Population structure and association analyses of the core collection of hexaploid accessions conserved *ex situ* in the Japanese gene bank NBRP-Wheat

Shotaro Takenaka1,2, Miyuki Nitta1 and Shuhei Nasuda1*

1Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University, Kitashirakawaoiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
2Department of Plant Life Science, Faculty of Agriculture, Ryukoku University, 1-5 Yokotani, Seta Ohe-cho, Otsu, Shiga 520-2194, Japan

(Received 9 August 2018, accepted 12 September 2018; J-STAGE Advance published date: 15 December 2018)

In this study, we investigated the genetic diversity and population structure of the core collection of hexaploid wheat accessions in the Japanese wheat gene bank NBRP-Wheat. The core collection, consisting of 188 accessions of *Triticum aestivum*, *T. spelta*, *T. compactum*, *T. sphaerococcum*, *T. macha* and *T. vavilovii*, was intensively genotyped by DArTseq markers and consisted of 20,186 SNPs and 60,077 present and absent variations (PAVs). Polymorphic markers were distributed in all chromosomes, with a tendency for smaller numbers on the D-genome chromosomes. We examined the population structure by Bayesian clustering and principal component analysis with a general linear model. Overall, the core collection was divided into seven clusters. Non-admixture accessions in each cluster indicated that the clusters reflect the geographic distribution of the accessions. Both structure analyses strongly suggested that the cluster consisting of *T. spelta* and *T. macha* is out-grouped from other hexaploid wheat accessions. We performed genome-wide association analysis pilot studies for nine quantitative and seven qualitative traits and found marker-trait associations for all traits but one, indicating that the current core collection will be useful for detecting uncharacterized QTLs associated with phenotypes of interest.

Key words: association analysis, core collection, genetic diversity, hexaploid wheat, population structure

INTRODUCTION

Along with maize and rice, wheat (*Triticum* spp.) is one of the most important staple crops that have supported human civilization. Hexaploid wheat (*T. aestivum* L.) originated from hybridization between domesticated tetraploid wheat (*T. turgidum* L.) and *Aegilops tauschii* Coss (Kihara, 1944; McFadden and Sears, 1944). Accumulated archeological and paleontological evidence indicates that the hybridization event occurred either south or west of the Caspian Sea approximately 8,000 years ago (Nesbitt and Samuel, 1996; Salamini et al., 2002; Giles and Brown, 2006). Today, hexaploid wheat is cultivated widely in subarctic, temperate and subtropical regions of both hemispheres, from 67°N in Norway, Finland and Russia to 45°S in Argentina, and accounts for more than 30% of the global cereal cultivation area (220 million ha in 2016, http://faostat.fao.org/). The production of wheat has increased since the release of modern high-yielding varieties and application of chemical fertilizers in the 1950s. The great success in raising wheat production has been referred to as a “green revolution”. Today, more than 700 million tonnes of wheat is harvested annually (749 million tonnes in 2016, http://faostat.fao.org/). However, the increase in wheat production has plateaued in recent decades and there is an urgent need to promote sustainable wheat production to meet the demands of the increasing world population amid global climate change (McCouch et al., 2013). Diverse genetic resources are essential for breeding new cultivars. The landraces cultivated globally are especially important as a genetic resource because they have adapted to various environments (Harlan, 1975; McCouch et al., 2013). The National Bioresources Project-Wheat (NBRP-Wheat), launched by the Japanese government in 2002, collects, stores and supplies wild species, landraces and experi-
Association mapping is a method of detecting quantitative trait loci (QTLs). This approach is based on correlations between genotypes and phenotypes, and is additionally affected by the structure of linkage disequilibrium (LD) within a test population (Buckler and Thornsberry, 2002; Zondervan and Cardon, 2004). In addition to QTL analysis, which is based on phenotypic and genotypic polymorphisms within a bi-parental mapping population, association mapping is also applicable to a much larger gene pool and a more diverse genetic structure (Remington et al., 2001). Whole-genome association mapping has been applied to common wheat for several agronomic traits (Breseghello and Sorrells, 2006; Crossa et al., 2007; Neumann et al., 2011; Le Gouis et al., 2012; Li et al., 2016). However, these studies focused on a common wheat (T. aestivum) population but not on other hexaploid Triticum species.

In this study, we provide intensive genotype information for the NBRP hexaploid wheat core collection to elucidate its genetic diversity and population structure, which is a prerequisite for association studies. We performed genotyping on the core collection lines using a DArT marker system (Jaccoud et al., 2001). The core collection includes modern cultivars frequently used for breeding in Japan, and landraces cover the geographical distribution and taxonomic groups of hexaploid wheat. Therefore, the core collection is a comprehensive hexaploid wheat collection in terms of systematics of the genus Triticum, i.e., it represents not only common wheat but also other hexaploid wheat species. We aimed to test the applicability of the core collection for association analyses of quantitative and qualitative traits.

### MATERIALS AND METHODS

**Plant materials and growth conditions**

Plant materials used in this study are the collection of 188 accessions of hexaploid wheat (Table 1), comprised of T. aestivum (T. aestivum ssp. aestivum L., 160 accessions), T. spelta (T. aestivum ssp. spelta (L.) Thell., 12 accessions).

