

TECHNICAL ADVANCE

Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A

Naser Poursarebani^{1,†,*}, Thomas Nussbaumer^{2,†}, Hana Šimková³, Jan Šafář³, Hanneke Witsenboer⁴, Jan van Oeveren⁴, Jaroslav Doležel³, Klaus F.X. Mayer², Nils Stein¹ and Thorsten Schnurbusch^{1,*}

¹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, D-06466 Stadt Seeland (OT) Gatersleben, Germany,

²MIPS/IBIS German Research Center for Environmental Health, D-85764 Neuherberg, Germany,

³Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, CZ-78371 Olomouc, Czech Republic, and

⁴Keygene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands

Received 3 March 2014; revised 25 April 2014; accepted 1 May 2014; published online 9 May 2014.

*For correspondence (e-mails thor@ipk-gatersleben.de and poursarebani@ipk-gatersleben.de).

International Wheat Genome Sequencing Consortium (IWGSC)—<http://www.wheatgenome.org/>; eversole@eversoleassociates.com).

†These authors contributed equally to this work.

SUMMARY

Bread wheat (*Triticum aestivum* L.) is the most important staple food crop for 35% of the world's population. International efforts are underway to facilitate an increase in wheat production, of which the International Wheat Genome Sequencing Consortium (IWGSC) plays an important role. As part of this effort, we have developed a sequence-based physical map of wheat chromosome 6A using whole-genome profiling (WGP™). The bacterial artificial chromosome (BAC) contig assembly tools FINGERPRINTED CONTIG (FPC) and LINEAR TOPOLOGICAL CONTIG (LTC) were used and their contig assemblies were compared. A detailed investigation of the contigs structure revealed that LTC created a highly robust assembly compared with those formed by FPC. The LTC assemblies contained 1217 contigs for the short arm and 1113 contigs for the long arm, with an L₅₀ of 1 Mb. To facilitate *in silico* anchoring, WGP™ tags underlying BAC contigs were extended by wheat and wheat progenitor genome sequence information. Sequence data were used for *in silico* anchoring against genetic markers with known sequences, of which almost 79% of the physical map could be anchored. Moreover, the assigned sequence information led to the 'decoration' of the respective physical map with 3359 anchored genes. Thus, this robust and genetically anchored physical map will serve as a framework for the sequencing of wheat chromosome 6A, and is of immediate use for map-based isolation of agronomically important genes/quantitative trait loci located on this chromosome.

Keywords: bread wheat chromosome 6A, whole-genome profiling, LINEAR TOPOLOGICAL CONTIGS, anchored physical map, bacterial artificial chromosome contigs, technical advance.

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) has been a constant staple food and major crop that has provided energy and protein for humankind for millennia. Today, it represents approximately 20% of all calories consumed by humans (<http://www.fao.org>). As the world population grows and the climate continuously changes, future generations may challenge the current food supplies, which could augment

the demand for wheat (Ortiz *et al.*, 2008; Sommer *et al.*, 2013). Therefore, accelerating wheat breeding and production by understanding the molecular basis of phenotypic variation and exploiting genetic diversity to improve performance are crucial. In this respect, the sequencing of the entire bread wheat genome has been considered to be a critical step for achieving these goals (Eversole *et al.*, 2009).

Accessing complete chromosomal sequences and the gene repertoire of hexaploid wheat ($2n = 6x = 42$) is quickly becoming a necessary but daunting task because of its large genome size (~17 Gb/1C), which has resulted from successive hybridization events of three diploid grasses with large and structurally similar genomes (A, B and D), populated with more than 80% of repetitive elements (Flavell *et al.*, 1977). To overcome these challenges, a strategy has been identified that aims to break down genomic analyses into manageable-sized tasks (i.e. individual chromosomes and/or chromosome arms). This strategy relies on flow-sorting of individual chromosome arms used for construction of BAC libraries and also for next-generation sequencing (Doležel *et al.*, 2007, 2014). Moreover, together with the advent of physical map building using whole-genome profiling (WGP™; van Oeveren *et al.*, 2011), chromosome flow cytometry represents an important technological development that provides easier access to large and polyploid genomes (Doležel *et al.*, 2014).

Whole-genome profiling (WGP™) produces sequence tags at terminal ends of enzymatic restriction fragments from individual bacterial artificial chromosome (BAC) clones using a short-read 'Next Generation' or 'High Throughput' sequencing device (van Oeveren *et al.*, 2011). Identification of BAC overlaps by pairwise WGP™ tag comparisons allows for the assembly of BAC clones into contigs using the FINGERPRINTED CONTIGS (FPC; Soderlund *et al.*, 1997) or LINEAR TOPOLOGICAL CONTIG (LTC) programs (Frenkel *et al.*, 2010). WGP™ is considered more robust, less laborious and more efficient at building physical contigs than conventional high-information content fingerprinting (HICF; Luo *et al.*, 2003; Philippe *et al.*, 2012). The WGP™ approach has been successfully used on a number of genomes (van Oeveren *et al.*, 2011; TTGC, 2012; Sierro *et al.*, 2013), and its application has also been demonstrated for BAC clone contig formation originating from only a small fraction of wheat chromosome 3B (Philippe *et al.*, 2012). Moreover, WGP™ short sequence tags connected to the sequence contigs obtained from whole-chromosome shotgun sequencing (WCS; or chromosome shotgun sequencing, CSS) may further expand the possibilities for using WGP™.

Substantially longer WCS contigs connected to BAC-associated WGP™ tags facilitate sequence homology searches against respective genetic markers with known sequences, thereby providing an *in silico*-anchored physical map. A physical map, aided by shotgun sequencing, can provide clear insight into the physically positioned and ordered gene repertoire before complete genomic sequences of the wheat chromosomes become available. Without a link to physical maps, individual wheat chromosome shotgun sequence data sets were previously used to estimate virtual gene order, composition and evolutionary chromosomal rearrangements (Vitulo *et al.*, 2011; Hernandez *et al.*, 2012; Akhunov *et al.*, 2013; Tanaka *et al.*, 2014).

Such studies were mainly performed by an analysis of extensive long-range conserved synteny with reference grass genomes [*Oryza sativa* (rice), *Sorghum bicolor* and *Brachypodium distachyon*]. Similarly, a fivefold genome coverage-based sequence assembly of the entire bread wheat genome shotgun sequence was previously completed and analyzed (Brenchley *et al.*, 2012) using a similar comparative genomics-based approach that involved comparison with sequences of diploid ancestral and progenitor genomes. Although these reports represent significant achievements in wheat genome biology, a more accurate wheat reference genome that avoids assumptions made by comparative genomics requires the establishment of genetically anchored physical maps.

Genes assigned to a physical map significantly facilitate positional gene cloning efforts and detection of regulatory elements. Herein, we constructed a physical map of wheat chromosome 6A linked to the annotated gene sequence information. We report a high-resolution gene map of chromosome 6A based on DNA sequences obtained from flow-sorted chromosome arms and by using the WGP™-based physical map approach. BAC assembly was performed using LTC software, and the assembly robustness was compared with FPC. WCS contigs from hexaploid wheat chromosome 6A [International Wheat Genome Sequencing Consortium (IWGSC), <http://www.wheatgenome.org>], together with available sequences from wheat ancestral diploid genomes (Jia *et al.*, 2013; Ling *et al.*, 2013), were aligned to the 6A physical map. Overall, we describe the development of a powerful resource for 6A that facilitates the study of its assigned genes and quantitative trait loci (QTLs).

RESULTS

BAC libraries of the 6A chromosome arms

Bacterial artificial chromosome libraries were constructed separately from short (6AS) and long (6AL) arms of chromosome 6A, which were purified separately by flow-cytometric sorting from two telosomic lines of wheat in which the arms originating from 6A of cv. Chinese Spring are stably maintained as telocentric chromosomes. Analysis of random samples of sorted chromosome fractions by fluorescence *in situ* hybridization (FISH) revealed an average purity of 89 and 86% for 6AS and 6AL, respectively. A total of 7.15 million and 6 million 6AS and 6AL arms were collected, which represented 4.91 and 4.53 µg DNA, respectively. The DNA was used to construct chromosome arm-specific BAC libraries TaaCsp6AShA (6AS; 49 152 clones, average insert size 125 Kb; *HindIII*) and TaaCsp6ALhA (6AL; 55 296 clones, average insert size 123 Kb; *HindIII*). Considering the molecular size of 336 Mbp for 6AS and 369 Mbp for 6AL (Šafář *et al.*, 2010), as well as contamination from other

Table 1 Chromosome 6A arm-specific BAC libraries used to construct the physical maps

No. of BAC clone	Short arm (6AS; 336 Mb) ^a	Long arm (6AL; 369 Mb) ^a	Library coverage
Complete BAC library	49 152	55 296	~16×
For WGP TM	22 656	24 576	~8×
As WGP TM output	19 289	18 660	~7×
WGP TM output used for BAC assembly ^b	18 820	17 309	~7× (6AS) & ~5× (6AL)

WGPTM, whole-genome profiling.

^aSafár *et al.* (2010).

^bBACs containing <6 or >68 tags were not entered into the physical map assembly pipeline.

chromosomes, the libraries represent ~16 equivalents of the 6A arms (Table 1).

