Theor Appl Genet (2014) 127:2029–2040
DOI 10.1007/s00122-014-2358-z

ORIGINAL PAPER

Utilization of deletion bins to anchor and order sequences along the wheat 7B chromosome

Tatiana Belova · Lars Grønvold · Ajay Kumar · Shahryar Kianian · Xinyao He · Morten Lillemo · Nathan M. Springer · Sigbjørn Lien · Odd-Arne Olsen · Simen R. Sandve

Received: 22 March 2014 / Accepted: 13 July 2014 / Published online: 19 August 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract

Key message **A total of 3,671 sequence contigs and scaffolds were mapped to deletion bins on wheat chromosome 7B providing a foundation for developing high-resolution integrated physical map for this chromosome.**

Abstract Bread wheat (*Triticum aestivum* L.) has a large, complex and highly repetitive genome which is challenging to assemble into high quality pseudo-chromosomes. As part of the international effort to sequence the hexaploid bread wheat genome by the international wheat genome sequencing consortium (IWGSC) we are focused on assembling a reference sequence for chromosome 7B. The successful completion of the reference chromosome sequence is highly dependent on the integration of genetic and physical maps. To aid the integration of these two types of maps, we have constructed a high-density deletion bin map of chromosome 7B. Using the 270 K Nimblegen comparative genomic hybridization (CGH) array on a set of cv. Chinese spring deletion lines, a total of 3,671 sequence contigs and scaffolds (~7.8 % of chromosome 7B physical length) were mapped into nine deletion bins. Our method of genotyping deletions on chromosome 7B relied on a model-based clustering algorithm (Mclust) to accurately predict the presence or absence of a given genomic sequence in a deletion line. The bin mapping results were validated using three different approaches, viz. (a) PCR-based amplification of randomly selected bin mapped sequences (b) comparison with previously mapped ESTs and (c) comparison with a 7B genetic map developed

Communicated by Hong-Qing Ling.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2358-z) contains supplementary material, which is available to authorized users.

T. Belova · L. Grønvold · M. Lillemo · O.-A. Olsen · S. R. Sandve (✉)
Department of Plant Sciences, Norwegian University of Life Sciences, Ås, Norway
e-mail: simen.sandve@nmbu.no

T. Belova
e-mail: tatiana.belova@nmbu.no

L. Grønvold
e-mail: lars.gronvold@nmbu.no

M. Lillemo
e-mail: morten.lillemo@nmbu.no

O.-A. Olsen
e-mail: odd-arne.olsen@nmbu.no

A. Kumar · S. Kianian
Department of Plant Sciences, North Dakota State University, Fargo, ND, USA
e-mail: Ajay.Kumar.2@ndsu.edu

S. Kianian
e-mail: S.Kianian@ndsu.edu

X. He
International Maize and Wheat Improvement Center (CIMMYT), Apdo.Postal 6-641, 06600 Mexico, DF, Mexico
e-mail: x.he@cgiar.org

N. M. Springer
Department of Plant Biology, Microbial and Plant Genomics Institute, University of Minnesota, Saint Paul, MN 55108, USA
e-mail: springer@umn.edu

S. Lien
Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences, Ås, Norway
e-mail: sigbjorn.lien@nmbu.no

in the present study. Validation of the bin mapping results suggested a high accuracy of the assignment of 7B sequence contigs and scaffolds to the 7B deletion bins.

Abbreviations

CGH	Comparative genomic hybridization
IWGSC	International wheat genome sequencing consortium
CSS	Chromosome survey sequencing
cv CS	Cultivar chinese spring
LDN	Langdon
LDN-DS 7D(7B)	Langdon 7B substitution line
ISBP	Insertion site-based polymorphism
RG	Random genomic probes
FL	Fraction length
RIL	Recombinant inbred lines

Introduction

The recent chromosome survey sequence (CSS) assembly of the hexaploid bread wheat genome (*Triticum aestivum* L.; $2n = 6 \times = 42$; AABBDD) by the international wheat genome sequencing consortium (IWGSC) (IWGSC, 2014) serves as an important first step towards a wheat reference genome sequence (www.wheatgenome.org). This chromosome-specific assembly allow for a deeper understanding of the wheat genome composition, organization, and evolution, as well as providing a resource for future research- and breeding efforts. However, due to the large chromosome size and extreme repeat content (>80 %), the wheat chromosome sequence assemblies are highly fragmented compared to for example barley (The International Barley Genome Sequencing Consortium 2012), rice (The International Rice Genome Sequencing Consortium 2005), potato (The Potato Genome Sequencing Consortium 2011) and sorghum (Paterson et al. 2009).

In order to move towards a complete genome assembly, physical contigs and scaffolds must be integrated with genetic maps at high density and high resolution. A major constraint for the genetic mapping in wheat is the non-uniform distribution of recombination events along the chromosomes, with recombination rates dropping dramatically towards the centromere (Devos et al. 1995; Werner et al. 1992; Akhunov et al. 2003). For instance, detailed analyses of recombination frequencies in bread wheat chromosome 3B show that 90 % of crossing overs occur in only 40 % of the chromosome (Saintenac et al. 2009). The same study also observed >85-fold differences for crossover frequency per physical distance (cM/Mb) for a centrometic bin (C-3BS1-0.33) compared to a sub-telomeric bin (3BS8-0.78-0.87) on chromosome 3B. This "recombination stiffness" makes it very difficult to place and order sequence contigs along a chromosome. One approach has therefore been to combine several independent and complementary mapping approaches with meiotic mapping, such as synteny-based mapping using closely related species and deletion bin mapping (e.g. 3B and 1BL) (Paux et al. 2008; Philippe et al. 2013). Although synteny-based mapping approaches can be powerful, inversions and translocations of genes and gene blocks in wheat relative to other grass genomes (like *Brachypodium*, rice and sorghum) is common (Kumar et al. 2012). Synteny-based mapping is therefore more reliable within smaller chromosomal blocks. Hence, assigning sequence contigs to smaller bins along the chromosome is of high value for the downstream synteny-based sequence ordering, but also an important source for independent verification of the meiotic mapping results.

Deletion bin mapping is a recombination independent mapping strategy and involves the use of a series of overlapping deletions to map markers to relatively short (range 20–155 Mb in size) chromosomal segments (deletion bins) (Qi et al. 2004). In bread wheat, aneuploid stocks have been extensively used to assign markers to chromosomes, chromosome arms, and bins within chromosome arms. Sears (1954) was the first to study and produce bread wheat aneuploids (cv. Chinese spring, CS), including 21 monosomics, 21 nullisomics and 21 tetrasomics (Sears 1954). In addition, more recently, using gametocidal genes to induce chromosome breaks, a set of 436 terminal chromosome deletions were identified in hexaploid wheat (Endo and Gill 1996). Later, using set of wheat aneuploids and deletion stocks, 16,000 ESTs were bin mapped (Qi et al. 2004), of which 549 ESTs (corresponding to ~0.08 % of the chromosome 7B physical length) were assigned to six bins on chromosome 7B (Hossain et al. 2004).