| Acc_ID* | Species            | Cultivar (strain no.) | Country (region) | Cluster_ID |
|---------|--------------------|-----------------------|------------------|------------|
| 6×001   | T. aestivum        | (KU-166)              | China (Inner Mongolia) | E          |
| 6×002** | T. aestivum        | Norin 10 (KU-265)    | Japan            | A          |
| 6×003** | T. aestivum        | Baart (KU-309)       | USA              | D          |
| 6×004** | T. aestivum        | Selkerk (KU-333)     | Canada           | E          |
| 6×005** | T. aestivum        | Wichita (KU-336)     | USA              | E          |
| 6×006** | T. aestivum        | April Bearded (KU-366)| UK               | F          |
| 6×007** | T. aestivum        | Renown (KU-370)      | UK               | F          |
| 6×008** | T. aestivum        | Wilhelmmina (KU-371) | UK               | F          |
| 6×009** | T. aestivum        | Little Joss (KU-372) | UK               | F          |
| 6×010** | T. aestivum        | Victor (KU-373)      | UK               | F          |
| 6×011** | T. aestivum        | Yeoman (KU-374)      | UK               | F          |
| 6×012   | T. aestivum        | (KU-405)             | Uzbekistan       | C          |
| 6×013   | T. aestivum        | (KU-479)             | China (Sichuan)  | B          |
| 6×014   | T. aestivum        | (KU-481)             | China (Shaanxi)  | E          |
| 6×015   | T. aestivum        | (KU-483)             | Tanzania         | E          |
| 6×016   | T. aestivum        | (KU-497)             | India            | E          |
| 6×017   | T. aestivum        | (KU-504)             | China            | B          |
| 6×018   | T. aestivum        | Gifu Wase Komugi (KU-601) | Japan | A |
| 6×019   | T. aestivum        | (KU-1002)            | Spain            | E          |
| 6×020   | T. aestivum        | (KU-1005)            | Spain            | F          |
| 6×021   | T. aestivum        | (KU-1011)            | Spain            | D          |
| 6×022   | T. aestivum        | (KU-1049)            | Spain            | F          |
| 6×023   | T. aestivum        | (KU-1143)            | Spain            | F          |
| No. | Species | Accession Number | Country | Code |
|-----|---------|------------------|---------|------|
| 6×024 | *T. aestivum* | Akita Zairai (KU-1215) | Japan | F |
| 6×025 | *T. aestivum* | Fukoku (KU-1230) | Japan | A |
| 6×026 | *T. aestivum* | Murasaki Aka (KU-1279) | Japan | F |
| 6×027 | *T. aestivum* | (KU-1302) | Greece | F |
| 6×028 | *T. aestivum* | (KU-1347) | Greece | D |
| 6×029 | *T. aestivum* | (KU-1392) | Romania | F |
| 6×030 | *T. aestivum* | (KU-1394) | Romania | F |
| 6×031 | *T. aestivum* | (KU-1421) | Romania | F |
| 6×032 | *T. aestivum* | (KU-1424) | Romania | F |
| 6×033 | *T. aestivum* | (KU-1521) | Azerbaijan | F |
| 6×034 | *T. aestivum* | (KU-1527) | Armenia | E |
| 6×035 | *T. aestivum* | (KU-1644) | Armenia | D |
| 6×036 | *T. aestivum* | (KU-1668) | Georgia | F |
| 6×037 | *T. aestivum* | (KU-1697) | Georgia | F |
| 6×038 | *T. aestivum* | (KU-1797) | Georgia | F |
| 6×039 | *T. aestivum* | (KU-3006) | Pakistan | C |
| 6×040 | *T. aestivum* | (KU-3010) | Pakistan | C |
| 6×041 | *T. aestivum* | (KU-3037) | Pakistan | C |
| 6×042 | *T. aestivum* | (KU-3045) | Afghanistan | C |
| 6×043 | *T. aestivum* | (KU-3054) | Afghanistan | C |
| 6×044 | *T. aestivum* | (KU-3062) | Afghanistan | C |
| 6×045 | *T. aestivum* | (KU-3083) | Afghanistan | C |
| 6×046 | *T. aestivum* | (KU-3089) | Afghanistan | C |
| 6×047 | *T. aestivum* | (KU-3097) | Iran | C |
| 6×048 | *T. aestivum* | (KU-3098) | Iran | C |
| 6×049 | *T. aestivum* | (KU-3121) | Iran | C |
| 6×050 | *T. aestivum* | (KU-3126) | Iran | C |
| 6×051 | *T. aestivum* | (KU-3136) | Iran | C |
| 6×052 | *T. aestivum* | (KU-3162) | Iran | C |
| 6×053 | *T. aestivum* | (KU-3184) | Iran | D |
| 6×054 | *T. aestivum* | (KU-3189) | Iran | D |
| 6×055 | *T. aestivum* | (KU-3202) | Iran | C |
| 6×056 | *T. aestivum* | (KU-3232) | Iran | D |
| 6×057 | *T. aestivum* | (KU-3236) | Iran | D |
| 6×058 | *T. aestivum* | (KU-3274) | Iran | C |
| 6×059 | *T. aestivum* | (KU-3289) | Iran | C |
| 6×060 | *T. aestivum* | (KU-3299) | Pakistan | B |
| 6×061 | *T. aestivum* | (KU-3351) | Pakistan | B |
| 6×062 | *T. aestivum* | (KU-3752) | Egypt | B |
| 6×063 | *T. aestivum* | (KU-3777) | Jordan | D |
| 6×064 | *T. aestivum* | (KU-3778) | Lebanon | E |
| 6×065 | *T. aestivum* | (KU-3780) | Syria | E |
| 6×066 | *T. aestivum* | (KU-3784) | Turkey | E |
| 6×067 | *T. aestivum* | (KU-3789) | Turkey | F |
| 6×068 | *T. aestivum* | (KU-3801) | Turkey | E |
| 6×069 | *T. aestivum* | (KU-3806) | Turkey | E |
| 6×070 | *T. aestivum* | (KU-3818) | Turkey | D |
| No.  | Species        | Variety       | Country       | Location |
|------|---------------|---------------|---------------|----------|
| 6×071| *T. aestivum*  | (KU-3834)     | Turkey        | D        |
| 6×072| *T. aestivum*  | (KU-3851)     | Turkey        | D        |
| 6×073| *T. aestivum*  | (KU-3857)     | Turkey        | F        |
| 6×074| *T. aestivum*  | (KU-3860)     | Turkey        | D        |
| 6×075| *T. aestivum*  | (KU-3868)     | Italy         | F        |
| 6×076| *T. aestivum*  | (KU-4703)     | Nepal         | E        |
| 6×077| *T. aestivum*  | (KU-4714)     | Nepal         | B        |
| 6×078| *T. aestivum*  | (KU-4734)     | Nepal         | B        |
| 6×079| *T. aestivum*  | (KU-4759)     | Nepal         | B        |
| 6×080| *T. aestivum*  | (KU-4769)     | Nepal         | B        |
| 6×081| *T. aestivum*  | (KU-4783)     | Nepal         | B        |
| 6×082| *T. aestivum*  | (KU-7001)     | Bhutan        | B        |
| 6×083| *T. aestivum*  | (KU-7041)     | Bhutan        | B        |
| 6×084| *T. aestivum*  | (KU-7113)     | Bhutan        | B        |
| 6×085| *T. aestivum*  | (KU-7180)     | Bhutan        | B        |
| 6×086| *T. aestivum*  | (KU-7356)     | Ethiopia      | E        |
| 6×087| *T. aestivum*  | (KU-7379)     | Ethiopia      | F        |
| 6×088| *T. aestivum*  | (KU-7406)     | Ethiopia      | F        |
| 6×089| *T. aestivum*  | (KU-7437)     | Afghanistan   | C        |
| 6×090| *T. aestivum*  | (KU-7459)     | Afghanistan   | B        |
| 6×091| *T. aestivum*  | (KU-7480)     | Afghanistan   | C        |
| 6×092| *T. aestivum*  | (KU-7624)     | Afghanistan   | C        |
| 6×093| *T. aestivum*  | (KU-7653)     | Afghanistan   | C        |
| 6×094| *T. aestivum*  | (KU-7669)     | Afghanistan   | C        |
| 6×095| *T. aestivum*  | (KU-9431)     | Ethiopia      | D        |
| 6×096| *T. aestivum*  | (KU-9460)     | Ethiopia      | D        |
| 6×097| *T. aestivum*  | (KU-9797)     | Ethiopia      | D        |
| 6×098| *T. aestivum*  | (KU-9820)     | Ethiopia      | D        |
| 6×099| *T. aestivum*  | (KU-9867)     | Ethiopia      | D        |
| 6×100| *T. aestivum*  | (KU-10001)    | Iraq          | D        |
| 6×101| *T. aestivum*  | (KU-10154)    | Iraq          | D        |
| 6×103| *T. aestivum*  | (KU-10393)    | Iran          | D        |
| 6×105| *T. aestivum*  | (KU-10439)    | Iran          | C        |
| 6×106| *T. aestivum*  | (KU-10480)    | Iran          | D        |
| 6×107| *T. aestivum*  | (KU-10510)    | Iran          | D        |
| 6×108| *T. aestivum*  | (KU-11201)    | Afghanistan   | D        |
| 6×109| *T. aestivum*  | (KU-11214)    | Afghanistan   | D        |
| 6×110| *T. aestivum*  | (KU-11240A)   | Afghanistan   | C        |
| 6×111| *T. aestivum*  | (KU-11351)    | Romania       | D        |
| 6×112| *T. aestivum*  | (KU-11702)    | Greece        | F        |
| 6×113| *T. aestivum*  | (KU-11809)    | Greece        | D        |
| 6×114| *T. aestivum*  | (KU-11829)    | Greece        | D        |
| 6×115| *T. aestivum*  | (KU-13501)    | China (Sichuan)| F     |
| 6×116| *T. aestivum*  | (KU-13506)    | China (Tibet) | F        |
| 6×117| *T. aestivum*  | (KU-13546)    | China (Tibet) | B        |
| 6×118| *T. aestivum*  | (KU-13631)    | China (Tibet) | B        |
| 6×119| *T. aestivum*  | (KU-13662)    | China (Tibet) | B        |
| Cross | Parent 1 | Parent 2 | Origin | Geographical Region |
|-------|----------|----------|--------|---------------------|
| 6×120 | *T. aestivum* | (KU-13708) | China (Tibet) | B |
| 6×121 | *T. aestivum* | (KU-13807) | China (Tibet) | B |
| 6×122 | *T. aestivum* | (KU-13891) | China (Tibet) | B |
| 6×123 | *T. spelta* | (KU-1020) | Spain | G |
| 6×124 | *T. spelta* | (KU-1062) | Spain | G |
| 6×125 | *T. spelta* | (KU-1137) | Spain | G |
| 6×126 | *T. spelta* | (KU-3377) | Iran | C |
| 6×127 | *T. spelta* | (KU-3401) | Germany | G |
| 6×128 | *T. spelta* | (KU-3413) | Germany | G |
| 6×129 | *T. spelta* | (KU-3416) | Germany | G |
| 6×130 | *T. spelta* | (KU-3417) | Germany | G |
| 6×131 | *T. spelta* | (KU-3421) | Germany | G |
| 6×132 | *T. spelta* | (KU-3443) | Germany | G |
| 6×133 | *T. spelta* | (KU-3444) | Germany | G |
| 6×134 | *T. spelta* | (KU-3445) | Germany | G |
| 6×135 | *T. compactum* | (KU-152) | China (Inner Mongolia) | F |
| 6×136 | *T. compactum* | (KU-1208) | Japan | A |
| 6×138 | *T. compactum* | (KU-3063) | Afghanistan | C |
| 6×139 | *T. compactum* | (KU-3242) | Iran | D |
| 6×140 | *T. compactum* | (KU-7350) | Turkey | C |
| 6×141 | *T. compactum* | (KU-9873) | Ethiopia | – |
| 6×142 | *T. macha* | (KU-197) | Turkey | G |
| 6×143 | *T. macha* | (KU-1812) | Georgia | G |
| 6×144 | *T. macha* | (KU-1814) | Georgia | G |
| 6×145 | *T. macha* | (KU-1817) | Georgia | G |
| 6×146 | *T. sphaeroecoccum* | (KU-161) | – | B |
| 6×147 | *T. sphaeroecoccum* | (KU-162-2) | Pakistan | B |
| 6×148 | *T. sphaeroecoccum* | (KU-3004) | Pakistan | B |
| 6×149 | *T. vavilovii* | (KU-192) | – | C |
| 6×150 | *T. vavilovii* | (KU-3848) | Turkey | C |
| 6×151** | *T. aestivum* | Opata 85 | Mexico | E |
| 6×152** | Synthetic | W7984 | – | G |
| 6×153 | *T. aestivum* | Chinese Spring (LPKU2269) | China | B |
| 6×154 | *T. spelta* | var. duhamerianum (KT019-001) | Germany | G |
| 6×155** | *T. aestivum* | Timstein | Australia | E |
| 6×156** | *T. aestivum* | Hope | USA | E |
| 6×157** | *T. aestivum* | Cheyenne | USA | F |
| 6×158** | *T. aestivum* | Hokkai 240 | Japan | F |
| 6×159** | *T. aestivum* | Kanto 107 | Japan | A |
| 6×160** | *T. aestivum* | Norin 26 (LPKU2270) | Japan | A |
| 6×161** | *T. aestivum* | Norin 61 (LPKU2271) | Japan | A |
| 6×163** | *T. aestivum* | Chihokukomugi | Japan | E |
| 6×164** | *T. aestivum* | Nanbukomugi | Japan | – |
| 6×165** | *T. aestivum* | Zenkowakamugi | Japan | A |
| 6×166** | *T. aestivum* | Sumai 3 | China | A |
| 6×167** | *T. aestivum* | Nobeokabouzu | Japan | A |
| 6×168** | *T. aestivum* | Tamaizumi | Japan | A |
sions), *T. compactum* (*T. aestivum* ssp. *compactum* (Host) MacKey, six accessions), *T. macha* (*T. aestivum* ssp. *macha* (Dekapr. & Menabde) MacKey, four accessions), *T. sphaerococcum* (*T. aestivum* ssp. *sphaerococcum* (Percival) MacKey, three accessions), *T. vavilovii* (*T. aestivum* var. *vavilovii* Jakubz., two accessions) and synthetic hexaploid wheat (one accession, W7984). The synthetic wheat W7984 (Altar84/Aegilops tauschii (219) CIGM86.940) is the parental accession of the mapping population distributed by the International Triticeae Mapping Initiative (ITMI). In this report, we used nomenclature according to the National BioResource Project (NBRP)-Wheat catalogue. Detailed information on accessions is available on the NBRP-Wheat database KOMUGI (http://www.shigen.nig.ac.jp/wheat/komugi/). The core collection includes both landraces and modern cultivars, and they were selected from the NBRP-Wheat hexaploid wheat collection based on their taxonomic and geographical distribution. In this paper, we refer to the accessions with evident breeding pedigrees as cultivars. Seeds were sown in January, 2011 and the seedlings were transplanted to pots (one plant per pot). The pots were placed in a glasshouse without air conditioning at Yoshida Campus, Kyoto University, Japan (N35.031, E135.79).

