Background: *Aegilops umbellulata* Zhuk. (2n = 14), a wild diploid wheat relative, has been the source of trait improvement in wheat breeding. Intraspecific genetic variation of *Ae. umbellulata*, however, has not been well studied and the genomic information in this species is limited.

Results: To develop novel genetic markers distributed over all chromosomes of *Ae. umbellulata* and to evaluate its genetic diversity, we performed RNA sequencing of 12 representative accessions and reconstructed transcripts by de novo assembly of reads for each accession. A large number of single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) were obtained and anchored to the pseudomolecules of *Ae. tauschii* and barley (*Hordeum vulgare* L.), which were regarded as virtual chromosomes of *Ae. umbellulata*. Interestingly, genetic diversity in *Ae. umbellulata* was higher than in *Ae. tauschii*, despite the narrow habitat of *Ae. umbellulata*. Comparative analyses of nucleotide polymorphisms between *Ae. umbellulata* and *Ae. tauschii* revealed no clear lineage differentiation and existence of alleles with rarer frequencies predominantly in *Ae. umbellulata*, with patterns clearly distinct from those in *Ae. tauschii*.

Conclusions: The anchored SNPs, covering all chromosomes, provide sufficient genetic markers between *Ae. umbellulata* accessions. The alleles with rarer frequencies might be the main source of the high genetic diversity in *Ae. umbellulata*.

Keywords: *Aegilops umbellulata*, *Aegilops tauschii*, Barley, DNA markers, RNA sequencing, Synteny

Background

*Aegilops umbellulata* Zhuk. (2n = 14), a wild diploid wheat relative, is distributed in West Asia and is known as the U-genome donor of *Ae. columnaris* and *Ae. triaristata* [1, 2]. *Ae. umbellulata* (UU genome) has cross-ability with tetraploid wheat (*T. turgidum* L.; AABB genome), which allows generation of synthetic hexaploids (AABBUU genome) through ABU triploids. Some combinations of interspecific crosses between *Ae. umbellulata* accessions and tetraploid wheat result in hybrid incompatibility, such as severe growth abortion and grass-clump dwarfness [3]. This observation suggests the existence of unrevealed genetic polymorphisms in *Ae. umbellulata* that potentially vary phenotypic traits.

*Ae. umbellulata* have been used for breeding of bread wheat and is a considerable resource of disease resistance genes [4–10]. Leaf rust and stripe rust resistance genes [6, 8, 11] and high-molecular weight glutenin subunits [5, 12] have been introduced into bread wheat cultivars. Chhuneja et al. (2008) [6] and Bansal et al. (2017)
[8] established introgression lines of leaf and stripe rust resistance genes derived from synthetic hexaploids (AAABBUU). The cross of the synthetic hexaploids (AAABBUU) with T. aestivum cv. Chinese Spring Ph1, which carries an epistatic inhibitor of Ph1 gene, induced homologous pairing and resulted in transfer of the leaf and stripe rust resistance genes of Ae. umbellulata into the bread wheat T. aestivum. Although Ae. umbellulata provides valuable genetic resources for breeding of bread wheat, it has not been well studied and information on its genome is limited. Evaluation of intraspecific genetic diversity based on genome-wide polymorphisms in Ae. umbellulata would provide valuable genetic resources for breeding of bread wheat, which has not been well studied and information on its genome is limited. Evaluation of intraspecific genetic diversity based on genome-wide polymorphisms in Ae. umbellulata would provide valuable genetic resources for breeding of bread wheat, which has not been well studied and information on its genome is limited.

Methods

Plant materials, library construction and RNA sequencing

Twelve accessions of Ae. umbellulata were chosen from the wheat genetic resources database of the National BioResource Project-Wheat (Japan, https://shigen.nig.ac.jp/wheat/komugi/top/top.jsp) to represent the diversity of this species (Fig. 1; Table 1). T. urartu KU-199-5 was used as the outgroup species for the comparative analysis between Ae. umbellulata and Ae. tauschii. Total RNA was extracted from leaves of Ae. umbellulata and T. urartu at the seedling stage using a Sepasol-RNA I Super G solution (Nacalai Tesque, Kyoto, Japan). The total RNA was treated with DNase I at 37 °C for 20 min to remove contaminating DNA. A total of 6 to 10 μg of RNA was used for constructing paired-end libraries. The libraries were constructed with TruSeq RNA Library Preparation Kit v2 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions, and were sequenced with 300-bp paired-end reads on an Illumina MiSeq sequencer.