BAC contig assembly of the 6A chromosome arms

Whole-genome profiling (WGPTM) data were produced for both chromosome arms 6AS and 6AL (Table 2; Appendix S1). Only BAC clones containing 6–68 tags entered the BAC assembly pipelines. A total of 18 820 BAC clones and 109 570 unique WGPTM tags were used as input for BAC assembly of 6AS (Table 2). For 6AL, 17 309 BAC clones containing 108 700 unique WGPTM tags were used to build

Table 2 Overview of general whole-genome profiling (WGPTM) parameters and sequence data processing

Parameter	6AS	6AL
Estimated size of chromosome	336 Mbp	369 Mbp
WGP TM tag length, incl. restriction site ^a	50 nt	50 nt
No. of BACs tested	22 656	24 576
Chromosome arm equivalent BACs tested	8.4	7.0
No. of high-quality reads produced (M)	99.8	165.4
No. of deconvolvable reads (M)	52.2	78.2
Percentage of deconvolvable reads (%)	52.30	47.30
No. of tagged BACs	19 289	18 660
No. of tagged BACs used in assembly	18 820	17 309
No. of unique WGP TM tags	109 611	108 811
No. of unique WGP TM tags used in assembly	109 570	108 700
Percentage of tagged BACs (%)	85	75.9
Average no. of WGP TM tags/BAC	29.7	28.6
Average no. of reads/tag	61.4	111.1

BAC, bacterial artificial chromosome.

^aThe enzyme combination applied was *HindIII/MseI* for both libraries.

the physical map (Table 2). This delivered an average of 29.7 WGPTM tags per BAC for 6AS and 28.6 WGPTM tags per BAC for 6AL, by considering all BACs for a given WGPTM tag (Table 2). We used LTC and FPC tools for assembly. LTC has been shown to outperform FPC because it can build longer, better ordered and more robust contigs, compared with FPC, employing HICF-based BAC fingerprints (Breen *et al.*, 2013; Lucas *et al.*, 2013; Philippe *et al.*, 2013; Raats *et al.*, 2013). The same set of BACs and WGPTM tags were used for both FPC and LTC analyses.

Automated LTC delivered 1217 contigs and 3136 singletons for 6AS and 1113 contigs and 2581 singletons for 6AL (Table 3; Appendices S2 and S3). All contigs contained at least two BAC clones. The 6AS average contig size was estimated to be 0.429 Mb with an L₅₀ contig size of 1 Mb, whereas the 6AL physical map had an average of 0.488 Mb and had an L₅₀ contig size of 0.945 Mb (Table 3). All contigs were validated for linear topology and represented as the net width value. Contigs with a width >1, which indicates the presence of clone(s) having no significant overlaps with clones from the selected minimal tiling path (MTP) of the contig, were considered questionable because their topological network representation contradicted the linear chromosomal structure. Almost 95% of the cumulative 6AS and 6AL contigs were initially linear, whereas the rest, having a width ≥2, were checked manually for linearity. In the latter group, nonlinearity was presumably caused by missing WGPTM tags in a particular single BAC. Thus, these contigs were kept as such and further break down was not performed. Using LTC, a total of 5139 (6AS) and 5621 (6AL) clones were selected for the MTP (Appendices S2 and S3). Another physical map was made by FPC (1e⁻¹¹ final cut-off; Appendix S4), in which a total of 640 and 620 contigs for 6AS and 6AL were formed, respectively. A total of 5045 and 3560 BACs were sorted out as singletons for 6AS and 6AL, respectively (Table S1; Appendix S5).

BAC WGPTM tags facilitated *in silico* genetic anchoring of the 6A physical map

To facilitate efficient *in silico* anchoring, the short WGPTM tags underlying 6A physical contigs were extended by connecting them to the available wheat sequence information in three ways (Figures S1 and S2; Appendix S6). Firstly, WGPTM tags were connected to the available shotgun sequence contigs (obtained from IWGSC) of the bread wheat 6A chromosome arms (Appendix S6). This resulted in assigning 165-Mb of sequences (out of 433.6 Mb) from WCS contigs to the wheat 6A physical map. Secondly, the previously assigned WCS contigs were connected to the progenitor genome sequences of *Triticum urartu* (an A-genome progenitor; Ling *et al.*, 2013), which resulted in assignments of 375 Mb of sequences along the physical map. Thirdly, in order to enrich the physical map with further sequence information, an additional layer of

Table 3 Chromosome 6A arm-separated assemblies via LINEAR TOPOLOGICAL CONTIGS (LTC), with at least two BAC clones

	6AS			6AL		
	All	LTC-specific contigs ^a	Contigs common to FPC ($1e^{-11}$) ^b	All	LTC-specific contigs ^a	Contigs common to FPC ($1e^{-11}$) ^b
Total no. of contigs (≥ 2 clones)	1217	539	678	1113	368	745
Singleton	3136	–	–	2581	–	–
Minimum Kb	42.3	42.3	98.7	47	47	65.8
Maximum Kb	9531.6	470	9531.6	3139.6	521.7	3139.6
Average Kb	428.526	131.033	665.029	487.887	123.336	667.961
L ₅₀ Kb	1090.4	136.3	1386.5	944.7	126.9	1005.8
N ₅₀	1106	351	595	921	249	577

FPC, FINGERPRINTED CONTIGS.

^aSmall contigs (average bacterial artificial chromosomes, BACs/contig = 2.6 for 6AS and 2.8 for 6AL) exclusively made by LTC. The average tag per BAC for the underlying clones was only 15 tags, equal to the value observed for singletons.

^bThe average number of tags per BAC for the underlying clones was 34 tags.

sequences was added to sequences already assigned to 6A WCS and *T. urartu* using *Aegilops tauschii* (a D-genome progenitor; Jia *et al.*, 2013; Figures S1 and S2; Appendix S6). This provided a total of 157 Mb of sequences from *Ae. tauschii* assigned to the 6A physical map. Considering the more distant relation of *Ae. tauschii* to bread wheat chromosome 6A, the effect of the *Ae. tauschii* sequence inclusion to our anchoring analysis was checked. We found that the incorporation of this sequence data set had a limited contribution to the overall anchoring of the physical map, with no negative effect on the accuracy of the genetic anchoring (Appendix S7). Altogether, the average cumulative sequence information per physical contig increased from 2144 nt using WGP™ tags to 11 067 nt, simply by adding the available wheat sequence resources (Figure S2; Appendix S8). This number of sequences directly assigned to the physical map enabled sequence homology searches against all available genetic markers with known sequences, and thus provided the basis for the integration of genetic and physical maps *in silico* (Appendix S6). Genetic anchoring was performed for both LTC and FPC physical map assemblies; however, only the result of genetic anchoring performed for LTC assembly is shown here. This assembly was selected as the final 6A assembly because we observed that LTC provided a more robust assembly than FPC (see the following section).

First, LTC contigs were anchored to two highly dense wheat genetic maps (Poland *et al.*, 2012; Cavanagh *et al.*, 2013), which allowed us to genetically anchor 298 6AS and 384 6AL LTC-derived contigs. This genetically anchored a total of 661 Mb out of the 1048 Mb, which represents the cumulative contig map length (Figure 1; Appendix S9). By considering only genetically anchored LTC-built contigs, we were able to genetically anchor 132 Mb of WCS contigs, 303 Mb of sequence data from *T. urartu* and 129 Mb from *Ae. tauschii* to the 6A genetic maps. The remaining

unanchored 6A physical contigs were subjected to a second round of anchoring using the publicly available barley genomic resource. This was performed to provide researchers with an additional layer of anchoring information for the respective physical map. The barley-based anchoring was kept independent (Appendix S10) of the wheat-based anchoring. Therefore, in our analysis, we used 15 719 high-confidence barley genes from the barley genome (IBSC, 2012), together with 723 499 anchored barley WCS contigs from barley POPSEQ data (Mascher *et al.*, 2013). Using high-confidence barley genes we were only able to anchor 26 additional 6A physical contigs (8 Mb), whereas 98 physical contigs from 6A (37 Mb) were exclusively anchored via barley population sequencing (POPS-SEQ) data (Appendix S10). Overall, 831 Mb (i.e. 79% of 6A physical contigs) were genetically anchored (Figure 1).

The large portion of the physical map anchored to the respective wheat genetic maps allowed for the analysis of recombination frequencies along the entire wheat 6A chromosome. For this purpose, an integrated wheat genetic map was constructed from the two aforementioned wheat genetic maps (Appendix S6). We then calculated the physical distance per 10-cM bins (Appendix S11). As expected, for plants with large genomes [i.e. *Zea mays* (maize; Anderson *et al.*, 2003), *Hordeum vulgare* (barley; Kunzel *et al.*, 2000) and wheat (Lukaszewski and Curtis, 1993)], we were able to reconfirm that recombination greatly increased from the centromere towards the telomeres (Figure 2). These detailed estimates related to recombination frequencies along 6A will be most useful for future map-based cloning and gene identification projects.