**Phenotyping** We scored nine quantitative traits for every accession: heading time (HT), flowering time (FT), maturation time (MT), number of spikelets per spike (SN), spike length (SL), spike density (SD), culm length (CL), 100-grain weight (HGW) and number of tillers (TN); and seven qualitative traits: wax-like deposits on culm (WC), ochrea (WO) and spike (WS), anthocyanin pigmentation on coleoptile (AC) and seedling (AS), glume hair (GH), and awnedness (AWN). HT, FT and MT are expressed as the number of days after November 1st. SD is obtained by dividing SN by SL. CL is the value of the longest culm of each individual. Every quantitative trait was measured for three individuals per accession and their means were used for analyses. The quantitative traits were rated on a three-point scale. Wax-like deposits and anthocyanin pigmentation on coleoptile (AC) and seedling (AS), glume hair (GH), and awnedness (AWN). HT, FT and MT are expressed as the number of days after November 1st. SD is obtained by dividing SN by SL. CL is the value of the longest culm of each individual. Every quantitative trait was measured for three individuals per accession and their means were used for analyses. The quantitative traits were rated on a three-point scale. Wax-like deposits and anthocyanin pigmentation on coleoptile (AC) and seedling (AS), glume hair (GH), and awnedness (AWN). We also scored glume hair and awnedness on three-point scales: hairless, short-hairy or hairy; and awnless, short-awn or long-awn, respectively.

**DNA extraction and genotyping** Total DNA was
Genetic diversity of a hexaploid core collection

extracted from young leaves of individual plants using a DNeasy Plant Maxi Kit (Qiagen Tokyo, Japan) according to the manufacturer’s protocol. DArTseq marker genotyping, a genotype-by-sequencing method, was carried out by Diversity Arrays Technology, Yarralumla, Australia. The DArTseq marker system generates two types of data: single-nucleotide polymorphisms (SNPs) and presence and absence variations (PAVs). We removed SNPs with call rates less than 0.8 or with heterozygous fractions less than 0.05 in the ensuing analyses. The marker system provides a pair of 69-bp sequence (tag) information; DNA polymorphisms therein were recognized as alleles. To select high-confidence markers and to reveal on which chromosome the markers were located, we performed a BLAST search (Zhang et al., 2000) for tag sequences on the IWGSC RefSeq ver. 1.0 (https://wheat-urgi.versailles.inra.fr/Seq-Repository) (International Wheat Genome Sequencing Consortium [IWGSC], 2018). We adopted only polymorphic markers with one of the tag sequences matched perfectly to a single position on a chromosome.

**Statistical analyses** Polymorphism information content (PIC) values were calculated for each marker according to Roldán-Ruis et al. (2000), as follows: $\text{PIC}_i = 2f_i(1 - f_i)$, where $\text{PIC}_i$ is the PIC of marker ‘$i’’, $f_i$ is the frequency of an allele and $1 - f_i$ is the frequency of another allele. Missing data were excluded from calculations.

We estimated the population structure by Bayesian clustering using the program ADMIXTURE 1.3 (Alexander et al., 2009). To estimate the optimal value of $K$, which describes the number of subpopulations making up the total population, we calculated a 10-fold cross-validation procedure for each $K$ value from 1 to 15. Principal component analysis (PCA) was also performed to estimate population structure using the pipeline implemented in the software TASSEL 5.0 (Bradbury et al., 2007). Missing data were imputed to the mean score for each locus (Jolliffe, 2002; Reich et al., 2008).

Inference on associations between genotypes and phenotypes was made using the pipeline implemented in TASSEL 5.0. We removed rare polymorphic markers that appeared in less than 5% of the accessions (PIC < 0.095) from analyses and used principal components (PC1 to 5) obtained from PCA to infer population structures. We performed analyses with an eigenvalue decomposition of the covariance matrix. We used a general linear model (GLM) for estimation of regression between genotypes and phenotypes. The permutation analysis was run 1,000 times.

**RESULTS**

Genotyping of NBRP-Wheat hexaploid wheat core collection by the DArT marker system The DArTseq genotyping system was applied for 188 accessions. We could not obtain genotyping data from one common wheat accession (6 × 164: T. aestivum ‘Nanbukomugi’), and obtained less than 80% data from one T. compactum accession (6 × 141: KU-9873). We removed these two accessions from the ensuing analyses. We obtained 34,234 SNP data sets by DArTseq genotyping. We removed low-confidence SNPs with call rates less than 0.8 or with heterozygous fractions greater than 0.05 from the SNP data sets. The mode of the frequency of missing data was 0.00–0.05 (Supplementary Fig. S1A). A total of 20,186 SNPs were selected and subjected to BLAST searches against IWGSC RefSeq ver. 1.0. A total of 13,394 tag sequences matched to single regions on the reference

| Chromosome | SNP Number | PIC | kbp/ SNP* | PAV Number | PIC | kbp/ PAV* |
|------------|------------|-----|-----------|------------|-----|-----------|
| chr1A      | 749        | 0.245 | 790       | 876        | 0.240 | 676       |
| chr2A      | 820        | 0.239 | 951       | 1,018      | 0.247 | 766       |
| chr3A      | 803        | 0.252 | 932       | 986        | 0.214 | 759       |
| chr4A      | 602        | 0.242 | 1,231     | 905        | 0.248 | 821       |
| chr5A      | 845        | 0.243 | 839       | 967        | 0.229 | 733       |
| chr6A      | 579        | 0.234 | 1,066     | 742        | 0.250 | 831       |
| chr7A      | 1,198      | 0.234 | 615       | 1,193      | 0.245 | 616       |
| chr1B      | 818        | 0.234 | 839       | 1,070      | 0.255 | 641       |
| chr2B      | 1,043      | 0.234 | 768       | 1,304      | 0.256 | 614       |
| chr3B      | 996        | 0.234 | 833       | 1,243      | 0.264 | 666       |
| chr4B      | 411        | 0.233 | 1,637     | 641        | 0.205 | 1,049     |
| chr5B      | 895        | 0.235 | 796       | 1,134      | 0.262 | 628       |
| chr6B      | 790        | 0.239 | 912       | 1,078      | 0.266 | 668       |
| chr7B      | 867        | 0.235 | 865       | 1,014      | 0.257 | 739       |
| chr1D      | 288        | 0.223 | 1,713     | 738        | 0.170 | 669       |
| chr2D      | 469        | 0.230 | 1,384     | 1,181      | 0.199 | 551       |
| chr3D      | 259        | 0.239 | 2,373     | 933        | 0.145 | 658       |
| chr4D      | 91         | 0.220 | 5,589     | 473        | 0.116 | 1,075     |
| chr5D      | 178        | 0.222 | 3,166     | 825        | 0.140 | 685       |
| chr6D      | 206        | 0.223 | 2,296     | 698        | 0.161 | 677       |
| chr7D      | 277        | 0.241 | 2,298     | 1,051      | 0.153 | 606       |
| chrUn**    | 210        | 0.239 | –         | 453        | 0.267 | –         |