De novo assembly of reads from RNA-seq

Low-quality bases (average quality score per 4 bp < 30), adapter sequences, and reads < 100 bp were removed using the Trimmomatic version 0.33 tool [25]. The paired-end reads were assembled with Trinity version 2.0.6 software to reconstruct transcripts for each accession [26, 27]. If a gene had multiple isoforms, the first transcript sequence designated by Trinity was chosen as a unigene. A set of unigenes was made for each accession according to our previous report [16], and was used as a reference transcript dataset. Paired-end reads from each accession were aligned to the reference transcripts using the Bowtie 2 [28]. SAMtools and Coval software were used for SNP and indel calling [29, 30]. SNPs and indels were called when over 95% of the aligned sequences were different from those of the reference transcripts at positions with read depth > 10. Sequence data have been deposited to DDBJ Sequence Read Archive DRA006404.

Mapping the assembled transcripts, SNPs and indels to barley and Ae. tauschii genome sequences

The transcripts were mapped to the barley (Hordeum vulgare L) reference genome “ASM32608v1 masked” [19] from the Ensembl Plants database [31] and to the Ae. tauschii genome “PRJNA341983” from the NCBI database [21] using Gmap software version 2014-12-31 [32] and bedtools [33]. Based on the transcripts mapping to the pseudomolecules of Ae. tauschii and barley, SNPs...
and indels were anchored to the chromosomes. The distribution of SNPs and indels on barley and *Ae. tauschii* chromosomes were visualized using CIRCOS software [34] (Krzywinski et al. 2009).

**Development of markers and genotyping**

Indel markers were designed using indels longer than 3 bp that were anchored to the barley chromosomes. Primer sets were constructed with Primer3plus software [35]. To validate marker alleles, we genotyped F₁ hybrid from a cross between *Ae. umbellulata* accessions KU-4017 and KU-4043. Total DNA was extracted from leaves of F₁ plants and their parents. PCR was conducted using Quick Taq HD DyeMix (TOYOBO, Osaka, Japan). PCR products were resolved in 17% acrylamide gels, and the products were visualized under UV light after staining by ethidium bromide.

**Comparison of genetic diversity between *Ae. umbellulata* and *Ae. tauschii***

The RNA-seq reads from the 10 *Ae. tauschii* accessions from the Transcriptome Shotgun Assembly division of DDBJ BioProject PRJDB4683 [16] were used for comparative analyses. We used the transcript sequences of *Ae. tauschii* KU-2075, which were constructed in our previous report [16], and *Ae. umbellulata* KU-4017 as reference transcripts. Quality control for the reads of *Ae. tauschii* and *T. urartu* was performed using Trimmomatic version 0.33 [25] in the same way as for *Ae. umbellulata*. The reads were aligned to the reference transcripts of *Ae. tauschii* KU-2075 and *Ae. umbellulata* KU-4017 using Bowtie 2 [28]. SNP calling was performed with SAMtools and Coval [29, 30] using the same criteria described above. SNPs that were assured of read depth > 10 and no ambiguous nucleotides in any accessions were selected as high-confidence SNPs and used for analyzing intra- and interspecific variation. The number of segregating sites, Tajima’s D statistic [36], and fixed nucleotide differences between species were estimated with DnaSP v5 software [37]. A neighbor-joining tree and a maximum likelihood tree were constructed based on the high-confidence SNPs. Bootstrap probability was calculated for 1000 replications.

**Estimation of orthologous transcripts of *Ae. umbellulata* and *Ae. tauschii***

Orthologous pairs of the reference transcripts of *Ae. tauschii* KU-2075 and *Ae. umbellulata* KU-4017 were estimated according to reciprocal best hits of BLAST analysis. A BLASTN search was performed using transcripts of *Ae. tauschii* KU-2075 as the queries against transcripts of *Ae. umbellulata* KU-4017, and vice versa.
When the same best hit was detected and query coverage was over 80% in both BLAST analyses, the transcripts from *Ae. umbellulata* KU-4017 and *Ae. tauschii* KU-2075 were judged an orthologous pair.

**Gene expression analysis**

The mapped reads that were concordantly aligned to the reference transcripts were chosen from the alignment file with SAMtools [29]. Fragments per kilobase per million mapped reads (FPKM) values were calculated based on the concordantly mapped reads [38].

