



# A High-Density Genetic Map with Array-Based Markers Facilitates Structural and Quantitative Trait Locus Analyses of the Common Wheat Genome

Iehisa, Julio Cesar Masaru ; Ohno, Ryoko ; Kimura, Tatsuro ; Enoki, Hiroyuki ; Nishimura, Satoru ; Okamoto, Yuki ; Nasuda, Shuhei ; Takumi...

---

(Citation)

DNA Research, 21(5):555-567

(Issue Date)

2014-10

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

© The Author 2014. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), ...

(URL)

<https://hdl.handle.net/20.500.14094/90002721>



# A High-Density Genetic Map with Array-Based Markers Facilitates Structural and Quantitative Trait Locus Analyses of the Common Wheat Genome

JULIO CESAR MASARU Iehisa<sup>1,†</sup>, RYOKO Ohno<sup>2,†</sup>, TATSURO Kimura<sup>3</sup>, HIROYUKI Enoki<sup>3</sup>, SATORU Nishimura<sup>3</sup>, YUKI Okamoto<sup>1</sup>, SHUHEI Nasuda<sup>4</sup>, and SHIGEO Takumi<sup>1,\*</sup>

Laboratory of Plant Genetics, Graduate School of Agricultural Science, Kobe University, Nada-ku, Kobe 657-8501, Japan<sup>1</sup>; Core Research Division, Organization of Advanced Science and Technology, Kobe University, Kobe, Japan<sup>2</sup>; Bio Research Laboratory, TOYOTA Motor Corporation, Toyota, Aichi 471-8572, Japan<sup>3</sup> and Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan<sup>4</sup>

\*To whom correspondence should be addressed. Tel. +81 78-803-5860. Fax. +81 78-803-5860.  
Email: takumi@kobe-u.ac.jp

Edited by Prof. Kazuhiro Sato  
(Received 5 March 2014; accepted 26 May 2014)

## Abstract

**The large genome and allohexaploidy of common wheat have complicated construction of a high-density genetic map. Although improvements in the throughput of next-generation sequencing (NGS) technologies have made it possible to obtain a large amount of genotyping data for an entire mapping population by direct sequencing, including hexaploid wheat, a significant number of missing data points are often apparent due to the low coverage of sequencing. In the present study, a microarray-based polymorphism detection system was developed using NGS data obtained from complexity-reduced genomic DNA of two common wheat cultivars, Chinese Spring (CS) and Mironovskaya 808. After design and selection of polymorphic probes, 13,056 new markers were added to the linkage map of a recombinant inbred mapping population between CS and Mironovskaya 808. On average, 2.49 missing data points per marker were observed in the 201 recombinant inbred lines, with a maximum of 42. Around 40% of the new markers were derived from genic regions and 11% from repetitive regions. The low number of retroelements indicated that the new polymorphic markers were mainly derived from the less repetitive region of the wheat genome. Around 25% of the mapped sequences were useful for alignment with the physical map of barley. Quantitative trait locus (QTL) analyses of 14 agronomically important traits related to flowering, spikes, and seeds demonstrated that the new high-density map showed improved QTL detection, resolution, and accuracy over the original simple sequence repeat map. **Key words:** array-based genotyping; chromosomal synteny; high-density genetic map; next-generation sequencing; QTL analysis**

## 1. Introduction

Common wheat (*Triticum aestivum* L., genome constitution AABBDD) is an allohexaploid species that originated from interspecific hybridization between tetraploid wheat (*Triticum turgidum* L., AABB) and the wild diploid relative, *Aegilops tauschii* Coss. (DD).<sup>1,2</sup>

In common wheat, its large genome (17 Gb) and allopolyploidy have complicated the construction of high-resolution genetic maps. Diversity arrays technology (DArT), a microarray hybridization-based technique, has been developed in common wheat for high-throughput genotyping without relying on sequence information.<sup>3</sup> DArT generates whole-genome fingerprints by scoring the presence versus absence of DNA fragments in genomic representations generated by the digestion of genomic DNA with a combination

<sup>†</sup> These authors contributed equally to this work.

of two restriction endonucleases.<sup>4</sup> Using the wheat *Pst*I(TaqI) v2.6 array, 1,348 DArT markers were mapped in a hexaploid Synthetic W7984 × Oyata M85 doubled haploid population.<sup>5</sup> The most recent common wheat *Pst*I(TaqI) v3.0 array comprises about 7,000 markers that are polymorphic in a wide range of wheat cultivars.<sup>6</sup>

Recent advances in next-generation sequencing (NGS) technology have made it possible to screen a huge number of polymorphic sites even in species without reference genome information. Single-nucleotide polymorphisms (SNPs) were discovered by sequencing of a complexity-reduced fraction of the tetraploid wheat genome<sup>7</sup> or cDNA of common wheat,<sup>8–10</sup> and have been used to develop high-throughput SNP-typing platforms such as BeadExpress,<sup>7</sup> KASPar,<sup>8</sup> and Infinium.<sup>10</sup> Using an Illumina 9K iSelect SNP assay, 7,504 polymorphic loci were positioned in a consensus map constructed based on integrating information on seven common wheat mapping populations.<sup>10</sup>

Improvements in the throughput of NGS technologies made it possible to conduct SNP typing of an entire mapping population by direct sequencing, combining the process of polymorphism discovery and genotyping in a single experiment. In tetraploid and hexaploid wheat, sequence-based genotyping has been performed by sequencing of complexity-reduced genomic DNA<sup>11–13</sup> or a targeted genomic region.<sup>14</sup> The former, called complexity reduction of polymorphic sequences (CRoPS), restriction site-associated DNA sequencing (RAD-seq), or genotyping-by-sequencing (GBS),<sup>15</sup> was applied to the development of high-density genetic maps ranging from >10,000<sup>11</sup> to >400,000 markers<sup>12</sup> in the hexaploid wheat mapping population. In the latter, the targeted genomic region (for example, exome) can be captured and re-sequenced.<sup>14</sup> Sequence-based genotyping is cost-effective and faster than array-based platforms, but often have a significant amount of missing data due to the low coverage of sequencing.<sup>15</sup>

To efficiently construct a high-resolution map in common wheat with few missing data points, we developed a microarray-based polymorphism detection system using NGS data obtained from complexity-reduced genomic DNA samples of two common wheat cultivars, Chinese Spring (CS) and Mironovskaya 808 (M808), and genotyped 210 recombinant inbred lines (RILs) between the two cultivars. Moreover, quantitative trait locus (QTL) analysis was performed for 14 morphological and agronomically important traits using the high-density map constructed.

## 2. Materials and methods

### 2.1. Plant material

CS and M808 were selected because of the difference in growth habit, geographical origin and abiotic stress

responses, and the availability of RILs derived from them. M808 was bred in Mironovska Institute, Ukraine. It is reported to be one of the most freezing tolerant wheat cultivars tested.<sup>16</sup> In contrast, CS is a spring-type wheat landrace from Sichuan Province in China. A mapping population of 210 RILs derived from a cross between two common wheat cultivars, CS and M808, developed by Kobayashi *et al.*,<sup>16</sup> was used in this study. The plants were grown individually in pots arranged randomly in a field of Kobe University (34°43'N, 135°13'E) in two seasons (2009–10 [Kobe10, indicating location and growing season] and 2011–12 [Kobe12]), and in a glasshouse of Kyoto University (35°2'N, 135°47'E) in the 2011–12 season (Kyoto12). Total DNA from each RIL and the parents was extracted from leaves using standard procedures.<sup>16</sup>

### 2.2. Library preparation and NGS

DNA samples (180 ng) of CS and M808 were digested first with *Pst*I (New England BioLabs, Inc., Ipswich, MA, USA) at 37°C for 1 h and then with *Bst*NI (New England BioLabs) at 60°C for 1 h. We selected this combination of restriction enzymes based on *in silico* analysis in rice, our experience in sugarcane,<sup>17,18</sup> and its use in barley.<sup>4</sup> Digested DNA was ligated to *Pst*I adapters (5'-CACGATGGATCCAGTGA-3' and 5'-CTGGATCCATCGTGCA-3') using T4 DNA ligase (Nippon Gene, Tokyo, Japan) and ATP (Wako Pure Chemical Industries, Osaka, Japan). The ligation reaction was run at 16°C for 16 h and the ligase was inactivated by holding at 60°C for 20 min.

