ORIGINAL ARTICLE

Genome-wide polymorphisms from RNA sequencing assembly of leaf transcripts facilitate phylogenetic analysis and molecular marker development in wild einkorn wheat

Asami Michikawa[1] · Kentaro Yoshida[1] · Moeko Okada[1] · Kazuhiro Sato[2] · Shigeo Takumi[1]

Abstract

A survey of genome-wide polymorphisms between closely related species is required to understand the molecular basis of the evolutionary differentiation of their genomes. Two wild diploid wheat species, namely *Triticum monococcum* ssp. *aegilopoides* and *T. urartu*, are closely related and harbour the A^m and A genomes, respectively. The A-genome donor of tetraploid and common wheat is *T. urartu*, and *T. monococcum* ssp. *monococcum* is the cultivated form derived from the wild einkorn wheat subspecies *aegilopoides*. Although subspecies *aegilopoides* has been a useful genetic resource in wheat breeding, genome-wide molecular markers for this subspecies have not been sufficiently developed. Here, we describe the detection of genome-wide polymorphisms such as single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels) from RNA sequencing (RNA-seq) data of leaf transcripts in 15 accessions of the two diploid wheat species. The SNPs and indels, detected using the A genome of common wheat as the reference genome, covered the entire chromosomes of these species. The polymorphism information facilitated a comparison of the genetic diversity of einkorn wheat with that of two related diploid *Aegilops* species, namely, *Ae. tauschii* and *Ae. umbellulata*. Cleaved amplified polymorphic sequence (CAPS) markers converted from the SNP data were efficiently developed to confirm the addition of *aegilopoides* subspecies chromosomes to tetraploid wheat in nascent allohexaploid lines with AABBA^{m}A^m genomes. In addition, the CAPS markers permitted linkage map construction in mapping populations of *aegilopoides* subspecies accessions. Therefore, these RNA-seq data provide information for further breeding of closely related species with no reference genome sequence data.

Communicated by Stefan Hohmann.

✉ Kentaro Yoshida
kentaro.yoshida@port.kobe-u.ac.jp

✉ Shigeo Takumi
takumi@kobe-u.ac.jp

1 Graduate School of Agricultural Science, Kobe University, Rokkodai 1-1, Nada, Kobe, Hyogo 657-8501, Japan

2 Institute of Plant Science and Resources, Okayama University, Kurashiki, Okayama, Japan

Introduction

Closely related species are useful resources in crop breeding, and efficient use of crop relatives by breeders is important for addressing recent and future climate change (Fita et al. 2015; Prohens et al. 2017). For the identification of agronomically important genes in crop relatives and the confirmation of introduced segments of the related species' chromosomes, genome-wide polymorphism information can be effective not only within the related species but also between the crop and its relatives (Rasheed et al. 2018). However, reference genome sequence information is not available for most crop relatives. Although cultivated wheat species have many wild relative species in *Triticum* and *Aegilops*, reference genome sequences have been published for only three wild relatives (Avni et al. 2017; Luo et al. 2017; Ling et al. 2018). However, in recent studies, next-generation sequencing data enabled in silico analysis of the

wild relative species' genomes based on the chromosomal synteny among wheat homoeologous genomes (Nishijima et al. 2016; Okada et al. 2018).

Einkorn wheat, *Triticum monococcum* L. (A^mA^m genome), includes two subspecies, with ssp. *monococcum* corresponding to the cultivated form and ssp. *aegilopoides* (Link) Thell. (syn. *T. boeoticum* Boiss) corresponding to the wild form. Wild einkorn wheat is widely distributed in countries of the eastern Mediterranean, and domestication from ssp. *aegilopoides* to ssp. *monococcum* in *T. monococcum* is inferred to have occurred once, in the Karacadag region of southeastern Turkey (Heun et al. 1997). *Triticum urartu* Tumanian ex Gandilyan (AA genome), which is closely related to einkorn wheat, is the A-genome donor of tetraploid wheat, namely *T. turgidum* L. (AABB genome) and *T. timopheevii* Zhuk (AAGG genome), and common wheat, *T. aestivum* L. (AABBDD genome) (Dvorak et al. 1988; Takumi et al. 1993). The distribution of *T. urartu* is restricted almost exclusively to the "Fertile Crescent" (Wang et al. 2017). Reproductive barriers, including abnormal hybrid seed formation and hybrid necrosis, underlie the separation of *T. monococcum* ssp. *aegilopoides* and *T. urartu* (Johnson and Dhaliwal 1976; Gill and Waines 1978; Fricano et al. 2014; Takamatsu et al. 2015). Interspecific hybrids between *T. monococcum* and *T. urartu* are almost always sterile, and many chromosomal rearrangements appear to exist between the A and A^m genomes (Dubcovsky et al. 1996; Fricano et al. 2014). Postzygotic reproductive isolation could result, at least in part, from nuclear genome differentiation between the two species, which was previously reported at the nuclear and organellar DNA levels (Castagna et al. 1994; Mizumoto et al. 2002; Brandolini et al. 2006).

Triticum urartu is evolutionarily important as the A-genome donor of polyploid wheat species, and the genome sequence of *T. urartu* has already been published (Ling et al. 2013, 2018). The *T. urartu* population harbours a large amount of variation in agronomically important traits such as disease resistance and grain quality (Qiu et al. 2005; Guzmán and Alvarez 2012). The recent accumulation of genome information has included surveys of genetic diversity and variation in *T. urartu* (Luo et al. 2015; Wang et al. 2017; Brunazzi et al. 2018). Cultivated einkorn wheat (ssp. *monococcum*) is an important resource for improving grain quality, resistance to various diseases, and resistance to abiotic stress in durum and common wheat (Vasu et al. 2001; Tranquilli et al. 2002; James et al. 2006). Domestication of einkorn wheat was achieved by early farmers in the Fertile Crescent, implying that cultivated einkorn wheat may retain a large amount of diversity and useful alleles for various traits (Jing et al. 2007). In recent years, cultivated einkorn wheat has served as a valuable model for wheat genetics and has been useful in the screening of mutant strains to identify mutant alleles (Yan et al. 2004; Murai et al. 2013; Gardiner

et al. 2014). Diverse accessions of cultivated einkorn wheat have also been useful for identifying agronomically important genes based on genome-wide association with molecular markers (Jing et al. 2007).

Similarly, wild einkorn wheat (ssp. *aegilopoides*) has been a useful resource for improving disease resistance and grain quality in common wheat (Rogers et al. 1997; Shi et al. 1998; Anker and Niks 2001; Hovhannisyan et al. 2011). However, in contrast to *T. urartu*, little genomic information based on next-generation sequencing (NGS) techniques is available for wild einkorn wheat. Whole-genome sequencing and exome sequencing approaches are convenient for multiple samples of wild wheat species because of the large genome size and sequencing costs of these species. RNA-seq is an effective approach for avoiding these problems. The RNA-seq approach enables us to find single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels) covering entire chromosomal regions in various diploid wheat relatives (Iehisa et al. 2012, 2014; Nishijima et al. 2016; Wu et al. 2018; Okada et al. 2018; Miki et al. 2019). To date, RNA-seq data from one accession of wild einkorn wheat have been compared with those from an einkorn wheat cultivar; a large number of polymorphisms were obtained in this comparative study (Fox et al. 2014). Some linkage maps have been constructed, mainly in two mapping populations, permitting comparisons of cultivated and wild accessions of einkorn wheat to find useful genes (Bullrich et al. 2002; Shindo et al. 2002; Hori et al. 2007; Jing et al. 2009; Yu et al. 2017). However, to the best of our knowledge, no genetic map has been constructed for any population derived from intra-subspecies crosses of wild einkorn wheat. For future use of genetic variation in wild einkorn wheat in breeding, genome-wide molecular markers need to be developed to permit efficient detection of DNA polymorphisms in specific chromosomal regions.

Recently, high-quality reference genome sequences have been reported for the A genomes of diploid, tetraploid, and hexaploid wheat (Avni et al. 2017; Ling et al. 2018; International Wheat Genome Sequencing Consortium (IWGSC) 2018). The pseudomolecules of the A genome sequences can be utilized as references for virtual anchoring of the NGS reads and polymorphisms to each chromosome of the A^m genome. The DNA polymorphisms from the NGS data showed a genome-wide distribution with high resolution, even between intra-lineage accessions of the wild diploid D-genome species *Aegilops tauschii* Coss (Nishijima et al. 2016). Thus, the RNA-seq approach has potential for the discovery of genome-wide SNPs and indels, permitting anchoring of these markers to each of the A^m-genome chromosomes in wild einkorn wheat. The objectives of the present study are (1) to detect genome-wide polymorphisms distributed across the entire chromosomes of wild einkorn wheat (ssp. *aegilopoides*), (2) to use the polymorphism data

to develop a large number of molecular markers distinguishing the A^m-genome chromosomes from A-genome chromosomes, and (3) to confirm the use of polymorphism data in the construction of linkage maps for wild einkorn wheat. To achieve these objectives, we performed RNA-seq analysis of 15 accessions of two diploid wheat species, namely *T. monococcum* and *T. urartu*. We provide two examples in which the polymorphism data are converted to PCR-based markers in wild einkorn wheat.