**Results**

**RNA sequencing of 12 *Ae. umbellulata* accessions**

To evaluate genetic diversity based on a large number of DNA polymorphisms in the U-genome species *Ae. umbellulata*, RNA-seq was performed on the 12 representative accessions, generating 3.5–6.1 million paired-end reads per one accession (Table 2). These reads were analyzed according to the workflow shown in Additional file 1: Figure S1. After filtering out reads with low quality, 2.2–3.9 million paired-end reads (56.2–74.1%) were obtained. Due to the absence of a reference genome for *Ae. umbellulata*, transcript sequences for each of the 12 accessions were constructed by de novo assembly of the filtered reads. For each accession, 20,996 to 59,253 transcripts with N50 values of 899 to 1365 bp were deduced. One isoform was chosen as a unigene if a transcript had multiple isoforms. Finally, 12 sets of unigenes composed of 20,675 to 55,831 representative isoforms were obtained (Table 2) and used as reference transcript datasets for pairwise alignments between the accessions.

**Genome-wide identification of SNPs and indels in *Ae. umbellulata***

To detect SNPs and indels among the accessions, the filtered reads of each accession were aligned to the reference transcripts of all other accessions, and SNPs and indels were called according to the thresholds with read depth > 10. SNPs and indels identified from comparisons of the same accessions were regarded as artifacts. After filtering to remove these putative artifacts, 2925–44,751 SNPs and 77–1389 indels were obtained among the accessions (Table 3). The maximum numbers of SNPs and indels were obtained between KU-4035 and KU-12180 (44,751 SNPs and 1389 indels), with the minimum between KU-4017 and KU-4026 (2925 SNPs and 77 indels).

**Table 2 Summary of RNA sequencing for 12 accessions of *Ae. umbellulata***

| Accession | Read pairs | Filtered read pairs (%) | (unigenes) | N50 (bp) | Median contig length (bp) | Total assembled bases (Mbp) |
|-----------|------------|-------------------------|------------|----------|---------------------------|----------------------------|
| KU-4017 | 3,738,403 | 2,515,683 (67.29%) | 39,359 (37640) | 1238 | 658 | 36 |
| KU-4026 | 4,935,634 | 3,429,252 (69.48%) | 20,096 (20675) | 900 | 578.5 | 15.6 |
| KU-4035 | 6,077,268 | 3,890,911 (64.02%) | 57,029 (52216) | 1350 | 654 | 52.9 |
| KU-4043 | 4,090,151 | 2,392,708 (58.50%) | 50,985 (48590) | 1095 | 518 | 38.9 |
| KU-4052 | 3,829,992 | 2,152,503 (56.20%) | 47,320 (44869) | 1179 | 577 | 39.4 |
| KU-4103 | 3,875,477 | 2,506,545 (64.68%) | 31,418 (30873) | 899 | 507 | 22.1 |
| KU-5934 | 5,114,283 | 3,507,234 (68.58%) | 57,466 (52751) | 1332 | 630 | 52.7 |
| KU-5954 | 3,686,807 | 2,694,916 (73.10%) | 45,164 (41780) | 1365 | 699 | 44.1 |
| KU-12180 | 3,455,952 | 2,560,779 (74.10%) | 46,323 (44000) | 1229 | 631 | 41.2 |
| KU-12198 | 3,623,492 | 2,666,981 (73.60%) | 50,069 (46178) | 1359 | 655 | 47.7 |
| KU-8-5 | 3,669,766 | 2,699,452 (73.56%) | 52,648 (48981) | 1259 | 654 | 48.2 |
| KU-8-7 | 3,798,153 | 2,524,378 (66.46%) | 59,253 (55831) | 1160 | 524 | 46.9 |

*Percentage of the number of filtered read pairs per the number of read pairs
Table 3 The number of SNPs and indels detected in each transcript-read pairing of 12 *Ae. umbellulata* accessions

