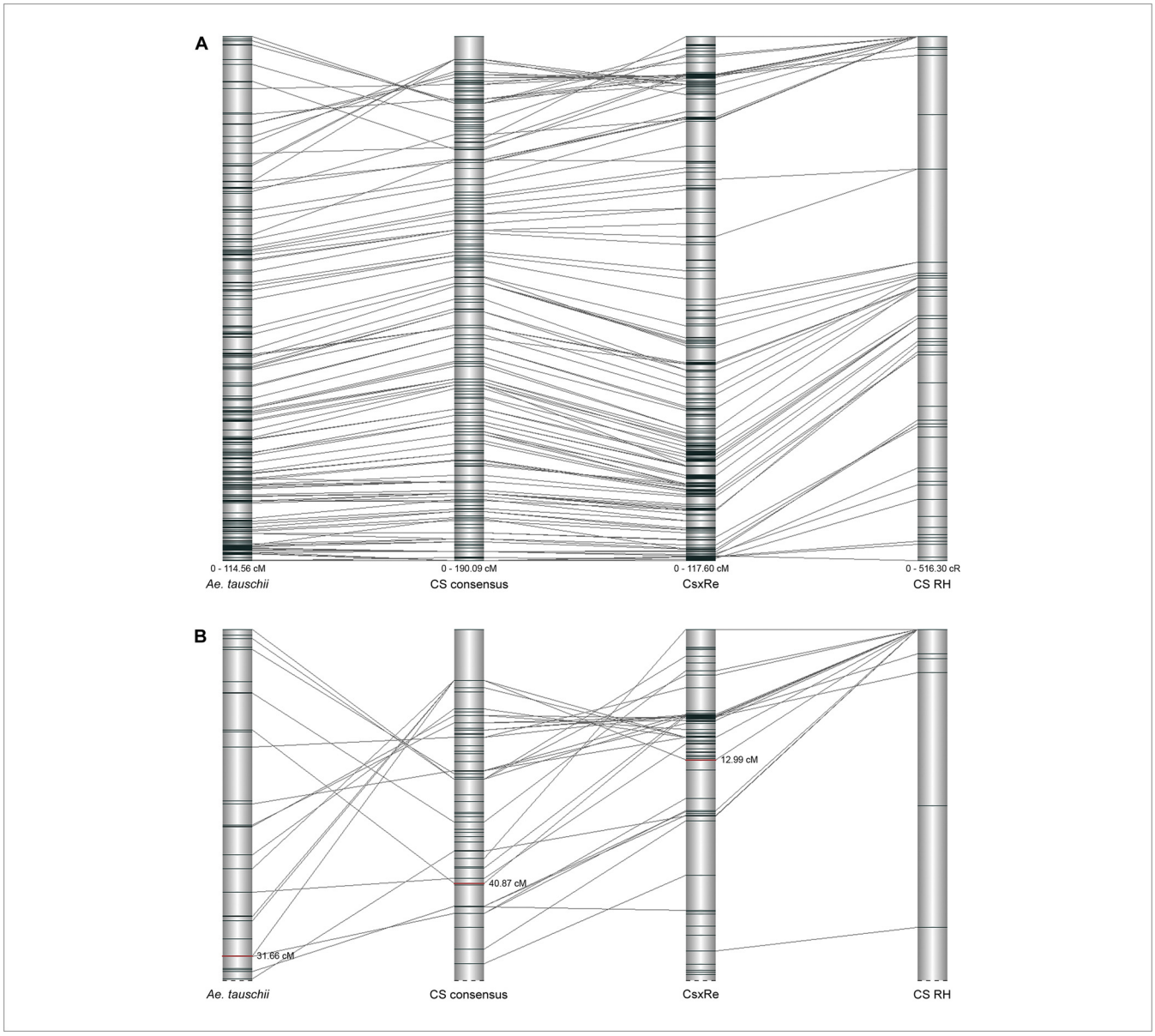

Fig. 2. Comparison of 7DS genetic and RH maps through 7DS physical map. Pairwise comparison of genetic/RH maps was performed after integrating particular marker datasets in contigs of the physical map. (**A**) Comparison of *Ae. tauschii* SNP map, CS consensus DArTseq map, CsxRe SNP map and CS RH map visualized by Strudel software. (**B**) Detail of collinearity disruption in CS consensus DArTseq map.