Materials and methods

Plant materials

Fifteen accessions of diploid wheat were used in this study, including three accessions of *T. urartu*, two of *T. monococcum* ssp. *monococcum*, and ten of *T. monococcum* ssp. *aegilopoides*. With the exception of DV92 and PI427634, diploid wheat seeds were obtained from the wheat genetic resource centre of the National BioResource Project-Wheat (Japan, https://shigen.nig.ac.jp/wheat/komugi/top/top.jsp) (Table 1). DV92 is an accession widely used as a parental line for population mapping and mutant panel construction (Bullrich et al. 2002; Murai et al. 2013).

An F_2 mapping population was generated from a cross between two wild einkorn wheat accessions: KU-3620 and KU-8276. Seeds of the F_2 population, with a population size of 103, were sown in November 2016 and the F_2 individuals as well as the two parental accessions were grown individually in randomly arranged pots during the 2016–2017 season in an experimental field at Kobe University (34°43′N, 135°13′E). The heading and flowering times of the F_2 individuals were recorded as days after sowing.

A synthetic hexaploid line was used to check the utility of the A^m-genome-specific markers developed in this study. For synthetic hexaploid production, a tetraploid wheat cultivar, Langdon (Ldn), was crossed with pollen of the wild einkorn wheat accession KU-3620, and one of the resulting F_1 plants was treated with 1 g L^{-1} colchicine and 2% dimethyl sulfoxide solution for 5 h to obtain selfed seeds. After confirmation of the somatic chromosome number (42) using root tips, a synthetic hexaploid with an $AABBA^mA^m$ genome was established and designated Ldn/KU-3620.

RNA sequencing

Total RNA was extracted using Sepasol-RNA I Super G (Nacalai Tesque, Kyoto, Japan) from the leaves of 10-day-old plants grown under conditions of 16-h light/8-h dark and 24 °C. Paired-end libraries for RNA-seq were constructed from 6 to 10 µg of total RNA using a TruSeq RNA Library Preparation kit v2 (Illumina, San Diego, CA, USA)

according to the manufacturer's procedure (Sato et al. 2016); the resulting libraries were then sequenced by 300-bp paired-end reads on an Illumina MiSeq sequencer. Five libraries per run were used for sequencing, and approximately 28 million reads were obtained. The sequenced reads were deposited in the DDBJ Sequence Read Archive under accession number DRA007574.

The quality of sequencing reads was evaluated using FASTQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmomatic software, version 0.33 (Bolger et al. 2014), was used to remove adapter sequences, low-quality bases with an average quality score per 4 bp of < 30, and reads of fewer than 50 bp. The filtered reads were aligned to the reference A genome sequence of *T. aestivum* cv. Chinese Spring (CS) version 1 (International Wheat Genome Sequencing Consortium (IWGSC) 2018) using HISAT2 software version 2.1.0 (Kim et al. 2015). The RNA-seq reads of 10 accessions of *Ae. tauschii* (Nishijima et al. 2016), 12 accessions of *Ae. umbellulata* (Okada et al. 2018), and 1 accession of *Ae. speltoides* Tausch (Miki et al. submitted) were also used for phylogenetic tree construction (Table 1). These RNA-seq reads were obtained from the DDBJ Sequence Read Archive as follows: DRA004604 for *Ae. tauschii*, DRA006404 for *Ae. umbellulata*, and DRA007097 for *Ae. speltoides*.

Phylogenetic analysis

SNPs and indels were called using SAMtools (Li et al. 2009) and Coval (Kosugi et al. 2013) with the same criteria as those described in Nishijima et al. (2016); the depth of read coverage was ≥ 10, and > 95% of the mapped reads included nucleotide sequences that differed from the reference sequence of the A genome. To obtain a high-confidence set of SNPs for construction of phylogenetic trees, we chose SNPs at positions for which the read depth was ≥ 10 and at which no ambiguous nucleotides were detected in any of the tested accessions. CIRCOS (Krzywinski et al. 2009) and R statistical software were used to visualize the distribution of SNPs/indels on the physical map of the A genome. Neighbour-joining (NJ) and maximum likelihood (ML) phylogenetic trees were constructed using Molecular Evolutionary Genetics Analysis (MEGA) software, version 7.0 (Kumar et al. 2016).

Marker development, map construction, and QTL analysis

De novo transcriptome assembly of the 12 accessions of einkorn wheat and the 3 accessions of *T. urartu* was performed using Trinity (Grabherr et al. 2011). If multiple isoforms were detected, the first isoform designated by Trinity was selected as a representative transcript. RNA-seq short

Table 1 List of the diploid wheat accessions used in this study

Species	Accession number	Origin	DRA number
T. monococcum ssp. *aegilopoides*	KU-101-3	Iran	DRA007574 (in this study)
	KU-3620	Turkey	
	KU-3646	Turkey	
	KU-8111	Iraq	
	KU-8120	Iraq	
	KU-8267	Iraq	
	KU-8276	Turkey	
	KU-8287	Turkey	
	KU-10725	Turkey	
	PI427634	Turkey	
T. urartu	KU-199-5	Turkey	DRA007574 (in this study)
	KU-199-11	Turkey	
	KU-199-16	Lebanon	
T. monococcum ssp. *monococcum*	DV92	Italy	DRA007574 (in this study)
	KT003-001	Unknown	
Ae. tauschii	AT76	China	DRA004604 (Nishijima et al. 2016)
	PI499262	China	
	KU-2627	Afghanistan	
	KU-2025	Afghanistan	
	KU-2087	Iran	
	KU-2003	Pakistan	
	KU-2078	Iran	
	KU-2124	Iran	
	KU-2075	Iran	
	KU-2093	Iran	
Ae. umbellulata	KU-4017	Iraq	DRA006404 (Okada et al. 2018)
	KU-4026	Iraq	
	KU-4035	Iraq	
	KU-4043	Iraq	
	KU-4052	Iraq	
	KU-4103	Turkey	
	KU-5934	Turkey	
	KU-5954	Turkey	
	KU-8-7	Turkey	
	KU-12180	Greece	
	KU-12198	Greece	
	KU-8-5	Syria	
Ae. speltoides	KU-7848	Iraq	DRA007097 (Miki et al. 2019)

KU Plant Germplasm Institute, Faculty of Agriculture, Kyoto University, Japan; *PI* National Small Grain Research Facility, USDA-ARS, USA; *KT* Kihara Institute for Biological Research, Yokohama City University, Japan

reads of *T. monococcum* ssp. *aegilopoides* KU-8276, one of the parental accessions of the F_2 mapping population, were aligned to the representative transcripts of another parental accession, *T. monococcum* ssp. *aegilopoides* KU-3620, using Bowtie2 (Langmead and Salzberg 2012). SNP calling was conducted with the same pipeline described above. SNPs and indels were called when the depth of read coverage was ≥ 10, and > 95% of the aligned reads included nucleotide sequences that differed from sequences of the assembled transcript. The representative transcripts of KU-3620 were anchored to the chromosomes of the A genome of CS (International Wheat Genome Sequencing Consortium (IWGSC) 2018) using GMAP (Wu and Watanabe 2005). According to the location of the anchored transcripts, the SNPs between KU-8276 and KU-3620 were placed on the chromosomes of the A genome.

Based on the SNP and indel information, primer sets were designed using Primer3Plus software (Untergasser et al. 2007). Some of the identified SNPs were converted to cleaved amplified polymorphic sequence (CAPS) markers (Supplementary Table S1). Information for four simple-sequence-repeat (SSR) markers and their respective annealing temperatures was obtained from the GrainGenes website (https://wheat.pw.usda.gov/GG2/index.shtml). PCR amplification, digestion of the PCR products, and visualization of the products were performed as described in our previous study (Sakaguchi et al. 2016).

Genetic mapping using the genotyping data was performed using the MAPMAKER/EXP version 3.0 package; the logarithm-of-odds (LOD) score threshold was set to 3.0 (Lander et al. 1987). Quantitative trait locus (QTL) analysis of heading and flowering times was conducted by single-marker analysis and composite interval mapping with Windows QTL Cartographer version 2.5 software and the backward regression method (http://statgen.acsu.edu/qtlcart/WQTLCart.htm). The LOD score threshold for QTL analysis was determined by a 1000-permutation test. The statistical significance of the QTL effect on the examined traits was estimated by Tukey–Kramer's HSD test.

Results

Single-nucleotide polymorphisms in the diploid wheat species

To estimate genome-wide DNA polymorphisms in einkorn wheat and *T. urartu*, 300-bp paired-end RNA sequencing of ten accessions of *T. monococcum* ssp. *aegilopoides*, two accessions of *T. monococcum* ssp. *monococcum*, and three accessions of *T. urartu* was performed, generating 220–410 million filtered paired-end reads for each accession (Table 2, Supplementary Table S2). The filtered reads were aligned to the reference sequence of the A genome of CS (International Wheat Genome Sequencing Consortium (IWGSC) 2018). Of the filtered reads, 86.07–95.00% for *T. monococcum* ssp. *aegilopoides*, 84.44–88.46% for *T. urartu*, and 91.43–97.01% for *T. monococcum* ssp. *monococcum* were aligned to the reference sequence.