In the present study we describe the development of a high-density deletion bin map of wheat chromosome 7B, placing ~7.8 % of the chromosome 7B physical length into nine bins using Nimblegen comparative genome hybridization (CGH). In addition, an F_6 recombinant inbred line (RIL) population containing 131 lines was assayed with the 90 K iSelect SNP chip (Wang et al. 2014), resulting in incorporation of 629 SNP markers into the 7B genetic map. This work is part of the IWGSC Norwegian 7B sequencing project and aid in anchoring and ordering of physical sequence contigs from MTP (Minimal Tiling Path) BAC sequencing, a critical step towards a complete 7B reference sequence.

Materials and methods

Oligonucleotide probe design

Two types of oligonucleotide probes were extracted from the shotgun sequence assembly of chromosome 7B (Belova

et al. 2013): random genomic (RG) probes and insertion site-based polymorphism (ISBP) probes (Fig. 1). In order to develop the RG probes, assemblies were first masked for repeats with RepeatMasker (Smit et al. 1996–2010) against an in-house repeat content database [TREP ten combined with the repeats identified in Choulet et al. (2010)]. Masked contigs were fragmented in non-overlapping sequences of 50 bp located ≥50 bp apart. ISBP finder (Paux et al. 2010) was used to identify ISBP sites with high and medium confidence levels from which sequences of 50 bp, 25 bp from each side of the junction, were selected as ISBP probes.

Subsequent to the identification of RG and ISBP probes, we used BLASTN (Altschul et al. 1990) to identify and remove probe sequences with high similarity (hit length >45 bp and identity >95 %) to contigs in the 7A and 7D assemblies (IWGSC; http://www.wheatgenome.org/). Probes carrying homopolymers longer than 8 bp were excluded from the analysis. We also excluded probes that did not pass the 'Cycle script' designed by Nimblegen or had a calculated oligonucleotide melting temperature outside the 66–86 °C range. A collection of wheat ESTs (Lazo et al. 2004) was used to design random control probes (50 bp long) that were not overrepresented with 7B sequences.

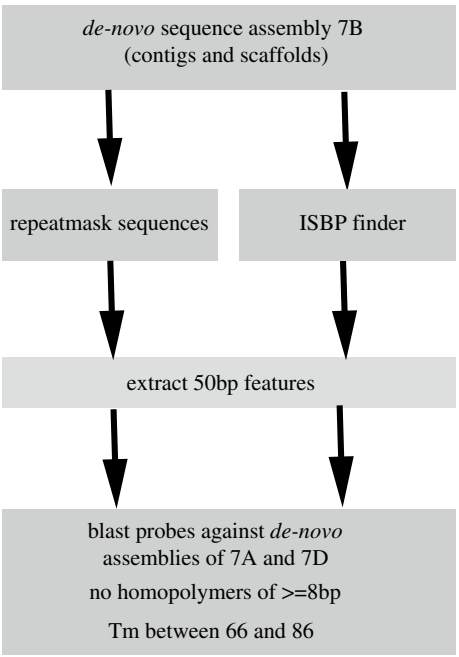

Fig. 1 Overview of the CGH Nimblegen probe design pipeline

Table 1 Set of deletion lines with their corresponding fragment length (FL), showing the percent of the chromosome arm present

Deletion stock	Fragment length	Nomenclature
Del7BS-2	0.27	FL-0.27
Del7BL-14	0.14	FL-0.14
Del7BL-2	0.33	FL-0.33
Del7BL-1	0.40	FL-0.40
Del7BL-9	0.45	FL-0.45
Del7BL-7Del1DS-3	0.63	FL-0.63
Del7BL-5	0.69	FL-0.69
Del7BL-13	0.79	FL-0.79
Del7BL-3	0.86	FL-0.86

Plant material

Two tetraploid wheat lines, Langdon (LDN; AABB; $2n = 4\times = 28{:}13'' + 7B'')$ and Langdon chromosome substitution line (LDN-DS 7D (7B), $2n = 4\times = 28{:}13'' + 7D''$, in which chromosomes 7B is substituted by 7D chromosomes of the hexaploid cultivar CS (Joppa and Williams 1977), were used initially for screening and identification of 7B specific probes and later as reference samples to estimate absence/presence (i.e. *M*-values) of probes in CS 7B deletion stocks (see sections below for details).

Among the deletion stocks of the hexaploid wheat cultivar CS (*T. aestivum*) (Endo and Gill 1996), lines with terminal deletions in chromosome 7B and its ditelosomic lines (DT7BL and DT7BS) were used in the CGH assays. Details of 7B deletion stocks used in this study are provided in Table 1. The fraction length (FL) reflects the position of the breakpoint from the centromere relative to the length of the complete arm. Seeds for deletion lines were kindly provided by Dr. Bikram S. Gill, Department of Plant Pathology, Kansas State University, Manhattan, KS, USA.

CGH sample preparation and hybridization

DNA from leaf tissue was isolated by the CTAB method (Springer 2010). Labeling and hybridization of samples were performed according to the Nimblegen protocol. Half a µg DNA of each sample was labeled using either Cy3 or Cy5-labeled Random Nonamers. Samples were denaturated at 98 °C for 10 min and chilled on ice for 2 min. The DNA was incubated for 2 h at 37 °C with 100 units Klenow Fragment (5′-3′ exo-) and dNTP mix (10 mM each). After adding stop solution (0.5 M EDTA), samples were precipitated with NaCl and isopropanol and centrifuged at 12,000 × *g* for 10 min. The pellets were re-suspended in 25 µl of H_20. Twenty µg of Cy3 and Cy5 labeled samples were mixed in a 1.5 ml tube and dried in a vacuum concentrator on low heat. Each sample pair was then re-suspended in unique sample tracking control and added to 8.7 µl of the hybridization solution mix. Tubes were first incubated at 95 °C for 5 min, and then at 42 °C for 5 min. Samples were hybridized to CGH array for 60–72 h at 42 °C. Slides were washed and immediately scanned using the MS 200

microarray scanner according to the array manufacturer's protocol. Probe fluorescence intensities were extracted with the NimbleScan 2.1 software. Raw data was normalized by two-dimensional loess spatial normalization followed by M-A loess normalization for each sample comparison using the control probes as training set (GEO submission GSE57461).