Ligated samples were PCR-amplified for 30 cycles using 5'-GATGGATCCAGTGCAG-3' with PrimeSTAR HS DNA Polymerase (Takara Bio, Shiga, Japan) under the following conditions: 10 s at 98°C, 15 s at 55°C, and 60 s at 72°C. PCR products were purified using a MinElute PCR Purification Kit (Qiagen, Hilden, Germany). The constructed libraries were sequenced as 2 × 100 nt paired-end reads on the Genome Analyzer IIx instrument (Illumina, San Diego, CA, USA). The sequences were deposited to the DNA Data Bank of Japan (DDBJ) Short Read Archive (DRA001257).

### 2.3. Processing of raw reads and design of probes

First, trimming of adaptor and primer sequences and low-quality 3' ends (quality score ≤ 2) from raw reads was performed using an in-house developed software. Reads with the length of < 70 nt were discarded including pairs. CS and M808 libraries were *de novo* assembled separately using Velvet version 1.1.05.<sup>19</sup> Velvet was run with a k-mer size of 65, and velvetg with expected coverage: 'auto', coverage cut-off: 'auto', insert length of 220 bp, minimum contig length of 100 bp, and allowing scaffolding. From contigs and unused reads

(in assembly), tiling probes were designed with a  $T_m$  of  $\sim 76^\circ\text{C}$  and the length of 50–70 bp.

To construct a primary array, the designed probes were first aligned to filtered reads of CS and M808 using Bowtie version 0.12.7,<sup>20</sup> where no more than three mismatches were allowed and up to 40 alignments were reported ( $-v\ 3$ ,  $-k\ 40$ ). CS-derived probes that met all the following five conditions were selected: (i) the number of aligned CS reads with no mismatch was  $\geq 5$ , (ii) no aligned CS reads with 1–3 mismatches, (iii) no aligned M808 reads with a perfect match, (iv) no aligned M808 reads with one or two mismatches, and (v) the number of aligned M808 reads with three mismatches was  $\leq 2$ . The last condition was included because this array-based polymorphism detection system has been demonstrated to accurately detect differences of more than three nucleotides.<sup>17,18</sup> A similar selection was performed with M808 probes. In total, 223,000 probes from both CS and M808 were spotted in triplicate on a NimbleGen  $3 \times 1.4\ \text{M}$  array (Roche Diagnostics, Mannheim, Germany).

#### 2.4. Hybridization and construction of secondary array

To select probes for the construction of a secondary array, CS, M808, and 14 randomly selected RILs were genotyped. After digestion of total DNA with *Pst*I and *Bst*NI, adapter ligation, PCR amplification and purification of PCR products as described in the library preparation section, they were labelled with Cy3 using the NimbleGen One-Color DNA Labeling Kit (Roche Diagnostics) according to the manufacturer's instructions. Hybridization was performed at  $42^\circ\text{C}$  for 72 h on a NimbleGen Hybridization System (Roche Diagnostics). Arrays were scanned using a NimbleGen MS200 Microarray Scanner (Roche Diagnostics) and genotype calling was performed based on the signal intensities.

For selection of probes, a  $\chi^2$  test was used to compute segregation of each probe in the 14 RILs. Those deviating from Mendelian 1 : 1 segregation at a 1% significance level were discarded. The secondary array was constructed using one selected probe per contig or read, consisting of 21,346 CS-derived probes, 21,205 M808-derived probes, and 2,303 probes for normalization. These 44,854 probes were spotted in triplicate on a NimbleGen  $12 \times 135\ \text{K}$  array (Roche Diagnostics). This array was used for genotyping of 210 RILs and the two parental lines.

#### 2.5. Genetic map construction

The genotyping data for 422 simple sequence repeat (SSR) markers were previously generated.<sup>16</sup> A total of 140 barc, 1 cfa, 34 cfd, 18 gdm, 109 gwm, 4 hbd, 15 hbe, 41 hbg, and 60 wmc loci were used for chromosomal assignment of wheat array-based markers

(WABMs) developed in this study. *VRN-D1* was mapped using primer sets 5'-CTGGTTGTCTGCCTCA TCA-3' and 5'-CTCTCTCCCCCTGCTAGT-3', and a second copy of *Ppd-B1*<sup>21</sup> using 5'-TAACTGCTCGTCA CAAGTGC-3' and 5'-CCGGAACCTGAGGATCATC-3'. Loci were assembled into linkage groups using Antmap version 1.2<sup>22</sup> with the 'all combination' method and different threshold distances (5.75, 12.0, 17.5, and 25.0 cM). Loci were ordered using MapDisto version 1.7.7,<sup>23</sup> and the genetic distances were calculated with the Kosambi function.<sup>24</sup>

#### 2.6. BLAST searches of the mapped contigs and reads

To search repeat elements (REs) in the mapped contigs and reads, they were used as queries against the Munich Information Center for Protein Sequences (MIPS) RE database version 9.0 for Poaceae (<ftp://ftp.mips.helmholtz-muenchen.de/plants/REdat/>) using the blastn algorithm. A blastn search was also performed against *Ae. tauschii*,<sup>25</sup> *Triticum urartu*,<sup>26</sup> and barley high-confidence genes and genomic sequences.<sup>27</sup> Contigs and reads mapping to the A-genome were searched by the blastn algorithm against the *T. urartu* draft genome sequence, while those mapping to the D-genome were searched against the *Ae. tauschii* draft genome sequence.<sup>25,26,28</sup> All BLAST hits were filtered with an *E*-value cut-off of  $10^{-5}$  and hit length of  $\geq 50\ \text{bp}$ . The genetic map and the results of the RE search were represented in a circular plot using the Circos version 0.63 software.<sup>29</sup>

#### 2.7. In silico digestion of the draft genome sequences of *T. urartu* and *Ae. tauschii*

Using an in-house PERL script, genomic scaffolds of *T. urartu* and *Ae. tauschii* were fragmented at *Pst*I and *Bst*NI sites, and those having only *Pst*I sites at both ends but not *Bst*NI sites were extracted. A blastn search was performed against the MIPS RE database version 9.0 for Poaceae using as a query the extracted *T. urartu* and *Ae. tauschii* genomic fragments with lengths between 65 and 6,000 bp (*E*-value  $< 10^{-5}$  and hit length  $\geq 50\ \text{bp}$ ).

#### 2.8. In silico mapping to the physical map of barley

For reads and contigs mapping to the A- and D-genomes, a blastn search was first conducted against draft genomic sequences (*T. urartu* for the A-genome and *Ae. tauschii* for the D-genome)<sup>25,26</sup> as described above. The obtained scaffolds were mapped based on a blastn search of the genes contained in these scaffolds against the barley genome (*E*-value  $< 10^{-5}$  and hit length  $\geq 50\ \text{bp}$ ). The A- and D-genome scaffolds without genes as well as reads and contigs mapping to the B-genome were searched against the barley genome using the same blastn parameters.



## 2.9. Phenotypic measurements

A list of traits, growing season, and location is presented in Supplementary Table S1. Heading time (HT) and flowering time (FT) were recorded as days after sowing. Maturation time (MT) was evaluated as the number of days that had passed when the ear neck turned yellow. HT, FT, and MT were measured for the three earliest tillers of each plant, and mean values were calculated. Grain-filling period (GFP) was defined as the number of days from flowering to maturation.