By conducting pairwise comparisons of all the tested accessions of the diploid wheat species and the A genome of CS, 21,057–109,314 SNPs and 315–1853 indels for *T. monococcum* ssp. *aegilopoides,* 45,101–59,378 SNPs and 758–1109 indels for *T. monococcum* ssp. *monococcum*, and 33,286–53,419 SNPs and 644–1286 indels for *T. urartu* were detected (Table 3). These SNPs and indels covered all the chromosomes of the A genome of CS (Fig. 1).

Phylogenetic relationships among the diploid wheat species

To clarify the phylogenetic relationships among *T. monococcum* ssp. *aegilopoides*, *T. monococcum* ssp. *monococcum, T. urartu*, and the other wild diploid wheat species, NJ and ML trees were constructed for 10 accessions of *Ae. tauschii* and 12 accessions of *Ae. umbellulata*; these trees were based on the high-confidence set of SNPs and included *Ae. speltoides* as an outgroup species (Fig. 2). The einkorn wheat species and *T. urartu* were separated from the other diploid wheat species, consistent with the findings of previous reports (Mizumoto et al. 2002; Fricano et al. 2014). The comparison based on the branch lengths of the trees

Table 2 Summary of RNA sequencing information for the 15 accessions of diploid wheat

Species	Accession	Read pairs	Filtered read pairs (%)	Alignment rate to the A genome of CS (%)
T. monococcum ssp. *aegilopoides*	KU-101-3	6,075,718	3,228,143 (53.14%)	88.99
	KU-3620	6,947,661	3,452,355 (49.69%)	88.24
	KU-3646	5,916,083	3,308,350 (55.92%)	86.07
	KU-8111	5,671,368	3,164,909 (55.81%)	89.79
	KU-8120	5,781,015	3,264,122 (56.46%)	89.91
	KU-8267	5,028,572	2,837,046 (56.42%)	93.51
	KU-8276	5,299,177	2,776,271 (52.39%)	92.75
	KU-8287	5,854,651	3,497,596 (59.74%)	95.00
	KU10725	5,608,746	2,931,313 (52.26%)	92.64
	PI427634	6,548,802	4,046,867 (61.80%)	88.10
T. urartu	KU-199-5	4,364,112	2,597,523 (59.52%)	87.87
	KU-199-11	5,248,913	3,186,090 (60.70%)	88.46
	KU-199-16	5,306,355	2,847,868 (53.67%)	84.44
T. monococcum ssp. *monococcum*	DV92	6,294,027	2,963,686 (47.09%)	91.43
	KT003-001	4,773,333	2,250,442 (47.15%)	97.01

Table 3 Identification of SNPs and indels between the diploid wheat accessions and the A genome of CS

Species	Accession	SNPs	Indels
T. monococcum ssp. *aegilopoides*	KU-101-3	68,978	1059
	KU-3620	73,593	1423
	KU-3646	91,749	1759
	KU-8111	62,474	799
	KU-8120	69,117	979
	KU-8267	70,678	1009
	KU-8276	68,765	1057
	KU-8287	109,314	1853
	KU-10725	45,805	772
	PI427634	21,057	315
T. urartu	KU-199-5	41,384	1145
	KU-199-11	53,419	1286
	KU-199-16	33,286	644
T. monococcum ssp. *monococcum*	DV92	45,101	758
	KT003-001	59,378	1109
Ae. tauschii	AT76	152,202	1624
	PI499262	202,613	2052
	KU-2627	231,625	2468
	KU-2025	109,404	1042
	KU-2087	242,384	2597
	KU-2003	215,883	2098
	KU-2078	209,773	2271
	KU-2124	223,990	2346
	KU-2075	233,817	2554
	KU-2093	212,025	2089
Ae. umbellulata	KU-4017	89,189	780
	KU-4026	23,518	232
	KU-4035	153,197	1600
	KU-4043	93,137	983
	KU-4052	100,660	956
	KU-4103	35,700	340
	KU-5934	146,081	1544
	KU-5954	111,307	927
	KU-8-7	114,241	1185
	KU-12180	106,179	1042
	KU-12198	116,340	1018
	KU-8-5	123,484	1257
Ae. speltoides	KU-7848	146,008	1268

The A genome of CS was used as a reference sequence (International Wheat Genome Sequencing Consortium (IWGSC) 2018)

indicated that *T. monococcum* ssp. *aegilopoides* had the highest genetic diversity among the einkorn wheat species and *T. urartu*. The genetic diversity in *T. monococcum* ssp. *aegilopoides* was higher than that in *Ae. umbellulata* but was lower than that in *Ae. tauschii*. *T. monococcum* ssp. *aegilopoides* diverged into three groups. The existence of three groups in wild einkorn wheat was consistent with a

previous observation (Pour-Aboughadareh et al. 2017). *T. monococcum* ssp. *monococcum* belonged to the group that contained *T. monococcum* ssp. *aegilopoides*; this group included accessions distributed throughout southeastern Turkey and Iran. The other two groups were geographically separated; members of one group were limited to areas surrounding Ankara, Turkey, while members of the other group were widely distributed throughout southeastern Turkey, Iran, and Iraq.

Although the genetic distance between the A^m- and A-genome species was smaller than the genetic distance between the U- and D-genome species, the A-genome species were clearly separated from the A^m-genome species (Fig. 2). The three accessions of the A-genome species *T. urartu* were genetically distinguished with 100% bootstrap probability and geographically separated. All of the tested accessions exhibited divergence from the A genome of CS. The genetic distance between *T. urartu* and the A genome of CS was larger than that between the wild and cultivated einkorn wheat species.

Development of genetic markers that discriminate between the sub-genomes

To develop genetic markers that distinguish between species or between the A^m and A genomes, fixed nucleotide differences (SNPs) that distinguished species or the sub-genomes were evaluated. *T. monococcum* ssp. *aegilopoides*, *T. monococcum* ssp. *monococcum*, and *T. urartu* were distinguished from the A genome of CS by 11,903, 14,021, and 5693 fixed nucleotide differences, respectively (Table 4). Of these fixed nucleotide differences, 25, 187, and 1892 were uniquely observed in *T. monococcum* ssp. *aegilopoides*, *T. monococcum* ssp. *monococcum*, and *T. urartu*, respectively. A total of 8309 fixed nucleotide differences were observed between the A^m genome of *T. monococcum* ssp. and the A genome of *T. urartu*; these differences were widely distributed over all of the chromosomes (Fig. 3a). A total of 228 fixed nucleotide differences were observed between *T. monococcum* ssp. *aegilopoides* and *T. monococcum* ssp. *monococcum*; these differences were frequently located in the distal chromosomal regions (Fig. 3b). The fixed nucleotide differences between *T. urartu* and the A genome of CS were also widely distributed over all the chromosomes (Fig. 3c). The right (long-arm) end of chromosome 4 lacked SNPs in the comparisons between the two subspecies of *T. monococcum* and between *T. urartu* and the A genome of CS.

Based on the fixed nucleotide differences between the A^m and A genomes, seven A^m-chromosome-specific CAPS markers were designed for each chromosome (Supplementary Table S1). To test the utility of these CAPS markers, the markers were applied to Ldn, CS, *T. urartu*, and *T. monococcum* ssp. *aegilopoides* KU-3620, and four individuals of a
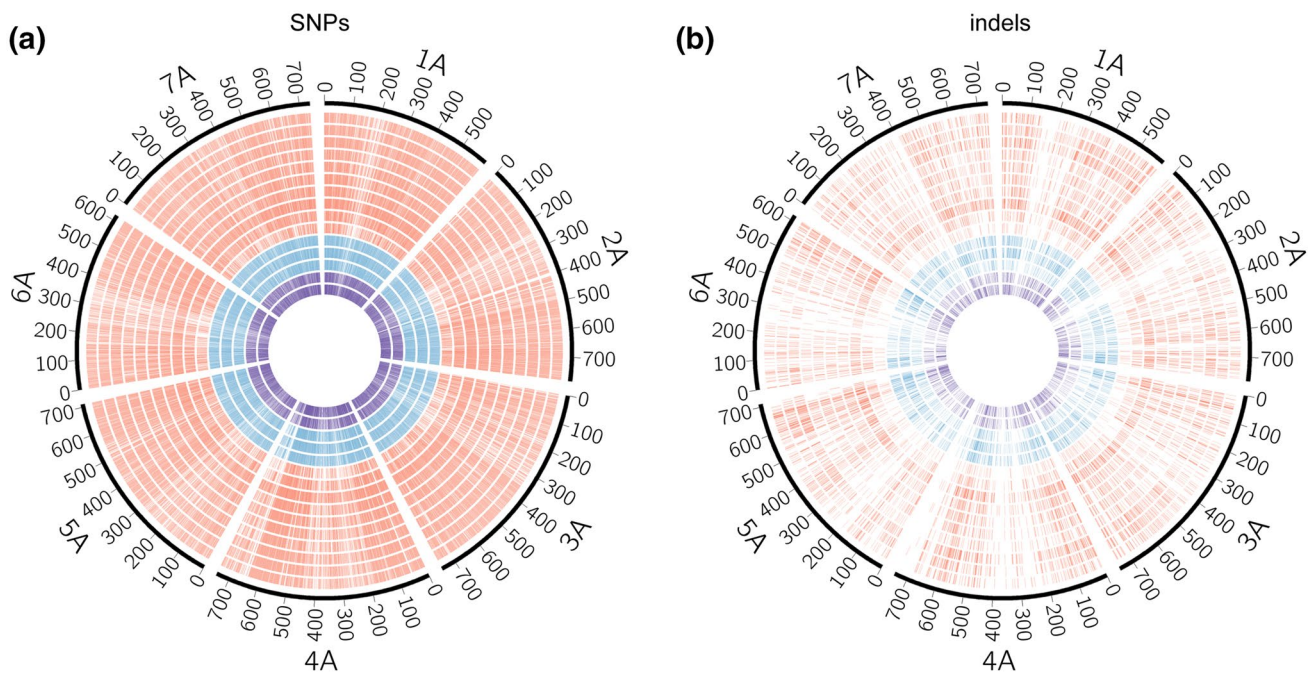
Fig. 1 Distribution of SNPs (**a**) and indels (**b**) between the A genome of CS and the examined accessions of einkorn wheat and *T. urartu* across the seven chromosomes of the A genome. Red, blue, and pur-ple circles indicate the ten *T. monococcum* ssp. *aegilopoides* accessions, three *T. urartu* accessions, and two *T. monococcum* ssp. *monococcum* accessions, respectively (color figure online)