SNP, PAV, and PIC are abbreviations for single-nucleotide polymorphism, presence and absence variation, and polymorphism information content, respectively. * Average distances between markers. ** A pseudomolecule composed of unanchored scaffolds (IWGSC RefSeq ver. 1.0).
sequences. We obtained 5,596 SNPs located on the A genome, 5,820 on the B genome, 1,768 on the D genome, and 210 on one pseudomolecule composed of unanchored scaffolds (chrUn, Table 2). The number of SNPs located on each chromosome ranged from 91 on chromosome 4D to 1,198 on chromosome 7A (Table 2). There were fewer SNPs located on homoeologous group-4 chromosomes than on other chromosomes in each genome (Table 2). The average PIC values for each chromosome ranged from 0.220 (chromosome 4D) to 0.252 (chromosome 3A) (Table 2).

A data set of 60,077 PAVs was obtained by DArTseq genotyping and the data call rates for every PAV were more than 0.8. Overall, the mode of frequency of missing data was 0.000–0.025. For the genomically tagged PAVs, the mode of frequency of missing data was 0.125–0.150 (Supplementary Fig. S1B). BLAST searches against IWGSC RefSeq ver. 1.0 resulted in the matching of a total of 20,523 tag sequences to unique reference sequences. We obtained 6,687, 7,484, 5,899 and 453 PAV tags located on genomes A, B, D and one pseudomolecule composed of unanchored scaffolds, respectively (Table 2). The PAVs assigned to each chromosome ranged from 473 on chromosome 4D to 1,304 on chromosome 2B (Table 2). The average PIC values of PAVs for each chromosome ranged from 0.116 on chromosome 4D to 0.266 on chromosome 6B (Table 2).

A total of 2,326 SNPs and 3,084 PAVs matched to at least one gene model sequence (Table 3) (IWGSC RefSeq annotation ver. 1.1 https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations). Due to the presence of multiple tag sequences in a gene model, we found a final

| Chromosome | No. SNPs within gene models | No. gene models including SNPs | No. PAVs within gene models | No. gene models including PAVs | No. markers within gene models | No. gene models including markers |
|------------|-----------------------------|--------------------------------|-----------------------------|-------------------------------|-------------------------------|---------------------------------|
| chr1A      | 126                         | 114                            | 129                         | 115                           | 255                           | 204                             |
| chr2A      | 145                         | 132                            | 185                         | 174                           | 330                           | 272                             |
| chr3A      | 107                         | 99                             | 119                         | 116                           | 226                           | 204                             |
| chr4A      | 100                         | 90                             | 101                         | 94                            | 201                           | 170                             |
| chr5A      | 130                         | 115                            | 133                         | 130                           | 263                           | 224                             |
| chr6A      | 131                         | 115                            | 127                         | 118                           | 258                           | 201                             |
| chr7A      | 165                         | 146                            | 134                         | 126                           | 299                           | 236                             |

A genome

| chr1B      | 133                         | 111                            | 166                         | 155                           | 299                           | 239                             |
| chr2B      | 193                         | 174                            | 213                         | 198                           | 406                           | 334                             |
| chr3B      | 190                         | 166                            | 174                         | 161                           | 364                           | 293                             |
| chr4B      | 92                          | 82                             | 98                          | 91                            | 190                           | 162                             |
| chr5B      | 148                         | 124                            | 166                         | 155                           | 314                           | 255                             |
| chr6B      | 156                         | 142                            | 168                         | 160                           | 324                           | 266                             |
| chr7B      | 130                         | 117                            | 134                         | 127                           | 264                           | 219                             |

B genome

| chr1D      | 48                          | 44                             | 124                         | 118                           | 172                           | 154                             |
| chr2D      | 115                         | 106                            | 185                         | 173                           | 300                           | 254                             |
| chr3D      | 45                          | 41                             | 161                         | 151                           | 206                           | 180                             |
| chr4D      | 20                          | 20                             | 90                          | 83                            | 110                           | 98                              |
| chr5D      | 25                          | 24                             | 134                         | 127                           | 159                           | 143                             |
| chr6D      | 39                          | 38                             | 135                         | 129                           | 174                           | 157                             |
| chr7D      | 44                          | 41                             | 149                         | 140                           | 193                           | 173                             |

D genome

| chrUn*     | 336                         | 314                            | 978                         | 921                           | 1,314                         | 1,159                           |

All

| 2,326      | 2,079                       | 3,084                          | 2,897                       | 5,410                         | 4,524                         |

SNP, and PAV are abbreviations for single-nucleotide polymorphism, and presence and absence variation, respectively.

* A pseudomolecule composed of unanchored scaffolds (IWGSC RefSeq ver. 1.0).
total of 4,524 gene models that contained at least one SNP or PAV (Table 3).

**Population structure analysis of the core collection** To investigate the population structure of the hexaploid wheat core collection, ADMIXTURE software

![Population structures based on DArTseq genotyping data. Analyses for K = 2 to 7 for the NBRP-Wheat hexaploid wheat core collection lines. The seven clusters, according to the largest Q value, are shown to the left of the bar plots. Boxes to the right indicate the non-aestivum exotic wheat accessions.](image-url)
was used for all the marker data. The optimal cluster number based on the cross-validation value is seven (Supplementary Fig. S2). Here, we name the seven clusters Pop-A to -G in the order of largest to smallest Q values (the ancestry fractions, Fig. 1 and Table 1).

Pop-A consisted of 28 Eastern Asian *T. aestivum* (27 Japanese and one Chinese) accessions and one *T. compactum* accession (KU-1208) (Table 1). The 25 Pop-B accessions originated from southwest China, Himalayan regions, and Southern Asian countries. The Chinese Spring standard cultivar was classified into this cluster. Notably, all *T. sphaerococcum* accessions belong to this cluster (Table 1). Pop-C contains 30 accessions including all *T. vavilovii* (KU-192 and KU-3848), two *T. compactum* (KU-3063 and KU-7350), one *T. spelta* (KU-3377) and 25 *T. aestivum*. They are derived from Western Asian regions (Pakistan, Afghanistan, Uzbekistan, Iran and Turkey) (Table 1). Pop-D, consisting of 28 *T. aestivum* and one *T. compactum* (KU-3242 from Iran) accessions, is characterized by richness of admixture accessions, with more than half being admixture accessions with Pop-C or Pop-F (Fig. 1). These admixture accessions are derived from Western Asian and European regions, respectively. The accessions with higher ancestry fractions in this cluster (more than 0.8) are from Turkey, Iran and Ethiopia. Pop-E contains 22 accessions, the majority of which are also admixture accessions. We noted that they are modern cultivars from various regions (Fig. 1 and Table 1). The accessions with higher ancestry fractions in this cluster (more than 0.8) are from Turkey, Syria and Lebanon. All but one of the accessions in Pop-F (35 accessions) are *T. aestivum*, the exception being a *T. compactum* accession derived from China (KU-152). Many modern European cultivars are included in this cluster and some accessions in this cluster are admixture accessions with Pop-D and Pop-E (Fig. 1). All 17 accessions classified into Pop-G are non-*aestivum* (*T. spelta* derived from Europe (Germany and Spain), *T. macha* and synthetic hexaploid) (Table

![Fig. 2. Scatter plots based on principal component analysis. Colors of each plot represent the seven clusters based on the population structure analysis illustrated in Fig. 1.](image-url)
Both *T. macha* and synthetic wheat accessions were admixtures with Pop-D (Fig. 1).