| Transcript model | KU-4017 | KU-4026 | KU-4035 | KU-4048 | KU-4052 | KU-4103 | KU-5934 | KU-5954 | KU-12,180 | KU-12,198 | KU-8–5 | KU-8–7 | total NR SNPs and indels |
|------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--------|--------|-------------------------|
| KU-12,180        | 2925    | 16,611  | 11,171  | 12,091  | 5316    | 24,975  | 14,064  | 29,726  | 23,244  | 13,400  | 24,601 | 85,758 |
| KU-12,198        | 77      | 524     | 348     | 381     | 169     | 854     | 401     | 869     | 434     | 443     | 446    | 846    | 3284 |
| KU-8–5           | 10,754  | 16,335  | 11,666  | 12,869  | 5428    | 19,463  | 11,889  | 23,915  | 17,452  | 11,478  | 20,185 | 63,233 |
| KU-8–7           | 350     | 612     | 429     | 420     | 153     | 762     | 405     | 787     | 565     | 412     | 754    | 2645   |
| KU-5934          | 9069    | 3639    | 11,969  | 13,141  | 5372    | 25,450  | 14,991  | 30,925  | 23,575  | 13,356  | 25,797 | 89,369 |
| KU-5954          | 274     | 94      | 426     | 423     | 167     | 889     | 387     | 901     | 651     | 396     | 847    | 3303   |
| KU-12,180        | 9748    | 2959    | 17,571  | 11,772  | 5325    | 24,871  | 15,017  | 30,191  | 22,942  | 13,557  | 26,117 | 85,656 |
| KU-12,198        | 287     | 83      | 611     | 404     | 155     | 867     | 423     | 883     | 662     | 459     | 809    | 3398   |
| KU-5934          | 10,355  | 3393    | 20,370  | 11,585  | 5172    | 25,332  | 14,713  | 30,020  | 23,340  | 14,824  | 26,197 | 88,716 |
| KU-5954          | 347     | 92      | 665     | 408     | 153     | 848     | 406     | 951     | 716     | 493     | 872    | 3525   |
| KU-12,180        | 443     | 97      | 854     | 582     | 527     | 187     | 385     | 932     | 663     | 540     | 855    | 3489   |
| KU-12,198        | 14,265  | 3465    | 25,574  | 16,432  | 16,766  | 6131    | 13,619  | 31,109  | 23,359  | 15,968  | 25,591 | 89,344 |
| KU-5934          | 5934    | 443     | 97      | 854     | 582     | 527     | 187     | 385     | 932     | 663     | 540    | 855    | 3489 |
| KU-5954          | 15,032  | 3585    | 28,423  | 18,003  | 18,391  | 6424    | 25,505  | 30,219  | 22,205  | 17,868  | 25,394 | 91,059 |
| KU-12,180        | 433     | 89      | 909     | 572     | 560     | 174     | 859     | 942     | 660     | 563     | 872    | 3723   |
| KU-12,198        | 24,232  | 6057    | 44,751  | 27,569  | 29,176  | 9695    | 42,984  | 27,022  | 24,142  | 30,721  | 34,445 | 99,411 |
| KU-5934          | 698     | 163     | 1389    | 901     | 896     | 296     | 1384    | 768     | 719     | 935     | 1140   | 4246   |
| KU-5954          | 21,096  | 5282    | 40,683  | 24,658  | 25,784  | 9259    | 39,208  | 22,641  | 26,570  | 27,036  | 31,815 | 100,683 |
| KU-12,180        | 12,822  | 3444    | 23,775  | 15,520  | 16,626  | 5893    | 24,993  | 15,271  | 30,526  | 23,495  | 25,746 | 91,080 |
| KU-12,198        | 8–5     | 375     | 100     | 784     | 558     | 577     | 164     | 907     | 439     | 952     | 689    | 860    | 3635 |
| KU-12,180        | 17,577  | 5041    | 33,501  | 20,936  | 21,374  | 7235    | 32,035  | 18,925  | 30,559  | 24,589  | 21,646 | 94,187 |
| KU-12,198        | 8–7     | 530     | 128     | 1068    | 694     | 650     | 227     | 1057    | 557     | 924     | 716    | 670    | 3728 |

Upper and lower numbers at each comparison indicate SNPs and indels, respectively. NR non-redundant.
For efficient use of the identified SNPs and indels as genetic markers, their chromosomal locations must be known. Here, we used the *Ae. tauschii* and barley pseudomolecules as virtual chromosomes of *Ae. umbellulata*, and mapped the unigene sequences of the *Ae. umbellulata* reference transcript datasets to the *Ae. tauschii* and barley chromosomes. In the reference transcripts, 75.87–85.35% of the unigenes were mapped to *Ae. tauschii* and 52.08–67.69% to barley chromosomes (Additional file 1: Table S1). Based on the positional information of the mapped unigenes, SNPs and indels were anchored to the chromosomes of both species. In any pairwise comparison between *Ae. umbellulata* accessions, 81.83–89.50% of SNPs and 75.28–89.26% of indels were anchored to *Ae. tauschii* chromosomes, while 63.17–75.16% of SNPs and 59.04–77.78% of indels were anchored to barley chromosomes (Additional file 1: Tables S2, S3). The distribution of SNPs over each chromosome of *Ae. tauschii* and barley was visualized with CIRCOS [34] for the *Ae. umbellulata* accession pairs with the maximum or minimum number of SNPs. The SNPs covered all chromosomes (Fig. 2).