Spike-related traits (such as spike length [SL], number of spikelets per spike [SpN], SL per number of spikelets [SLperSpN], and top five spikelets length [T5SpL]) and culm length (CL) were assessed for the three earliest tillers of each plant before anthesis. Seed-related traits such as seed length (SdL), seed width (SdW), seed height (SdH), and seed length-to-width ratio (SdLperW) were measured using seeds on the central spikelets of the first, second, and third spikes. For each spikelet, only the components of the first and second florets were used. The mean of the replicated measurements over three spikes was calculated for each trait and used as the trait value in subsequent analyses. SdL, SdW, SdH, and SdLperW were evaluated using at least 10 seeds per line. For tiller number (TN), the number of tillers bearing ears was counted at harvest.

## 2.10. QTL analysis

QTLs were analysed by composite interval mapping using R/qtl package version 1.21-2<sup>30</sup> for the new map and for that constructed by Kobayashi *et al.*<sup>16</sup> using SSR markers. This SSR map was constructed using 412 SSR markers. First, QTL genotype probability was calculated using the function *calc.genoprob* with a step size of 1 cM and Kosambi map function. QTL analysis was performed with *cim* function using three marker covariates and Kosambi map function. The log-likelihood (LOD) score threshold was determined by computing a 1,000 permutation test for all traits. The percentage of phenotypic variation (PV) explained by a QTL for a trait was estimated using the *fitqtl* function in R/qtl.

## 3. Results

### 3.1. Construction of a microarray

To design probes for the development of a microarray, we sequenced complexity-reduced genomic DNA samples from two common wheat cultivars, CS and M808. Genome complexity was reduced using a combination of two restriction enzymes, *Pst*I and *Bst*NI. After ligation of adapters with sticky ends complementary to the 3'-overhang created by *Pst*I, only the DNA fragments having a *Pst*I site at both ends, with no *Bst*NI site, were amplified by PCR. After trimming and filtering of raw reads, 63,563,396 (6.04 Gb, genome

coverage of  $0.36 \times$  [Supplementary Table S2]) and 63,961,698 (6.09 Gb,  $0.36 \times$ ) reads were generated for CS and M808, respectively (Fig. 1). Both libraries were independently assembled using Velvet,<sup>19</sup> yielding 84,985 contigs and 39,324,373 unused reads during assembly for CS and 123,680 contigs and 33,006,701 unused reads for M808.

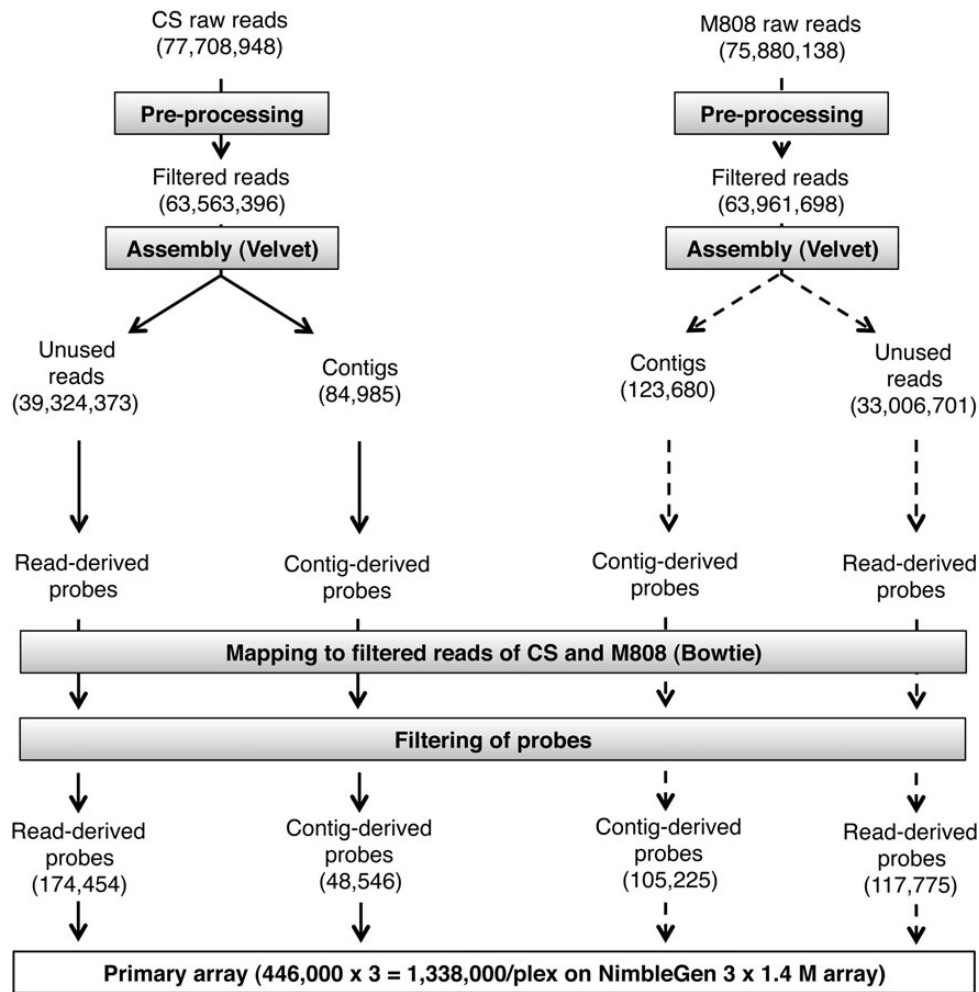
From contigs and unused reads, tiling probes were designed. The probes were aligned with the filtered reads of CS and M808, and only the probes that predicted detection of 'presence' or 'absence' alleles were selected (see Section 2.3.). Thus, the primary array contained 1,338,000 probes per plex (446,000 probes in triplicate) on a  $3 \times 1.4$  M array (Fig. 1) with a genome coverage of  $4.51 \times 10^{-3}$  fold.

The primary arrays were used for genotyping of CS, M808, and their 14 derived RILs, and of the 446,000 probes, 303,387 were polymorphic. Probes showing segregation distortion from a 1 : 1 ratio at a 1% significance level were discarded and only one probe per contig or read was selected. The final array contained 44,854 probes (including 2,303 probes for normalization, genome coverage of  $7.45 \times 10^{-4}$  fold) in triplicate on a  $12 \times 135$  K array and was used for genotyping of CS, M808, and 210 RILs. We named these microarray probes as WABMs.

### 3.2. Construction of a high-density linkage map

In a previous study, a linkage map was constructed for 210 RILs of CS and M808 using 422 SSR loci.<sup>16</sup> Of the 42,551 microarray probes, polymorphism was observed for 15,530 (36.5%) with a minor allele frequency of  $\geq 10\%$  among the 210 RILs. Missing data points were counted in each of the RILs and ranged from 17 to 1,074, with an average of  $159.0 \pm 153.0$  (Supplementary Fig. S1A). To use high-quality data in further analyses, nine RILs with  $\geq 465.1$  (mean + 2 SD) missing data points were excluded. Missing data points were also counted for each marker and ranged from 0 to 42, with an average of  $2.49 \pm 5.84$  (Supplementary Fig. S1B). Markers with >20 missing data points were excluded, and 14,464 markers were used for genetic map construction.

In total, 13,480 markers, including 13,056 WABMs, 422 SSRs, *VRN-D1*, and *Ppd-B1*, were grouped into 21 linkage groups. The remaining markers did not form linkage group or the linkage group was small (<10 cM) without chromosomal assignment. Of the mapped WABMs, 7,253 (55.6%) were probes derived from contigs. The probe sequences for these markers are presented in Supplementary Table S3. The number of total markers per chromosome ranged from 1,256 on chromosome 2B to 85 on chromosome 4D (Table 1). The number of markers with a non-redundant (nr) genotype pattern was 5,243, comprising 2,744 loci



**Figure 1.** Schematic of primary array development. Raw reads obtained from sequencing complexity-reduced CS and M808 genomic DNA were pre-processed for quality filtering prior to assembly with Velvet. Tiling probes were designed from contigs and unused reads. To predict non-polymorphic probes, alignment to filtered CS and M808 reads was performed. The primary array was constructed by spotting 446,000 filtered probes in triplicate. See Materials and methods for more details.