Comparison between LTC and FPC assembly at different cut-off values

To date, physical maps of four arms of wheat chromosomes, including 1AS, 1AL, 1BS, 1BL and 3B, have been

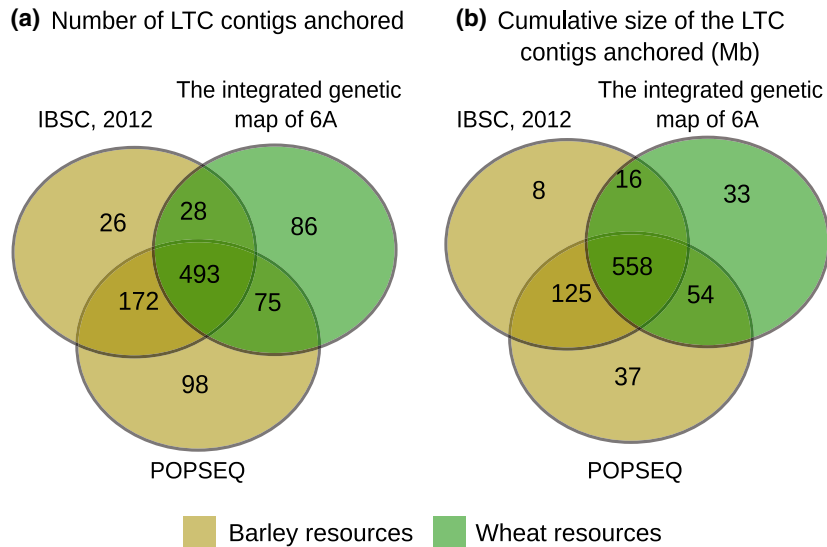


Figure 1. Contribution of different genetic map resources to the *in silico*-based anchoring of the wheat chromosome 6A physical map. (a) The number of LTC contigs that were anchored using one or a combination of different map resources: e.g. 493 LTC contigs were anchored by all three resources. (b) The corresponding cumulative size of the LTC contigs that were anchored using one or a combination of different map resources: e.g. 558 Mb was anchored by all three resources. The integrated genetic map of 6A (constructed in the current study; Appendix S6) refers to the combined genetic map derived from two highly dense wheat genetic maps, as previously described (Poland *et al.*, 2012; Cavanagh *et al.*, 2013). IBSC (2012) and POPSEQ (Mascher *et al.*, 2013) refer to the publicly available *Horedum vulgare* (barley) resources, including data sets from the International Barley Genome Sequencing Consortium (IBSC, 2012) and population sequencing (Mascher *et al.*, 2013), respectively.

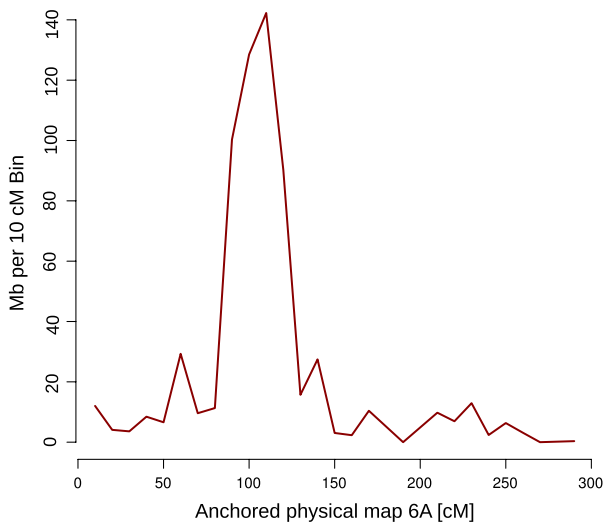


Figure 2. Estimated recombination frequency along wheat chromosome 6A. The underlying genetic map used for 6A genetic anchoring was divided into 26 bins, each of 10 cM in size. The physical map anchored to each bin measured the recombination pattern (Mb/cM) along the 6A chromosome. The genetic map is derived from integrating two wheat genetic maps (Appendix S6) developed previously (Poland *et al.*, 2012; Cavanagh *et al.*, 2013).

reported (Paux *et al.*, 2008; Breen *et al.*, 2013; Lucas *et al.*, 2013; Philippe *et al.*, 2013; Raats *et al.*, 2013). With the exception of chromosome 3B, both LTC and FPC tools have been used to assemble wheat chromosome physical maps. A comparison was performed between the two tools for

wheat chromosome arm 1AL and showed that the LTC assembly provided significantly higher accuracy (Breen *et al.*, 2013); however, the previous study only described the advantages of LTC by comparing the LTC assembly with a single FPC assembly constructed at $1e^{-45}$ and by using HICF data (and not WGP™) as the input (Breen *et al.*, 2013). In our study, we compared contigs obtained by LTC as the reference assembly with individual FPC assemblies (Table S1) generated at different stringencies to assess the differences between the LTC and the FPC assemblies employing WGP™. This allowed for the identification of the most robust physical map for 6A. Our comparison distributed an assortment of LTC-made contigs into five groups (Figure 3; Tables S2 and S3; Appendix S12). The LTC ≥ 2 FPC group (LTC-derived contigs for which BACs were assembled into two or more different contigs via FPC; cases of conflict; Figure 4) was considered the most suitable group for comparing the two assembly platforms.

The number of such LTC-built contigs (LTC ≥ 2 FPC) was found to decrease with declining FPC assembly cut-off values. This was explained by end-to-end merging during FPC assembly. To check whether contig merges were accurately formed via FPC while decreasing assembly cut-off, all LTC ≥ 2 FPC cases were visualized and inspected (Figure 4). This inspection was only performed for cases of conflict between LTC and FPC assemblies at a $1e^{-11}$ FPC cut-off for each chromosome arm (Figure 4; Appendices S13 and S14). This cut-off was selected because it was closest to the LTC cut-off value ($1e^{-10}$), and thus was considered to

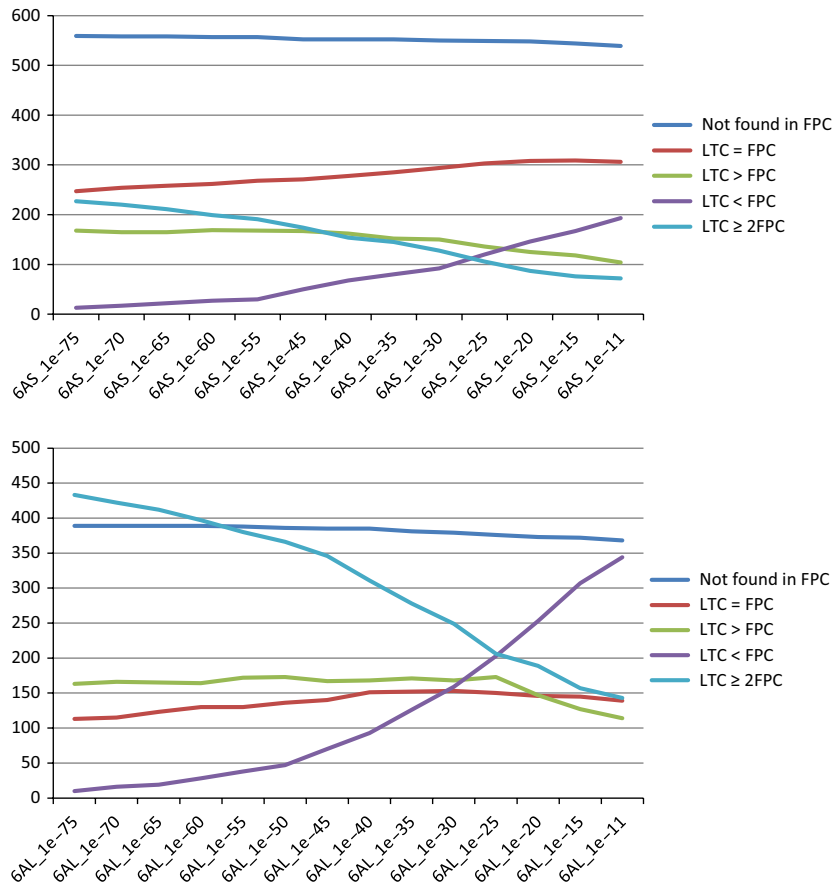


Figure 3. Comparison of LINEAR TOPOLOGICAL CONTIG (LTC)-derived physical contigs with FINGERPRINTED CONTIGS (FPC) at different stringencies. Different classes of LTC-derived contigs were identified when compared with FPC assemblies that include the first class, which are contigs that were exclusively made by LTC (not found in FPC). The number of such LTC-specific contigs was relatively constant for both of the 6A arms when compared with FPC assemblies at various stringencies. The second class contained contigs that were identical between assemblies made by both tools (LTC = FPC, i.e. BAC composition and order were the same in corresponding LTC and FPC contigs). This class showed a slight increase when the FPC stringency was lowered. The third class includes LTC > FPC (i.e. LTC contigs longer than FPC contigs), which refers to contigs with the same backbone, whereas more BACs were added to the end of corresponding contigs via LTC. This class also showed a slight decrease in number by lowering the stringency in FPC. The two remaining classes include LTC < FPC (i.e. LTC contigs that are shorter than FPC contigs) and LTC ≥ 2FPC (i.e. LTC-made contigs, the BACs of which were assembled into two or more different contigs via FPC). Both classes showed higher differences in number across the FPC assemblies that are mainly explainable by end-to-end merging during FPC assembly. For example, a decrease in the number of LTC ≥ 2FPC from 1e⁻³⁰ to 1e⁻²⁵ is likely to have resulted from the merging of the two corresponding FPC contigs at 1e⁻²⁵. Therefore, contigs of 1LTC = 2FPC at a higher level of FPC stringency later became 1LTC < FPC (or 1LTC = 1FPC) at a lower level of stringency. In both chromosome arms, identical classes show a relatively similar trend of variation across FPC assemblies, although they may contain different numbers of contigs in each of the arms.