synthetic hexaploid line (Ldn/KU-3620) with AABBA^{m}A^m genomes (Fig. 4). These markers successfully discriminated between the A^m and A genomes for each chromosome. The four individuals of the Ldn/KU-3620 line retained all seven chromosomes derived from the A^m genome.

Application of molecular markers for genetic analysis of wild einkorn wheat

In addition to SNPs that were detected based on the alignment of RNA-seq reads to the reference sequence of the A genome of CS (International Wheat Genome Sequencing Consortium (IWGSC) 2018), we also detected SNPs between the F$_2$ parents, *T. monococcum* ssp. *aegilopoides* KU-3620 and KU-8276 based on de novo assembled transcripts, as described below. De novo transcriptome assembly of the tested diploid accessions generated 25,422–75,573 transcripts; the N50 values of these assemblies ranged from 914 to 1325 bp. Alignment of the RNA-seq data of short reads of KU-8276 to the representative transcripts of KU-3620 revealed 18,360 SNPs, 8925 of which were shared with those detected based on the A genome of CS (Fig. 5). Thus, the approach using the de novo transcriptome was able to detect a larger number of SNPs between accessions of *T. monococcum* ssp. *aegilopoides*. The identified SNPs and indels anchored to chromosomal positions spanning the entire A genome of CS (Table 5).

Genetic markers for the F$_2$ mapping population were developed based on the SNPs and indels obtained by integrating both approaches. To convert the identified SNPs and indels to PCR-based genetic markers, ten CAPS and two indel markers were designed on chromosome 7A^m and applied to the F$_2$ mapping population (Supplementary Table S1). The fragments digested by the restriction enzymes clearly distinguished between the F$_2$ parental accessions of *T. monococcum* ssp. *aegilopoides* KU-3620 and KU-8276, and the resulting polymorphisms were used to genotype the F$_2$ individuals. Based on the genotyping data, a linkage map was constructed for chromosome 7A^m (Fig. 6a). The chromosome 7A^m map included 4 SSR markers (cfa2028, wmc405, barc174, and gwm573) and the 12 markers developed in the present study. To confirm the chromosomal positions of the 12 novel markers, they were anchored to chromosome 7A of CS (Supplementary Table S1). Chromosomal synteny was well conserved in the mapping regions shared between chromosomes 7A and 7A^m that were defined in the present work.

To easily confirm the usefulness of the constructed map, we conducted QTL analysis of heading and flowering times in the F$_2$ mapping population. The F$_2$ individuals in the KU-3620/KU-8276 population were grown under field conditions. The heading and flowering times of the F$_2$ individuals varied, and transgressive segregation was observed (Fig. 6b, c). Through single-marker analysis using SSR markers from each chromosomal arm of the A genome,
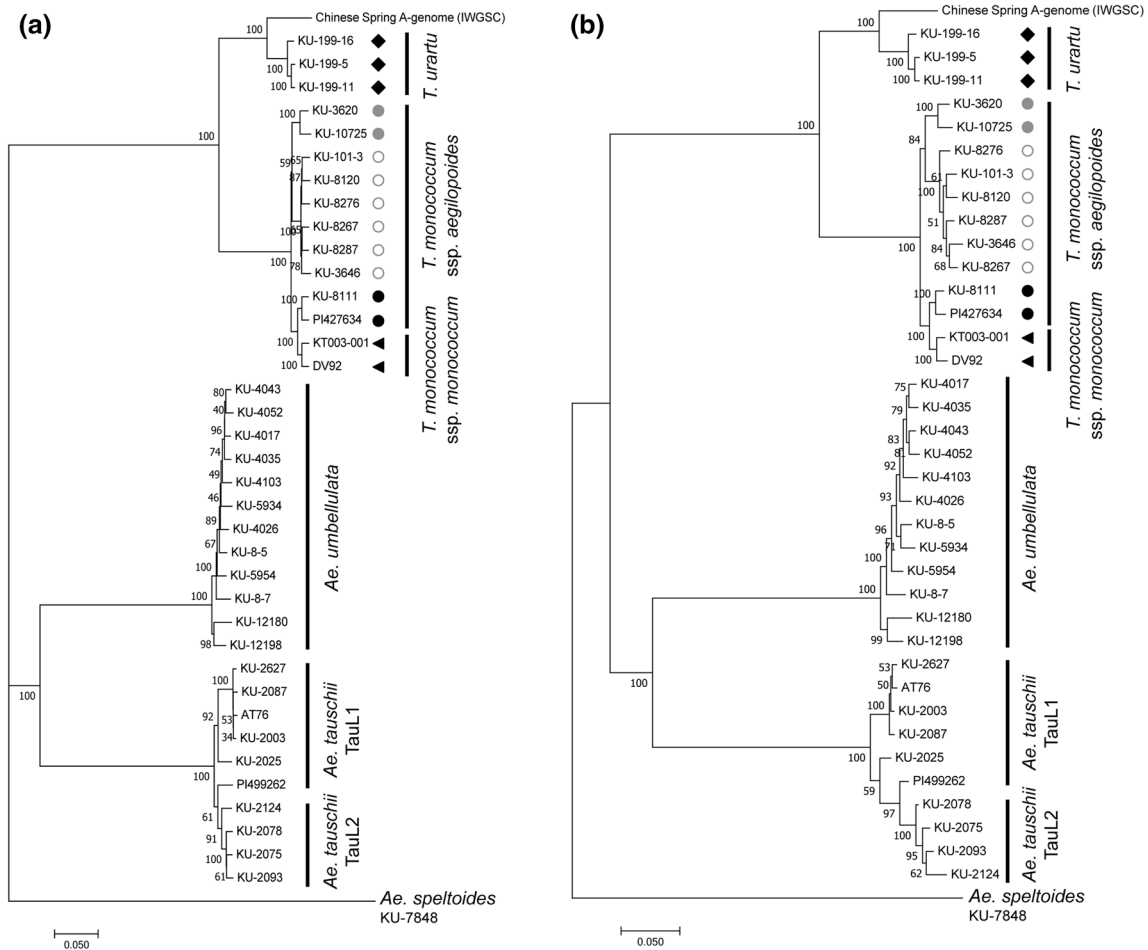
Fig. 2 Phylogenetic relationships among the diploid *Triticum*/*Aegilops* accessions. The neighbour-joining tree (**a**) and maximum likelihood tree (**b**) were constructed based on RNA-seq-derived SNPs. Bootstrap probabilities are shown on the branches (number of bootstrap replicates = 1000). A scale bar for genetic distance is shown on the left side of each phylogenetic tree

SSR markers on the short arm of chromosome 7A^m were found to be associated with the phenotypic variation in heading and flowering times in the mapping population. Using the 7A^m linkage map constructed in the present study, QTL analysis of these traits was performed by composite interval mapping. Significant QTLs for heading and flowering times were found with LOD scores of 5.17 and 2.69, respectively, within the same region on chromosome 7A^m (Fig. 6a). The 7A^m QTLs mapped between *Xcfa2028* and *Xwmc405*, and the peaks of QTL likelihood curves of the LOD scores were located in the *Xcfa2028–Xa24482* region. The 7A^m QTLs explained 23.0% and 11.1% of the combined heading and flowering time variation in the KU-3620/KU-8276 population, respectively. The genotypic effects of the 7A^m QTLs on heading and flowering times were examined under field conditions using F$_2$ individuals selected from the mapping population. The F$_2$ individuals carrying homozygous KU-3620 alleles in the 7A^m QTL region (*Xcfa2028–Xa24482*) showed significantly earlier heading and flowering times than those with homozygous KU-8276 alleles (Fig. 6d,e). The heading and flowering times of the heterozygous F$_2$ individuals were intermediate to those of the homozygous individuals.