Non-redundant SNPs and indels were estimated for each of the 12 sets of reference transcripts. A total of 63,233–100,683 non-redundant SNPs and 2645–4246 non-redundant indels were detected in the tested *Ae. umbellulata* accessions (Table 3). On average, 73,075 non-redundant SNPs (85.07%) were anchored to *Ae. tauschii* chromosomes, and 58,247 (70.40%) non-redundant SNPs to barley chromosomes (Additional file 1: Tables S4, S5). The smallest number of anchored non-redundant SNPs was observed on chromosomes 4D in *Ae. tauschii* and 4H in barley (Fig. 3). Each chromosome of *Ae. tauschii* and barley had an average of 10,439 and 8321 non-redundant SNPs, respectively. The anchored non-redundant SNPs were distributed over all seven chromosomes of *Ae. tauschii* and barley (Fig. 2).

We estimated the percentages of non-redundant SNPs anchored to the *Ae. tauschii* chromosomes overlapped those on barley chromosomes (Fig. 4). Venn diagrams showed that 69.18% of non-redundant SNPs were anchored to both *Ae. tauschii* and barley chromosomes. The percentage of non-redundant SNPs uniquely anchoring to *Ae. tauschii* chromosomes was 24.96%. Only 5.86% of non-redundant SNPs were uniquely anchored to barley chromosomes. After integration of these anchored non-redundant SNPs, 77,625 non-redundant SNPs were placed on the chromosomes.

**Application of indel markers to confirmation of F1 formation**

To confirm usefulness of the identified polymorphisms as genetic markers, primer sets for 27 indels were designed. The indel markers were applied to genotype F1 hybrid from a cross between two *Ae. umbellulata* accessions, KU-4017 and KU-4043; nine markers enabled detection of the genetic differences between the accessions and confirmed their F1 formation (Additional file 1: Figure S2). The difference in amplicon size between the parents was observed in the five markers. Presence/absence of amplicons between the parents was detected in the two markers. In the other two markers, the parents were distinguished by an extra band.

**Comparison of genetic diversity in *Ae. umbellulata* and *Ae. tauschii***

*Ae. tauschii* is widely distributed over central Eurasia and has three divergent lineages, TauL1, TauL2 and TauL3 [39]. On the other hand, the habitat of *Ae. umbellulata* is limited to West Asia. To examine how differences in geographic distribution and evolutionary history of these species affected the extent of DNA polymorphisms and the distribution of allele frequency, genetic diversity in *Ae. umbellulata* and *Ae. tauschii* was evaluated with SNPs deduced using the same RNA-seq platform. To compare intraspecific diversity of the two *Aegilops* species, reads from RNA-seq of the 12 *Ae. umbellulata* accessions, the 10 *Ae. tauschii* accessions [16] and T. urartu KU-199-5 were aligned to the reference transcripts of *Ae. umbellulata* KU-4017. T. urartu KU-199-5 was used as an outgroup species. To elucidate the phylogenetic relationship of *Ae. umbellulata* and *Ae. tauschii* accessions, maximum likelihood and neighbor-joining trees were constructed based on the high-confidence SNPs (Fig. 5; Additional file 1: Figure S3a). The three species were clearly separated, with the *Aegilops* species more closely related than *T. urartu*, with fixed nucleotide differences between *Ae. umbellulata* and *Ae. tauschii* smaller than those between *Ae. tauschii* and *T. urartu* or between *Ae. umbellulata* and *T. urartu* (Additional file 1: Table S6).

The external branches of *Ae. umbellulata* were longer than those of *Ae. tauschii*. *Ae. umbellulata* KU-12180 was isolated from the other accessions, supporting observations from the phylogenetic trees constructed based on nucleotide polymorphisms in a small number of genes [3]. However, the clear divergent lineages observed in *Ae. tauschii* were not found in the *Ae. umbellulata* accessions (Fig. 5). When the reference transcripts of *Ae. tauschii* KU-2075 was used for the alignments and SNP calling, similar results were obtained (Additional file 1: Table S6, Figures S3b, S4).

The number of segregating sites in *Ae. umbellulata* was larger than in *Ae. tauschii* (Table 4), indicating that *Ae. umbellulata* has relatively high genetic diversity. To test how differences in habitat and evolutionary history between *Ae. umbellulata* and *Ae. tauschii* affected allele frequency distribution in these two species, the derived allele frequency distribution for each species was estimated using *T. urartu* as an outgroup species (Fig. 6). At
a polymorphic site, a nucleotide that is inconsistent with that of outgroup species is defined as a derived allele, because this allele is considered to be newly generated by a mutation in population of the tested species [40]. The derived allele frequency distributions of the two species showed distinct patterns. Alleles with intermediate frequency were predominantly detected in *Ae. tauschii*, while alleles with rarer frequency were more common in *Ae. umbellulata*. As expected from the difference in the allele frequency distributions, Tajima’s D statistic [36] for *Ae. tauschii* and *Ae. umbellulata* respectively gave positive and negative values (Table 4).