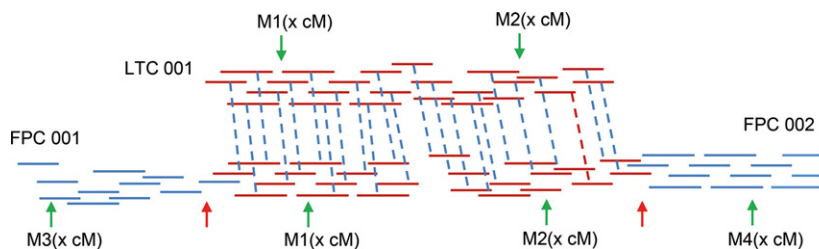


Figure 4. Schematic representation of a case of conflict (LTC ≥ 2FPC) in FINGERPRINTED CONTIGS (FPC) versus LINEAR TOPOLOGICAL CONTIGS (LTC), as a reference. Anchored genetic markers (M) were used to confirm the corresponding bacterial artificial chromosome (BAC) contig structure. Three situations were observed. Status 1: concordant M₁ and M₂ genetic positions, with the genetic position for the corresponding marker in at least one FPC contradictory; LTC 001 correctly assembled. Status 2: non-concordant M₁ and M₂ genetic positions, with the genetic position of the corresponding marker in both FPC contigs reconciled; LTC 001 not correctly assembled. Status 3: concordant M₁ and M₂ genetic positions, with the genetic position for the corresponding marker in both FPC contigs reconciled; all contigs correctly assembled. Red/blue bars, BAC clones underlying BAC contigs; dashed bars, connect shared BACs; green arrowheads, genetic marker position; red arrowheads, region of low BAC coverage.

Table 4 Cases of conflict ($LTC \geq 2FPC$) for which marker data allowed further confirmation of contig structure

Chromosome arm	6AS	6AL
No. of conflicts	72	143
No. of conflicts for which informative markers were available ^a	22	26
Engaged FPC contig(s) is chimeric	17	20
Engaged LTC contig is chimeric	4	4
The LTC contig corresponds to the ends of the respective FPC contigs ^b	1	2

FPC, FINGERPRINTED CONTIGS; LTC, LINEAR TOPOLOGICAL CONTIGS.

^aGenetic markers from the same map were anchored to different parts of the same contig.

^bNeither the LTC contig nor the respective FPC contigs could be flagged as chimerical on the basis of anchored genetic markers.

represent the most comparable cut-off value. In the case of 6AS, 72 conflicts were found for which a particular set of overlapping BACs were always shared between LTC and corresponding FPC counterparts. These BACs were also similarly ordered along the respective contigs (Figure 4, red-colored BACs; Table 4).

We then analyzed whether the conflict could be resolved by consulting genetic markers anchored to corresponding BAC clones. For 22 of the 72 6AS cases of conflict, the corresponding affected LTC- and FPC-made contigs were provided with a sufficient number of anchored genetic markers (informative markers). Such markers allowed us to ascertain whether the different parts of affected contigs had been correctly overlapped/merged and, if so, whether the conflict could be resolved. In the majority of cases (17 out of 22), genetic markers revealed a false assembly of different FPC-derived contig parts (e.g. a chimerical FPC contig was built; Figure 4, status 1). For the remaining five cases, either the LTC-built contigs were chimeric (four cases; Figure 4, status 2) or the respective BACs were correctly assembled by both FPC and LTC tools (one case), and thus corresponding conflicts were resolved (Figure 4, status 3). We achieved a similar result for 143 conflicts identified for 6AL (Table 4). Therefore, we concluded that the ability of FPC to form longer and consequently lower numbers of contigs, while decreasing the assembly cut-off value, could potentially produce higher numbers of chimerical contigs as a result of false BAC contig end-to-end merges.

We also observed a highly consistent number of LTC-specific contigs (small contigs, average BAC/contig = 2.6–2.9) compared with FPC assemblies. These LTC-derived contigs (539 for 6AS and 368 for 6AL, compared with FPC at $1e^{-11}$) were composed of BACs that were left out as singletons in the initial FPC build at $1e^{-75}$ (Table S1), where BACs were assembled into contigs only if they shared >70% of WGP™ tags (Paux *et al.*, 2008). BAC clones containing fewer tags (an average of 15 tags/BAC, compared with 34 tags/BAC for the rest of the assembly) were incorporated into contigs via LTC because the initial assembly was performed at

a lower stringency ($1e^{-2}$ cut-off). Therefore, such LTC-specific small contigs resulted in a higher number of LTC-made contigs, and consequently had higher chromosomal arm coverage, as they most likely cover the same regions as larger contigs. Although LTC had an artificially generated larger number of small contigs, the robustness of BAC order and overlap identification for the remaining contigs was significantly higher in LTC versus FPC assemblies (see cases of conflict above). For this reason, we considered the entire LTC-based assembly to be a more reliable physical map for subsequent analyses, including synteny analysis, MTP selection and future BAC-based sequencing of wheat chromosome 6A.

Efficacy of FPC at improving LTC-made assembly

To test whether gaps between LTC-derived contigs could be closed via FPC-assembled BACs, LTC-derived contigs were aligned against FPC assemblies using a $1e^{-50}$ cut-off as a reference. This was in contrast to the aforementioned comparisons, where LTC was used as a reference. We selected this stringently formed FPC assembly to ensure the robustness of overlap among BACs of the corresponding contigs. Using this approach, we identified 25 FPC-built contigs for 6AS and 45 FPC-built contigs for 6AL in which BAC clones were represented in more than two LTC-made contigs (Table 5). These LTC-derived contigs were thus considered potentially mergable or potential scaffolds because the gap between two LTC-made contigs could be bridged using a robustly formed FPC-built contig. All of these scaffolds were depicted as individual images according to BAC position to enable a visual inspection of their structure (Figure S3). Only scaffolds for which the corresponding FPC-made contig was composed of BACs from the ends of two different LTC-made contigs were flagged as potentially true scaffolds (10 for 6AS; 17 for 6AL); however, those FPC-derived contigs that, for example, contained BACs from the end of

Table 5 Scaffolding LINEAR TOPOLOGICAL CONTIGS (LTC) using FINGERPRINTED CONTIGS (FPC) assembled at a cut-off of $1e^{-50}$, for which the same set of bacterial artificial chromosomes (BACs) was employed

Chromosome arm	6AL	6AS
No. of potential scaffolds	45	25
Scaffolding was structurally allowed ^a	17	10
Structurally allowed scaffolds for which genetic markers were available ^b	6	4
Structurally allowed scaffolds were supported by genetic markers	3	3
Structurally allowed scaffolds were not supported by genetic markers	3	1

^aThe corresponding FPC contig was only composed of BACs from the end of two different LTC contigs, and not from the end of one LTC and the middle of another (not allowed).

^bThe corresponding LTC contigs were provided by genetic anchoring information.

one LTC-made contig and BACs from the middle of another were rejected (almost 60% for each arm; Figure S3). These rejected scaffolds most likely represent falsely assembled BACs via FPC (chimerical contigs).

Using these scaffolds we then tested whether any anchoring information could support scaffold accuracy. For 6AS, genetic markers were available for four such scaffolds, of which one was confirmed genetically; for 6AL, we found three cases out of six (Table 5). This result indicated that the efficacy of FPC assemblies (at $1e^{-50}$) at improving LTC-derived assemblies was relatively low. In general, this might indicate that the complexity of the wheat sequence restricted the ability of FPC, using WGP™ data as input, to construct robust contigs, even at the initial higher stringency (in this case $1e^{-50}$).

Scaffolding the LTC assembly using WGP™ tags and shotgun sequence contigs

Publicly available *T. urartu* sequence contigs (Tu contigs) were used to determine whether they allow for the bridging of LTC-made contigs and building of LTC scaffolds. We checked whether a single Tu contig exclusively matched terminal parts of two different LTC-derived contigs (Appendix S15). Using Tu contigs as a proxy, we identified 84 6AS and 65 6AL potential scaffolds (Table 6; Appendix S16). Those LTC-made contigs engaged in scaffolding were relatively small (average BAC/contig = 5 6AS and 17 6AL), and therefore could not have a significantly positive effect on the L_{50} of the overall BAC assembly; however, we asked whether LTC-derived contigs involved in such scaffolds were supported by genetically anchored markers to validate our approach (with a similar approach as described in Figure 4). For 6AS, we found only four of these scaffolds, of which two could be genetically validated; for 6AL, 12 scaffolds with genetic markers were provided, of which eight were genetically confirmed (Table 6). Although genetic markers were not available for structural confirma-

Table 6 Scaffolding LINEAR TOPOLOGICAL CONTIGS (LTC) using the underlying whole-genome profiling (WGP™) tag and whole-genome shotgun sequence contigs of *Triticum urartu*^a

Chromosome arm	6AL	6AS
No. of potential scaffolds ^b	65	84
No. of scaffolds for which genetic markers were available ^c	12	4
No. of genetically confirmed scaffolds	8	2
No. of not genetically confirmed scaffolds	4	2

^aThis investigation was performed using only a subset of *T. urartu* sequence info (*T. urartu* scaffolds): 90.7 Mb, with an L_{50} of 64 532 nt.