**Principal component analysis of the core collection** We used PCA as an alternative method for investigating the population structure of the hexaploid wheat core collection. The first, second and third principal components (PC1 to PC3) explained respectively 8.1, 5.6 and 4.1% of their variation. Accessions plotted on the PC1-PC2 plane formed a triangular shape (Fig. 2A). Accessions classified as Pop-A and Pop-C clustered separately at two vertices (Pop-A at the bottom and Pop-C at the top right in Fig. 2A) of the triangle. Accessions identified as Pop-B were plotted between Pop-A and Pop-C (Fig. 2A). Accessions classified as Pop-F and Pop-G clustered together at the third vertex of the triangle (Fig. 2A). Accessions of Pop-D and Pop-E were plotted along the line between Pop-C and Pop-F (Fig. 2A).

Accessions identified as Pop-G were divided into two sub-clusters according to PC1 and PC3 (Fig. 2B). These two sub-clusters correspond to a taxonomical discrimination between *T. spelta* and *T. macha*. Synthetic hexaploid wheat accession W7984 belongs to the sub-cluster with *T. macha*. On the other hand, scatter plots on the PC1-PC3 as well as on the PC2-PC3 planes show that Pop-G accessions are distinguishable from other accessions by PC3 (Figs. 2B, 2C).

**Phenotype diversity and association mapping of the hexaploid wheat core collection** We measured nine quantitative traits, namely heading time (HT), flowering time (FT), maturation time (MT), number of spikelets per spike (SN), spike length (SL), spike density (SD), culm length (CL), 100-grain weight (HGW) and number of tillers (TN); and seven qualitative traits, namely wax-like deposits on culm (WC), ochrea (WO) and spike (WS), anthocyanin pigmentation on coleoptile (AC) and seedling (AS), glume hair (GH), and awnedness (AWN) (Tables 4 and 5). We performed marker-trait association analyses for these traits using both SNP and PAV data sets. We detected a total of 183 significant marker-trait associations (MTAs) for 127 SNPs and 140 for 96 PAVs at a significance level of 1% after Bonferroni multiple test correction. Fourteen SNP and eight PAV marker-trait sequences were within gene model sequences (Supplementary Table S1).

The 68 MTAs (38 SNPs and 30 PAVs) with HT were distributed on 20 chromosomes except for 4D (Fig. 3A and Table 6). Five SNPs and five PAVs are within gene model sequences (Supplementary Table S1). We detected 66 significant MTAs (37 SNPs and 29 PAVs) with FT on all chromosomes except 4D and 6A (Supplementary Fig. S3A and Table 6). Four SNPs and three PAVs reside in gene model sequences (Supplementary Table S1). We detected 50 MTAs (32 SNPs and 18 PAVs) with MT on 17 chromosomes (Supplementary Fig. S3B and Table 6). Six SNPs and one PAV were within gene model sequences (Supplementary Table S1). MTA analyses with the three flowering-related traits (HT, FT and MT) resulted in detection of a similar set of markers: 26 markers (17 SNPs and nine PAVs) were commonly associated with all three flowering-related traits, 29 markers (13 SNPs and 16 PAVs) with the HT and FT traits, one SNP with the HT and MT traits, and three markers (two SNPs and one PAV) with the FT and MT traits (Supplementary Table S1). As expected, the phenotypic scores of the three flowering-related traits are highly correlated with each other (Supplementary Fig. S4).

A total of 47 MTAs with CL were found on all the chromosomes except 3D, 4D, 5D, 6D, 7A and 7D (Fig. 3B and Table 6). Three markers (two SNPs and one PAV) were within gene model sequences (Supplementary Table S1).

### Table 4. Summary of quantitative traits of the NBRP-Wheat hexaploid wheat core collection lines

| Trait | Max. | Min. | Mean | S.D. |
|-------|------|------|------|------|
| HT (days) | 230.0 | 155.7 | 187.3 | 14.2 |
| FT (days) | 232.0 | 164.3 | 190.5 | 12.9 |
| MT (days) | 248.0 | 209.7 | 224.4 | 8.2 |
| CL (cm) | 145.5 | 54.9 | 100.8 | 20.7 |
| TN | 34.0 | 7.0 | 16.6 | 4.9 |
| SL (mm) | 178.6 | 51.7 | 108.8 | 26.5 |
| SN | 29.0 | 10.7 | 18.2 | 3.3 |
| HGW (g) | 5.56 | 1.53 | 3.55 | 0.73 |
| SD (SN/SL) | 0.426 | 0.105 | 0.174 | 0.044 |

HT: heading time; FT: flowering time; MT: maturation time; CL: culm length; TN: number of tillers; SL: spike length; SN: number of spikelets per spike; HGW: 100-grain weight; SD: spike density.

### Table 5. Summary of qualitative traits of NBRP-Wheat hexaploid wheat core collection lines

| Trait – | + | ++ |
|--------|---|---|
| WC | 8 | 95 | 87 |
| WO | 5 | 51 | 134 |
| WS | 14 | 114 | 62 |
| AC | 130 | 47 | 12 |
| AS | 151 | 34 | 4 |
| GH | 164 | 14 | 12 |
|AWN | 5 | 51 | 134 |

WC, WO and WS: wax-like deposits on culm, ochrea and spike, respectively; AC and AS: anthocyanin pigmentation on coleoptile and seedling, respectively; GH: glume hair; AWN: awnedness.

–: Non-waxed, non-pigmentated, hairless or awnless.
+: Few waxed, few pigmented, short-hairy or short-awn.
++: Waxed, pigmented, hairy or long-awn.
Fig. 3. Manhattan plots of marker-trait associations. Blue, red and green circles indicate markers on the A, B and D genomes, respectively. Purple circles indicate markers on a pseudomolecule composed of unanchored scaffolds (IWGSC RefSeq ver. 1.0). In each chromosome, mapped markers are sorted according to genetic positions (from the tip of the short arm to the tip of the long arm, according to IWGSC RefSeq ver. 1.0). The vertical axis shows the negative logarithm of the association $P$-value. (A) Association with heading time (HT). (B) Association with culm length (CL). (C) Association with awnedness (AWN). (D) Association with anthocyanin pigmentation on coleoptile (AC).
Table 6. Summary of marker-trait associations (MTAs) detected in this study and of contigs, including significant marker tags

| Chromosome | HT | FT | MT | CL | TN | SL | SN | HGW | SD | WO | AC | AS | GH | AWN | Total | Within gene model |
|------------|----|----|----|----|----|----|----|-----|----|----|----|----|----|-----|-------|------------------|
| chr1A      | 3 | 1 | 2 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 1 | 1 | 0 | 1 | 14 | 3 | 0 |
| chr1B      | 3 | 0 | 3 | 1 | 2 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 10 | 3 |
| chr1D      | 1 | 1 | 1 | 0 | 1 | 0 | 2 | 0 | 1 | 4 | 6 | 5 | 0 | 1 | 11 | 6 |
| chr2A      | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 14 | 5 |
| chr2B      | 3 | 1 | 3 | 0 | 3 | 0 | 1 | 1 | 0 | 2 | 1 | 0 | 1 | 0 | 14 | 5 |
| chr2D      | 2 | 1 | 2 | 1 | 2 | 0 | 3 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 14 | 5 |
| chr3A      | 2 | 3 | 2 | 3 | 1 | 2 | 1 | 0 | 0 | 1 | 0 | 3 | 1 | 1 | 9 | 12 |
| chr3B      | 2 | 3 | 1 | 4 | 2 | 2 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 9 | 12 |
| chr3D      | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 5 |
| chr4A      | 1 | 6 | 1 | 5 | 1 | 1 | 1 | 2 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 14 |
| chr4B      | 2 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 11 |
| chr4D      | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 11 |
| chr5A      | 1 | 2 | 2 | 3 | 0 | 2 | 5 | 1 | 0 | 1 | 0 | 3 | 3 | 1 | 0 | 13 |
| chr5B      | 2 | 0 | 4 | 1 | 2 | 0 | 0 | 2 | 1 | 0 | 0 | 1 | 1 | 0 | 12 | 4 |
| chr5D      | 1 | 2 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 11 |
| chr6A      | 1 | 0 | 0 | 0 | 4 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 11 |
| chr6B      | 3 | 2 | 5 | 2 | 6 | 1 | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 19 | 5 |
| chr6D      | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 11 |
| chrUn*     | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| All        | 3 | 8 | 30 | 37 | 29 | 32 | 18 | 26 | 21 | 6 | 7 | 3 | 3 | 6 | 10 | 1 |
| within gene model | 5 | 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 22 |

HT: heading time; FT: flowering time; MT: maturation time; CL: culm length; TN: number of tillers; SL: spike length; SN: number of spikelets per spike; HGW: 100-grain weight; SD: spike density; WC, WO and WS: wax-like deposits on culm, ochrea and spike, respectively; AC and AS: anthocyanin pigmentation on coleoptile and seedling, respectively; GH: glume hair; AWN: awnedness.

Number of (SNPs) | (PAVs) detected with significant MTAs (a significance level of 1% after Bonferroni multiple test correction).