Selection of 7B chromosome specific probes

In order to select a subset of 7B specific probes as well as a set of control probes which do not hybridize to the 7B genomic sequence we first performed CGH between two tetraploid Langdon wheat lines that differ only by the presence of the 7B chromosome in the genome (LDN contains 7B, while LDN-DS 7D (7B) lacks 7B). The experiment was carried out with a 3*720 K CGH microarray using a dye swap design where each sample was labeled with both Cy5 and Cy3. The selected set of 7B specific probes and control probes was then printed on a 12*270 K CGH chip (Roche, NimbleGen Inc.) and hybridized with CS deletion lines.

Genotyping presence absence variation in CS deletion lines

The CS deletion lines have various sized terminal overlapping deletions, usually >10 % of the chromosome arm (Endo and Gill 1996). The distribution of $\log_2$ ratios of hybridization signal intensities between deletion lines and wild type (referred to as *M*-values) is therefore expected to be a combination of two underlying distributions, representing probes being deleted (i.e. absent) and those that are present (Fig. 2).

To determine the probability for a probe to belong either to the "present" or "absent" classes, we used Gaussian mixture model clustering [P_{abs}, $(1 - P_{abs}) = P_{pres}$] as implemented in the R package *Mclust* (Fraley and Raftery 2007). The parameter '*G*' (number of groups) was fixed to 2, while all other parameters were estimated by the Mclust software. The number of absent probes for each deletion line was estimated by intersecting results from two different Mclust analyses using different LDN line hybridizations as a reference. A sequence was only assigned as absent or present if both Mclust analyses supported the same classification with ≥80 % probability. Probes that did not meet this criterion were assigned to the NA class (i.e. not possible to classify). Long *M*-value distribution tails in combination with a limited separation of the absent and present distribution peaks sometimes lead to erroneously assignment of probes with high *M*-values to the absent class. Probes with *M*-values higher than the mean *M*-value in the present class were therefore given a probability of 0 for belonging to the absent class.

Assignment of sequences to deletion bins

Assignment of sequences to deletion bins was based on a two-step strategy using the absence/presence classifications from each deletion line. In the first step, we compared each deletion line (X) to the deletion line with an incrementally smaller deletion (Y) and identified sequences present in Y but absent in X. To assign sequences to the most distal deletion bins on the two 7B arms, we compared lines with the smallest deletions to ditelosomic lines carrying a complete copy of that chromosome arm. In the second step, we used this initial bin assignment and confirmed that each bin mapped sequence was present in all other deletion lines with smaller deletions.

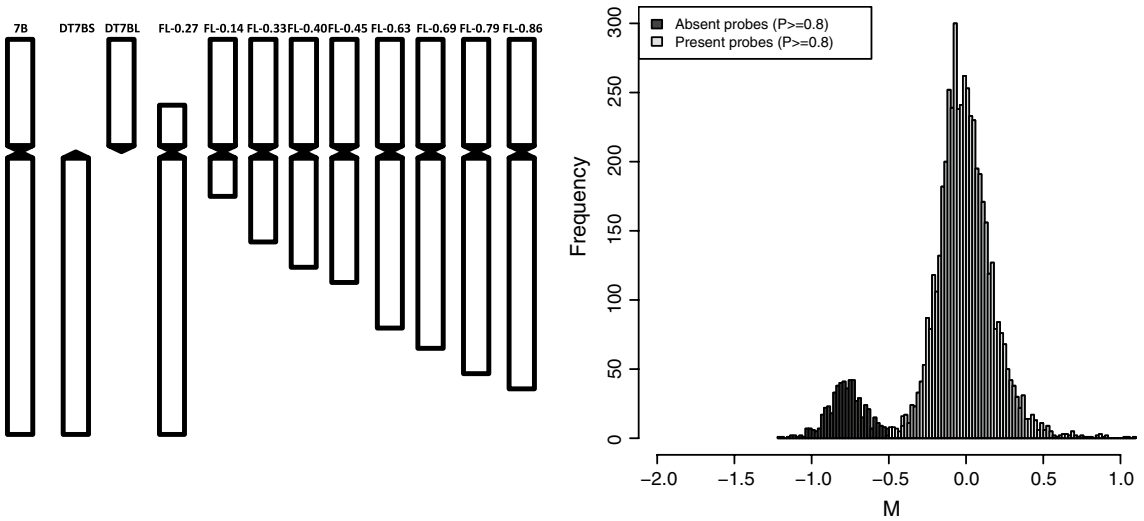

Fig. 2 Schematic explanation of the two groups of probes (present and absent) when comparing *deletion line* and *reference line*

SNP-based genetic map of chromosome 7B

A mapping population of 131 RILs was developed from a cross between the CIMMYT breeding line 'SABUF/5/BCN/4/RABI//GS/CRA/3/AE.SQUARROSA (190)' (selection history CASS94Y00042S-32PR-1B-0 M-0Y) and the German spring wheat cv. 'Naxos' (pedigree Tordo/St.Mir808-Bastion//Miranet). The population was advanced from F_2 to F_6 through the single seed descent (SSD) method. DNA was extracted from F_6 plants using DNeasy plant DNA extraction kit (Qiagen). The population was genotyped with the iSelect 90 K wheat chip from Illumina, which contains a total of 81,587 SNP markers (Wang et al. 2014). Genotypes were called using Genome Studio V2011.1. Due to the hexaploid genome constitution of wheat, the automatic clustering algorithm identified only 3,117 polymorphic markers that fit the expected segregation ratio for a diploid locus in the F_6 population. An additional 7,255 polymorphic markers with skewed clustering patterns due to signal noise from the other two genomes were called manually giving a total of 10,372 SNP markers for further analysis. Genetic linkage groups were created using the program MST map (Wu et al. 2008) with a cutoff *p* value of 1e-6, maximum distance of 15 cM between markers, minimum size of linkage group being 2 cM. MST map linkage groups were then assigned to chromosomes based on the BLASTN results of SNP sequences against survey sequences of A-, B- and D genomes (IWGSC data repository at http://wheat-urgi.versailles.inra.fr/). Only markers giving a hit to a single chromosome with ≥99 % sequence identity and 100 % coverage were assigned to a chromosome. Finally, the JoinMap v.4 Maximum Likelihood algorithm was used to estimate marker order for the 7B linkage group.

Verification of the CGH bin mapping results

We used three independent methods to verify the bin mapping approach and estimate the error rate. In the first approach, we performed PCR-based verification of the mapped sequences. Primers were designed from bin mapped sequences with ISBP finder, tested for 7B specificity using the 7B CS ditelosomic lines and then used for PCR amplification in CS deletion lines to identify the bin location of the markers. PCR reactions were carried out in 10 µl total reaction using 60 ng of genomic DNA containing 1 µl 10 × PCR buffer, 0.2 µl 10 mM dNTPs and 0.1 µl of 5 units/µl of AmpliTaq DNA polymerase (Applied Biosystems). The PCR conditions used were as follows: 94 °C for 3 min, 45 cycles of: 45 s at 94 °C, 45 s at 59 °C, 90 s at 72 °C, followed by a final extension at 72 °C for 10 min. The PCR products were separated on a 1.5 % agarose gel and visualized using ethidium bromide staining.