Table 3
Physical map of the genetic centromere (100 7DS physical-map contigs associated with *Ae. tauschii* centromeric marker cluster).

	No. 7DS contigs anchored	Total anchored length	No. scaffolds/ map positions	No. contig joins
Ae. tauschii co-assembly	95	55,481 kb	28	75
BNG map	66	49,755 kb	54	22
CsxRe map[a]	22	19,222.4 kb	5	NA
CS consensus map[a]	12	10,916 kb	4	NA
Combined[b]	**100**	**59,581 kb**	**16**	**86**

[a] Excluding markers mapping outside the most proximal map positions.
[b] *Ae. tauschii* co-assembly plus BNG map.

to five positions of the CS consensus map and 28 markers assigned to six positions of the CsxRe map (Suppl. Table C). Seventeen of the DArTseq markers mapped to 185.522–190.086 cM, which is the most proximal interval on the CS consensus map of 7DS. DArTseq marker 7D_3222174 assigned to 7DS ctg833 was allocated to 175.571 cM, suggesting a more distal position of the contig. Excluding this marker, we could anchor twelve of 7DS centromeric contigs to four positions of the CS consensus map (Table 3) and propose their tentative order. Out of 28 SNP markers from the CsxRe map, 26 mapped to interval 117.187-117.588, neighbouring the last 7DS position identified in the CsxRe map (117.597). While these 26 markers mapped to a total of five positions, they did not allow unambiguous ordering of 22 contigs containing them (Suppl. Table C). The remaining two SNP markers allocated in 7DS ctg3991 mapped to position 5.79 cM, which indicated a subtelomeric rather than centromeric location of the contig. None of the 7DS contigs associated with the centromeric marker cluster contained a SNP marker from the CS RH map and thus this resource could not be used for ordering contigs in the centromere.

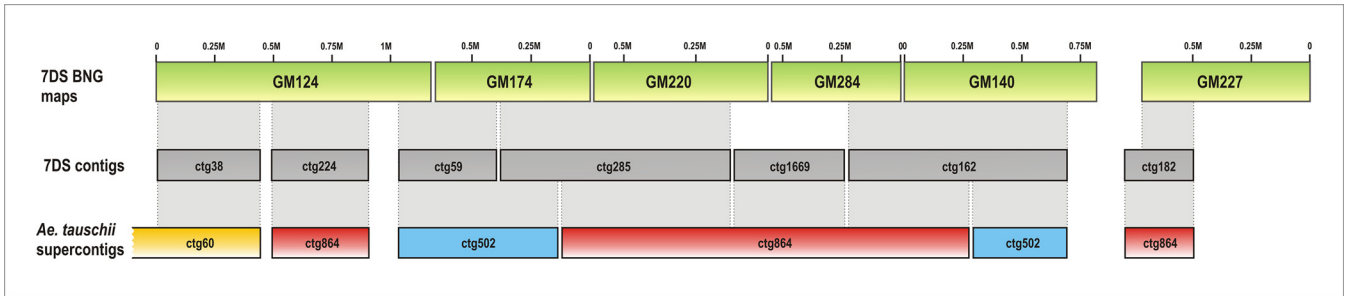

Fig. 3. Rearrangement in the pericentromeric region of Chinese Spring 7DS. 7DS contigs (grey boxes) assigned to *Ae. tauschii* supercontigs 60, 864 and 502 (yellow, red and blue boxes) by co-assembly were aligned through sequences of their BAC clones to 7DS BNG maps (green boxes), which indicated their different arrangement in 7DS of bread wheat compared to its ancestor. 7DS ctg1669 could not be aligned to any BNG map due to the lack of significant sequence match.