(markers with the same position on genetic map). The total map length was 5,093.12 cM, with an average marker interval of 0.38 cM when considering all markers, 0.97 cM per nr marker, and 1.86 cM per locus.

Segregation distortion was computed for each marker and  $P$ -values of  $< 1.82 \times 10^{-5}$  (Bonferroni corrected  $P$ -value: 0.05/2,744 loci) from a  $\chi^2$  test (1 : 1 ratio) were observed for markers of chromosomes 2B, 2D, 5A, 5B, 6B, and 7B (Fig. 2, Tracks 6 and 7; Supplementary Table S4). The greatest distortion was observed at the telomeric region of chromosome 6BL (235.8 cM), where the allele frequency was 89.5% for CS and 10.5% for M808.

### 3.3. Annotation of mapped contigs and reads

To annotate mapped contigs or reads, blastn searches were conducted against protein-encoding genes of *Ae. tauschii*, *T. urartu*, and barley, and against the RE database of the Poaceae of the MIPS. Of the 13,056

contigs/reads, 32.5, 27.0, and 23.9%, respectively, had hits to the *Ae. tauschii*, *T. urartu*, and barley genes (Fig. 3 and Supplementary Table S5). Overall, 39.8% of the contigs/reads had at least one hit in one of these three databases. These contigs/reads derived from genic regions were distributed all across the wheat genome and chromosomes (Fig. 2, Track 2; Supplementary Table S6).

REs were found in 11.4% of the mapped contigs/reads (Fig. 3). The proportion of DNA transposons and retroelements was similar (Table 2), and no bias in the distribution was found for these REs (Fig. 2, Tracks 3 and 4; Supplementary Table S6). The CACTA superfamily (present in 38.7% of the probes) was mainly found among the DNA transposons followed by the miniature inverted-repeat transposable element (MITE) family (25.6%) and *Mariner* superfamily (17.5%). Among the retroelements, *Gypsy* (30%), *Copia* (23%), and other (29%) long terminal repeat (LTR) retrotransposons were the predominant classes.

**Table 1.** Number of markers mapped to each chromosome of common wheat

Chromosome	Map length (cM)	Number of loci	Number of nr marker	Number of total marker
1A	239.87	168	241	571
1B	277.28	214	428	1,026
1D	137.82	76	131	353
2A	245.01	147	259	574
2B	309.67	201	428	1,256
2D	272.18	112	280	854
3A	228.25	128	233	548
3B	310.09	231	408	939
3D	256.95	70	161	401
4A	221.29	141	317	1,013
4B	158.60	77	145	506
4D	121.46	31	52	85
5A	319.07	129	206	473
5B	315.32	197	376	984
5D	292.74	73	142	272
6A	179.32	112	219	667
6B	241.01	167	363	940
6D	193.48	67	112	207
7A	288.54	178	338	771
7B	269.17	168	289	851
7D	216.00	57	115	189
Total	5,093.12	2,744	5,243	13,480
A genome	1,721.35	1,003	1,813	4,617
B genome	1,881.14	1,255	2,437	6,502
D genome	1,490.63	486	993	2,361

To test whether digestion with *Pst*I and *Bst*NI and selection of DNA fragments having only *Pst*I sites at both ends reduced the genomic fraction containing REs, *T. urartu* and *Ae. tauschii* genomic scaffolds were digested *in silico*. Overall, 136,189 and 106,724 fragments were, respectively, generated from the *T. urartu* and *Ae. tauschii* scaffolds. The size of A-genome fragments ranged from 6 to 23,779 bp with an average of  $57.2 \pm 551.7$  bp, and that of D-genome fragments varied from 6 to 28,677 bp with an average of  $44.9 \pm 346.7$  bp. The majority of these fragments were  $\leq 80$  bp (Supplementary Fig. S2). Because contigs and reads obtained from the NGS data of CS and M808 ranged from 65 bp to  $<6,000$  bp, a blastn search of the *in silico* digested genomic sequences against the MIPS RE database was performed using as query the A- and D-genome-derived fragments between 65 and 6,000 bp. REs were found in 84,410 (76.1%) of the 110,984 A-genome fragments, and in 60,522 (69.2%) of the 87,466 D-genome fragments. In contrast to the sequences mapping to the linkage map, retroelements

predominated on DNA transposons in the fragments derived from the *in silico* digestion of both *T. urartu* and *Ae. tauschii* genomes (Supplementary Table S7). REs found were mainly members of the CACTA superfamily of DNA transposons or *Gypsy* retroelements.

### 3.4. *In silico* mapping of the contigs/reads to the physical map of barley

To determine how many of the mapped sequences could be used for the analysis of synteny between wheat and barley (*Hordeum vulgare* L., H-genome) chromosomes, they were mapped *in silico* to the physical map of barley.<sup>27</sup> For sequences mapping to the wheat A- and D-genomes, a blastn search was first conducted, respectively, against the *T. urartu* and *Ae. tauschii* draft genomic sequences;<sup>25,26</sup> the scaffolds obtained were mapped based on a blastn search of the genes contained in these scaffolds against the barley genome. Of the 4,462 WABMs mapping to the A-genome, hits against the *T. urartu* genomic scaffolds were obtained in 3,918 (87.8%) sequences, and 2,856 (64.0%) were anchored to the barley physical map (Table 3). In total, 1,658 (37.2%) markers were mapped to the same homoeologous group chromosomes (e.g. between 1A and 1H), accounting for  $\sim 40\%$  for most of the chromosomes but only 8.5% for chromosome 4A. Collinearity between the A- and H-genomes was observed for many of these markers (Fig. 4), except on chromosome 4A, where inversions and translocations have been reported.<sup>31,32</sup> Translocations among chromosomes 4A, 5A, and 7B could also be observed, as previously reported (Supplementary Fig. S3).

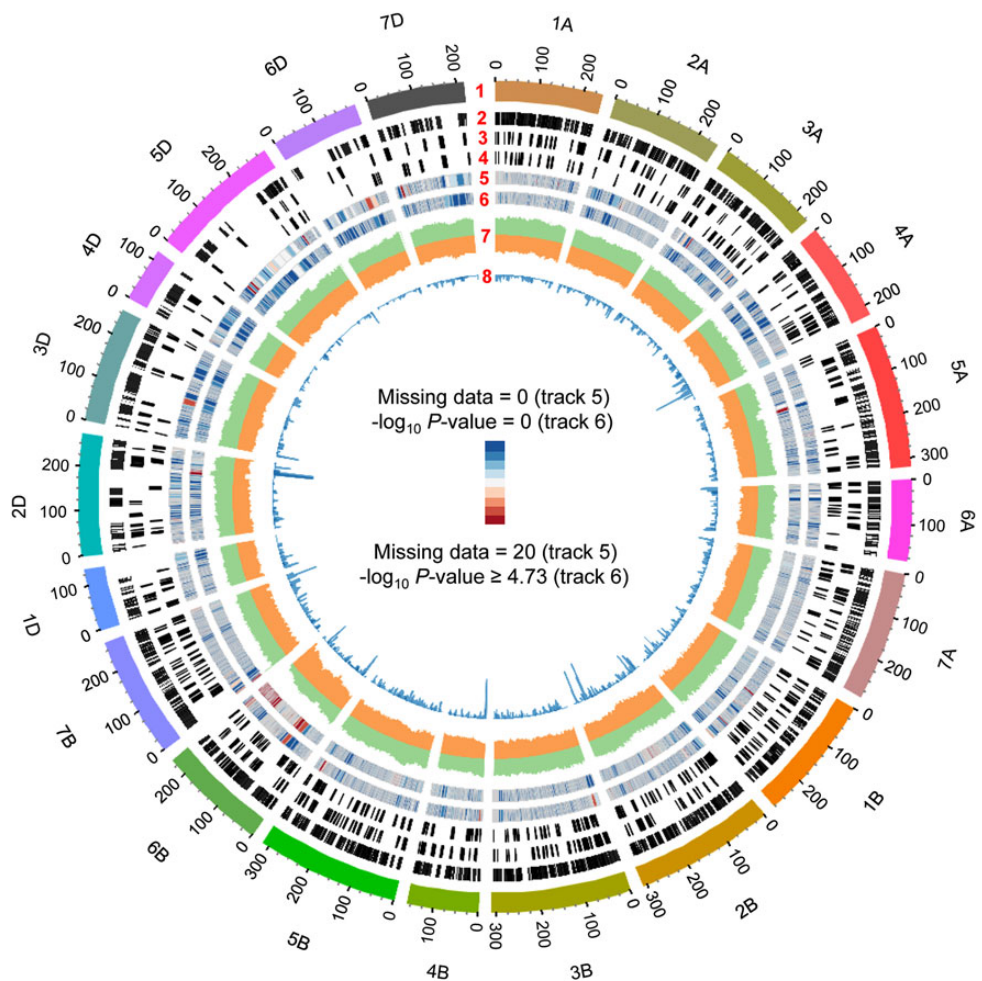
The proportion of D-genome markers with hits against the *Ae. tauschii* genomic sequence (1,319/2,228), and consequently, those anchored on the barley physical map (1,006/2,228), was much lower than the A-genome markers (Table 3). However, a similar percentage (41.4%) was observed for markers mapping to the same homoeologous group chromosomes.