In the second verification approach, we took advantage of the fact that some of the bin mapped sequences in this study have sequence homology with the previously bin mapped ESTs (Hossain et al. 2004). The sequences that we bin map in the present study were used in a BLASTN search against sequences of the previously 7B bin mapped ESTs. BLAST hits were filtered based on ≥99 % identity and 100 % coverage. Redundant ESTs were not considered in this analysis. Redundant ESTs are defined as ESTs giving a hit to the same bin mapped sequence with the identical start and end position, identical mismatches, identical gap length and identical hit length.

The final validation of deletion bin mapping results was done by integrating genetically mapped SNPs into the deletion bin map. In order to assign SNP markers to deletion bins, BLASTN search of SNP sequences against bin mapped sequences was performed. Only hits with ≥99 % identity and 100 % coverage of the marker locus were considered in this study.

Distribution of genes along deletion bins

The 7B gene calls from the wheat CSS (IWGSC data repository at http://wheat-urgi.versailles.inra.fr/Seq-Repository/Genes-annotations) were used in a BLASTX (Altschul et al. 1990) search to estimate gene content of the bin mapped sequences. BLAST result filtering was carried out in the following way: (a) Only query hits with a minimum sequence identity of ≥99 % and a minimum length of 30 amino acid were considered in the analyses (b) Duplicated gene hits in one scaffold were removed from the analyses. Duplicated hits were defined as hits belonging to the same gene ID. The gene density in a bin was calculated by dividing the number of gene hits with the total scaffold length in that bin.

Results

CGH and selection of chromosome 7B specific CGH probes

In order to identify probe sequences which detect presence/absence variation (PAV) between LDN and LDN 7D (7B) genotypes we conducted a pilot experiment using a 720 K CGH microarray chip. The *M*-values of LDN versus LDN-DS 7D (7B) comparisons, was used as probe selection criterium. Probes with large difference in hybridization intensity (*M*-values >0.35) and high reproducibility between replicates were classified as chromosome 7B specific. Non-polymorphic control probes were selected from the subset of probes with an *M*-value close to zero ($-0.02 < M < +0.02$). From this experiment, a set

of 49,500 7B probes (11 % ISBP and 89 % RG probes) and 18,000 control probes were selected and printed on a 12*270 K CGH chip with each probe replicated four times per chip. Using BLASTN against 7B IWGSC gene calls we estimate that 0.9 % of RG probes on the 270 K CGH chip are derived from coding genes. This is comparable to the total percentage of coding sequence in the 7B CSS assembly (0.7 %). Low quality CGH hybridizations were excluded from the dataset based on the experimental metrics reports (NimbleScan 2.1 software). In total, we hybridized 17 CS cytogenetic stocks out of which 11 yielded high quality CGH results and were used for the bin mapping (Table 1).

Effect of probe type on *M*-value distribution

In hexaploid wheat, ISBP markers have provided high level of sub-genome specificity compared to DNA probes designed from the coding regions (Choulet et al. 2010). Generally, probes will have a better signal to background ratio when there is less cross hybridization to other regions of the genome. In order to investigate the relationship between the type of the probe and its hybridization properties, we first generated ten *M*-value distributions between different deletion lines, calculated the proportion of ISBP and RG probes in the 10 % lowest range of $\log_2$ distributions, and then compared this with the total proportions of ISBP and RG probes on the array. Mean proportion of ISBP probes in the lowest range of $\log_2$ distributions were 10 % (range 7–14 %) (data not shown), comparable to the proportion of ISBPs among the total number of the probes (11 %) indicating similar hybridization properties of both types of probes.

Effect of combining signals from multiple probes in presence/absence genotyping

Our approach to genotyping presence/absence variation on chromosome 7B relies on the ability to accurately predict which sub-distribution of *M*-values a particular probe belongs to (Fig. 2). Thus, a good separation of the two underlying *M*-values distributions is expected to result in more robust probe classification. We have used 7B assembly sequence to design 7B specific probes. In many cases multiple probes were derived from the same contig or scaffold (combination of several contigs). In total, we designed 49,500 probes from 33,286 contigs, giving an average of 1.49 probes per contig (range 1–17). Furthermore, many contigs belong to larger scaffolds (1.43 contigs per scaffold, range 1–18). We therefore investigated the relationship between the number of probes used to estimate *M*-values and the peak separation in the bimodal *M*-value distribution. This showed that the separation between *M*-value distributions significantly improved as the number of probes used for *M*-value calculation increased (Fig. 3a). However, merging signals from several probes comes at a cost, since it leads to fewer data points. We chose to conduct all following analyses using *M*-values from sequences containing at least 3 probes, giving a total of 5,577 sequences for the analysis. The ratio of sequences derived from 7BS to 7BL was 40:60 % (2267/3310) and a total of 5,177 probe sequences could be classified as either absent or present in at least one deletion line. Ninety-two percent of the sequences were assigned to a class in at least 50 % of all CGH analyses of deletion lines. Among these, a strong positive correlation was observed between the number of probe observations and the power to classify a sequence as present/absent (Fig. 3b). Thus, the parameters chosen represent a reasonable trade-off between the number of probes used per sequence for presence/absence calling and mapping accuracy.

Determining PAV in the deletion lines

The estimated number of deleted sequences in the CS deletion lines is shown in Table 2. With a few exceptions, our results fit well with the expected deletion sizes based on cytological evidence (Endo and Gill 1996). Mean discrepancy between expected and observed deletion size was 5.2 %. Three lines with deletions in the 7BL (FL-0.86, FL-0.79 and FL-0.69) showed significant deviations from the expectations. The FL-0.86 had a 7 % higher proportion of absent probes than expected, while FL-0.79 and FL-0.69 had fewer absent probes than expected (Table 2).

To further describe the nature of deletions in FL-0.86, FL-0.79 and FL-0.69 lines, we plotted $\log_2$ distributions of FL-0.86 vs. FL-0.79, FL-0.79 vs. FL-0.69, and FL-0.86 vs. FL-0.69 (Fig. 4). If FL-0.86, FL-0.79 and FL-0.69 represent incrementally larger deletions, we expect set of *M*-values to occur on the lower right side of the plot (i.e. present probes in the smaller deletion which are absent in the larger deletion). From the plots it appears that FL-0.79 has presence and absence variation relative to both FL-0.86 and FL-0.69. Moreover, FL-0.86 and FL-0.69 have virtually identical *M*-values across the 7BL sequences (Fig. 4). This result shows that cytological estimation of the deletion length and type in the FL-0.86, FL-0.79 and FL-0.69 lines-most likely is wrong.