^bScaffolding was allowed if the corresponding sequence contigs hit at least three WGP™ from the ends of only two LTC contigs.

^cThe corresponding LTC contigs were provided by genetic anchoring information.

tion of all the scaffolds constructed, we kept and reported all formed scaffolds (Appendix S16) because they could potentially support and/or guide future BAC-based sequencing and sequence assembly of the respective physical contigs.

'Gene decoration' of the newly formed anchored physical map of wheat chromosome 6A

Chromosome 6A shotgun sequences (6A WCS contigs) have already been annotated for genes (IWGSC, In press), which established 5024 genes grouped into four confidence classes (HC1–HC4). Of these, 2531 genes were assigned to the HC1 category because $\geq 70\%$ of coding sequences overlapped with a reference gene in *B. distachyon*, rice or *S. bicolor*. The remaining confidence classes (HC2–HC4) had less overlap with reference genes (IWGSC, In press); however, in our analysis of the total 5024 HC1–HC4 genes, 3359 genes (1667 6AS; 1692 6AL) were assigned to the anchored portion of the 6A physical map (i.e. genetically positioned genes). This was accomplished by assigning the corresponding WCS contigs to the WGP™ tags (Appendix S6). We then calculated the gene density on this newly formed physical map along 6A by dividing the chromosome length in bins of 5 Mb (Figures 5 and S4). In more telomeric bins we found a maximum of 84 genes, whereas in the centromeric region with low recombination the number of genes decreased dramatically to less than 20 (Figure 5). Moreover, this analysis revealed a general correlation between the recombination rate pattern and distribution of genes along the chromosome (Figures 2 and 5).

Syntenic-based approach addressed 6A evolutionary relations with model grass genomes

To analyse the completeness of the anchored physical map and to gain insight into the evolutionary origin of

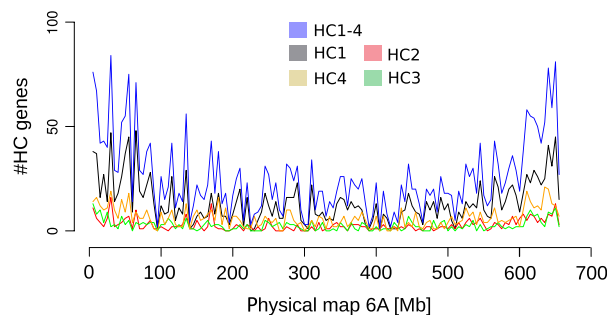


Figure 5. Estimated gene distribution along wheat chromosome 6A. Chromosome 6A gene-annotated whole chromosome survey sequencing contigs were linked to whole-genome profiling tags underlying bacterial artificial chromosome (BAC) contigs. High-confidence (HC) genes represent annotated genes, wherein HC1 are genes with $>70\%$ coding sequence–reference gene overlap in *Brachypodium distachyon*, *Oryza sativa* (rice) or *Sorghum bicolor*, whereas the other HC2–HC4 classes have a smaller overlap (IWGSC, In press).

chromosome 6A, physical contig-associated sequences were compared against reference genomes (Appendix S6). Associated sequence information, including WGP™ tags, WCS, *T. urartu* and *Ae. tauschii* contigs, were compared with coding sequences from other grass genomes of *Hordeum vulgare* (Hv), *Brachypodium distachyon* (Bd), *Oryza sativa* (Os), and *Sorghum bicolor* (Sb). Anchored sequences showed sequence homology to 2799 Hv, 2455 Bd, 2539 Os and 2465 Sb genes. Of these genes, 40.0–48.5% matched syntenic chromosomes of the three genomes, including Bd3, Os2 and Sb4. These results are

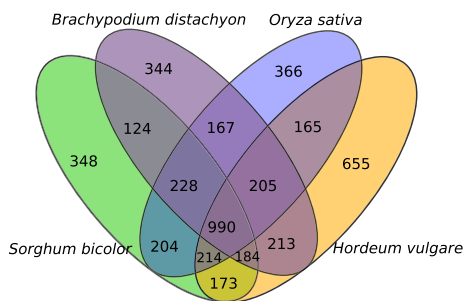


Figure 6. Venn diagram of wheat chromosome 6A physical map-associated gene distribution with significant similarity to *Brachypodium distachyon*, *Oryza sativa* (rice) and *Sorghum bicolor*. Chromosome 6A physical map-associated sequence information, including whole chromosome survey sequencing, *Triticum urartu* and *Aegilops tauschii* contigs, were compared with coding sequences from model grass genomes. In *Hordeum vulgare*, all genes including genetically anchored and unanchored genes were considered (IBSC, 2012).

comparable with those reported previously (40.2–59.7%) for syntenic genes of 6B (Tanaka *et al.*, 2014).

Figure 6 depicts the 6A gene distribution among model grass genomes. We observed that the number of wheat coding sequences shared with at least three species (1994 genes) was higher than the genes presented in two (873 genes) or a single species (1713 genes). Proteins with significant similarity (seed length of 20, identity $\geq 75\%$; <http://www.vmatch.de>) to wheat sequence information were plotted along the chromosomes (Figure 7). The average syntenic gene content was calculated as the number of matching wheat proteins in a window of 1 Mb, without overlap. In this analysis, 6A sequence information showed homology with genes/proteins located on chromosome 6H of Hv (Hv6H), Bd3, Os2 and Sb4, but not with any other chromosomes of these genomes (Figure 7), which is consistent with previous reports (Salse *et al.*, 2008; TIBI, 2010; Brenchley *et al.*, 2012; IBSC, 2012).

DISCUSSION

The current study is the German contribution to the IWGSC. This international consortium has aimed to sequence wheat chromosomes and/or chromosome arms individually, with the construction of the respective BAC-based physical maps as a necessary intermediate step. In this context, we report here the successful application of WGP™ together with the contig assembly tool LTC to robustly assemble BAC clones into 1113 contigs for the long arm and 1217 contigs for the short arm, representing the wheat chromosome 6A physical map. To date, physical

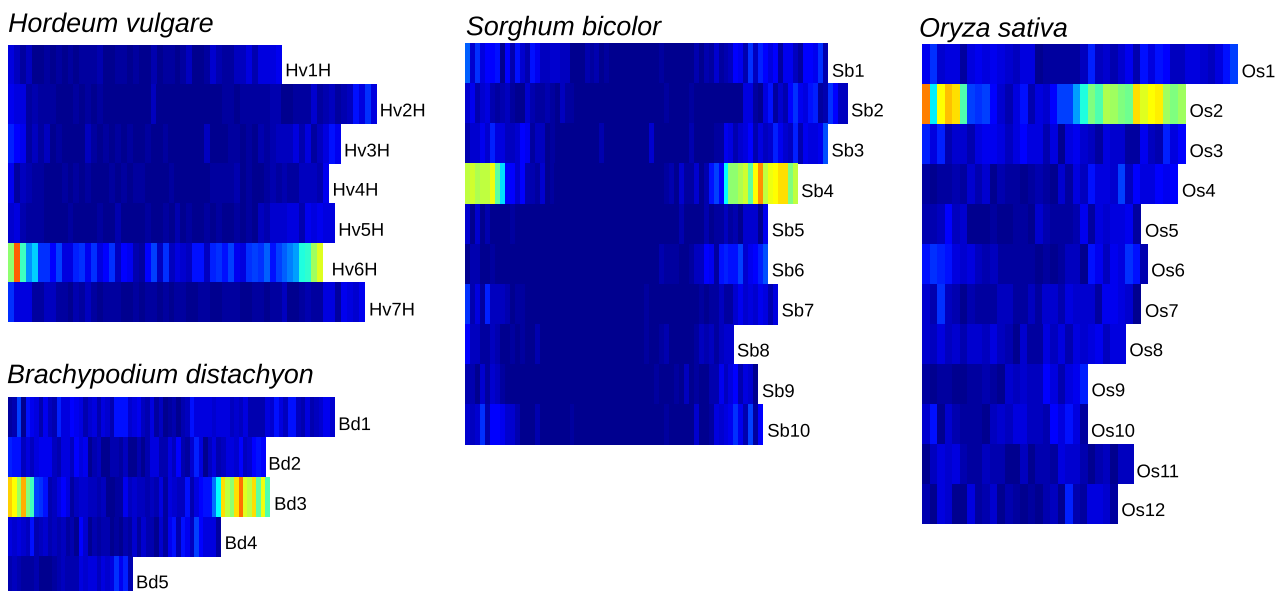


Figure 7. Wheat chromosome 6A physical map-associated genes mapped to barley (*Hordeum vulgare*; Hv), *Brachypodium distachyon* (Bd), rice (*Oryza sativa*; Os) and *Sorghum bicolor* (Sb) genomes. In barley, only genetically anchored genes are depicted (IBSC, 2012). Gene overlaps ranged from 0 (dark blue) to the maximum, shown in dark red.

maps of four arms of wheat chromosomes, including 1AS, 1AL, 1BS and 1BL, as well as the entire chromosome of 3B, have been reported (Paux *et al.*, 2008; Breen *et al.*, 2013; Lucas *et al.*, 2013; Philippe *et al.*, 2013; Raats *et al.*, 2013). The 6AS chromosome arm (336 Mb) has the closest estimated size to that of 1BS (314 Mb; Šafář *et al.*, 2010), for which an initial assembly (before manual end-to-end merging) of 254 LTC-derived contigs (with six or more clones) was reported. Excluding small contigs (with five or fewer clones) from our 6A assemblies would result in 293 (368 Mb) and 545 (459 Mb) contigs for the short and long arms, respectively. Therefore, the number of contigs obtained for 6AS is comparable with the total number of contigs in the initial assembly of 1BS; however, in contrast to the final assembly (after manual end-to-end merging) of 1BS that resulted in 57 scaffolds, no further contig merging or scaffold construction was considered in the final 6A assemblies, as this must be guided by highly reliable and robust genetic maps. Otherwise, the complexity and repeat content of the wheat genome could potentially hamper any manual contig merging or super-contig construction. Excluding small contigs, we obtained more than 100% coverage for 6AS (109%) and 6AL (124%) assemblies. This is most likely because the original chromosomal arm sizes have been underestimated.