Discussion

RNA sequencing is among the technologies used to efficiently detect SNPs and indels in the large genome of wheat species, a genome that is composed primarily (~85%) of repeat sequences such as transposons (International Wheat Genome Sequencing Consortium (IWGSC) 2018), facilitating the design of genetic makers covering the whole genome. In the present study, the RNA-seq approach facilitated an efficient search for genome-wide polymorphisms in the diploid wheat species with an A or A^m genome; a number of SNPs and indels were detected by comparison with the A genome of CS (Table 3), and the positions of these polymorphisms were confirmed to be distributed throughout the

Table 4 Summary of fixed nucleotide differences between species

Species being compared (1 vs. 2)		Number of fixed differences
1	2	
T. monococcum ssp. *aegilopoides*	CS	11,903
T. urartu	CS	5693
T. monococcum ssp. *monococcum*	CS	14,021
T. monococcum ssp. *aegilopoides*, *T. monococcum* ssp. *monococcum*, *T. urartu*	CS	3371
T. monococcum ssp. *aegilopoides*	CS, *T. urartu*, *T. monococcum* ssp. *monococcum*	25
T. urartu	CS, *T. monococcum* ssp. *aegilopoides*, *T. monococcum* ssp. *monococcum*	1892
T. monococcum ssp. *monococcum*	CS, *T. urartu* *T. monococcum* ssp. *aegilopoides*	187
T. monococcum ssp. *aegilopoides*, *T. monococcum* ssp. *monococcum*	CS, *T. urartu*	8309
T. monococcum ssp. *aegilopoides*	*T. monococcum* ssp. *monococcum*	228
T. urartu	CS	5694

The A genome of CS was used as a reference sequence (International Wheat Genome Sequencing Consortium (IWGSC) 2018)

genome (Fig. 1). Similar results have been reported in the wheat D-genome-containing progenitor *Ae. tauschii* (Iehisa et al. 2012, 2014; Nishijima et al. 2016), the U-genome-containing diploid species *Ae. umbellulata* (Okada et al. 2018), and section Sitopsis of the genus *Aegilops* (Miki et al. 2019). The positions of the genome-wide polymorphisms detected by RNA-seq were efficiently determined based on chromosomal synteny conserved among the genomes of wheat relatives (Mayer et al. 2011). Of course, structural rearrangements accumulated on the chromosomes of the wild diploid species during genome differentiation (Wicker et al. 2003; Molnár et al. 2016; Danilova et al. 2017), and inter-chromosomal translocations occurred after allotetraploid wheat speciation (Devos et al. 1995; Dvorak et al. 2018). Such rearrangements sometimes disturb chromosomal synteny; thus, the predicted chromosomal positions of the RNA-seq-derived polymorphisms are not necessarily precise. To solve this problem, the construction of genetic maps using molecular markers based on RNA-seq-derived polymorphism information in the target species is still important. Nonetheless, this putative positional information is expected to be useful for further genetic studies because local synteny is well conserved in each restricted chromosomal region among closely related species (Lu and Faris 2006).

The RNA-seq approach does have a limitation: it detects only the SNPs present in the exons of expressed genes. Polymorphism information can be collected from the exons of unexpressed genes but not from the introns and promoter regions of the expressed genes and intergenic chromosomal regions. Nonetheless, RNA-seq identified many genome-specific SNPs distributed across all the chromosomes in the Aᵐ genome, permitting the successful development of genome-specific CAPS markers that distinguished the Aᵐ and A genomes. These markers were useful for confirming the alien addition of the Aᵐ-genome chromosomes to tetraploid wheat (AABB genome) in synthetic hexaploid lines with AABBAᵐAᵐ genomes (Fig. 4). Using a similar strategy, U-genome-specific markers were successfully developed to distinguish the U-genome copy from the A- and B-genome-derived copies in our recent study (Okada et al. 2018). U-genome-specific markers are available for validation of interspecific hybridization in crosses between tetraploid wheat and *Ae. umbellulata* (Okada et al. 2018). Phylogenetic trees constructed in the present study showed that the A and Aᵐ genomes are evolutionarily more closely related than the A and U genomes (Fig. 2). The RNA-seq approach allowed the efficient detection of genome-wide polymorphisms between closely related genomes, permitting the development of genome-specific markers. This result also indicated that RNA-seq analysis is appropriate for elucidation of the evolutionary relationships among homoeologous genomes based on genome-wide exon sequences.

Similarly, RNA-seq-derived polymorphisms are an efficient source of information for the development of genome-wide genetic markers; these markers are expected to facilitate linkage map construction, gene mapping, and identification of QTLs for agronomic traits in mapping populations derived from intra-subspecies crossed with wild einkorn wheat. The CAPS markers obtained via the RNA-seq-derived polymorphisms permitted the

Fig. 3 Chromosomal distribution of fixed nucleotide differences **a** between the A and A^m genomes, **b** between *T. monococcum* ssp. *aegilopoides* and *T. monococcum* ssp. *monococcum*, and **c** between *T. urartu* and the A genome of CS
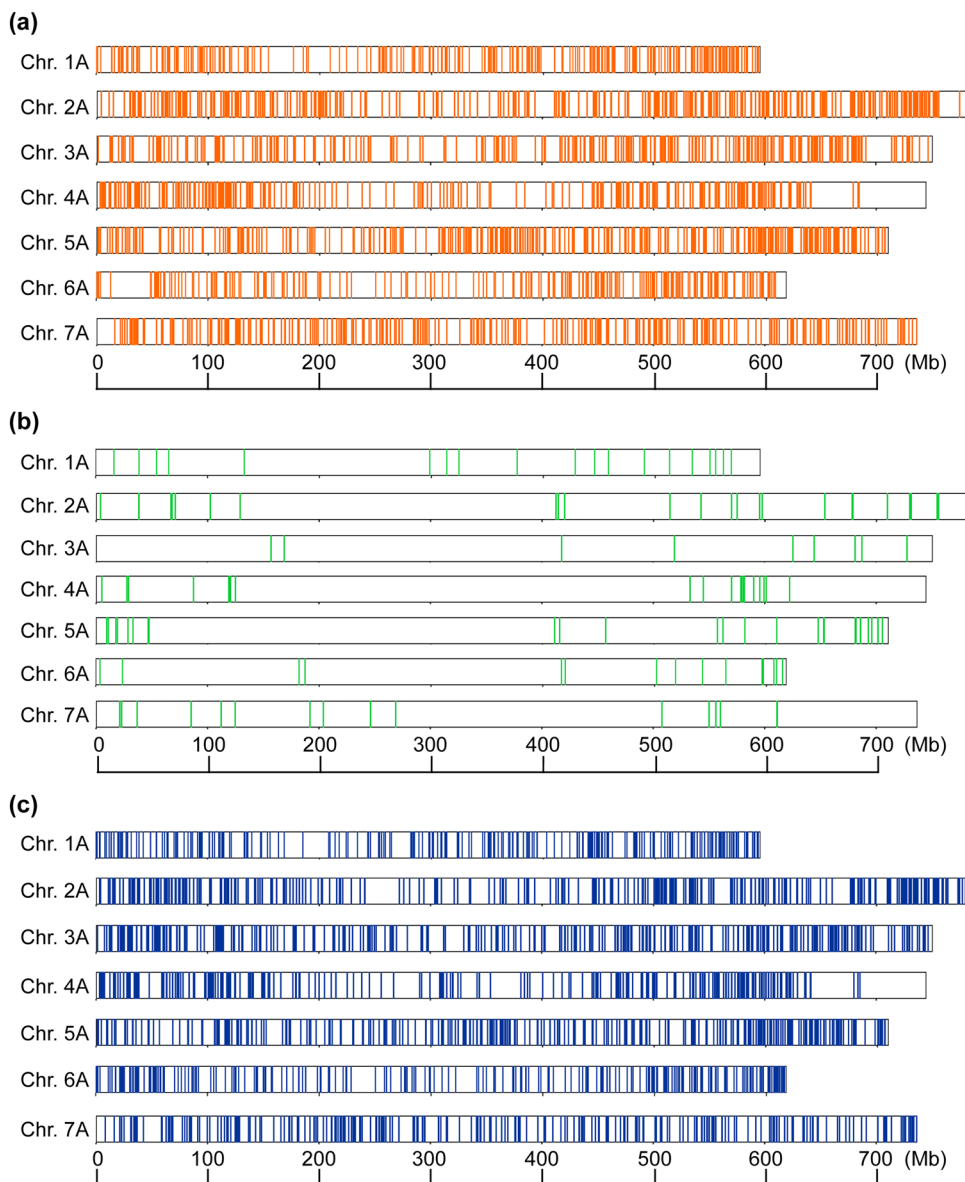
construction of a genetic map, and the constructed map facilitated the identification of QTLs for agronomic traits in wild einkorn wheat (Fig. 6). Heading and flowering time QTLs have been detected on the short arms of the homoeologous group 7 chromosomes in common wheat (Kuchel et al. 2006; Lin et al. 2008), einkorn wheat (Yu et al. 2017), and *Ae. tauschii* (Koyama et al. 2017). The heading time QTL reported previously in einkorn wheat is located near *Xbarc174* on chromosome 7A^m (Yu et al. 2017), whereas the 7A^m QTL found in the present study is distal relative to that reported by Yu et al. (2017). The 7A^m QTLs were detected here in an F$_2$ population of wild einkorn wheat; thus, the significance of the QTLs for heading and flowering time should be validated using their progeny in further studies. Therefore, although genome information, including the transcriptome, is lacking for

many species of wheat relatives, RNA-seq data facilitate genetic analyses of target traits in these wheat relatives.