Bin mapping of 7B sequences

In total we bin mapped ~74,130 Kbp (3,671 sequence contigs and scaffolds), representing ~7.8 % of 7B chromosome sequence. Using the 7B gene models generated in the CSS project we estimated the gene density (genes/Kbp) along

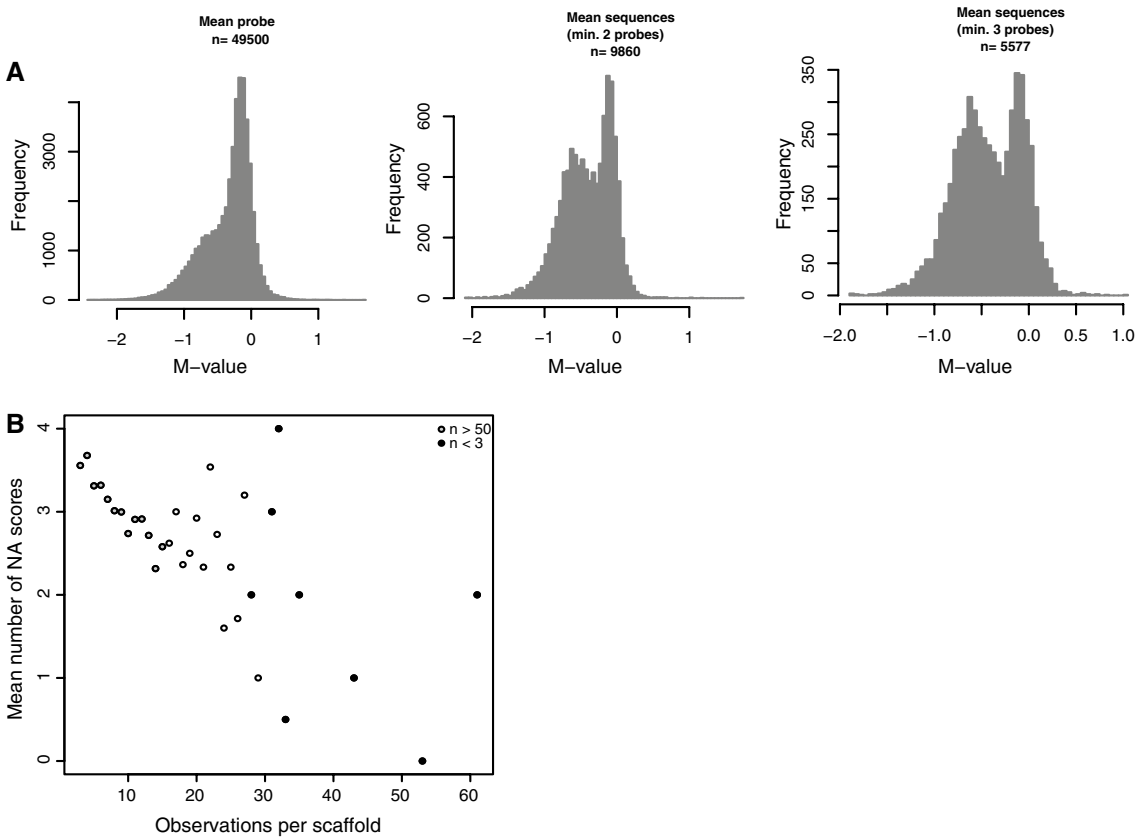

Fig. 3 **a** *M*-value distributions with *M*-values averaged per probe and per sequence containing 2 and 3 probes. **b** Relationship between the propensity of sequences to be assigned a class (present or absent) and the number of probes contained in sequences

Table 2 Deleted sequences in 7B deletion lines

Arm	Deletion line	Expected proportion deleted	Absent probes		
			7BS	7BL	Proportion[a]
7BS	DT7BS	–	1,680	7	–
	FL-0.27	0.73	1,294	4	0.57
7BL	FL-0.14	0.86	1	2,227	0.67
	FL-0.33	0.67	2	2,257	0.68
	FL-0.40	0.6	1	1,855	0.56
	FL-0.45	0.55	0	1,770	0.53
	FL-0.63	0.37	2	1,453	0.44
	FL-0.69	0.31	0	658	0.20
	FL-0.79	0.21	0	563	0.17
	FL-0.86	0.14	8	709	0.21
	DT7BL	–	7	2,902	–

[a] Proportion is calculated based on predicted absent scaffolds on the correct arm only

the bins on 7B (Table 3). Gene density was distributed unevenly along the 7B chromosome with an increase in gene density from the centromere to the telomere (from 0.01 to 0.02 for 7BL and from 0.01 to 0.02 for 7BS). The average gene density for the centromeric region (bins 7BL_0 - 0.14 and 7BS_0 - 0.27) was 1 gene per 107 Kb. The gene density increased by ~2-fold for distal bins.

In constructing a precise bin map, single terminal deletions are preferred over multiple or interstitial deletions (Hohmann et al. 1995). Since our hybridization data support an aberrant nature of deletions types (i.e. not single terminal deletions) in the FL-0.69, FL-0.79 and FL-0.86 lines, all sequences deleted in any of these lines were grouped into one pseudo bin (7BL_0.69* – 1.00). For each of these three lines bin mapping was first performed by comparison with ditelosomic line carrying a complete 7BL arm. Three hundred and nine, 283 and 351 sequences were mapped to 7BL_0.69 – 1.00, 7BL_0.79 – 1.00 and 7BL_0.86 – 1.00, respectively. Among them, 253 sequences (4,429 Kbp) were mapped to all three bins. Bin 7BL_0.79 – 1.00 had 25 unique mapped sequences compared to 7BL_0.86 – 1.00. Eight out of 25 were mapped to the bin BL_0.69 – 1.00 as well. Bins 7BL_0.69 – 1.00 and 7BL_0.86 – 1.00 shared additional 63 mapped sequences. 5, 17 and 30 sequences were uniquely mapped to 7BL_0.69 – 1.00, 7BL_0.79 – 1.00 and 7BL_0.86 – 1.00. In total, 381 sequences were mapped to 7BL_0.69* – 1.00.