The co-assembly with *Ae. tauschii* aligned 95 7DS centromeric contigs with the *Ae. tauschii* genome, thus proposing 75 joins between separated 7DS contigs and combining the contigs in 28 physical-map scaffolds, corresponding to individual *Ae. tauschii* supercontigs (Table 3, Suppl. Table C). In parallel, projection of MTP sequences from centromeric contigs on 7DS BNG map aligned 66 contigs to particular genome maps, which proposed 22 contig joins. As a result, the centromeric contigs were ordered in 54 BNG map-based scaffolds. The 7DS BNG map confirmed the majority of scaffolds proposed by *Ae. tauschii* co-assembly and suggested 13 joins of *Ae. tauschii* contigs or their parts. Another two joins were indicated by 7DS contigs, which were split in two *Ae. tauschii* supercontigs each (Suppl. Table D). Combining both scaffolding approaches, we were able to organize the 7DS region corresponding to the *Ae. tauschii* centromeric marker cluster, spanning over 59,581 kb and comprising 100 7DS physical-map contigs, into 16 superscaffolds. A region delimited by *Ae. tauschii* supercontigs 864 and 502 appears to be rearranged in bread wheat with respect to *Ae. tauschii* since parts of the two *Ae. tauschii* contigs are interspersed in the CS 7DS, as proved by the BNG map (Fig. 3, Suppl. Table D).

Data accessibility

The 7DS physical-map assembly can be accessed through the GBrowse interface at https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_7DS_v2. The 7DS – *Ae. tauschii* co-assembly and 7DS BNG map are available on request. The BNG map is also deposited at the URGI repository https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.0/iwgsc_refseqv1.0_optical_maps_group7.zip. The 7DS MTP scaffolds have been deposited at DDBJ/ENA/GenBank under the accession No. PKRY00000000 (BioSample SAMN07709812). The version described in this paper is version PKRY01000000 and can be assigned to particular clones or BAC pools using Suppl. File B.

Discussion

Physical map assembly and anchoring

The construction of a 7DS physical map involved the most widely used methods, such as contig fingerprinting by SNaPShot-based HICF technology and contig building by FPC. Alternative approaches, used in other wheat physical-map projects, included whole-genome profiling (WGP) to generate fingerprints of BAC clones [16,17] and application of LTC software for contig assembly. WGP was not compatible with our approach, which employed co-assembly with previously fingerprinted *Ae. tauschii* clones, since the co-assembly was conditioned by identical processing of clones from CS 7DS and *Ae. tauschii*. This also determined the use of FPC as the primary assembly software. LTC was shown to outperform FPC in terms of contig number and length (N50 value) in several wheat chromosomal projects, but the LTC assemblies generally covered smaller portions of the estimated chromosome arm lengths [12,13,15,19]. In the present work, LTC was only used for contig validation. Despite relying on FPC, the parameters of our 7DS assembly were comparable or superior to those obtained for other chromosomes. In our assembly, we considered all contigs > 2 clones, which were subject to MTP selection, whereas other physical maps excluded contigs with less than five or six clones, mainly due to difficulty in anchoring them. BNG map applied in our project resolved this problem as it reliably anchors contigs with only two sequenced clones [32]. Thanks to the inclusion of the short contigs, we achieved one of the highest chromosome-arm coverages (94%) among all wheat physical maps published to date. In comparison with other physical maps assembled by FPC, our full assembly's N50 value (555 kb) is higher than that obtained for 3DS (445 kb) [20], 1AL (460 kb) [13] or 5A (296 kb and 252 kb for 5AS and 5AL) [19] but lower than the N50 value for 1BS (1033 kb) [14] or 5DS (1141 kb) [18]. Excluding contigs < 6 clones, the N50 of the 7DS assembly increased up to 616 kb, which brings the value closer to that obtained for 3 B (783 kb) [10].

In order to position contigs of the 7DS physical map on the 7DS, we combined several approaches and genomic resources, including three genetic maps, one RH map, the 7DS BNG map and co-assembly with the *Ae. tauschii* physical map. Using 1713 markers with distinct genomic positions, we ordered 506 contigs, totaling 74% of the assembly, along the chromosome arm. The most efficient means of allocation was *in silico* anchoring, based on searching homology between marker and MTP-clone sequences, which positioned 460 contigs on the chromosome. Another 45 marker-free contigs were joined to those anchored by BNG map, which poses a highly reliable and recombination-independent tool to scaffold assemblies. The resulting high rate of assembly anchoring was achieved without employing high-throughput platforms such as NimbleGen UniGene microarray, used in other wheat projects [12–15]. Several projects [12,14,19,20] also exploited synteny-based approaches and anchored physical map assemblies to so called genome zippers [42], which predict a virtual gene order on the basis of synteny with three grass genomes (rice, brachypodium and sorghum). The limitation of this approach can be a relatively high proportion of genes (or non-recognized pseudogenes) in non-syntenic positions [12] and also low quality or different origin of genetic map used to build the zipper [18].