Our study showed that the cultivated species *T. monococcum* ssp. *monococcum* belonged to one of the three groups of *T. monococcum* ssp. *aegilopoides* (Fig. 2), suggesting that *T. monococcum* ssp. *monococcum* originated recently from a limited subpopulation of *T. monococcum* ssp. *aegilopoides*. The recent divergence between *T. monococcum* ssp. *aegilopoides* and ssp. *monococcum* can make it difficult to identify genetic markers that distinguish between these two subspecies. However, using the RNA-seq approach, the present work identified 228 SNPs that differentiate these subspecies. Interestingly, these SNPs were patchily distributed over the distal chromosomal region (Fig. 3). This result suggests that in the process of domestication, nucleotide substitutions were not distributed across the entirety of
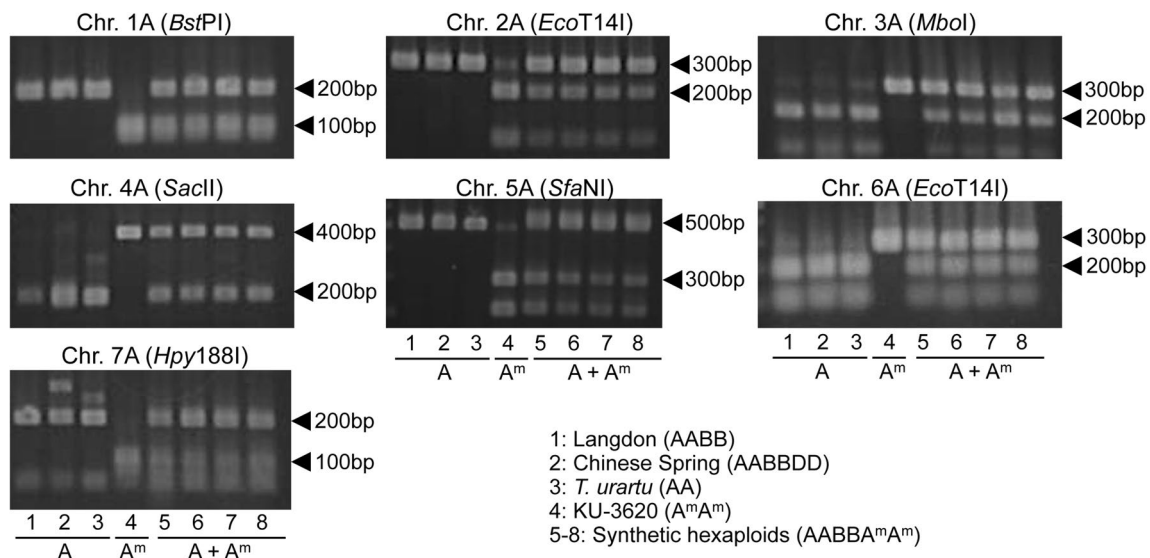
Fig. 4 A^m-genome-specific CAPS markers derived from the RNA-seq data. Each PCR product was digested with the indicated restriction enzyme. Lanes 1–4: Langdon, Chinese Spring, *T. urartu* KU-199-16, and *T. monococcum* ssp. *aegilopoides* KU-3620, respectively; lanes 5–8: four individuals of the synthetic hexaploid line Ldn/KU-3620
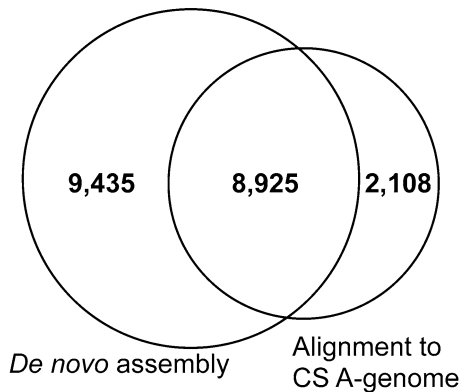

Fig. 5 Venn diagram of the SNPs detected between two accessions (KU-3620 and KU-8726) of *T. monococcum* ssp. *aegilopoides*. The SNPs were called based on the alignment of RNA-seq reads of KU-8726 to the non-redundant transcripts constructed with de novo assembly of the RNA-seq reads of KU-3620. The SNPs between these two accessions were detected based on the alignment of reads to the A-genome chromosomes of CS

the chromosomes, although only two accessions of cultivated einkorn wheat were examined. Specific chromosomal regions may have contributed to the domestication of *T. monococcum* ssp. *monococcum*. The association of certain chromosomal regions with species or subspecies differentiation was also observed between two *Aegilops* species, *Ae. longissima* and *Ae. sharonensis*, both of which belong to section Sitopsis and contain the S^l genome (Miki et al. 2019). In the present study, the diploid wheat accessions examined were limited. Recent studies using a large number of diploid wheat accessions showed that two species, namely *T. urartu* and *T. monococcum*, accumulated abundant genetic diversity (Heun et al. 1997; Jing et al. 2007). The genome-wide polymorphism data obtained in the present study could be utilized to further evaluate the genetic diversity in the two diploid wheat species and determine domestication-related chromosomal regions in einkorn wheat. These observations indicate that RNA-seq-derived genome-wide polymorphisms can contribute to genetic studies on the evolutionary differentiation of closely related species and subspecies.

Table 5 SNPs and indels between two wild einkorn wheat accessions, namely KU-3620 and KU-8276, that were detected based on assembly of the KU-3620 transcripts

Chromosome	1A	2A	3A	4A	5A	6A	7A	Unanchored	Total
SNP	2692	3307	2870	2189	3198	2676	2957	3547	23,436
Indel	78	87	68	57	84	85	88	99	646

Chromosome positions of the SNPs and indels were determined by alignment of the KU-3620 transcripts to the A genome of CS (International Wheat Genome Sequencing Consortium (IWGSC) 2018)
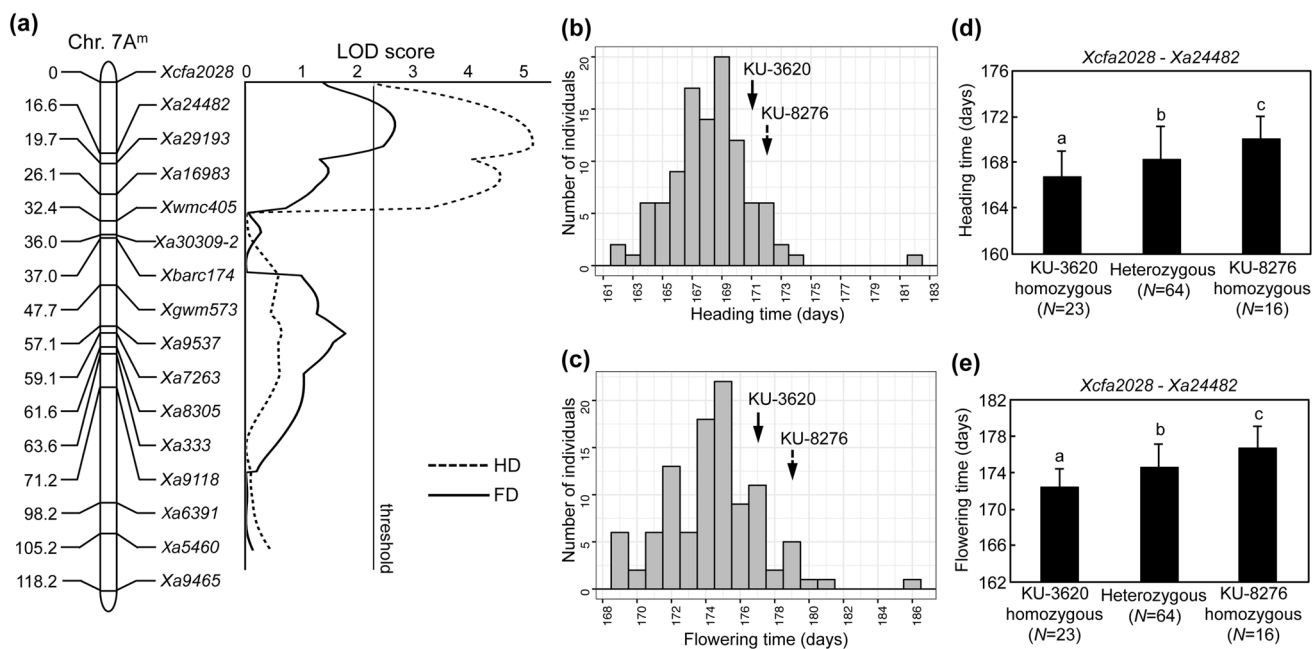
Fig. 6 Genetic map construction and QTL analysis of the F$_2$ population of wild einkorn wheat. **a** A linkage map of chromosome 7A^m and QTL likelihood curves of LOD scores showing the locations of QTLs for heading and flowering times. Genetic distances are shown in centimorgans to the left of chromosome 7A^m. **b, c** Frequency distribution of **b** heading time and **c** flowering time in the KU-3620/KU-8276 mapping population. **d, e** Comparison of **d** heading time and **e** flow-ering time among putative genotypes at the 7A^m QTL. The KU-3620/KU-8276 individuals were classified based on genotypes of markers between *Xcfa2028* and *Xa24482*. The statistical significance of the QTL effect on the examined traits was estimated by a two-tailed one-way analysis of variance with a post hoc Tukey–Kramer HSD test. Means with the same letters do not differ significantly ($p > 0.05$)