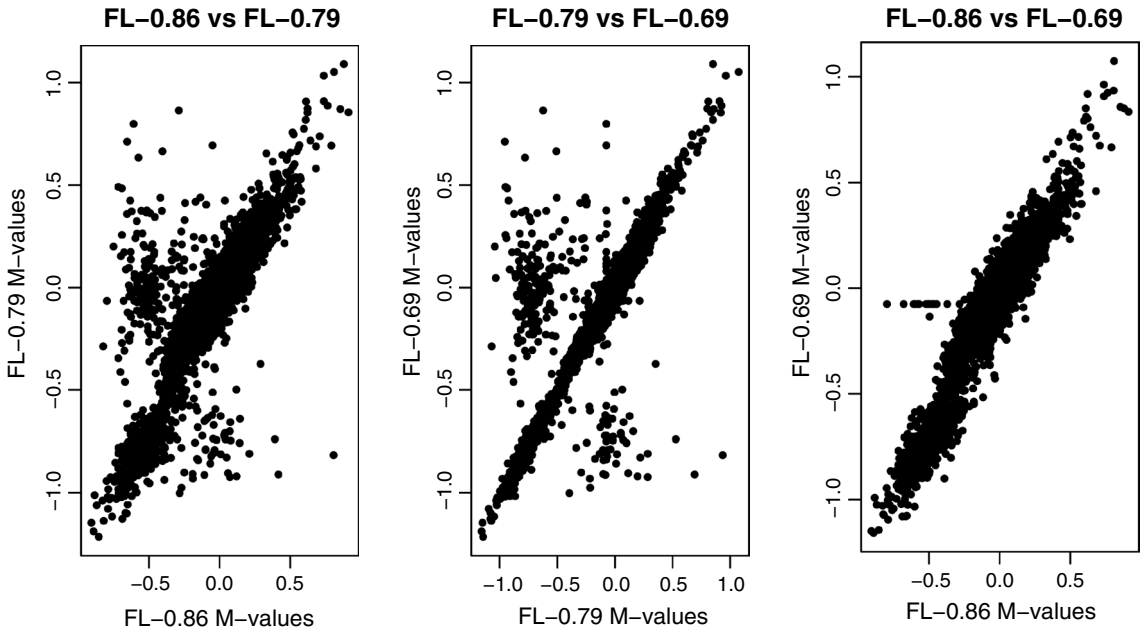

Fig. 4 *M*-value correlations between different deletions lines

Table 3 Bin mapping of sequences with proportion of gene hits in deletion bins

Arm	Bin	No of sequences		Mapped Kbp	Gene hits per Kbp	Number of unique gene hits
		7BS	7BL			
7BS	BS	1,643	0			
	BS_0.27 – 1.00	1,262	0	26,807	0.023	607
	BS_0 – 0.27	211	0	5,318	0.011	60
7BL	BL_0 – 0.14	1	407	8,222	0.010	83
	BL_0.14 – 0.33	0	64	1,118	0.016	18
	BL_0.33 – 0.4	0	233	4,494	0.019	85
	BL_0.4 – 0.45	0	127	2,268	0.020	46
	BL_0.45 – 0.63	0	312	6,420	0.022	138
	BL_0.63 - 0.69*	0	674	12,909	0.018	241
	BL_0.69* - 1.00	0	381	6,574	0.021	136
	BL	0	1,906			

BL_0.69* - 1.00 contains sequences mapped to BL_0.69 – 1.00, BL_0.79 - 1.00 and BL_0.86 - 1.00

Validation of the bin mapping results

In order to verify the accuracy of our bin mapping results we used three different approaches. First, PCR-based ISBP markers were designed from randomly selected bin mapped 7B sequences and used to validate the bin assignment. Out of 12 markers screened, ten were absent in relevant deletion lines. Two markers were mapped to all deletion lines tested, however, there were clear band intensity difference (Supplementary material 1). In the second validation approach we compared our bin mapping results with previously mapped ESTs (Hossain et al. 2004). Seventy-one percent of the sequences were assigned to the same bin as reported earlier (Hossain et al. 2004). The remaining 29 % of sequences (20 sequences out of 69 tested) were mapped to different deletion bin than previously reported. In order to determine if this discrepancy represented error in our CGH-based bin mapping, PCR-based ISBP markers were designed from sequence scaffolds from which the mapped ESTs were derived (determined by BLASTN). Twelve out of the 20 ISBP markers were 7B specific and used in PCR reactions with CS deletion lines. Eleven of these assays supported our CGH-based results, while only one supported the results from Hossain et al. (2004) (Fig. 5). The third approach to confirm the results of CGH-based bin mapping involved a comparison of the positions of marker sequences of bin maps with a genetic map. A RIL mapping population was used to construct genetic map of

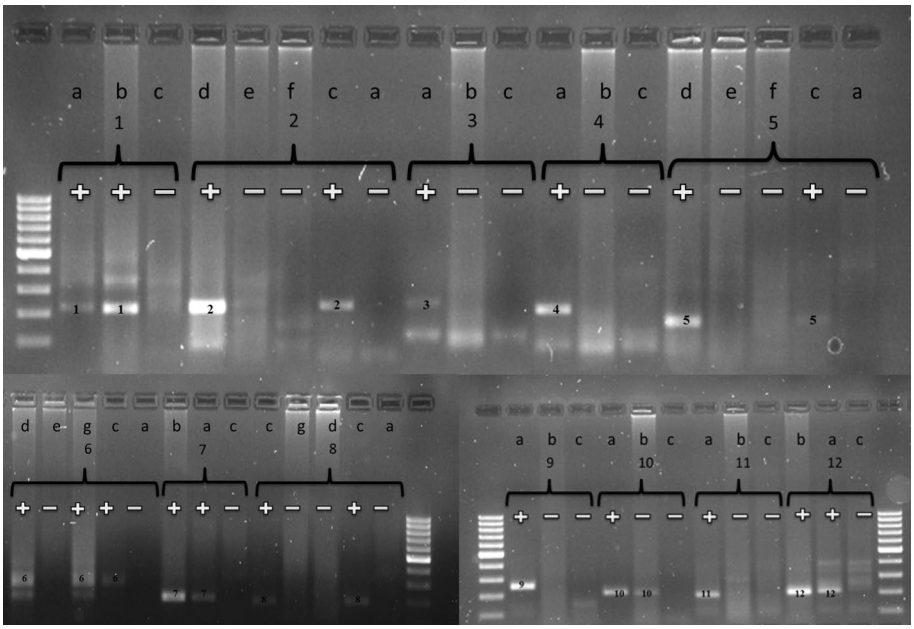

Fig. 5 Agarose gel profile of PCR products amplified by ISBP primers designed from sequences assigned to deletion bins. *Deletion lines* used are: *a*-DT7BS, *b*-FL-0.27, *c*-DT7BL, *d*-FL-0.63, *e*-FL-0.45, *f*-FL-0.33, *g*- FL-0.79. *Numbers* indicate the primer pairs (supplementary material 3). Plus and minus indicate expected presence or absence of bands in the corresponding CS deletion lines according to the deletion bin mapping results

chromosome 7B consisting of 629 SNP markers representing 225 unique loci (total map length 180 cM). A total of 116 markers on the genetic map could be assigned to 70 sequences mapped across five deletion bins of the wheat 7B (Fig. 6). An almost perfect relationship between the order of markers on the genetic map and the order of deletion bin assignment was identified by visual inspection. The only exception was marker wsnp_BE443010B_Ta_2_2 (see dotted line with aestrix in Fig. 6), which was mapped to the deletion bin BL_0.45 − 0.63 (Supplementary material 2), while the surrounding markers were assigned to BL_0.63 − 0.69*. Therefore, the comparative analysis with previously bin mapped ESTs, 7B genetic map, and PCR verification suggest high reliability of our approach of the contig and scaffold sequence assignation to deletion bins.