* A pseudomolecule composed of unanchored scaffolds (IWGSC RefSeq ver. 1.0).

Table S1). Notably, three markers significantly associated with CL were also associated with HT and FT, and two with HT (Supplementary Table S1). We detected MTAs with SN with sixteen markers (six SNPs and 10 PAVs) on chromosomes 2D, 3B, 4A, 4B, 5A, 5B, 7A and 7B (Supplementary Fig. S3E and Table 6). Four markers (one SNP and three PAVs) were commonly associated with SN as well as HT, FT and MT (Supplementary Table S1). Moreover, one PAV marker associated with SN on chromosome 4A was also detected in the MTA analysis of HT and FT (Supplementary Table S1).

The results of other MTAs for the traits TN, SL, SD, AWN, AC, AS, WO, WC, WS and GH are depicted in Supplementary Figs. S3 and S5, and summarized in Table 6. We found significant MTAs for the traits TN, SL, SD, AWN, AC, AS, WO and GH, but not for WC or WS. The number of significant MTAs for these traits varied from zero (WC and WS) to 26 (AS, Supplementary Table S1).

**DISCUSSION**

Diversity of the NBRP-Wheat hexaploid wheat core collection as revealed by DArT markers. The 187 accessions collected from various geographical regions were genotyped by the DArTseq marker system (34,234 SNPs and 60,077 PAVs). For analyses, we selected 13,394 SNPs and 20,523 PAVs as the data sets that could be anchored to the IWGSC RefSeq ver. 1.0 by BLAST.
The number of markers was highest in the B genome (5,820 SNPs and 7,484 PAVs) and lowest in the D genome (1,768 SNPs and 5,899 PAVs) (Table 2). These differences in the number of markers among genomes are coincident with those observed in previous studies (Akbari et al., 2006; Semagn et al., 2006; Zhang et al., 2011; Nielsen et al., 2014). The mean PIC value was also lowest in the D genome (SNPs: 0.230; PAVs: 0.159) (Table 2). These results show that the diversity is lower in the D genome than in the A and B genomes. The fewer polymorphisms present in the D genome may reflect the genetic bottleneck that occurred through hexaploidization and that brought only a limited amount of genetic diversity to Ae. tauschii (Caldwell et al., 2004; Chao et al., 2009; Mizuno et al., 2010).

Population structure and genetic relationships among common wheat Bayesian clustering analysis divided common wheat accessions into six clusters (Pop-A, -B, -C, -D, -E and -F) (Table 1) and most of them were admixtures (Fig. 1). However, there are some non-admixture accessions that help to infer the characteristics of each cluster (Fig. 1). Accessions classified as Pop-A belong to Eastern Asian lines present at the eastern limit of wheat distribution (mostly Japanese cultivars; Table 1). Landraces derived from the Himalayan region to Southern Asia are categorized into Pop-B (Table 1). Accessions included in Pop-C are mostly Western Asian lines (Afghanistan and Iran; Table 1). Landraces derived from the Middle East are categorized into Pop-D and Pop-E (Iran, Turkey, Lebanon and Syria) (Table 1). Most European landraces are classified as Pop-F (Table 1). Overall, the common wheat landraces of the NBRP-Wheat hexaploid wheat core collection are genetically divergent and form at least six groups each consisting of accessions from close geographical regions.

The accessions analyzed in this study form a triangular shape on the PC1-PC2 plane (Fig. 2). Accessions classified as Pop-A and Pop-C form two vertices of the triangle, and accessions of Pop-B are plotted on the side of the triangle between Pop-A and Pop-C (Fig. 2). Accessions classified as Pop-F form the third vertex of the triangle, and accessions of Pop-D and Pop-E were plotted on the side of the triangle between Pop-C and Pop-F (Fig. 2). The configurations indicate that the Himalayan region group (Pop-B) is situated between Eastern Asian (Pop-A) and Western Asian (Pop-C) groups, which may represent the dispersal pathway of hexaploid wheat through the Silk Road. Middle Eastern groups (Pop-D and Pop-E) are situated between Western Asian and European groups (Pop-F, Fig. 2), likely indicating that the latter two groups originated from the Middle Eastern groups. The finding of accessions that connect the vertices of the triangle (Fig. 2A) suggests that the differentiation of the six groups among common wheat landraces was genetically and geo-
and northwestern India (Percival, 1921; Ellerton, 1939; Mori et al., 2013). Together with Pakistani common wheat, all *T. sphaerococcum* accessions were classified as Pop-B according to population structure analysis (Fig. 1). This result suggests a single origin for *T. sphaerococcum* in Southern Asia.

*Triticum vavilovii* is endemic to the southern Caucasus. All *T. vavilovii* accessions were classified into Pop-C, together with *T. aestivum* accessions. They are admixed with *T. vavilovii* accessions derived from Western Asia: whereas common wheat is free-threshing, *T. vavilovii* is non-free-threshing. The Iranian *T. spelta* accession in Pop-C is non-free-threshing (Fig. 1), which may indicate that Asian spelt wheat contributed to the origin of *T. vavilovii*.

Club wheat (*T. compactum*) differs from other hexaploid varieties principally due to its compact spike, which results from the activity of the dominant *Ppd-1* locus on the short arms of chromosomes 2B and 2D and on the long arms of chromosomes 5A and 5D (Supplementary Table S1), indicating that HT was not associated with the *Ppd-1* locus on the long arm of *T. vavilovii*.

Club wheat accessions originated independently in various regions through mutations in different loci.

Association analyses We performed association analyses for nine quantitative traits and seven qualitative traits and found a total of 322 significant marker-trait associations (MTAs, Table 6). Here, we take MTAs for heading time (HT), culm length (CL), awnedness (AWN) and anthocyanin pigmentation on coleoptile (AC) as examples to compare our results to previous studies and test the potential usefulness of the NBRP-Wheat hexaploid wheat core collection for association mapping studies.

Eight QTLs for ear emergence are located on chromosomes 2B, 2D, 4A, 5A, 5D and 7D (for a summary of the wheat gene catalogue, see Gene Catalogue 2013, https://shigen.nig.ac.jp/wheat/komugi/genes/download.jsp). Meta-QTL analysis in elite European winter wheat germplasms found QTLs for ear emergence on chromosomes 1B, 1D, 2A, 3A, 3B, 4B, 4D, 5A, 6A, 6B, 7A, 7B and 7D (Griffiths et al., 2009). Neumann et al. (2011) detected MTAs for HT on chromosomes 1B, 1D, 2D, 4A, 4B, 5D, 6A, 6B and 7A by genome-wide association analysis of 96 winter wheats with 874 DArT markers. Edae et al. (2014) reported MTAs for HT on chromosomes 1D, 2A, 2B, 3A, 3B, 4B and 7D by genome-wide association analyses in 285-294 spring wheats grown in five different environments with 1,863 DArT markers. Genome-wide association analysis of a core collection of 227 bread wheats with 760 DArT markers revealed 62 markers corresponding to 33 chromosomal regions individually associated with earliness components (Le Gouis et al., 2012). These studies suggest that the HT trait is strongly related to the photoperiod response gene (*Ppd-1*) on the short arms of homeologous group-2 chromosomes and the vernalization response gene (*Vrn-1*) on the long arms of homeologous group-5 chromosomes. We found 68 MTAs (30 SNPs and 28 PAVs) for HT on 20 chromosomes, except for 4D (Fig. 3A and Table 6). The significant markers were located on the short arms of chromosomes 2B and 2D and on the long arms of chromosomes 5A and 5D (Supplementary Table S1), which may indicate the detection of the *Ppd-1* and *Vrn-1* homeologs in the association study. The significant markers on chromosome 2A were all located on the long arms (Supplementary Table S1), indicating that HT was too low to detect MTAs due to the predominance of the semi-dwarf accessions, as was also indicated by Neumann et al. (2011). Wheat plant height is controlled by genes located on at least 17 of the 21 chromosomes (Snape et al., 1977). Previously reported QTLs and genes for plant height (culm length; CL in this study) are summarized in Gene Catalogue 2013. Neumann et al. (2011) detected MTAs for plant height on chromosomes 1A, 1B, 2B, 4A, 6A, 6B, 7A and 7B; meanwhile, Edae et al. (2014) reported MTAs for plant height on chromosomes 3B, 5B, 6A, 7A and 7B. In this study, we found 47 MTAs (26 SNPs and 21 PAVs) for CL on chromosomes 1A, 1B, 1D, 2A, 2B, 2D, 3A, 3B, 4A, 4B, 5A, 5B, 6A, 6B and 7B (Fig. 3B and Table 6). The semi-dwarfing genes *Rht8*, *Rht-B1* and *Rht-D1* on chromosomes 2D, 4B and 4D, respectively, which have been derived from Japanese common wheat, are widely used in modern cultivars (Gale and Marshall, 1976; Börner et al., 1996; Korzun et al., 1998; Worland et al., 1998; Peng et al., 1999; Singh et al., 2001; Pearce et al., 2011). Accordingly, both Japanese wheat accessions and modern cultivars selected from the NBRP-Wheat hexaploid wheat core collection are expected to carry these semi-dwarfing genes. However, we did not detect MTAs on chromosome 4D. We suspect that allelic variation at the *Rht-D1* locus was too low to detect MTAs due to the predominance of the semi-dwarf accessions, as was also indicated by Neumann et al. (2011).