Nucleotide diversity (*θ*) [41] in *Ae. umbellulata* and *Ae. tauschii* was estimated for each transcript. The *θ* value for each transcript of *Ae. umbellulata* was weakly correlated with that of *Ae. tauschii* (Figs. 7a, b: Kendall’s rank correlation $\tau = -0.026$ and 0.043). To avoid the possibility of bias due to differences in the accuracy and efficiency of short read alignments between intra- and interspecies, we compared *θ* for the 6062 orthologous

---

**Fig. 2** Distribution of transcripts and SNPs detected in pairwise comparisons of *Ae. umbellulata* accessions on the physical map of *Ae. tauschii* (a) and barley (b); scale in Mb. The three circles of the same color show the number of transcripts and SNPs from outer to inner circles. The red circles indicate the richest SNP pairs (KU-4035 read-mapped to KU-12180 transcript). The blue circles indicate the least rich pairs (KU-4026 read-mapped to KU-4017 transcript) in each combination of lineages. The green circles indicate the non-redundant SNPs mapped against KU-4043 transcripts. Arrowheads represent centromeric positions of each chromosome.

---

**Fig. 3** Boxplots of non-redundant SNPs anchored to chromosomes of *Ae. tauschii* (gray) and barley (black). The number of non-redundant SNPs on the chromosomes for each reference transcript dataset is plotted. Ave.D and Ave.H respectively indicate average number of non-redundant SNPs per chromosome when genomes of *Ae. tauschii* and barley were used.
pairs between *Ae. umbellulata* and *Ae. tauschii*. These pairs were retrieved by reciprocal best hits of BLAST analysis between the reference transcript datasets of *Ae. umbellulata* KU-4017 and *Ae. tauschii* KU-2075. This approach enables evaluation of genetic diversity using only the θ value based on SNPs derived from the intraspecies alignments of reads. Although gene expression of the orthologous pairs showed a relatively strong correlation (Fig. 7c: τ = 0.577), the values of θ between the pairs designated a weak correlation (Fig. 7d: τ = 0.049). Taken together, the reproducible observations from different approaches underpin the distinct extent
of nucleotide polymorphisms between *Ae. umbellulata* and *Ae. tauschii* at the gene level.

**Discussion**

**RNA-seq is a powerful approach to identify novel genetic markers for Triticeae species**

To identify genome wide polymorphisms (SNPs and indels) and to develop novel genetic markers, we conducted 300-bp paired-end RNA sequencing of leaf tissues from 12 representative *Ae. umbellulata* accessions using the Illumina MiSeq platform. By using *Ae. tauschii* and barley pseudomolecules as the virtual chromosomes of *Ae. umbellulata* due to the conserved synteny between Triticeae species [23, 24], an average of 73,075 and 58,247 non-redundant SNPs in *Ae. umbellulata* were successfully anchored to the chromosomes of *Ae. tauschii* and barley, respectively (Fig. 2). When a genetic map is constructed using these anchored SNPs and indels, changes in the marker order should be considered carefully due to the existence of chromosomal rearrangements in *Ae. umbellulata*. Structural rearrangements have been observed for *Ae. umbellulata* chromosomes when the order of genetic markers was compared among *Ae. umbellulata*, *Ae. tauschii* and common wheat [9, 44–46]. For example, chromosome 4 U has segmental homoeology to the group 6 chromosomes of common wheat [46]. Similarly, partial segments of chromosome 6 U have homoeology to hexaploid wheat group 4 and 5 chromosomes [9]. These observations support the occurrence of structural rearrangements such as translocation in *Ae. umbellulata*.

The power of indel detection with RNA-seq is not as high as that of SNPs, because indels in exons often have functionally deleterious effects on proteins and are purged from the genome by purifying selection. Notwithstanding this disadvantage, RNA-seq still provides useful indel markers for genetic mapping [47]. The indel markers were effective for validating detection of F1 alleles between *Ae. umbellulata* KU-4017 and KU-4043 (Additional file 1: Figure S2). These markers would allow rough map construction.