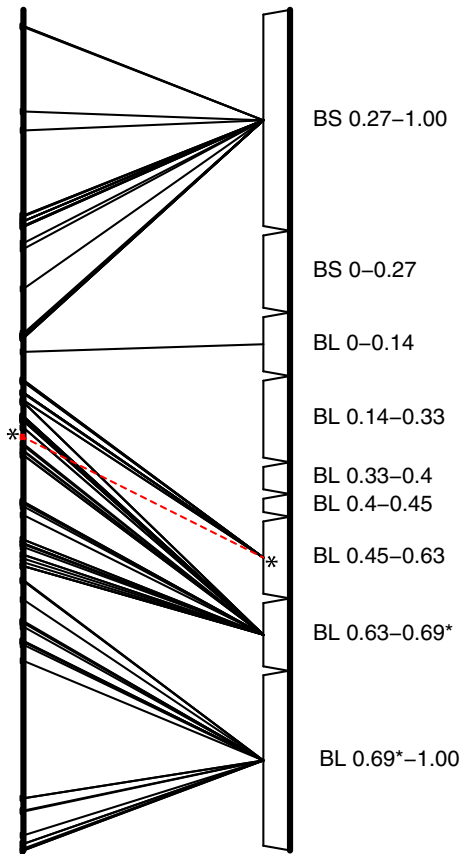

Fig. 6 Comparison of the 7B genetic map and deletion bin map

Discussion

The development of a high quality and high resolution integrated physical and genetic map for the allohexaploid genome of bread wheat represents a significant challenge. Various mapping data, including recombination based mapping, radiation hybrid mapping, synteny-based mapping and deletion bin mapping is deemed necessary to anchor and order physical contigs (Paux et al. 2008; Philippe et al. 2013). In this study a high-density deletion bin map of chromosome 7B placing ~7.8 % of the 7B sequence into nine deletion bins was developed using high-throughput Nimblegen CGH microarray platform. This work represents an important step towards a physically ordered 7B reference sequence. In addition, high-density deletion bin map of 7B may serve as an excellent resource of new markers for fine mapping and map based cloning of genes/QTL located on chromosome 7B.

CGH design, PAV genotyping, and error rates

Deletion calling in a hexaploid genome with hybridization based methods is inherently difficult due to background signals from highly similar homeologous DNA sequences. It is therefore critical to maximize the specificity of probe hybridization. As a first step we used the CSS assemblies to perform a BLAST-based filtering step to remove probe sequences with high sequence similarity with 7A and 7D chromosomes. Furthermore, we included different classes of genomic probes on the CGH array with potential differences in sub-genome specificity. Interestingly, the ISBP-based probes known to be highly sub-genome specific (Choulet et al. 2010) did not differ significantly in hybridization properties compared to RG probes. Thus both probe types can be successfully used in CGH experiments in hexaploid wheat.

Mclust method was used to classify probes based on their *M*-values as a DNA sequence that was either "present" or "absent" (i.e. deleted). It was evident that this method had low power when applied to single probes (i.e. 50 bp replicated four times per CGH array) (Fig. 3). However, when multiple co-localized probes were used to estimate *M*-values for whole sequence contigs and/or scaffolds we could successfully assign DNA sequences to chromosome arms and deletion bins (Table 3) with low error rates. Frequencies of incorrect assignment of sequences to chromosome arms were very low, but slightly higher when we used two (0.9 %) compared to three probes per estimated *M*-value (0.03 %) (data not shown). Furthermore, three different CGH-independent verification approaches also supported a high level of accuracy for bin assignment. In the PCR-based validation experiments, 83 % of the markers were unambiguously mapped to the expected deletion bins. The remaining markers mapped to all deletion lines tested, however, a clear band intensity difference between deletion lines indicates a difference in DNA content (i.e. deletion) as predicted from the CGH results (Supplementary material 1). Next, a comparison between CGH results and previous bin mapped ESTs revealed 29 % discordance; however, additional PCR based assays revealed that only 8 % of this discordance suggested error in our CGH results. The comparison with previous EST bin mapping therefore suggests an error rate of approximately 2.5 % ($0.29 \times 0.08 = 0.023$). Another validation of the accuracy of our results was provided by comparison with a genetic map of 7B. The comparison of sequence order from a genetic map with the bin map showed that only 1 out of 116 (0.8 %) sequences in the genetic map did not concur with the bin map order. In conclusion, the results from validation experiment suggest an error rate of <2.5 %.

Discrepancy between cytological and genetic estimates of deletion sizes

A strong overall correlation between the observed and expected proportion of deleted probes was found for the aneuploid lines (Table 2). However, three deletion lines (FL_0.69, FL_0.79 and FL_0.86) showed discrepancies in the ranking of deletion size and that these lines have both presence and absence variation relative to each other (Table 2; Fig. 4). These results agree with earlier reports that these three lines contain terminal deletions combined with interstitial deletions rather than single terminal deletions (Hohmann et al. 1995).

Conclusion

The high-density deletion bin map of wheat chromosome 7B was successfully constructed by genotyping aneuploid wheat stocks using 270 K CGH Nimblegen microarray. Using the most recently published chromosome survey sequences of bread wheat A, B and D sub-genomes (IWGSC, 2014) we could design 7B specific CGH probes, and accurately assign a total of 3,671 sequence contigs and scaffolds (~8 % of the chromosome 7B) to nine chromosomal bins. This map is the highest density deletion bin map for 7B so far, representing a ~100× increase in the bin mapped 7B sequence compared to previous studies (Hossain et al. 2004) and represents an important step towards high-resolution physical map of 7B.

Author contributions TB was responsible for carrying out experimental work, participated in data analysis and writing the manuscript, LG performed the normalization of hybridization data, AK and SK participated in drafting the manuscript, XH performed the clustering of SNP data in the RIL population, ML developed the RIL mapping population, worked on clustering SNP data in the RIL population, NS participated in the design of the experiment, data analysis and drafting the manuscript, SL and OA helped coordinate the study and draft the manuscript, SS performed data analysis, helped to draft the manuscript and was responsible for final version of the manuscript.