Our results indicate that the RNA-seq approach is quite useful for the discovery of genome-wide polymorphisms in wild wheat relatives. The resulting polymorphism data are available for subsequent genetic studies such as phylogenetic analysis, genetic map construction, and target gene mapping in wheat relatives for which sufficient genomic information is lacking. Wheat relatives are believed to share homoeologous chromosomes derived from a predicted common ancestor (Mayer et al. 2011). Therefore, the RNA-seq approach should contribute to genetic analyses and introgression of agriculturally important phenotypes from wild wheat relatives. The wheat relatives carrying genomes other than A, B, and D are useful for wheat breeding, serving as a tertiary gene pool. In wheat breeding processes, target genes can be transferred from chromosomes of wheat relatives to the A-, B- or D-genome chromosomes of common wheat by homoeologous recombination (Qi et al. 2007). In contrast, gene transfer from the tertiary gene pool using homologous recombination is not expected. Therefore, cytogenetic techniques to enhance homoeologous recombination, such as the use of *ph* (pairing homoeologous) gene mutants, will be helpful in accelerating gene transfer from the A^m-genome chromosome to the A-genome chromosome (Dubcovsky et al. 1995).

Genome-specific molecular markers are expected to enable confirmation of the introgression of target chromosomal segments from wheat relatives. Moreover, markers that are spread over all of the chromosomes could help determine the precise positions of the chromosomal regions transferred from wild relatives. Indeed, RNA-seq-derived PCR markers have been used to validate alien gene transfer from rye to common wheat (Wu et al. 2018). Thus, the genome-wide polymorphism information obtained from RNA-seq is expected to enlarge the genetic variation in the tertiary gene pool available for future wheat breeding.

Compliance with ethical standards

References

Anker CC, Niks RE (2001) Prehaustorial resistance to the wheat leaf rust fungus, *Puccinia triticina*, in *Triticum monococcum* (ss). Euphytica 117:209–215

Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K, Jordan KW, Golan G, Deek J, Ben-Zvi B, Ben-Zvi G, Himmelbach A, MacLachlan RP, Sharpe AG, Fritz A, Ben-David R, Budak H, Fashima T, Korol A, Faris JD, Hernandez A, Mikel MA, Levy AA, Steffenson B, Maccaferri M, Tuberosa R, Cattivelli L, Faccioli P, Ceriotti A, Kashkush K, Purkheirandish M, Komatsuda T, Eilam T, Sela H, Sharon A, Ohad N, Chamovitz DA, Mayer KFX, Stein N, Ronen G, Peleg Z, Pozniak CJ, Akhunov ED, Distelfeld A (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357:93–97

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Brandolini A, Vaccino P, Boggini G, Özkan H, Kilian B, Salamini F (2006) Quantification of genetic relationships among A genomes of wheats. Genome 49:297–305

Brunazzi A, Scaglione D, Talini RF, Miculan M, Magni F, Poland J, Pe ME, Brandolini A, Dell'Acqua M (2018) Molecular diversity and landscape genomics of the crop wild relative *Triticum urartu* across the Fertile Crescent. Plant J 94:670–684

Bullrich L, Appendino ML, Tranquilli G, Lewis S, Dubcovsky J (2002) Mapping of a thermo-sensitive earliness *per se* gene on *Triticum monococcum* chromosome 1A^m. Theor Appl Genet 105:585–593

Castagna R, Maga G, Perenzin M, Heun M, Salamini F (1994) RFLP-based genetic relationships of einkorn wheats. Theor Appl Genet 88:818–823

Danilova TV, Akhunova AR, Akhunov ED, Friebe B, Gill BS (2017) Major structural genomic alterations can be associated with hybrid speciation in *Aegilops markgrafii* (Triticeae). Plant J 92:317–330

Devos KM, Dubcovsky J, Dvorak J, Chinoy CN, Gale MD (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor Appl Genet 91:282–288

Dubcovsky J, Luo M, Dvorak J (1995) Differentiation between homoeologous chromosomes 1A of wheat and 1A^m of *Triticum monococcum* and its recognition by the wheat *Ph1* locus. Proc Natl Acad Sci USA 92:6645–6649

Dubcovsky J, Luo MC, Zhong GY, Brasterter R, Desai A, Killian A, Kleinhofs A, Dvorák J (1996) Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps *Hordeum vulgare* L. Genetics 143:983–999

Dvorak J, McGuire PE, Cassidy B (1988) Apparent sources of the A genome of wheat inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 30:680–689

Dvorak J, Wang L, Zhu T, Jorgensen CM, Luo MC, Deal KR, Gu YQ, Gill BS, Distelfeld A, Devos KM, Qi P, McGuire PE (2018) Reassessment of the evolution of wheat chromosomes 4A, 5A, and 7B. Theor Appl Genet 131:2451–2462

Fita A, Rodríguez-Burruezo A, Boscaiu M, Prohens J, Vicente O (2015) Breeding and domesticating crops adapted to drought and salinity: a new paradigm for increasing food production. Front Plant Sci 6:978

Fox SE, Geniza M, Hanumappa M, Naithani S, Sullivan C, Preece J, Tiwari VK, Elser J, Leonard JM, Sage A, Gresham C, Kerhornou A, Bolser D, McCarthy F, Kersey P, Lazo GR, Jaiswal P (2014) *De novo* transcriptome assembly and analyses of gene expression during photomorphogenesis in diploid wheat *Triticum monococcum*. PLoS One 9:e96855

Fricano A, Brandolini A, Rossini L, Sourdille P, Wunder J, Effgene S, Hidalgo A, Erba D, Piffanelli P, Salamini F (2014) Crossability of *Triticum urartu* and *Triticum monococcum* wheats, homoeologous recombination and description of a panel of interspecific introgression lines. G3 4:1931–1941

Gardiner LJ, Gawronski P, Olohan L, Schnurbusch T, Hall N, Hall A (2014) Using genic sequence capture in combination with a syntenic pseudo genome to map a deletion mutant in a wheat species. Plant J 80:895–904

Gill BS, Waines JG (1978) Paternal regulation of seed development in wheat hybrids. Theor Appl Genet 51:265–270

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 29:644–652

Guzmán C, Alvarez JB (2012) Molecular characterization of a novel waxy allele (*Wx-A^u1a*) from *Triticum urartu* Thum. Ex Gandil. Genet Res Crop Evol 59:971–979

Heun M, Schäfer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278:1312–1314

Hori K, Takehara S, Nankaku N, Sato K, Sasakuma T, Takeda K (2007) Barley EST markers enhance map saturation and QTL mapping in diploid wheat. Breed Sci 57:39–45

Hovhannisyan NA, Dulloo MH, Yesayan AH, Knüpffer H, Amri A (2011) Tracking of powdery mildew and leaf rust resistance genes in *Triticum boeoticum* and *T. urartu*, wild wheat relatives of common wheat. Czech J Genet Plant Breed 47:45–57

Iehisa JCM, Shimizu A, Sato K, Nasuda S, Takumi S (2012) Discovery of high-confidence single nucleotide polymorphisms from large-scale de novo analysis of leaf transcripts of *Aegilops tauschii*, a wild wheat progenitor. DNA Res 19:487–497

Iehisa JCM, Shimizu A, Sato K, Nishijima R, Sakaguchi K, Matsuda R, Nasuda S, Takumi S (2014) Genome-wide marker development for the wheat D genome based on single nucleotide polymorphisms identified from transcripts in the wild wheat progenitor *Aegilops tauschii*. Theor Appl Genet 127:261–271

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

James RA, Davenport RJ, Munns R (2006) Physiological characterisation of two genes for Na^+ exclusion in durum wheat: *Nax1* and *Nax2*. Plant Physiol 142:1537–1547

Jing HC, Kornyukhin D, Kanyuka K, Orford S, Zlatska A, Mitrofanova OP, Koebner R, Hammond-Kosack K (2007) Identification of variation in adaptively important traits and genome-wide analysis of trait-marker associations in *Triticum monococcum*. J Exp Bot 58:3749–3764