**Contrasting patterns of nucleotide diversity between *Ae. umbellulata* and *Ae. tauschii***

Differences in the habitats, morphology, population structure and phenological traits between *Ae. tauschii* and *Ae. umbellulata* may result in differences in the pressures of natural selection and the effect of genetic drift on genes, shaping the extent of DNA polymorphisms and allele frequency distribution between the species. In spite of the limited habitats of *Ae.
umbellulata, the present study showed that *Ae. umbellulata* has higher genetic diversity than the more widely distributed species *Ae. tauschii* (Fig. 5; Table 4). This observation is consistent with a previous report [48], in which intra- and interspecific genetic variation in seven diploid *Aegilops* species was evaluated using amplified fragment length polymorphisms, also concluding that genetic diversity in *Ae. umbellulata* is higher than in *Ae. tauschii*. Our comparative analyses showed no clear lineage differentiation in *Ae. umbellulata* (Fig. 5; Additional file 1: Figures S3, S4) and the prevalence of alleles with rarer frequencies (Fig. 6; Additional file 1: Figure S5), implying that the alleles with rarer frequencies are the main source of the genetic diversity observed in *Ae. umbellulata*.

The longer external branches of the phylogenetic tree in *Ae. umbellulata* suggest higher genetic differentiation of each *Ae. umbellulata* accession than *Ae. tauschii* (Fig. 5; Additional file 1: Figures S3, S4). Generally self-pollination inhibits gene flow via pollen, increasing genetic differentiation among local populations [49]. Since *Ae. umbellulata*...
is a self-fertilizing plant, this general view could be applicable to the observed genetic differentiation between the accessions of *Ae. umbellulata*. Considering *Ae. tauschii* is also a self-fertilizing species, another factor may contribute to shaping the distinct patterns of nucleotide polymorphism in these two species. If the time of expansion and colonization into the modern habitats differed between species, neutral mutations are expected to have accumulated more within a local population of the species with the earlier expansion and colonization, generating genetic differentiation between local populations under the limited gene flow. If this hypothesis is accepted, the time of expansion and colonization into the modern habitat of *Ae. umbellulata* is presumed to be older than that of *Ae. tauschii*. These different evolutionary scenarios and habitats of *Ae. tauschii* and *Ae. umbellulata* are likely to have shaped distinct genetic diversity for each gene from their common ancestor. The scatter plots of nucleotide diversity in the transcripts of *Ae. umbellulata* and *Ae. tauschii* show weaker correlations between the orthologous pairs (Fig. 7), suggesting that genes of *Ae. umbellulata* were subjected to natural selection pressure and effects of genetic drift that were distinct from those of *Ae. tauschii*. Future larger-scale population genomic analyses in both species will disclose population dynamics with higher resolution and more powerfully detect footprints of natural selection in each gene.

**Conclusion**

The RNA-seq-based approach is efficient for development of a large number of molecular markers and for conducting population genetic analyses for a large number of genes in wheat wild relatives such as *Ae. umbellulata* lacking genomic information. In addition, *Ae. umbellulata*, harboring relatively high genetic diversity, has considerable potential as a genetic resource for breeding of common wheat.

**Additional file**

**Additional file 1** · Table S1. Summary of the number of unigenes anchored to barley and *Ae. tauschii* genome. **Table S2.** The number of SNPs and indels anchored to the chromosomes of *Ae. tauschii* out of the SNPs and indels detected in each transcript-read pairing of 12 *Ae. umbellulata* accessions. **Table S3.** The number of SNPs and indels anchored to the barley chromosomes out of the SNPs and indels detected in each transcript-read pairing of 12 *Ae. umbellulata* accessions. **Table S4.** The number of non-redundant SNPs anchored to each *Ae. tauschii* chromosome. **Table S5.** The number of non-redundant SNPs anchored to each barley chromosome. **Table S6.** Summary of nucleotide polymorphism and divergence in *Ae. umbellulata*, *Ae. tauschii* and *T. urartu*. **Figure S1.** The workflow of RNA-seq analysis. **Figure S2.** Images of polyacrylamide gel electrophoresis for indel markers. **Figure S3.** Phylogenetic relationship between 12 *Ae. umbellulata* accessions, 10 *Ae. tauschii* accessions and one *T. urartu* accession based on SNPs that was estimated using the *Ae. tauschii* KU-4017 reference transcript dataset (a) and the *Ae. tauschii* KU-2075 reference transcript dataset (b). **Figure S4.** Phylogenetic relationship between the 12 *Ae. umbellulata* accessions, 10 *Ae. tauschii* accessions and one *T. urartu* accession based on SNPs estimated using the *Ae. tauschii* KU-2075 reference transcript dataset. The tree was constructed by the maximum-likelihood method. **Figure S5.** Derived allele frequency distribution in *Ae. umbellulata* (*n = 12*) (a) and *Ae. tauschii* (*n = 10*) (b), respectively. *Ae. tauschii* KU-2075 transcripts were used as the reference. Derived alleles were estimated using the outgroup species *T. urartu*. (PDF 824 kb)