Overall, we were able to anchor 79% of the physical contigs into corresponding genetic maps, which is greater than the aforementioned wheat chromosomal physical maps, including 1BL (74%), 1AS (74%), 1AL (~75%) and the first version of 3B with a 56% anchored physical map. Nevertheless, the 1BS physical map contained 83% of the contigs anchored to the respective genetic maps because of the low number of contigs (Raats *et al.*, 2013). The availability of extended WGP™ tags allowed for the direct placement of 3843 genes into a physical map, which was comparable with the recent physical map for barley, where an average of ~3700 genes could be assigned to each chromosome (IBSC, 2012). This high level of genetically anchored physical maps and their respective genes provides a more efficient way to clone agronomically important genes/QTLs located on 6A. Fine-mapping and the identification of genes underlying these important QTLs have been inhibited in wheat, mainly by technical constraints linked to its genome size (17 Gb), repeat content (>80%) and genomic redundancy (presence of three highly homologous genomes: A, B and D). These limitations may explain why very few wheat genes have been cloned (Krattinger *et al.*, 2009). Therefore, chromosomal BAC-based physical maps are of utmost importance to promote and simplify positional cloning in this large genome.

In this study, the comprehensive and integrated 6A physical map localized some genetic determinants to the corresponding physical map, and provided information

required for the development of tightly linked genetic markers (Figure 8). Such loci include one resistance QTL that is important against the wheat disease *Fusarium* head blight (Schmolke *et al.*, 2005; Holzapfel *et al.*, 2008), the stem rust resistance gene *Sr13* (Dubcovsky *et al.*, 2011; Simons *et al.*, 2011), an anti-xenosis gene against a new aphid biotype (Castro *et al.*, 2005) and QTLs involved in adult plant resistance to powdery mildew (Muranty *et al.*, 2009), as well as greater seedling vigor (Spielmeyer *et al.*, 2007). All of these QTLs/genes had already been genetically mapped to 6A; however, in the current study, only two corresponding gene intervals were successfully localized to the 6A physical map. For the remaining genes, either the corresponding primer/marker sequence information was not publicly available or the marker sequence could not be detected in our 6A-connected sequence data set. For the identified intervals, the respective information, including the number of genes and physical contigs assigned to a respective region, were identified (Figure 8). This analysis shows the usefulness of our physical map and represents an unprecedented opportunity to accelerate detailed gene studies, including positional cloning, and ultimately wheat breeding programs.

In the absence of the aforementioned wheat physical maps, comparative genomics and collinearity between

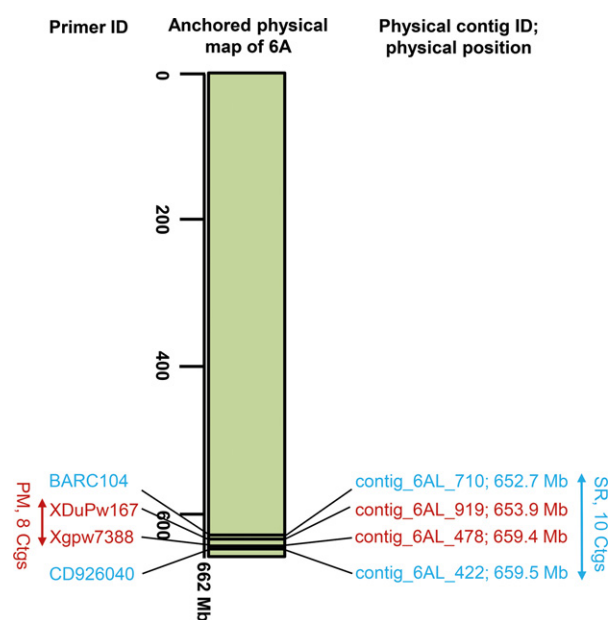


Figure 8. Physical map intervals containing agronomically important genes on wheat chromosome 6A. Primers or corresponding PCR-amplified sequences were used to define the corresponding interval. An *in silico* sequence homology search was performed to connect primer sequences or corresponding PCR-amplified sequences to sequence information underlying the 6A physical map; PM, powdery mildew-resistant gene (Muranty *et al.*, 2009); SR, stem rust resistance (Dubcovsky *et al.*, 2011; Simons *et al.*, 2011). Bi-directional arrows indicate gene containing intervals; corresponding genetic and physical information of each interval have been highlighted with the same color.

wheat and related grass genomes have been the method of choice for map-based cloning in wheat. Previous strategies have been shown to be very arduous, costly and to require a variety of genomic resources. For example, initial attempts at cloning the wheat *Lr34/Yr18* locus failed because of a lack of sufficient collinearity between wheat and the small rice and *B. distachyon* genomes (Spielmeyer *et al.*, 2008). Although conserved synteny was of great support in narrowing down respective gene-containing intervals, the region carrying *Lr34/Yr18* is absent in both rice and *B. distachyon* syntenic segments, validating the hypothesis that grass resistance genes have less conserved micro-collinearity (Leister *et al.*, 1998). Access to robustly assembled hexaploid wheat genome physical maps is the key to expediting future wheat genome sequencing efforts.

Together with the 6A physical map, we were also able to estimate the rate of recombination along 6A. A general agreement between the pattern of recombination rate and gene distribution along chromosomes was observed, similar to an earlier report (Erayman *et al.*, 2004). This report revealed a correlation between recombination gradient and gene distribution by physically mapping 3025 genes/QTLs to 334 deletion break points that spanned all seven wheat chromosomal groups. This eventually enabled the co-localization of gene- and recombination-rich regions along wheat chromosomes. In our study, the highest density of genes was found for more distal regions of 6A and where recombination rates reached their maximum. In the centromeric area of 6A, however, where recombination rates were the lowest, the number of genes per physical unit also declined dramatically. Such suppression of recombination limits our genetic anchoring resolution for the (peri)centromeric area in which a high number of physical contigs are anchored with unclear order. Therefore, different mapping approaches (e.g. radiation hybrid mapping that is independent of recombination; Kalavacharla *et al.*, 2006) should be used in the future to precisely order contigs in this region. In any case, the recombination estimates along 6A provided are imperative for accomplishing a more efficient isolation of important genes/QTLs via map-based cloning, as genes located in regions with high recombination rates are more accessible to map-based isolation (Jander *et al.*, 2002).

The efficiency of WGP™ technology and the advantage of the LTC assembly tool over FPC

This study reports the application of WGP™ on individual wheat chromosome arms and effectively confirms how well WGP™ technology works on arms with highly repetitive sequences (>80%) to accurately form a physical map. The advantage of WGP™ over conventional HICF-BAC fingerprinting assembly in wheat has previously been discussed (van Oeveren *et al.*, 2011; Philippe *et al.*, 2012).

In this respect, the robustness and quality of a BAC assembly may be justified with the quality and quantity of information (e.g. tag length and density in WGP™ or band size and similarity in HICF) provided for pairwise comparison of BACs and the subsequent establishment of respective contigs. In our study, both tag length and density were significantly improved by 64% compared with the values obtained in the WGP™-based pilot study performed on wheat chromosome 3B (Philippe *et al.*, 2012). Our improved WGP™ results were achieved by applying a different combination of restriction enzymes with higher cut-site frequency (*HindIII/MseI*), and by increasing the initial sequence read length to 100 nt, as previously suggested (Philippe *et al.*, 2012). Therefore, by using additional sequence information per BAC, together with higher stringency (tolerance value = 0) during assembly and a more efficient contig formation tool (LTC), we propose that a highly accurate BAC-based physical map of 6A has been developed.