For B-genome markers, the proportion of hits against the barley genome was similar to that in the A-genome (62.4%), but the number of markers mapping to the same homoeologous group was lower in the B-genome (27.8%, Table 3). Although collinearity with the barley genome was observed for many of these markers, the number of non-syntenic markers tended to be higher in the B-genome (Fig. 4). In general, synteny with barley chromosomes at the centromeric region was clearly observed only in a few chromosomes such as 1A and 2A.

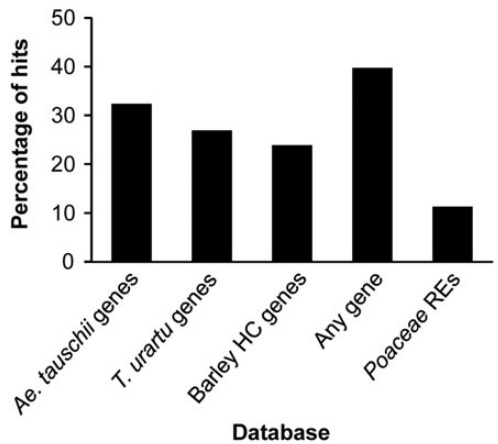
### 3.5. QTL analysis for 14 morphological and agricultural traits

To compare the efficiency of the newly constructed map with the original SSR map developed by Kobayashi *et al.*<sup>16</sup> for analysing complex traits, QTL





**Figure 2.** Distribution of WABMs, REs, missing data, and allele frequency on the constructed genetic map. From the outer to the inner circle, Track 1 shows the genetic map (scale in cM); Track 2 represents the distribution of sequence derived from genic regions; Tracks 3 and 4 represent, respectively, the distribution of retroelements and DNA transposons; Tracks 5 and 6, respectively, a heat map of missing data and  $P$ -values for segregation distortion; Track 7, the allele frequency of M808 (green) and CS (orange); and Track 8, a histogram of the number of markers ( $y$ -axis range 0–100).



**Figure 3.** Histogram of annotated WABMs revealed by blastn analyses of genome databases. Reads/contigs anchored to the high-density genetic map were searched using the blastn algorithm against genes of *Ae. tauschii*, *T. urartu*, and barley and against Poaceae REs.

analysis of 14 agronomically important traits was performed using both maps. Flowering-related traits such as HT, FT, MT, and GFP were evaluated over two growing seasons at Kobe10 and Kobe12 (see Section 2.1.), and at Kyoto12 (Supplementary Table S1). Of the spike-related traits, SpN, SLperSpN (indicative of spike compactness), and T5SpL were evaluated for Kobe12; SL was also evaluated for Kyoto12. Seed-related traits such as SdL, SdW, SdH, and SdLperW were measured for Kobe12. Phenotype data for CL and TN were obtained for Kyoto12. The mean values of the parental lines and RILs are shown in Supplementary Table S8. Strong positive correlations were observed among HT, FT, and MT, and between SL and SLperSpN (Supplementary Table S9). Relatively, strong correlation was also found between T5SpL and SLperSpN, SdH and SdW, and SdLperW and SdW.



**Table 2.** REs found in the mapped contigs and reads

Repeat elements	Number of contigs (%)
DNA transposons	679 (45.8)
CACTA	263 (38.7)
MITE	174 (25.6)
<i>Tc1/Mariner</i>	119 (17.5)
Other DNA transposons	57 (8.39)
<i>Mutator</i>	37 (5.4)
<i>Harbinger</i>	29 (4.3)
Retroelements	659 (44.4)
LTR/ <i>Gypsy</i>	196 (29.7)
Other LTRs	193 (29.3)
LTR/ <i>Copia</i>	152 (23.1)
LINE	116 (17.6)
SINE	2 (0.3)
Others	146 (9.8)
Total	1,484

QTL analysis was performed using composite interval mapping, and only the QTLs with peaks above the LOD threshold computed by a 1,000-permutation test were selected. The LOD threshold values were significantly higher ( $P < 0.001$ ) in the new high-density map, with an average of  $4.76 \pm 0.15$ , than in the original SSR map ( $3.76 \pm 0.08$ ). In total, 29 QTLs were identified using the SSR map and 32 QTLs using the high-density map, of which 25 were identified in both maps (Supplementary Table S10 and Fig. S4); seven QTLs were identified only in the high-density map and four QTLs only in the SSR map.

The 1.5-LOD support interval, meaning the interval in which the LOD score was within 1.5 units of its maximum, tended to be smaller in the high-density map (average of  $14.62 \pm 18.03$  cM) than in the SSR map ( $22.01 \pm 19.85$  cM), but the difference was not significant ( $P = 0.129$ ) (Table 4). Excluding the D-genome QTLs, where the map resolution was still low, the difference became significant ( $9.45 \pm 3.29$  cM for the high-density map and  $17.21 \pm 11.29$  cM for the SSR map,  $P < 0.01$ ). The number of markers within the 1.5-LOD intervals was also significantly higher ( $P < 0.05$ ) in the high-density map (average of  $16.40 \pm 22.97$ ) than in the SSR map (average of  $4.20 \pm 1.98$ ) (Table 4).

Most of the identified QTLs were located in a similar chromosomal region to those previously reported (Supplementary Table S10). Of the flowering-related traits, the early flowering effect of the CS *Ppd-B1* allele was observed in all the datasets and explained  $>30\%$  of the PV. The early flowering effect of CS at 5B and 5D QTL for HT and FT was observed only in one environment (Kyoto12 and Kobe12, respectively), indicating a possible genotype  $\times$  environment interaction.

Of the spike-related traits, QTLs for SL were identified on chromosomes 3A, 4A, and 6A, and explained  $\sim 10\%$  of the total PV. The CS allele increased SL at the 3A QTL, but decreased it at the 4A and 6A QTLs. The 3A QTL also exhibited an effect on T5SpL (explaining 7.77% of PV) and SLperSpN (9.82% of PV). Similarly, the 4A QTL also affected T5SpL (15.39% of PV) and SLperSpN (10.39% of PV) and might explain the positive correlation between these traits (Supplementary Table S9). The 6A QTL had an effect on SLperSpN explaining 15.01% of the PV. QTLs for SpN were identified on homoeologous group 2 chromosomes, of which 2B QTL (linked to the *Ppd-B1* locus) exhibited a major effect and explained 22.16% of the PV.

A QTL for SdL was detected on chromosome 3A (6.25% of PV), and another was detected for SdH on chromosome 1B (9.60% PV). QTLs for SdW and SdLperW were detected on the long arm of chromosome 2D, and explained, respectively, 5.94 and 8.78% of the total PV. This QTL also might explain the significant correlation between SdW and SdLperW (Supplementary Table S9).