Ordering the centromeric region

Ordering physical-map contigs and sequences in non-recombining regions around genetic centromeres has been the most challenging part of genome-sequencing projects. In our study, the genetic centromere was delimited as a cluster of contigs associated with markers of the *Ae. tauschii* map co-segregating at the most proximal position of the 7DS arm. This corresponded to a cumulative length of 70,882.5 kb and 59,581 kb for 7DS-assigned *Ae. tauschii* and Chinese Spring contigs, respectively. These values are with a high probability underestimates of the real size of the non-recombining region since we considered only contigs anchored to *Ae. tauschii* markers and those added to them by

BNG map-based scaffolding. Contigs anchored to the most proximal positions of the other two linkage maps used in our study are candidates for expanding the region of the 7DS genetic centromere. The 59,581 kb region, representing 16.6% of the 7DS assembly length, is only a part of the difficult-to-order low-recombining region, which was reported to span over more than 40% of the chromosome length in wheat chromosomes [10,15].

As expected, integrating information from additional two genetic maps did not aid in resolving the order of contigs around the centromere (Table 3, Suppl. Table C). The wheat radiation hybrid map, not relying on genetic recombination and showing a high resolution in the pericentromeric region of 7D chromosome [38], appeared a promising genomic resource. Unfortunately, we failed to allocate SNP markers from the RH map to the 100 7DS centromeric contigs. This may reflect the fact that the iSelect 90 K SNP array [39] used to genotype the CS RH panel mostly comprises markers from genic/low-copy regions that are relatively sparsely represented in the centromeric part of RH maps [43]. Repeat junction markers (RJM) or ISBP markers, derived from junctions of transposable elements and dispersed throughout the genome, were shown to be a favorable complement to the gene-based markers in RH mapping [44]. In our physical map project, we tested 380 RJM markers intended for mapping on the Chinese Spring 7D RH panel (V. Tiwari, personal communication) and managed to align six of them to the centromeric contigs (data not shown), thus indicating a modest potential of the anticipated RH map for the centromeric region.

Alternative methods for ordering contigs/BAC clones in low-recombining regions include cytogenetic mapping [45], which is laborious and cannot be used in a high-throughput manner, or synteny-based anchoring to genes virtually ordered in genome zipper [42]. The synteny-based approach did not prove to be beneficial for centromeric regions because of low representation of genes [12], which are mostly in non-syntenic positions around the centromeres [44]. The most efficient approach to assembling non-recombining regions was recently reported by Mascher et al. [6] who applied an innovative method of chromosome conformation capture-based ordering to complete BAC-based assembly of barley. This approach is applicable only at final stages of sequencing projects, as its resolution in large genomes is relatively limited (1 Mb in the barley project) and thus is conditional on availability of large sequence scaffolds.

BNG map-based ordering applied in our study also relies on the availability of BAC clone sequences. However, as we demonstrated earlier [32], assemblies of 90 kb or even less are sufficient to align to the BNG map if used in the context of a physical map. Anchoring 100 centromeric contigs to BNG maps proposed 22 joins and split the region into 54 physical-map scaffolds. The potential of the BNG map could not be fully exploited in the pericentromere due to a high content of DNA repeats, which hampered sequence assembly and deconvolution and thus reduced the number of sequences that could be aligned to the BNG map. Overall, the best clue to ordering contigs in the genetic centromere was the co-assembly with the *Ae. tauschii* physical map, which indicated 75 contig joins and distributed the 7DS centromeric contigs into 28 scaffolds. Nevertheless, the *Ae. tauschii*-based ordering must be taken with caution as the fingerprint co-assembly was done with less stringent parameters and might have resulted in erroneous clone joins. Furthermore, structural variation between bread wheat and its ancestor should be considered. Nevertheless, the co-assembly with *Ae. tauschii* provided a valuable clue for ordering contigs of the CS 7DS physical map and this beneficial effect was reciprocal, since the co-assembly, combined with alignment of the 7DS BNG map, added three marker-free *Ae. tauschii* contigs to this region and provided a clue for scaffolding a large part of *Ae. tauschii* centromeric contigs. The CS 7DS – *Ae. tauschii* co-assembly was previously used to validate the *Ae. tauschii* physical map assembly [31] and here we show further extension of this approach, that can be used to support *Ae. tauschii* genome assembly in the centromeric region of 7DS. To summarize, by combining both physical-mapping approaches, we could organize the region of 59,581 kb, associated with the 7DS genetic centromere, into 16 validated physical map scaffolds, thus outperforming outcomes of other wheat physical map projects.