The superior performance of LTC using HICF data compared with FPC was recently illustrated in wheat (Breen *et al.*, 2013; Raats *et al.*, 2013). Here, similar conclusions were derived while applying WGP™ tags for BAC assembly of both 6A arms. LTC-derived contigs could be classified into five groups when compared with FPC-made contigs obtained at a given cut-off value. The striking difference between LTC and FPC is reflected in LTC-built contigs (i.e. LTC ≥ 2 FPC; cases of conflicts) in which underlying BAC clones were assembled in more than one contig when compared with FPC assemblies. In a randomly selected sample (LTC ≥ 2 FPC, FPC at $1e^{-11}$), genetic anchoring revealed that in 71% of such cases, the corresponding FPC-derived contigs were incorrectly merged, whereas for LTC, this was only 13%. Moreover, by visually inspecting mis-assembled FPC-made contigs, we often observed that inconsistent cases had highly degenerate BAC coverage compared with other contig parts (Figure 4, red arrowheads). Low-coverage regions were rarely detectable in corresponding LTC-derived contigs for which multiple FPC contigs were available. Low-coverage regions in FPC-built contigs are most likely the result of false end-merging during the stepwise reduction of assembly stringency in FPC.

Our observations further demonstrate the advantages of the LTC-based BAC assembly for large genomes containing large numbers of repetitive elements. This is in agreement with a previous report that showed LTC was more efficient at forming BAC contigs using HICF (Breen *et al.*, 2013; Raats *et al.*, 2013). In addition, we showed that by applying a more reliable technology (i.e. WGP™), FPC performance is still considerably lower than that of LTC in physical map assembly. Therefore, we highly recommend combining WGP™ fingerprint methodology together with LTC assembly software for future physical mapping efforts in wheat.

EXPERIMENTAL PROCEDURES

Chromosome sorting and construction of BAC libraries

The 6AS arm was flow-sorted from a double ditelosomic line of wheat carrying both arms of chromosome 6A as telosomes ($2n = 40 + 2t6AS + 2t6AL$), whereas the 6AL arm was purified from a ditelosomic line carrying only the 6AL arm as a telocentric chromosome ($2n = 40 + 2t6AL$), according to Vrana *et al.* (2000). Both flow-sorted telosomes were derived from cv. Chinese Spring. The identity and purity in the sorted fractions was checked by fluorescence *in situ* hybridization using probes for the telomeric repeat and for the GAA repeat (Janda *et al.*, 2006). Chromosome arm-specific BAC libraries were constructed as described by Šimková *et al.* (2011). In order to estimate the average insert size, 160 BAC clones were randomly selected from each of the libraries and analyzed as described in Janda *et al.* (2006).

WGP™ data production

A 3D format of BAC pools was made for each BAC library (Appendix S17). High-concentration BAC DNA was subsequently isolated from pooled BACs, followed by WGP™ sample preparation, as previously described (van Oeveren *et al.*, 2011). Briefly, restriction ligation templates were prepared from pooled BAC DNA by digestion using *HindIII* and *MseI*, followed by ligation of adaptor sequences containing sample identification tags (barcodes), PCR amplification and the pooling of respective PCR products. Sequencing of the resulting amplified cluster was performed using the Illumina HiSeq with a 100-nt read length. Sequence reads were used for WGP™ tag production (see Appendix S1), which included barcode and restriction site identification, deconvolution of reads as WGP™ tags to the individual BACs and the filtering of WGP™ tags using various quality controls. This filtering pipeline was used to eliminate tags matching vectors, *E. coli* or chloroplast sequences. Tags containing homopolymer sequences ≥ 5 nt were considered uninformative (i.e. with a high chance of being present in more than one BAC in a particular plate). Moreover, tags potentially introducing ambiguities were also eliminated (i.e. those present in >12 BACs).

BAC contig assembly

To operate the BAC assembly, we used LTC (Frenkel *et al.*, 2010) and FPC 9.4 (<http://www.agcol.arizona.edu/software/fpc/>). Both employ the same metric or so-called Sulston score. In the case of LTC, as the tolerance could be set at the best stringency (tolerance = 0), the initial net of significant clone overlaps obtained at a $1e^{-2}$ cut-off was considered. Corresponding subnets were then obtained at a $1e^{-10}$ cut-off and used for contig formation. All contigs with at least two clones were exported into FPC format and checked for linear topology. All contigs with a width ≥ 2 were checked and split manually to obtain linear contigs. If only one clone explained the nonlinearity, the contigs were left as such because this nonlinearity was likely to be caused by a lack of WGP™ tags in the corresponding clone (Philippe *et al.*, 2013). Further parameters required to establish the LTC physical map were as follows: a tolerance of 0; gel length of 111 000; N_bands_Sulston (number of bands for Sulston score calculation) equal to gel length; and a minimum contig size of two clones. Adaptive clustering was performed using the following criteria: a $1e^{-3}$ cut-off (while the initial value to make the net of significant clone overlaps was set to $1e^{-2}$) and a step size of 1, with seven steps. MTP clones were selected applying the aforementioned parameters in the LTC program. Because of the low number of genetic markers

and the lack of adequate anchoring, no further contig merging or supercontig construction was performed.

In addition to LTC, contig assembly was also performed via FPC to compare the performance of each tool in physical map construction. Briefly, the initial FPC assembly was performed with a $1e^{-75}$ cut-off. This was subsequently run through single-to-end and end-to-end merging (Match, 1; From End, 13) at 13 sequentially higher cut-offs (thus, lower stringency) that ended up at $1e^{-11}$, as was suggested for a WGP™-based strategy in wheat (Appendix S4; Philippe *et al.*, 2012).

ACKNOWLEDGEMENTS

We thank B.S. Gill and Adam Lukaszewski for providing seeds of the wheat ditelosomic lines. We are grateful to Marie Kubaláková, Jarmila Čihalíková, Jan Vrána, Romana Šperková, Jana Dostálová and Zdeňka Dubská for assistance with chromosome sorting and BAC library construction. We would like to acknowledge Zeev Frankel and Abraham Korol for technical advice and Jenny Knibbiche for excellent technical support. The work was financially supported by a grant from the German Ministry of Education and Research (BMBF) in the framework of the TRITEX grant number FKZ 0315954A to NS and to TS, and partly by the GABI-FUTURE Start Program grant number FKZ 0315071 to TS. The WGP™ technology is protected by patents and patent applications owned by Keygene N.V. WGP™ is a trademark of Keygene N.V.

CONFLICT OF INTEREST

The authors have declared that they have no conflicts of interest.

Data availability

The URL to view the wheat 6A physical map, to blast against assigned LTC contig sequences, and to download the respective sequences is: http://seacow.helmholtz-muenchen.de/cgi-bin/gb2/gbrowse/Wheat_PhysMap_6A.

SUPPORTING INFORMATION

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

Figure S1. Steps for the elongation of WGP™ tags by connecting them to the available 6A related sequence information.

Figure S2. Different quantities of publicly available sequence information were connected with the physical contigs using the underlying WGP™ tags.

Figure S3. An example of homology between LTC and FPC contigs (as reference). LTC contigs were aligned against FPC contigs.

Figure S4. Different gene classes assigned to the physical contigs.

Table S1. Reduction in number of contigs and singleton assembled using FPC as a result of decreasing cut-off value.

Table S2. Comparison of 1214 LTC-assembled physical contigs of 6AS with FPC at different stringencies.

Table S3. Comparison of 1108 LTC-assembled physical contigs of 6AL with FPC at different stringencies.

Appendix S1. Whole-genome profiling (WGP™) of 6A chromosome arms.

Appendix S2. 6AS LTC-derived physical map.

Appendix S3. 6AL LTC-derived physical map.

Appendix S4. The FPC-based assembly of the WGP™-based BAC fingerprints.

Appendix S5. The FPC-based BAC assembly.

Appendix S6. Construction of an *in silico* anchored, gene-decorated physical map.

Appendix S7. The effect of incorporating *Ae. tauschii* sequences to the overall anchoring of the 6A physical map.

Appendix S8. 6A LTC contigs and the associated sequence information.

Appendix S9. 6A LTC-derived contigs genetically anchored to the genetic maps of wheat.

Appendix S10. 6A LTC-derived contigs anchored to barley genomic resources.

Appendix S11. An estimation of the recombination frequencies along the chromosome.

Appendix S12. Complementary description of five different classes of LTC contigs when compared with individual assemblies obtained via FPC.

Appendix S13. 6AS final FPC-derived physical map with a cut-off at $1e^{-11}$.

Appendix S14. 6AL final FPC-derived physical map with a cut-off at $1e^{-11}$.

Appendix S15. Scaffolding the LTC contig.

Appendix S16. 6A LTC-derived contigs mergable via Tu contigs (LTC scaffolds).

Appendix S17. BAC pooling scheme for WGP™ data production.