Acknowledgments The project was funded by grants from the Norwegian Research Council (project no. 199387/I99) and Graminor A/S to Odd-Arne Olsen.

Conflict of interest The authors declare they have no conflict interests.

References

Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao SM, Anderson OD, Linkiewicz AM, Dubcovsky J, La Rota M, Sorrells ME, Zhang DS, Nguyen HT, Kalavacharla V, Hossain K, Kianian SF, Peng JH, Lapitan NLV, Gonzalez-Hernandeiz JL, Anderson JA, Choi DW, Close TJ, Dilbirligi M, Gill KS, Walker-Simmons MK, Steber C, McGuire PE, Qualset CO, Dvorak J (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13(5):753–763. doi:10.1101/Gr.808603

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2

Belova T, Zhan B, Wright J, Caccamo M, Asp T, Simkova H, Kent M, Bendixen C, Panitz F, Lien S, Dolezel J, Olsen OA, Sandve SR (2013) Integration of mate pair sequences to improve shotgun assemblies of flow-sorted chromosome arms of hexaploid wheat. BMC Genom 14:222. doi:10.1186/1471-2164-14-222

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier MC, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22(6):1686–1701. doi:10.1105/tpc.110.074187

Devos KM, Dubcovsky J, Dvorak J, Chinoy CN, Gale MD (1995) Structural evolution of wheat chromosomes 4a, 5a, and 7b and its impact on recombination. Theor Appl Genet 91(2):282–288. doi:10.1007/Bf00220890

Endo TR, Gill BS (1996) The deletion stocks of common wheat. J Hered 87(4):295–307

Fraley C, Raftery AE (2007) Bayesian regularization for normal mixture estimation and model-based clustering. J Classif 24(2):155–181. doi:10.1007/S00357-007-0004-5

Hohmann U, Endo TR, Herrmann RG, Gill BS (1995) Characterization of deletions in common wheat induced by an Aegilops cylindrica Chromosome: detection of multiple chromosome rearrangements. Theor Appl Genet 91(4):611–617

Hossain KG, Kalavacharla V, Lazo GR, Hegstad J, Wentz MJ, Kianian PM, Simons K, Gehlhar S, Rust JL, Syamala RR, Obeori K, Bhamidimarri S, Karunadharma P, Chao S, Anderson OD, Qi LL, Echalier B, Gill BS, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Akhunov ED, Dvorak J, Miftahudin, Ross K, Gustafson JP, Radhawa HS, Dilbirligi M, Gill KS, Peng JH, Lapitan NL, Greene RA, Bermudez-Kandianis CE, Sorrells ME, Feril O, Pathan MS, Nguyen HT, Gonzalez-Hernandez JL, Conley EJ, Anderson JA, Choi DW, Fenton D, Close TJ, McGuire PE, Qualset CO, Kianian SF (2004) A chromosome bin map of 2148 expressed sequence tag loci of wheat homoeologous group 7. Genetics 168(2):687–699. doi:10.1534/genetics.104.034850

International Barley Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491(7426):711–716. doi:10.1038/nature11543

International Rice Genome Sequencing Consortium (2005) The map-based sequence of the rice genome. Nature 436(7052):793–800. doi:10.1038/nature03895

International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345(6194):1251788

Joppa LR, Williams ND (1977) D-genome substitution-monosomics of durum-wheat. Crop Sci 17(5):772–776

Kumar S, Balyan HS, Gupta PK (2012) Comparative DNA sequence analysis involving wheat, brachypodium and rice genomes using mapped wheat ESTs. Triticeae Genomic Genetic 3(3):25–37

Lazo GR, Chao S, Hummel DD, Edwards H, Crossman CC, Lui N, Matthews DE, Carollo VL, Hane DL, You FM, Butler GE, Miller RE, Close TJ, Peng JH, Lapitan NL, Gustafson JP, Qi LL, Echalier B, Gill BS, Dilbirligi M, Randhawa HS, Gill KS, Greene RA, Sorrells ME, Akhunov ED, Dvorak J, Linkiewicz AM, Dubcovsky J, Hossain KG, Kalavacharla V, Kianian SF, Mahmoud AA, Miftahudin, Ma XF, Conley EJ, Anderson JA, Pathan MS, Nguyen HT, McGuire PE, Qualset CO, Anderson OD (2004) Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L.): EST generation, unigene analysis, probe selection and bioinformatics for a 16,000-locus bin-delineated map. Genetics 168(2):585–593. doi:10.1534/genetics.104.034777

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457(7229):551–556. doi:10.1038/nature07723

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322(5898):101–104. doi:10.1126/science.1161847

Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant JP, Sourdille P, Balfourier F, Le Paslier MC, Chauveau A, Cakir M, Gandon B, Feuillet C (2010) Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol J 8(2):196–210. doi:10.1111/j.1467-7652.2009.00477.x

Philippe R, Paux E, Bertin I, Sourdille P, Choulet F, Laugier C, Simkova H, Safar J, Bellec A, Vautrin S, Frenkel Z, Cattonaro F, Magni F, Scalabrin S, Martis MM, Mayer KF, Korol A, Berges H, Dolezel J, Feuillet C (2013) A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. Genome Biol 14(6):R64. doi:10.1186/gb-2013-14-6-r64

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475(7355):189–195. doi:10.1038/nature10158

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorak J, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Bermudez-Kandianis CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrells SF, Sorrells ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma XF, Miftahudin, Gustafson JP, Conley EJ, Nduati V, Gonzalez-Hernandez JL, Anderson JA, Peng JH, Lapitan NLV, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi DW, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 168(2):701–712. doi:10.1534/genetics.104.034868

Saintenac C, Falque M, Martin OC, Paux E, Feuillet C, Sourdille P (2009) Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). Genetics 181(2):393–403. doi:10.1534/genetics.108.097469

Sears ER (1954) The aneuploids of common wheat. College of Agriculture, Agricultural Experimental Station. Res Bull 572:1–58

Smit AFA, Hubley R, Green P (1996–2010) RepeatMasker Open-3.0. http://www.repeatmasker.org.

Springer NM (2010) Isolation of plant DNA for PCR and genotyping using organic extraction and CTAB. Cold Spring Harb Protoc 2010(11):pdb prot5515. doi:10.1101/pdb.prot5515

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, International Wheat Genome Sequencing C, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796. doi:10.1111/pbi.12183

Werner JE, Endo TR, Gill BS (1992) Toward a cytogenetically based physical map of the wheat genome. Proc Natl Acad Sci USA 89(23):11307–11311

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet 4(10):e1000212. doi:10.1371/journal.pgen.1000212