Comparison of bread wheat and Ae. tauschii

Ae.tauschii is a donor of the youngest component of the bread wheat genome, which implies relatively low diversity between the D-genome of bread wheat and its ancestor [1]. Several rearrangements on the genic level and a small duplication were indicated by physical mapping on chromosome 5DS [18]. Structural variability between wheat and its ancestor was also proposed in chromosome 2D [46]. Chromosome arm 7DS seems highly collinear in wheat cv. Chinese Spring and *Ae. tauschii*, as indicated by inter-map comparison (Fig. 2A). The disruption of collinearity at the distal end of 7DS, apparent from the comparison of *Ae. tauschii* and CS consensus map (Fig. 2B), was not confirmed by the CsxRe map (Suppl. Figure A) and is likely due to an error in building the map consensus or to a structural variability in one of consensus-map components. The analysis of the distribution of CS 7DS contigs in the co-assembled supercontigs indicated several regions of potential variability between bread wheat and *Ae. tauschii* at a BAC-contig level. Validation of supercontigs from the centromeric cluster by the BNG map confirmed local rearrangements in a region spanning over more than 5 Mb (Fig. 3, Suppl. Table D), which could not be detected by the genetic maps.

Conclusions

This work resulted in a GBrowse-accessible 7DS physical map, which will facilitate gene cloning and comparative genome analyses in wheat. The physical map interconnects a total of 1713 markers from one radiation-hybrid map and three genetic maps with ordered 7DS BAC clones. All clones are available from library depositories at IEB (http://olomouc.ueb.cas.cz/dna-libraries/cereals) or CNRGV (http://cnrgv.toulouse.inra.fr/en/library/wheat), thus providing an opportunity to access a region of interest with advanced sequencing technologies in a focused and affordable manner. Mate-pair Illumina assemblies of MTP BAC clones generated in this study are available from NCBI and we anticipate their utilization in the envisaged version 2 of the bread wheat reference sequence. The co-assembly of 7DS with *Ae. tauschii* indicated regions of potential structural variation between bread wheat and its ancestor, which can be validated using BNG map, as demonstrated for the pericentromeric region of 7DS.

Author contributions

Z.T. performed physical-map anchoring, final stage of map assembly and centromere analyses. She drafted the manuscript. M.C.L. carried out fingerprinting of 7DS-specific BAC library and generated *Ae. tauschii* – 7DS co-assembly. H.T. participated in the construction of physical map and MTP sequencing. P.Visendi and D.E. assembled MTP sequences. S.H., J. Batley, P.Vojta and M.H. were engaged in MTP sequencing. E.P. and A.K. provided CS consensus and CsxRe maps, respectively. M.A. was involved in map integration and comparison. J. Bartoš supervised 7DS physical map assembly. J.D. participated in designing the study and revised the manuscript. H.Š. designed and coordinated the study and completed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank to Dr. Vijay Tiwari from the University of Maryland for providing prepublication access to Chinese Spring RH map and marker sequences. We acknowledge excellent technical assistance of Radka Tušková, Marie Seifertová and Helena Tvardíková. This work was supported by the Czech Science Foundation (award No. P501/12/

2554), and by the Ministry of Education, Youth and Sports of the Czech Republic (grant LO1204 from the National Program of Sustainability I). The research leading to these results has received funding from the French Government managed by the Research National Agency (ANR) under the Investment for the Future programme (BreedWheat project ANR-10-BTBR-03), from FranceAgriMer, French Funds to support Plant Breeding (FSOV) and from INRA.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.nbt.2018.03.003.