Jing HC, Bayo C, Kanyuka K, Berry S, Wenzel O, Huttner E, Kilian A, Hammond-Kosack K (2009) DArT markers: diversity analyses, genomes comparison, mapping and integration with SSR markers in *Triticum monococcum*. BMC Genom 10:458

Johnson BL, Dhaliwal HS (1976) Reproductive isolation of *Triticum boeoticum* and *Triticum urartu* and the origin of the tetraploid wheats. Am J Bot 63:1088–1094

Kim D, Langmead B, Salzbergk SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360

Kosugi S, Natsume S, Yoshida K, MacLean D, Cano L, Kamoun S, Terauchi R (2013) Coval: improving alignment quality and variant calling accuracy for next-generation sequencing data. PLoS One 8:e75402

Koyama K, Okumura Y, Okamoto E, Nishijima R, Takumi S (2017) Natural variation in photoperiodic flowering pathway and identification of photoperiod-insensitive accessions in wild wheat *Aegilops tauschii*. Euphytica 214:3

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genetics. Genome Res 19:1639–1645

Kuchel H, Hollamby G, Langridge P, Williams K, Jefferies SP (2006) Identification of genetic loci associated with ear-emergence in bread wheat. Theor Appl Genet 113:1103–1112

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) Subgroup 1000 genome project data processing. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Lin F, Xue SL, Tian DG, Li CJ, Cao Y, Zhang ZZ, Zhang CQ, Ma ZQ (2008) Mapping chromosomal regions affecting flowering time in a spring wheat RIL population. Euphytica 164:769–777

Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Lii Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, Zn X, Dong Z, Zhang K, Zhang X, Luo MC, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87–90

Ling HQ, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, Yu Y, Du H, Qi M, Li Y, Lu H, Yu H, Cui Y, Wang N, Chen C, Wu H, Zhao Y, Zhang J, Li Y, Zhou W, Zhang B, Hu W, van Eijk MJT, Tang J, Witsenboer HMA, Zhao S, Li Z, Zhang A, Wang D, Liang C (2018) Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. Nature 557:424–428

Lu H, Faris JD (2006) Macro- and microcolinearity between the genomic region of wheat chromosome 5B containing the *Tsn1* gene and the rice genome. Funct Integr Genom 6:90–103

Luo G, Zhang X, Zhang Y, Yang W, Liu Y, Sun J, Zhan K, Zhang A, Liu D (2015) Composition, variation, expression and evolution of low-molecular-weight glutenin subunit genes in *Triticum urartu*. BMC Plant Biol 15:68

Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, McGuire PE, Liu S, Long H, Ramasamy RK, Rodriguez JC, Van SL, Yuan L, Wang Z, Xia Z, Xiao L, Anderson OD, Ouyang S, Liang Y, Zimin AV, Pertea G, Qi P, Bennetzen JL, Dai X, Dawson MW, Müller HG, Kugler K, Rivarola-Duarte L, Spannagl M, Mayer KFX, Lu FH, Bevan MW, Leroy P, Li P, You FM, Sun Q, Liu Z, Lyons E, Wicker T, Salzberg SL, Devos KM, Dvorák J (2017) Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature 551:498–502

Mayer KFX, Martis M, Hedley PE, Simkova H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Miki Y, Yoshida K, Mizuno N, Nasuda S, Sato K, Takumi S (2019) Origin of wheat B-genome chromosomes inferred from RNA sequencing analysis of leaf transcripts from section Sitopsis species of *Aegilops*. DNA Res 26:171–182

Mizumoto K, Hirosawa S, Nakamura C, Takumi S (2002) Nuclear and chloroplast genome genetic diversity in the wild einkorn wheat, *Triticum urartu*, revealed by AFLP and SSLP analyses. Hereditas 137:208–214

Molnár I, Vrána J, Buresová V, Cápal P, Farkas A, Darkó É, Cseh A, Kubaláková M, Molnár-Láng M, Dolezel J (2016) Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their synteny with wheat. Plant J 88:452–467

Murai K, Nishiura A, Kazama Y, Abe T (2013) A large-scale mutant panel in wheat developed using heavy-ion beam mutagenesis and its application to genetic research. Nucl Instrum Methods Phys Res B 314:59–62

Nishijima R, Yoshida K, Motoi Y, Sato K, Takumi S (2016) Genome-wide identification of novel genetic markers from RNA sequencing assembly of diverse *Aegilops tauschii* accessions. Mol Genet Genom 291:1681–1694

Okada M, Yoshida K, Nishijima R, Michikawa A, Sato K, Takumi S (2018) RNA-seq analysis reveals considerable genetic diversity and provides genetic markers saturating all chromosomes in the diploid wild wheat relative *Aegilops umbellulata*. BMC Plant Biol 18:271

Pour-Aboughadareh A, Mahmoudi M, Moghaddam M, Ahmadi J, Mehrabi AA, Alavikia SS (2017) Agro-morphological and molecular variability in *Triticum boeoticum* accessions from Zagros Mountains Iran. Genet Res Crop Evol 64:545–556

Prohens J, Gramazio P, Plazas M, Dempewolf H, Kilian B, Díez MJ, Fita A, Herraiz FJ, Rodríguez-Burruezo A, Soler S, Knapp S, Vilanova S (2017) Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. Euphytica 213:158

Qi L, Friebe B, Zhang P, Gill BS (2007) Homoeologous recombination, chromosome engineering and crop improvement. Chromosome Res 15:3–19

Qiu YC, Zhou RH, Kong XY, Zhang SS, Jia JZ (2005) Microsatellite mapping of a *Triticum urartu* Tum. derived powdery mildew resistance gene transferred to common wheat (*Triticum aestivum* L.). Theor Appl Genet 111:1524–1531

Rasheed A, Mujeeb-Kazi A, Ogbonnaya FC, He Z, Rajaram S (2018) Wheat genetic resources in the post-genomics era: promise and challenges. Ann Bot 121:603–616

Rogers WJ, Miller TE, Payne PI, Seekings JA, Sayers EJ, Holt LM, Law CN (1997) Introduction to bread wheat (*Triticum aestivum* L.) and assessment for bread-making quality of alleles from *T. boeoticum* Boiss. ssp. *thaoudar* at *Glu-A1* encoding two high-molecular-weight subunits of glutenin. Euphytica 93:19–29

Sakaguchi K, Nishijima R, Iehisa JCM, Takumi S (2016) Fine mapping and genetic association analysis of *Net2*, the causative D-genome locus of low temperature-induced hybrid necrosis in interspecific crosses between tetraploid wheat and *Aegilops tauschii*. Genetica 144:523–533

Sato K, Tanaka T, Shigenobu S, Motoi Y, Wu J, Itoh T (2016) Improvement of barley genome annotations by deciphering the Haruna Nijo genome. DNA Res 23:21–28

Shi AN, Leath S, Murphy JP (1998) A major gene for powdery mildew resistance transferred to common wheat from wild einkorn wheat. Phytopathology 88:44–147

Shindo C, Sasakuma T, Watanabe N, Noda K (2002) Two-gene systems of vernalization requirement and narrow-sense earliness in einkorn wheat. Genome 45:563–569

Takamatsu K, Iehisa JCM, Nishijima R, Takumi S (2015) Comparison of gene expression profiles and responses to zinc chloride among inter- and intraspecific hybrids with growth abnormalities in wheat and its relatives. Plant Mol Biol 88:487–502

Takumi S, Nasuda S, Liu YG, Tsunewaki K (1993) Wheat phylogeny determined by RFLP analysis of nuclear DNA. 1. Einkorn wheat. Jpn J Genet 68:73–79

Tranquilli G, Cunberti M, Gianibelli MC, Bullrich L, Larroque OR, MacRitchie F, Dubcovsky J (2002) Effect of *Triticum monococcum* glutenin loci on cookie making quality and on predictive tests for bread making quality. J Cereal Sci 36:9–18

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acid Res 35:W71–W74

Vasu K, Singh H, Singh S, Chhuneja P, Dhaliwal HS (2001) Microsatellite marker linked to a leaf rust resistance gene from *Triticum monococcum* L. transferred to bread wheat. J Plant Biochem Biotechnol 10:127–132

Wang X, Luo G, Yang W, Li Y, Sun J, Zhan K, Liu D, Zhang A (2017) Genetic diversity, population structure and marker-trait associations for agronomic and grain traits in wild diploid wheat *Triticum urartu*. BMC Plant Biol 17:112

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15:1186–1197

Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875

Wu N, Li M, Sun H, Cao Z, Liu P, Ding T, Xu H, Chu C, Zhuang L, Qi Z (2018) RNA-seq facilitates development of chromosome-specific markers and transfer of rye chromatin to wheat. Mol Breed 38:6

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644

Yu K, Liu D, Wu W, Yang W, Sun J, Li X, Zhan K, Cui D, Ling H, Liu C, Zhang A (2017) Development of an integrated linkage map of einkorn wheat and its application for QTL mapping and genome sequence anchoring. Theor Appl Genet 130:53–70