#### 4. Discussion

The construction of high-resolution genetic maps in common wheat has required extensive effort due to its large genome and allohexaploid nature. More than 60% of wheat and barley genomes are composed of REs, with retroelements found at least four times more frequently than DNA transposons.<sup>25–27,33</sup> In this study, a microarray-based polymorphism detection system was developed using NGS data obtained from complexity-reduced genomic DNA samples of CS and M808. Although the reduction in complexity using a combination of two restriction enzymes, *Pst*I and *Bst*NI, did not affect the composition of REs (Supplementary Table S7), most of the array-based markers used for construction of the genetic map were derived from genic regions (Supplementary Table S6). In addition, retroelements were under-represented in the array-based markers, indicating that the selection of useful polymorphic probes decreased the number of sequences derived from highly repetitive regions.

SNP typing of the entire mapping population using NGS technology has also been applied in polyploid wheat species.<sup>11–13</sup> Although the GBS approach is cost-effective and fast, it also often leads to a significant amount of missing data due to the low coverage of sequencing.<sup>15</sup> These missing data points have mostly been replaced with estimated genotypes through statistical imputation.<sup>15</sup> An advantage of the array-based genotyping approach in wheat is the low generation of

**Table 3.** Summary of blastn results against genome sequences

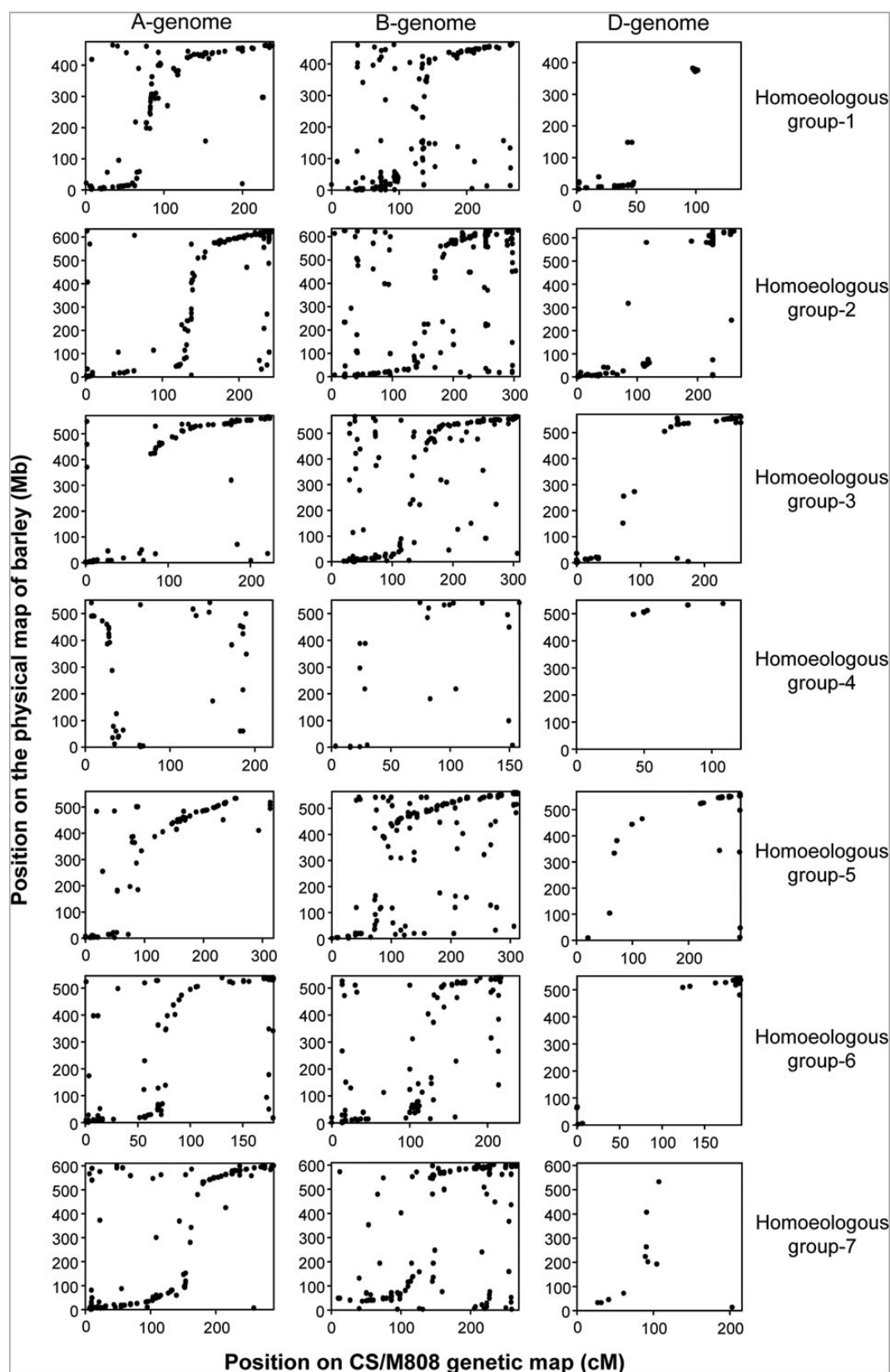
Mapped sequences	Hits against draft genome	Hits against barley genome	Hits on the same homoeologous chromosomes
A genome	3,918 (87.8%)	2,856 (64.0%)	1,658 (37.2%)
1A	464 (84.5%)	379 (69.0%)	265 (48.3%)
2A	488 (91.6%)	338 (63.4%)	238 (44.7%)
3A	458 (87.2%)	333 (63.4%)	238 (45.3%)
4A	851 (85.2%)	608 (60.9%)	85 (8.5%)
5A	416 (91.4%)	315 (69.2%)	184 (40.4%)
6A	555 (85.6%)	409 (63.1%)	293 (45.2%)
7A	686 (91.1%)	474 (62.9%)	355 (47.1%)
B genome	–	3,972 (62.4%)	1,772 (27.8%)
1B	–	629 (63.4%)	279 (28.1%)
2B	–	793 (64.4%)	389 (31.6%)
3B	–	568 (61.3%)	249 (26.9%)
4B	–	290 (58.7%)	150 (30.4%)
5B	–	588 (60.9%)	268 (27.8%)
6B	–	573 (61.9%)	207 (22.4%)
7B	–	531 (63.8%)	230 (27.6%)
D genome	1,319 (59.2%)	1,006 (45.2%)	923 (41.4%)
1D	216 (64.7%)	150 (44.9%)	110 (32.9%)
2D	514 (61.6%)	436 (52.3%)	417 (50.0%)
3D	230 (59.3%)	181 (46.6%)	176 (45.4%)
4D	30 (41.7%)	23 (31.9%)	23 (31.9%)
5D	131 (52.6%)	90 (36.1%)	89 (35.7%)
6D	90 (50.3%)	68 (38.0%)	64 (35.8%)
7D	108 (62.8%)	58 (33.7%)	44 (25.6%)
Total	5,237 (78.3%)	7,834 (60.0%)	4,353 (33.3%)

Sequences mapped on A- and D-genomes were first blast searched against *T. urartu* and *Ae. tauschii* genome sequences and then the obtained scaffolds were assigned to the barley physical map based on their genes. B-genome sequences were directly aligned to the barley genome.

missing data, in contrast to the GBS approach. In 9 of the 210 RILs ( $F_7$  generation) of CS and M808, the number of missing data points was higher than the mean + 2 SD, and they were thus excluded. These missing data points can be explained in part by the presence of heterozygous regions, because data of ambiguous intensity were converted into missing values. In the remaining 201 RILs, no missing data were observed for most of the markers (Supplementary Fig. S1B), indicating a high yield of data from this array-based genotyping system.