References

[1] Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Rogers J, et al. Ancient hybridizations among the ancestral genomes of bread wheat. Science 2014;345:1250092.
[2] Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. Plant Mol Biol Rep 1991;9(3):208–18.
[3] AGI. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000;408:796–815.
[4] IRGSP – International Rice Genome Sequencing Project. The map-based sequence of the rice genome. Nature 2005;436:793–800.
[5] Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. Science 2009;326(5956):1112–5.
[6] Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, et al. A chromosome conformation capture ordered sequence of the barley genome. Nature 2017;544:427–33.
[7] Doležel J, Kubaláková M, Paux E, Bartoš J, Feuillet C. Chromosome-based genomics in the cereals. Chromosome Res 2007;15:51–66.
[8] Feuillet C, Eversole K. 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 2007;55:307–13.
[9] Šafář J, Bartoš J, Janda J, Bellec A, Kubaláková M, Valárik M, et al. Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. Plant J 2004;39:960–8.
[10] Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, et al. A physical map of the 1-gigabase bread wheat chromosome 3B. Science 2008;322:101–4.
[11] Šafář J, Šimková H, Kubaláková M, Čihalíková J, Suchánková P, Bartoš J, et al. Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenet Genome Res 2010;129:211–23.
[12] Breen J, Wicker T, Shatalina M, Frenkel Z, Bertin I, Philippe R, et al. A physical map of the short arm of wheat chromosome 1A. PLoS One 2013;8(11):e80272.
[13] Lucas SJ, Akpinar BA, Kantar M, Weinstein Z, Aydınoğlu F, Šafář J, et al. Physical mapping integrated with syntenic analysis to characterize the gene space of the long arm of wheat chromosome 1A. PLoS One 2013;8(4):e59542.
[14] Raats D, Frenkel Z, Krugman T, Dodek I, Sela H, Šimková H, et al. The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. Genome Biol 2013;14:R138.
[15] Philippe R, Paux E, Bertin I, Sourdille P, Choulet F, Laugier Ch, et al. A high density physical map of chromosome 1BL supports evolutionary studies, map-based clonig and sequencing in wheat. Genome Biol 2013;14:R64.
[16] Poursarebani N, Nussbaumer T, Šimková H, Šafář J, Witsenboer H, van Oeveren J, et al. Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A. Plant J 2014;79:334–47.
[17] Kobayashi F, Wu J, Kanamori H, Tanaka T, Katagiri S, Karasawa W, et al. A high-resolution physical map integrating an anchored chromosome with the BAC physical maps of wheat chromosome 6B. BMC Genom 2015;16:595.
[18] Akpinar BA, Magni F, Yuce M, Lucas SJ, Šimková H, Šafář J, et al. The physical map of wheat chromosome 5DS revealed gene duplications and small rearrangements. BMC Genom 2015;16:453.
[19] Barabaschi D, Magni F, Volante A, Gadaleta A, Šimková H, Scalabrin S, et al. Physical mapping of bread wheat chromosome 5A: an integrated approach. Plant Genome 2015;8:1–24.
[20] Holušová K, Vrána J, Šafář J, Šimková H, Balcárková B, Frenkel Z, et al. Physical map of the short arm of bread wheat chromosome 3D. Plant Genome 2017;10:1–11.
[21] Krattinger S, Wicker T, Keller B. Map-based cloning of genes in *Triticeae* (wheat and barley). In: Feuillet C, Muehlbauer GJ, editors. Genetics and genomics in the Triticeae. New York: Springer; 2009. p. 337–58.
[22] Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, et al. Structural and functional partitioning of bread wheat chromosome 3B. Science 2014;345(6194):1249721.
[23] Avni R, Nave M, Barad O, Naruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 2017;357(6346):93–7.
[24] Schneider LM, Adamski NM, Christensen CE, Stuart DB, Vatrin S, Hansson M, et al. The *Cer-cqu* gene cluster determines three key players in a β-diketone synthase polyketide pathway synthesizing aliphatics in epicutilar waxes. J Exp Bot 2016;67(9):2715–30.
[25] Roselli S, Olry A, Vautrin S, Coriton O, Ritchie D, Galati G, et al. A bacterial artificial chromosome (BAC) genomic approach reveals partial clustering of the furanocoumarin pathway genes in parsnip. Plant J 2017;89(6):1119–32.
[26] Röder MS, Huang XQ, Börner A. Fine mapping of the region on wheat chromosome 7D controlling grain weight. Funct Integr Genomics 2008;8:79–86.