Awning of wheat is controlled by three genes, *Hd*, *B1* and *B2* (Gene Catalogue 2013), which are located on the short arm of chromosome 4A (Sears, 1954; Rao, 1981; Sourdille et al., 2002), the long arm of chromosome 5A
(Sears, 1954; Sourdille et al., 2002) and the long arm of chromosome 6B (Sears, 1954; Sourdille et al., 2002), respectively. Wheat varieties homozygous recessive for _hd_, _b1_ and _b2_ alleles are fully awned, while those homozygous dominant for either _Hd_ _B2_ or _B1_ _B2_ are awnless; those with _Hd_ _B2_ are expected to be awnless (Sourdille et al., 2002). We detected significant MTAs on chromosomes 2A, 4D and 5A (Fig. 3C and Table 6), and MTAs on chromosome 5A likely correspond to _B1_. However, we could not find MTAs corresponding to _Hd_ or _B2_. QTL analysis for three traits related to awning (awn length at the base, middle and top of spike) in a doubled-haploid population derived from the cross between _T. aestivum_ ‘Courtot’ (awned) and _T. aestivum_ ‘Chinese Spring’ (awnless) revealed two QTLs on chromosomes 4A and 6B that cosegregated with _Hd_ and _B2_, respectively. The QTL on chromosome 4A (_Hd_) had a stronger effect on the awns at both ends of the spike and that on chromosome 6B (_B2_) on the awns in the middle of the spike (Sourdille et al., 2002). Recently, Yoshioka et al. (2017) reported finding a variety of awned and awnless spikes in a segregating population between _T. aestivum_ ‘Mironovskaya 808’ (awnless) and _T. aestivum_ ‘Chinese Spring’ (awnless). Additionally, they found that all phenotypic variation could be explained by the three loci _Hd_, _B1_ and _B2_ by a combination of close observation and genotype prediction. In this study, however, we rated awnness as a qualitative trait with a three-point scale, which would have resulted in poor phenotypic resolution in detecting significant MTAs in other predicted genomic regions.

Anthocyanin pigments in wheat are found in the anthers (gene symbol _Pan_), auricle (_Ra_), leaf blade (_Plb_), leaf sheath (_Pls_), coleoptile (_Rc_), culm (_Pc_) or pericarp (_Pp_) (Gene Catalogue 2013). The _Pan_, _Plb_, _Pls_, _Rc_ and _Pc_ genes were closely linked on the short arms of homoeologous group-7 (Khlestkina et al., 2002, 2009). Ahmed et al. (2006) noted that the _Rc_ gene likely encodes a transcriptional activator of late biosynthesis genes involved in light-regulated anthocyanin synthesis. We detected significant MTAs with _AC_ on the short arm of chromosome 7A, as expected for the _Rc_ gene (Fig. 3D and Table 6).

The measurements for each trait in this study were obtained from a single cultivation season at one location. Therefore, we are not taking account of phenotypic variance due to environmental effects. However, as discussed above, some MTAs detected in this study are consistent with genes, QTLs and MTAs that have been reported previously. The genotype information acquired from the NBRP-Wheat haploid wheat core collection in this study provides suitable research tools to identify novel QTLs by marker-trait association analyses. The collection has previously been used in forward genetics studies (Yokota et al., 2016; Yoshioka et al., in press). The public release of the core collection with intensive genotype information should promote gene discovery in hexaploid wheat.

We thank the staff of NBRP-Wheat for their excellent technical support. This work was supported by the National BioResource Project-Wheat from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The accessions in the core collection are available from NBRP-Wheat. The genotype data including SNPs and PAVs are also to be available from the NBRP-Wheat Database KOMUGI (https://shigen.nig.ac.jp/wheat/komugi/). This manuscript is given the contribution number 620 by the Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University.

REFERENCES

Ahmed, N., Maekawa, M., Utsugi, S., Rikiishi, K., Ahmad, A., and Noda, K. (2006) The wheat _Rc_ gene for red coleoptile colour codes for a transcriptional activator of late anthocyanin biosynthesis genes. _J. Cereal Sci._ 44, 54–58.

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H., et al. (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. _Theor. Appl. Genet._ 113, 1409–1420.

Alexander, D. H., Novembre, J., and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. _Genome Res._ 19, 1655–1664.

Börner, A., Plaschke, J., Korzun, V., and Worland, A. J. (1996) The relationships between the dwarfing genes of wheat and rye. _Euphytica_ 89, 69–75.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. _Bioinformatics_ 23, 2633–2635.

Breseghelli, F., and Sorrells, M. E. (2006) Association mapping of kernel size and milling quality in wheat (_Triticum aestivum_ L.) cultivars. _Genetics_ 172, 1165–1177.

Buckler IV, E. S., and Thornberry, J. M. (2002) Plant molecular diversity and applications to genomics. _Curr. Opin. Plant Biol._ 5, 107–111.

Caldwell, K. S., Dvorak, J., Lagudah, E. S., Akhunov, E., Luo, M., Wolters, P., and Powell, W. (2004) Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. _Genetics_ 167, 941–947.

Cao, W., Hucl, P., Scoles, G., and Chibbar, R. N. (1998) Genetic diversity within spelta and macha wheats based on RAPD analysis. _Euphytica_ 104, 181–189.

Chao, S., Zhang, W., Akhunov, E., Sherman, J., Ma, Y., Luo, M. C., and Dubcovsky, J. (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (_Triticum aestivum_ L.) cultivars. _Mol. Breed._ 23, 23–33.

Crossa, J., Burgueño, J., Dreisigacker, S., Vargas, V., Herrera-Foessel, S. A., Lillemo, M., Singh, R. P., Trehovan, R., Warburton, M., Franco, J., et al. (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. _Genetics_ 177, 1889–1913.

Dvorak, J., Deal, K. R., Luo, M. C., You, F. M., von Borstel, K., and Dehghani, H. (2012) The origin of spelt and free-threshing hexaploid wheat. _J. Hered._ 103, 426–441.

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014) Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. _Theor. Appl. Genet._ 127, 791–807.

Ellerton, S. (1939) The origin and geographical distribution of
Genetic diversity of a hexaploid core collection

_Triticum sphaerococcum_ Perc. and its cytogenetical behaviour in crosses with _T. vulgare_ VILL. J. Genet. 58, 307–324.

Gale, M. D., and Marshall, G. A. (1976) The chromosomal location of _Gat1_ and _Rht1_, genes for gibberellin insensitivity and semi-dwarfism, in a derivative of Norin 10 wheat. Heredity 37, 283–289.

Giles, R. J., and Brown, T. A. (2006) _GluD_ allele variations in _Aegilops tauschii_ and _Triticum aestivum_: implications for the origins of hexaploid wheats. Theor. Appl. Genet. 112, 1563–1572.

Griffiths, S., Simmonds, J., Leverington, M., Wang, Y., Fish, L., Sayers, L., Alibert, L., Orford, S., Wingen, L., Herry, L., et al. (2009) Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. Theor. Appl. Genet. 119, 383–395.

Harlan, J. R. (1975) Our vanishing genetic resources. Science 188, 617–621.

International Wheat Genome Sequencing Consortium (IWGSC) (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361, eaar7191.

Jaccoud, D., Peng, K., Feinstein, D., and Kilián, A. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res. 29, e25.

Johnson, E. B., Nalam, V. J., Zemreta, R. S., and Kerra-Lizarazu, O. (2008) Mapping the compactum locus in wheat (_Triticum aestivum_ L.) and its relationship to other spike morphology genes of the Triticaceae. Euphytica 163, 193–201.

Jolliffe, I. T. (2002) Principal Component Analysis, Second Edition. Springer, New York, USA.

Khlestkina, E. K., Pestsova, E. G., Röder, M. S., and Börner, A. (2002) Molecular mapping, phenotypic expression and geographical distribution of genes determining anthocyanin pigmentation of coleoptiles in wheat (_Triticum aestivum_ L.). Theor. Appl. Genet. 104, 632–637.

Khlestkina, E., Pshenichnikova, T., Röder, M., and Börner, A. (2009) Clustering anthocyanin pigmentation genes in wheat group 7 chromosomes. Cereal Res. Commun. 37, 391–398.

Kihara, H. (1944) Discovery of the DD-analysre, one of the ancestors of _Triticum vulgare_. Agric. Hortic. 19, 889–890.