The high-density genetic map was composed of 13,480 SSR marker and WABMs, and the markers were distributed across the allohexaploid wheat genome, especially in the A- and B-genomes (Fig. 2). The mapped marker number was similar to that constructed using the GBS approach.<sup>11</sup> As reported in previous studies, the number of markers in the D-genome was much lower than in the A- and B-genomes of

common wheat (Table 1).<sup>8,10</sup> The reads and contigs anchored to the high-density map were also mapped *in silico* against the physical map of barley, because common wheat lacks any physical map. The draft genome sequences of *T. urartu*<sup>26</sup> and *Ae. tauschii*<sup>25</sup> were used for the anchoring of A- and D-genome markers, whereas B-genome markers were directly mapped to the barley genome. The use of *T. urartu* and *Ae. tauschii* genome information greatly increased the number of markers aligned on the same homoeologous group chromosomes of barley compared with the B-genome markers (Table 3). Because of the lower number of D-genome markers, collinearity with barley chromosomes was difficult to observe (Fig. 4). In most cases, collinearity was not observed at the centromeric region. This might be due to the lower gene density and higher number of repetitive sequences present in the proximal region of barley and wheat chromosomes.<sup>27,28</sup>



**Figure 4.** Relationship between the wheat genetic and barley physical maps. Based on blastn search sequences anchored to A- (left column), B- (centre), and D-genomes (right column) were aligned to the barley genome. Only the sequences mapped on the same homoeologous group are presented in each row. The y-axis indicates the position on the physical map of barley in Mb and the x-axis the position on the genetic map of wheat in cM.

**Table 4.** Comparison of 1.5-LOD support intervals and number of markers within this interval between the newly constructed and original SSR maps

QTL <sup>a</sup>	Chr	Newly constructed map		Original SSR map <sup>b</sup>	
		1.5-LOD interval (cM)	Number of markers	1.5-LOD interval (cM)	Number of markers
HT_Kobe10	2B	6.49	4	18.18	3
FT_Kobe10	2B	6.49	4	18.18	3
MT_Kobe10	2B	6.49	4	18.18	3
GFP_Kobe10	1A	9.53	14	9.47	4
HT_Kobe12	2B	9.14	7	18.18	3
HT_Kobe12	5D	73.64	5	76.43	5
FT_Kobe12	2B	9.14	7	18.18	3
FT_Kobe12	5D	73.64	5	76.43	5
MT_Kobe12	2B	6.49	4	18.18	3
HT_Kyoto12	2B	6.49	4	18.18	3
FT_Kyoto12	2B	6.49	4	18.18	3
SL_Kobe12	3A	14.68	10	31.42	4
SL_Kobe12	6A	10.98	23	22.34	7
SL_Kyoto12	4A	10.99	28	12.81	6
SpN_Kobe12	2A	14.68	9	5.80	8
SpN_Kobe12	2B	9.14	7	11.01	2
SLperSpN_Kobe12	4A	9.80	21	12.81	6
SLperSpN_Kobe12	6A	9.67	22	12.00	5
T5SpL_Kobe12	2B	2.31	5	11.01	2
T5SpL_Kobe12	3A	14.68	10	60.77	5
T5SpL_Kobe12	4A	10.99	28	5.34	3
SdH_Kobe12	1B	8.06	45	15.61	10
SdLperW_Kobe12	2D	10.41	115	41.19	3
CL_Kyoto12	2B	15.06	10	11.01	2
TN_Kyoto12	1A	10.03	15	11.74	4
Total average		14.62 ± 18.03	16.40 ± 22.97*	22.91 ± 19.85	4.20 ± 1.97
Average of AB-genome		9.45 ± 3.29**	12.95 ± 10.75***	17.21 ± 11.29	4.18 ± 2.08

<sup>a</sup>Only the 25 QTLs identified in both maps were included.<sup>b</sup>Kobayashi *et al.*<sup>16</sup>Student's *t*-test was used to test for statistical significance (\**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001).

In the high-density map, segregation distortion was observed in six chromosomal regions, with the long arm of chromosome 6B showing the greatest distortion (Fig. 2 and Supplementary Table S4). This might have been caused by the presence of the pollen killer gene, *Ki*, on the long arm of chromosome 6B of CS.<sup>34,35</sup> *Ki* would affect marker segregation in a large chromosomal region of 6BL, though its position on 6BL has not been accurately determined yet.<sup>35</sup> Distortions on chromosomes 2B and 5B are due to an epistatic interaction of hybrid necrosis alleles, *Ne1<sup>w</sup>* on the long arm of chromosome 5B of CS, and *Ne2* on the short arm of chromosome 2B of M808.<sup>35</sup> *Ne1* and *Ne2* have been roughly mapped to 5BL and 2BS, respectively.<sup>36</sup> The distorted regions of 2BS and 5BL were more limited around the causal genes than

on 6BL. The three other segregation distortions on chromosomes 2A, 5A, and 7B were not observed in our previous study using the original SSR map of CS and M808.<sup>35</sup> Thus, the increased map resolution should clarify the presence of additional regions showing segregation distortion. However, the causal genes for the segregation distortion on chromosomes 2A, 5A, and 7B are unknown. In general, segregation distortion is considered to occur due to the presence of a reproductive barrier gene,<sup>37–39</sup> suggesting that reproductive barrier genes might be located in these three chromosomal regions.

QTL analysis using the high-density and SSR maps identified a similar number of QTLs for both maps (Supplementary Table S10). Although most of the QTLs were commonly identified in the two maps, a



higher number of QTLs was identified in the high-density map. In the high-density map, the LOD threshold values computed by a 1,000-permutation test were significantly higher, as reported in rice.<sup>40</sup> Thus, the four QTLs identified only in the SSR map might be masked in the high-density map due to these higher threshold values. The increased map resolution resulted in identification of QTLs in a narrower chromosomal region and in a more accurate position than for the SSR map (Table 4 and Supplementary Fig. S4).

In the high-density map, the QTL associated with flowering-related traits corresponded well to the *Ppd-B1* locus on chromosome 2B and the *VRN-D1* locus (Supplementary Fig. S4). The early flowering phenotype of CS (Supplementary Table S8) can be explained mainly by the effect of the photoperiod-insensitive allele of *Ppd-B1* (probably due to its higher copy number)<sup>41,42</sup> and dominant spring allele of *VRN-D1*.<sup>43</sup> The spike compactness (SL/SpN) and short SL of CS can be explained by the 6A and 4A QTLs. The 4A QTL also exhibited an effect on T5SpL, which might contribute to the square-head like spike morphology of CS. On the other hand, the CS allele at the 3A QTL contributed to a less compact spike, longer SL, and longer T5SpL. Grain size and shape are associated with milling quality and yield in wheat. Large and spherical grains have been predicted to be the optimum grain morphology to increase milling yield.<sup>44</sup> The CS allele at the 2D QTL, with an increasing effect on SdW and reduced SdLperW, and the M808 allele at the 1B QTL (increased SdH) contributed in seed roundness. Favourable alleles for increasing yield were also found in CS at the 2B QTL for CL (reduced plant height) and at the 1A QTL for TN (increased number of tillers with ears). Therefore, the results of QTL analyses indicate that the high-density map efficiently and precisely functions to identify QTLs for morphological and agricultural traits.

Because of the difference in geographical origin and growth habit of the common wheat cultivars CS (spring habit Chinese variety) and M808 (winter habit Ukrainian variety) used for the development of our array-based system, it is expected to work better between spring and winter wheat and/or Chinese/Asian and European variety. Further study is needed to test the usefulness of this genotyping system using a large wheat population set.

A number of tightly linked WABMs were found in most of the QTLs, which could facilitate marker-assisted selection or map-based cloning of genes. For this application, the marker of interest should be converted into a PCR marker. Our results demonstrated that this array-based polymorphism detection system can yield high-quality genotype data useful for high-density genetic map construction and QTL analysis of agronomically important traits.

**Supplementary Data:** Supplementary data are available at [www.dnaresearch.oxfordjournals.org](http://dnaresearch.oxfordjournals.org).

## Funding

This work was supported by a grant from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan [Grant-in-Aid for Scientific Research (B) no. 25292008] to S.T., by cooperative research funds from TOYOTA Motor Co. Ltd., and by Special Coordination Funds for Promoting Science and Technology, Creation of Innovation Centers for Advanced Interdisciplinary Research Areas (Innovative BioProduction Kobe), MEXT, Japan. Funding to pay the Open Access publication charges for this article was provided by a grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan [Grant-in-Aid for Scientific Research (B) No. 25292008] to S.T.