[27] Hirao K, Nishijima R, Sakaguchi K, Takumi S. Fine mapping of Hch1, the causal D-genome gene for hybrid chlorosis in interspecific crosses between tetraploid wheat and *Aegilops tauschii*. Genes Genet Syst 2015;90:283–91.
[28] Reddy INBL, Chandrasekhar K, Zewdu Y, Dinoor A, Keller B, Ben-David R. Identification and genetic mapping of PmAF7DS a powdery mildew resistance gene in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 2016;129(6):1127–37.
[29] Staňková H, Valárik M, Lapitan N, Berkman PJ, Batley J, Edwards D, et al. Chromosomal genomics facilitates fine mapping of a Russian wheat aphid resistance gene. Theor Appl Genet 2015;128(7):1373–83.
[30] Šimková H, Šafář J, Kubaláková M, Suchánková P, Čihalíková J, Robert-Quatre H, et al. BAC libraries from wheat chromosome 7D: efficient tool for positional cloning of aphid resistance genes. J Biomed Biotechnol 2011:302543.
[31] Luo MC, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, et al. A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. PNAS 2013;110:7940–5.
[32] Staňková H, Hastie A, Chan S, Vrána J, Tulpová Z, Kubaláková M, et al. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembling in complex plant genomes. Plant Biotechnol J 2016;14:1523–31.
[33] Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buel CR, et al. High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labelling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 2003;82:378–89.
[34] Soderlund C, Humphray S, Dunham A, French L. Contig built with fingerprints, markers and FPC V4.7. Genome Res 2000;10:1772–87.
[35] Frenkel Z, Paux E, Mester D, Feuillet C, Korol A. LTC a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. BMC Bioinf 2010;11:584–601.
[36] Visendi P, Berkman PJ, Hayashi S, Golicz AA, Bayer PE, Ruperao P, et al. An efficient approach to BAC based assembly of complex genomes. Plant Methods 2016;12:2.
[37] Rimbert H, Darrier B, Navarro J, Kitt J, Choulet F, Leveugle M, et al. High throughput SNP discovery and genotyping in hexaploid wheat. PLoS One 2017;13(1):e0186329.
[38] Tiwari VK, Heesacker A, Riera-Lizarazu O, Gunn H, Wang SC, Wang Y, et al. A whole-genome radiation hybrid mapping resource of hexaploid wheat. Plant J 2016;86(2):195–207.
[39] Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol 2014;12(6):787–96.
[40] IWGSC. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 2014;345:1251788.
[41] Bayer I, Milne I, Stephen G, Shaw P, Cardle L, Wright F, et al. Comparative visualization of genetic and physical maps with Strudel. Bioinformatics 2011;27(9):1307–8.
[42] Mayer KF, Taudien S, Martis M, Šimková H, Suchánková P, Gundlach H, et al. Gene content and virtual gene order of barely chromosome 1H. Plant Physiol 2009;151:496–505.
[43] Balcárková B, Frenkel Z, Škopová M, Abrouk M, Kumar A, Chao SM, et al. A high resolution radiation hybrid map of wheat chromosome 4A. Plant Sci 2017;7:2063.
[44] Kumar A, Simons K, Iqbal MJ, Michalak de Jimenéz M, Bassi FM, Ghavami F, et al. Physical mapping resources for large plant genomes: radiation hybrids for wheat D-genome progenitor *Aegilops tauschii*. BMC Genom 2012;13:597.
[45] Aliyeva-Schnorr L, Beier S, Karafiatova M, Schmutzer T, Scholz U, Doležel J, et al. Cytogenetic mapping with centromeric bacterial artificial chromosomes contigs shows that this recombination-poor region comprises than half of barley chromosome 3H. Plant J 2015;84(2):285–394.
[46] Thind AK, Wicker T, Šimková H, Fossati D, Moullet O, Brabant C, et al. Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. Nat Biotechnol 2017;35:793–6.

Glossary

BAC-based physical map: Constructed from overlapping clones of large-insert libraries, maintained in bacterial-artificial-chromosome (BAC) vector. The overlaps are identified by pairwise comparison of fingerprints generated from particular BAC clones.

Bionano genome (BNG) map: Also called optical map. Physical map of short sequence motifs recognition sites of a nicking enzyme – along hundreds to thousands kilobases-long stretches of DNA. It is created by imaging labelled DNA molecules on nanochannel arrays.