**Abbreviations**

FPKM: Fragments per kilobase per million mapped reads; indels: Insertions and deletions; RNA-seq: RNA-sequencing; SNPs: Single nucleotide polymorphisms

**Acknowledgments**

The *Ae. umbellulata* seeds used in this study were supplied by the National BioResource Project-Wheat, Japan (www.nbrp.jp). Computations for RNA sequence assembly and alignments of reads were performed on the NIG supercomputer at the RIOS National Institute of Genetics, Japan.

**Funding**

This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas No. 17H03842, by Grant-in-Aid for Scientific Research (B) No. 16H04862 to ST from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, and by MEXT as part of a Joint Research Program implemented at the Institute of Plant Science and Resources, Okayama University, Japan. KY was supported by JST, PRESTO (No. JPMJPR15Q8).

**Availability of data and materials**

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Authors’ contributions**

KY, KS and ST designed the whole project. MO, KY, KS and ST wrote the manuscript. MO, AM, and YM performed experiments. MO, RN, and KY conducted RNA-sequencing analyses. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Competing interests**

The authors declare that they have no conflicts of interest.

**Publisher’s Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

1Graduate School of Agricultural Science, Kobe University, Rokkodai 1-1, Nada-ku, Kobe 657-8501, Japan. Institute of Plant Science and Resources, Okayama University, Kurashiki, Japan.

Received: 29 January 2018 Accepted: 25 October 2018

Published online: 08 November 2018

**References**

1. Lillenfeld FA, H. Khizar: genome-analysis in Triticum and Aegilops. X. Concluding review. Cytologia. 1951;16:101–23.
2. Wang GZ, Miyashita NT, Tsunewaki K. Plasmon analyses of Triticum (wheat) and Aegilops: PCR-single strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs. Proc Natl Acad Sci U S A. 1997;94:14570–7.
3. Okada M, Yoshida K, Takami S. Hybrid incompatibilities in interspecific crosses between tetraploid wheat and its wild diploid relative Aegilops umbellulata. Plant Mol Biol. 2017;95:625–45.
40. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics. 2000;155:1405–13.
41. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975;7:256–76.
42. Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, KFX M, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Wang J. International wheat genome sequencing consortium, Yang H, Liu X, He Z, Mao L, Wang J. Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. Nature. 2013;496:91–5.
43. Luo M-C, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, Jorgensen CM, Zhang Y, McGuire PE, Pasternak S, Stein JC, Ware D, Kramer M, McCombie WR, Kianian SF, Martis MM, Mayer KFX, Sehgal SK, Li W, Gill BS, Bevan MW, Smkóvá H, Doležel J, Weining S, Lazo GR, Anderson OD, Dvorak J. A 4-gigabase physical map unlocks the structure and evolution of the complex genome of Aegilops tauschii, the wheat D-genome progenitor. Proc Natl Acad Sci U S A. 2013;110:7940–5.
44. Zhang H, Jia J, Gale MD, Devos KM. Relationships between the chromosomes of Aegilops umbellulata and wheat. Theor Appl Genet. 1998;96:69–75.
45. Devos KM, Gale MD. Genome relationships: the grass model in current research. Plant Cell. 2000;12:637–46.
46. Molnár I, Vrána J, Burešová V, Cápal P, Farkas A, Darkó É, Cseh A, Kubaláková M, Molnár-Láng M, Doležel J. Dissecting the U, M, S and C genome of wild relatives of bread wheat (Aegilops spp.) into chromosomes and exploring their synteny with wheat. Plant J. 2016;84:452–67.
47. Nishijima R, Okamoto Y, Hatano H, Takumi S. Quantitative trait locus analysis for spikelet shape-related traits in wild wheat progenitor Aegilops tauschii: implications for intraspecific diversification and subspecies differentiation. PLoS One. 2017;12:e0173210.
48. Sasanuma T, Chabane K, Endo TR, Valkoun J. Characterization of genetic variation in and phylogenetic relationships among diploid Aegilops species by AFLP: incongruity of chloroplast and nuclear data. Theor Appl Genet. 2004;108:612–8.
49. Wright SI, Kalisz S, Slotte T. Evolutionary consequences of self-fertilization in plants. Proc R Soc B. 2013;280:20130133.