## References

1. Kihara, H. 1944, Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*, *Agric. Hortic.*, **19**, 889–90 [in Japanese].
2. McFadden, E.S. and Sears, E.R. 1944, The artificial synthesis of *Triticum spelta*, *Rec. Genet. Soc. Am.*, **13**, 26–7.
3. Akbari, M., Wenzl, P., Caig, V., et al. 2006, Diversity arrays technology (DART) for high-throughput profiling of the hexaploid wheat genome, *Theor. Appl. Genet.*, **113**, 1409–20.
4. Wenzl, P., Carling, J., Kudrna, D., et al. 2004, Diversity arrays technology (DART) for whole-genome profiling of barley, *Proc. Natl. Acad. Sci. USA*, **101**, 9915–20.
5. Sorrells, M.E., Gustafson, J.P., Somers, D., et al. 2011, Reconstruction of the Synthetic W7984 × Opata M85 wheat reference population, *Genome*, **54**, 875–82.
6. Zhang, L.Y., Liu, D.C., Guo, X.L., et al. 2011, Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DART markers, *BMC Genetics*, **12**, 42.
7. Trebbi, D., Maccaferri, M., de Heer, P., et al. 2011, High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.), *Theor. Appl. Genet.*, **123**, 555–69.
8. Allen, A.M., Barker, G.L.A., Berry, S.T., et al. 2011, Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.), *Plant Biotechnol. J.*, **9**, 1086–99.
9. Lai, K., Duran, C., Berkman, P.J., et al. 2012, Single nucleotide polymorphism discovery from wheat next-generation sequence data, *Plant Biotechnol. J.*, **10**, 743–9.
10. Cavanagh, C.R., Chao, S., Wang, S., 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**, 8057–62.
11. 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.
12. Saintenac, C., Jiang, D., Wang, S. and Akhunov, E. 2013, Sequence-based mapping of the polyploid wheat genome, *G3-Genes Genomes Genet.*, **3**, 1105–14.
  13. van Poecke, R.M.P., Maccaferri, M., Tang, J., et al. 2013, Sequence-based SNP genotyping in durum wheat, *Plant Biotechnol. J.*, **11**, 809–17.
  14. Winfield, M.O., Wilkinson, P.A., Allen, A.M., et al. 2012, Targeted re-sequencing of the allohexaploid wheat exome, *Plant Biotechnol. J.*, **10**, 733–42.
  15. Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L. 2011, Genome-wide genetic marker discovery and genotyping using next-generation sequencing, *Nat. Rev. Genet.*, **12**, 499–510.
  16. Kobayashi, F., Takumi, S. and Handa, H. 2010, Identification of quantitative trait loci for ABA responsiveness at the seedling stage associated with ABA-regulated gene expression in common wheat, *Theor. Appl. Genet.*, **121**, 629–41.
  17. Enoki, H., Nishimura, S. and Murakami, A. 2011, Method for designing probe in DNA microarray, and DNA microarray provided with probe designed thereby, International publication number: WO 2011/074510.
  18. Enoki, H., Nishimura, S. and Murakami, A. 2012, Method for designing probe in DNA microarray, and DNA microarray provided with probe designed thereby, US patent application 20120190582A1.
  19. Zerbino, D.R. and Birney, E. 2008, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Res.*, **18**, 821–9.
  20. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.*, **10**, R25.
  21. Beales, J., Turner, A., Griffiths, S., Snape, J.W. and Laurie, D.A. 2007, A *pseudo-response regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.), *Theor. Appl. Genet.*, **115**, 721–33.
  22. Iwata, H. and Ninomiya, S. 2006, AntMap: constructing genetic linkage maps using an ant colony optimization algorithm, *Breed. Sci.*, **56**, 371–7.
  23. Lorieux, M. 2012, MapDisto: fast and efficient computation of genetic linkage maps, *Mol. Breed.*, **30**, 1231–5.
  24. Kosambi, D.D. 1944, The estimation of map distance from recombination values, *Ann. Eugen.*, **12**, 172–5.
  25. Jia, J., Zhao, S., Kong, X., et al. 2013, *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation, *Nature*, **496**, 91–5.
  26. Ling, H.Q., Zhao, S., Liu, D., et al. 2013, Draft genome of the wheat A-genome progenitor *Triticum urartu*, *Nature*, **496**, 87–90.
  27. International Barley Genome Sequencing Consortium. 2012, A physical, genetic and functional sequence assembly of the barley genome, *Nature*, **491**, 711–6.
  28. Luo, M.C., Gu, Y.Q., You, F.M., et al. 2013, A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor, *Proc. Natl. Acad. Sci. USA*, **110**, 7940–5.
  29. Krzywinski, M., Schein, J., Birol, I., et al. 2009, Circos: an information aesthetic for comparative genomics, *Genome Res.*, **19**, 1639–45.
  30. Broman, K.W., Wu, H., Sen, S. and Churchill, G.A. 2003, R/qtl: QTL mapping in experimental crosses, *Bioinformatics*, **19**, 889–90.
  31. Devos, K.M., Dubcovsky, J., Dvořák, J., Chinoy, C.N. and Gale, M.D. 1995, Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination, *Theor. Appl. Genet.*, **91**, 282–8.
  32. Ma, J., Stiller, J., Berkman, P.J., et al. 2013, Sequence-based analysis of translocations and inversions in bread wheat (*Triticum aestivum* L.), *PLoS ONE*, **8**, e79329.
  33. Brenchley, R., Spannagl, M., Pfeifer, M., et al. 2012, Analysis of the bread wheat genome using whole-genome shotgun sequencing, *Nature*, **491**, 705–10.
  34. Sears, E.R. and Loegering, W.Q. 1961, A pollen-killing gene in wheat, *Genetics*, **46**, 897.
  35. Takumi, S., Motomura, Y., lehisa, J.C.M. and Kobayashi, F. 2013, Segregation distortion caused by weak hybrid necrosis in recombinant inbred lines of common wheat, *Genetica*, **141**, 463–70.
  36. Chu, C.G., Faris, J.D., Friesen, T.L. and Xu, S.S. 2006, Molecular mapping of hybrid necrosis genes *Ne1* and *Ne2* in hexaploid wheat using microsatellite markers, *Theor. Appl. Genet.*, **112**, 1374–81.
  37. Bikard, D., Patel, D., Le Mett  , C., et al. 2009, Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*, *Science*, **323**, 623–6.
  38. Mizuta, Y., Harushima, Y. and Kurata, N. 2010, Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes, *Proc. Natl. Acad. Sci. USA*, **107**, 20417–22.
  39. Yang, J., Zhao, X., Cheng, K., et al. 2012, A killer-protector system regulates both hybrid sterility and segregation distortion, *Science*, **337**, 1336–40.
  40. Yu, H., Xie, W., Wang, J., et al. 2011, Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers, *PLoS ONE*, **6**, e17595.
  41. D    , A., Zikhali, M., Turner, A.S., Isaac, P. and Laurie, D.A. 2012, Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*), *PLoS ONE*, **7**, e33234.
  42. Nguyen, A.T., lehisa, J.C.M., Mizuno, N., Nitta, M., Nasuda, S. and Takumi, S. 2013, Differential contribution of two *Ppd-1* homoeoalleles to early-flowering phenotype in Nepalese and Japanese varieties of common wheat, *Breed. Sci.*, **63**, 374–83.
  43. Yan, L., Helguera, M., Kato, K., Fukuyama, S., Sherman, J. and Dubcovsky, J. 2004, Allelic variation at the *VRN-1* promoter region in polyploid wheat, *Theor. Appl. Genet.*, **109**, 1677–86.
  44. Evers, A.D., Cox, R.I., Shaheedullah, M.Z. and Withey, R.P. 1990, Predicting milling extraction rate by image analysis of wheat grains, *Asp. Appl. Biol.*, **25**, 417–26.