Korzn, V., Röder, M. S., Ganal, M. W., Worland, A. J., and Law, C. N. (1998) Genetic analysis of the dwarfing gene (_Rht8_) in wheat. Part I. Molecular mapping of _Rht8_ on the short arm of chromosome 2D of bread wheat (_Triticum aestivum_ L.). Theor. Appl. Genet. 98, 1104–1109.

Kosuge, K., Watanabe, N., Melnik, V. M., Laikova, L. I., and Goncharov, N. P. (2012) New sources of compact spike morphotype determined by the genes on chromosome 5A in hexaploid wheat. Genet. Resour. Crop Evol. 59, 1115–1124.

Le Gouis, J., Bordes, J., Ravel, C., Heumez, E., Faure, S., Prad, S., Galic, N., Remoué, C., Balfourier, F., Allard, V., et al. (2012) Genome-wide association study to identify chromosomal regions determining components of earliness in wheat. Theor. Appl. Genet. 124, 597–611.

Li, G., Xu, X., Bai, G., Carver, B. F., Hunger, B., Bonman, J. M., Kolmer, J., and Dong, H. (2016) Genome-wide association mapping reveals novel QTL for seedling leaf rust resistance in a worldwide collection of winter wheat. Plant Genome 9, plantgenome2016.06.0051.

Matsuoka, Y. (2011) Evolution of polyploid _Triticum_ wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. Plant Cell Physiol. 52, 750–764.

McCouch, S., Baute, G. J., Braddeen, J., Bramel, P., Brettign, P. K., Buckler, E., Burke, J. M., Charest, D., Cloutier, S., Cole, G., et al. (2013) Agriculture: Feeding the future. Nature 499, 22–24.

McFadden, E. S., and Sears, E. R. (1944) The artificial synthesis of _Triticum spelta_. Rec. Genet. Soc. Am. 13, 26–27.

Mizuno, N., Yamasaki, M., Matsuoka, Y., Kawahara, T., and Takumi, S. (2010) Population structure of wild wheat _D_ genome progenitor _Aegilops tauschii_ Cass.: implications for intraspecific lineage diversification and evolution of common wheat. Mol. Ecol. 19, 999–1013.

Mori, N., Ohta, S., Chiba, H., Takagi, T., Nimi, Y., Shinde, V., Kajale, M. D., and Osada, T. (2013) Rediscovery of Indian dwarf wheat (_Triticum aestivum_ L. ssp. _sphaerococcum_ (Perc.) MK.) an ancient crop of the Indian subcontinent. Genet. Resour. Crop Evol. 60, 1771–1775.

Nesbitt, M., and Samuel, D. (1996) From staple crop to extinction? The archaeology and history of the hulled wheats. In Hulled wheats, Proceedings of the First International Workshop on Hulled Wheats, 21. (eds.: Padulosi, S., Hammer, K., and Heller, J.), pp. 41–100, International Plant Genetic Resources Institute, Rome, Italy.

Neumann, K., Kobiljksi, B., Denić, S., Varshney, R. K., and Börner, A. (2011) Genome-wide association mapping: a case study in bread wheat (_Triticum aestivum_ L.). Mol. Breed. 27, 37–58.

Nielsen, N. H., Backes, G., Stougaard, J., Andersen, S. U., and Jahoor, A. (2014) Genetic diversity and population structure analysis of European hexaploid bread wheat (_Triticum aestivum_ L.) varieties. PLoS One 9, e94000.

Pearce, S., Saville, R., Vaughan, S. P., Chandler, P. M., Wilhelm, E. P., Sparks, C. A., Al-Kaff, N., Korolev, A., Boulton, M. L., Phillips, A. L., et al. (2011) Molecular characterization of _Rht-1_ dwarfing genes in hexaploid wheat. Plant Physiol. 157, 1820–1831.

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., et al. (1999) ‘Green revolution’ genes encode mutant gibberellin response modulators. Nature 400, 256–261.

Percival, J. (1921) The Wheat Plant: A Monograph. Duckworth and Co, London, UK.

Rao, M. V. P. (1972) Mapping of the _compactum_ gene C on chromosome 2D of wheat. Wheat Inf. Serv. 35, 9.

Rao, M. V. P. (1981) Telocentric mapping of the awn inhibitor gene _Hd_ on chromosome 4B of common wheat. Cereal Res. Commun. 9, 335–337.

Reich, D., Price, A. L., and Patterson, N. (2008) Principal component analysis of genetic data. Nat. Genet. 40, 491–492.

Remington, D. L., Thornberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Krosovich, S., Goodman, M. M., Edward, S., and Buckler, E. S. IV. (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA 98, 11479–11484.

Roldán-Ruiz, I., Dendauw, J., Van Bockstaele, E., Depicker, A., and De Loose, M. (2000) AFLP markers reveal high polymorphic rates in ryegrasses (_Lolium spp._). Mol. Breeding. 6, 125–134.

Salamin, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., and Martin, W. (2002) Genetics and geography of wild cereal domestication in the Near East. Nat. Rev. Genet. 3, 429–441.

Sears, E. R. (1954) The aneuploids of common wheat. Mo. Agr. Expt. Stat. Res. Bull. 572, 1–58.

Semagn, K., Bjørnstad, Å., Skinnes, H., Marøy, A. G., Tarkegne, Y., and Williams, M. (2006) Distribution of DARf, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid...
254  

S. TAKENAKA et al.

hexaploid wheat population. Genome 49, 545–555.
Singh, R. P., Huerta-Espino, J., Rajaram, S., and Crossa, J. (2001) Grain yield and other traits of tall and dwarf isolines of modern bread and durum wheats. Euphytica 119, 241–244.
Snpae, J. W., Law, C. N., and Worland, A. J. (1977) Whole chromosome analysis of height in wheat. Heredity 38, 25–36.
Sourdille, P., Cadalen, T., Gay, G., Gill, B., and Bernard, M. (2002) Molecular and physical mapping of genes affecting awning in wheat. Plant Breed. 121, 320–324.
Wang, G., Leonard, J. M., von Zitzewitz, J., Peterson, C. J., Ross, A. S., and Riera-Lizarazu, O. (2014) Marker–trait association analysis of kernel hardness and related agronomic traits in a core collection of wheat lines. Mol. Breed. 34, 177–184.
Worland, A. J., Korzun, V., Roëder, M. S., Ganal, M. W., and Law, C. N. (1998) Genetic analysis of the dwarfing gene Rht8 in wheat. Part II. The distribution and adaptive significance of allelic variants at the Rht8 locus of wheat as revealed by microsatellite screening. Theor. Appl. Genet. 96, 1110–1120.
Yokota, H., Iehisa, J. C. M., Nishijima, R., Nitta, M., Takenaka, S., Nasuda, S., and Takumi, S. (2016) Variation in abscisic acid responsiveness at the early seedling stage is related to line differences in seed dormancy and in expression of genes involved in abscisic acid responses in common wheat. J. Cereal Sci. 71, 167–176.
Yoshioka, M., Iehisa, J. C. M., Ohno, R., Kimura, T., Enoki, H., Nishimura, S., Nasuda, S., and Takumi, S. (2017) Three dominant awnless genes in common wheat: fine mapping, interaction and contribution to diversity in awn shape and length. PLoS One 12, e0176148.
Yoshioka, M., Takenaka, S., Nitta, M., Li, J., Mizuno, N., and Nasuda, S. Genetic dissection of grain morphology in hexaploid wheat by analysis of the NBRP-Wheat core collection. Genes Genet. Syst. (in press)
Zanke, C. D., Ling, J., Pieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, K., et al. (2014) Whole genome association mapping of plant height in winter wheat (Triticum aestivum L.). PLoS One 9, e113287.
Zanke, C. D., Ling, J., Pieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, F., et al. (2015) Analysis of main effect QTL for thousand grain weight in European winter wheat (Triticum aestivum L.) by genome-wide association mapping. Front. Plant Sci. 6, 644.
Zanke, C. D., Rodemann, B., Ling, J., Muqaddasi, Q. H., Pieske, J., Polley, A., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., et al. (2017) Genome-wide association mapping of resistance to eyespot disease (Pseudocercosporella herpotrichoides) in European winter wheat (Triticum aestivum L.) and fine-mapping of Pch1. Theor. Appl. Genet. 130, 505–514.
Zhang, L., Liu, D., Guo, X., Yang, W., Sun, J., Wang, D., Sourdille, P., and Zhang, A. (2011) Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. BMC Genet. 12, 42.
Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7, 203–214.
Zohary, D., and Hopf, M. (2000) Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe, and the Nile Valley. Oxford University Press, Oxford, UK.
Zondervan, K. T., and Cardon, L. R. (2004) The complex interplay among factors that influence allelic association. Nat. Rev. Genet. 5, 89–100.