REFERENCES

- Akhunov, E.D., Sehgal, S., Liang, H.Q. *et al.* (2013) Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. *Plant Physiol.* **161**, 252–265.
- Anderson, L.K., Doyle, G.G., Brigham, B., Carter, J., Hooker, K.D., Lai, A., Rice, M. and Stack, S.M. (2003) High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. *Genetics*, **165**, 849–865.
- Breen, J., Wicker, T., Shatalina, M. *et al.* (2013) A physical map of the short arm of wheat chromosome 1A. *PLoS ONE*, **8**, e80272.
- Brenchley, R., Spannagl, M., Pfeifer, M. *et al.* (2012) Analysis of the bread-wheat genome using whole-genome shotgun sequencing. *Nature*, **491**, 705–710.
- Castro, A.M., Vasicek, A., Manifiesto, M., Gimenez, D.O., Tacaliti, M.S., Dobrovolskaya, O., Roder, M.S., Snape, J.W. and Borner, A. (2005) Mapping antixenosis genes on chromosome 6A of wheat to greenbug and to a new biotype of Russian wheat aphid. *Plant Breeding*, **124**, 229–233.
- Cavanagh, C.R., Chao, S.M., Wang, S.C. *et al.* (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl Acad. Sci. USA*, **110**, 8057–8062.
- Doležel, J., Kubaláková, M., Paux, E., Bartoš, J. and Feuillet, C. (2007) Chromosome-based genomics in the cereals. *Chromosome Res.* **15**, 51–66.
- Doležel, J., Vrána, J., Čápal, P., Kubaláková, M., Burešová, V. and Šimková, H. (2014) Advances in plant chromosome genomics. *Biotechnol. Adv.* **32**, 122–136.
- Dubcovsky, J., Ordon, F., Perovic, D., Admassu, B., Friedt, W., Abate, Z., Zhang, W. and Chao, S. (2011) Conflicting mapping results for stem rust resistance gene Sr13. *Theor. Appl. Genet.* **122**, 659.
- Erayman, M., Sandhu, D., Sidhu, D., Dilbirligi, M., Baenziger, P.S. and Gill, K.S. (2004) Demarcating the gene-rich regions of the wheat genome. *Nucleic Acids Res.* **32**, 3546–3565.
- Eversole, K., Graner, A. and Stein, N. (2009) Wheat and barley genome sequencing. In *Genetics and Genomics of the Triticeae* (Feuillet, C. and Muehlbauer, G.J., eds). Berlin: Springer, pp. 713–742.
- Flavell, R.B., Rimpau, J. and Smith, D.B. (1977) Repeated sequence DNA relationships in 4 cereal genomes. *Chromosoma*, **63**, 205–222.
- Frenkel, Z., Paux, E., Mester, D., Feuillet, C. and Korol, A. (2010) LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. *BMC Bioinformatics*, **11**, 584–601.
- Hernandez, P., Martis, M., Dorado, G. *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.
- Holzapfel, J., Voss, H.H., Miedaner, T., Korzun, V., Haberer, J., Schweizer, G., Mohler, V., Zimmermann, G. and Hartl, L. (2008) Inheritance of resistance to Fusarium head blight in three European winter wheat populations. *Theor. Appl. Genet.* **117**, 1119–1128.
- IBSC (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711–716.
- IWGSC (In press) A chromosome-based draft sequence of the hexaploid bread wheat genome. *Science*.
- Janda, J., Safar, J., Kubaláková, M. *et al.* (2006) Advanced resources for plant genomics: a BAC library specific for the short arm of wheat chromosome 1B. *Plant J.* **47**, 977–986.
- Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M. and Last, R.L. (2002) Arabidopsis map-based cloning in the post-genome era. *Plant Physiol.* **129**, 440–450.
- Jia, J.Z., Zhao, S.C., Kong, X.Y. *et al.* (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, **496**, 91–95.
- Kalavacharla, V., Hossain, K., Gu, Y., Riera-Lizarazu, O., Vales, M.I., Bhamidimarri, S., Gonzalez-Hernandez, J.L., Maan, S.S. and Kianian, S.F. (2006) High-resolution radiation hybrid map of wheat chromosome 1D. *Genetics*, **173**, 1089–1099.
- Krattinger, S., Wicker, T. and Keller, B. (2009) Map-based cloning of genes in Triticeae (wheat and barley). In *Genetics and Genomics of the Triticeae* (Feuillet, C. and Muehlbauer, G.J., eds). Berlin: Springer, pp. 337–357.
- Kunzel, G., Korzun, L. and Meister, A. (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. *Genetics*, **154**, 397–412.
- Leister, D., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K., Graner, A. and Schulze-Lefert, P. (1998) Rapid reorganization of resistance gene homologues in cereal genomes. *Proc. Natl Acad. Sci. USA*, **95**, 370–375.
- Ling, H.Q., Zhao, S.C., Liu, D.C. *et al.* (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*, **496**, 87–90.
- Lucas, S.J., Akpinar, B.A., Kantar, M. *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.
- Lukaszewski, A.J. and Curtis, C.A. (1993) Physical distribution of recombination in B-genome chromosomes of tetraploid wheat. *Theor. Appl. Genet.* **86**, 121–127.
- Luo, M.C., Thomas, C., You, F.M., Hsiao, J., Shu, O.Y., Buell, C.R., Malandro, M., McGuire, P.E., Anderson, O.D. and Dvorak, J. (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.
- Mascher, M., Muehlbauer, G.J., Rokhsar, D.S. *et al.* (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.* **76**, 718–727.
- Muranty, H., Pavoine, M.T., Jaudeau, B., Radek, W., Doussinault, G. and Barloy, D. (2009) Two stable QTL involved in adult plant resistance to powdery mildew in the winter wheat line RE714 are expressed at different times along the growing season. *Mol. Breeding*, **23**, 445–461.
- van Oeveren, J., de Ruiter, M., Jesse, T. *et al.* (2011) Sequence-based physical mapping of complex genomes by whole genome profiling. *Genome Res.* **21**, 618–625.
- Ortiz, R., Sayre, K.D., Govaerts, B., Gupta, R., Subbarao, G.V., Ban, T., Hodson, D., Dixon, J.A., Ortiz-Monasterio, J.I. and Reynolds, M. (2008) Climate change: can wheat beat the heat? *Agric. Ecosyst. Environ.* **126**, 46–58.
- Paux, E., Sourdille, P., Salse, J. *et al.* (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science*, **322**, 101–104.
- Philippe, R., Choulet, F., Paux, E. *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**, 47–61.
- Philippe, R., Paux, E., Bertin, I. *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.

- Poland, J.A., Brown, P.J., Sorrells, M.E. and Jannink, J.L. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.
- Raats, D., Frenkel, Z., Krugman, T. *et al.* (2013) The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. *Genome Biol.* **14**, R138.
- Šafář, J., Šimková, H., Kubaláková, M., Čihalíková, J., Suchánková, P., Bartoš, J. and Doležel, J. (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet. Genome Res.* **129**, 211–223.
- Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegou, B., Quraishi, U.M., Calzagno, T., Cooke, R., Delseny, M. and Feuillet, C. (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell*, **20**, 11–24.
- Schmolke, M., Zimmermann, G., Buerstmayr, H., Schweizer, G., Miedaner, T., Korzun, V., Ebmeyer, E. and Hartl, L. (2005) Molecular mapping of Fusarium head blight resistance in the winter wheat population Dream/Lynx. *Theor. Appl. Genet.* **111**, 747–756.
- Sierro, N., van Oeveren, J., van Eijk, M.J., Martin, F., Stormo, K.E., Peitsch, M.C. and Ivanov, N.V. (2013) Whole genome profiling physical map and ancestral annotation of tobacco Hicks Broadleaf. *Plant J.* **75**, 880–889.
- Šimková, H., Šafář, J., Kubaláková, M. *et al.* (2011) BAC libraries from wheat chromosome 7D: efficient tool for positional cloning of aphid resistance genes. *J. Biomed. Biotechnol.* **2011**, 302543.
- Simons, K., Abate, Z., Chao, S.M., Zhang, W.J., Rouse, M., Jin, Y., Elias, E. and Dubcovsky, J. (2011) Genetic mapping of stem rust resistance gene Sr13 in tetraploid wheat (*Triticum turgidum* ssp *durum* L.). *Theor. Appl. Genet.* **122**, 649–658.
- Soderlund, C., Longden, I. and Mott, R. (1997) FPC: a system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.* **13**, 523–535.
- Sommer, R., Glazirina, M., Yuldashev, T. *et al.* (2013) Impact of climate change on wheat productivity in Central Asia. *Agric. Ecosyst. Environ.* **178**, 78–99.
- Spielmeier, W., Hyles, J., Joaquim, P., Azanza, F., Bonnett, D., Ellis, M.E., Moore, C. and Richards, R.A. (2007) A QTL on chromosome 6A in bread wheat (*Triticum aestivum*) is associated with longer coleoptiles, greater seedling vigour and final plant height. *Theor. Appl. Genet.* **115**, 59–66.
- Spielmeier, W., Singh, R.P., McFadden, H., Wellings, C.R., Huerta-Espino, J., Kong, X., Appels, R. and Lagudah, E.S. (2008) Fine scale genetic and physical mapping using interstitial deletion mutants of Lr34/Yr18: a disease resistance locus effective against multiple pathogens in wheat. *Theor. Appl. Genet.* **116**, 481–490.
- Tanaka, T., Kobayashi, F., Joshi, G.P. *et al.* (2014) Next-generation survey sequencing and the molecular organization of wheat chromosome 6B. *DNA Res.* **21**, 103–114. first published online October 1, 2013 doi: 10.1093/dnares/dst041.
- TIBI (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.
- TTGC (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.
- Vitolo, N., Albiero, A., Forcato, C. *et al.* (2011) First survey of the wheat chromosome 5A composition through a next generation sequencing approach. *PLoS ONE*, **6**, e26421.
- Vrana, J., Kubaláková, M., Šimková, H., Čihalíková, J., Lysak, M.A. and Doležel, J. (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). *Genetics*, **156**, 2033–2041.