



Full Paper

Origin of wheat B-genome chromosomes inferred from RNA sequencing analysis of leaf transcripts from section Sitopsis species of *Aegilops*

Yuka Miki¹, Kentaro Yoshida ^{1*}, Nobuyuki Mizuno², Shuhei Nasuda², Kazuhiro Sato³, and Shigeo Takumi ^{1*}

¹Graduate School of Agricultural Science, Kobe University, Kobe, Japan, ²Graduate School of Agriculture, Kyoto University, Kyoto, Japan, and ³Institute of Plant Science and Resources, Okayama University, Kurashiki, Japan

*To whom correspondence should be addressed. Tel. +81 78-803-5860. Fax. +81 78-803-5860.

Email: kentaro.yoshida@port.kobe-u.ac.jp; takumi@kobe-u.ac.jp

Edited by Dr. Sachiko Isobe

Received 26 July 2018; Editorial decision 9 December 2018; Accepted 22 December 2018

Abstract

Dramatic changes occasionally occur in intergenic regions leading to genomic alterations during speciation and will consequently obscure the ancestral species that have contributed to the formation of allopolyploid organisms. The S genome of five species of section Sitopsis of genus *Aegilops* is considered to be an origin of B-genome in cultivated tetraploid and hexaploid wheat species, although its actual donor is still unclear. Here, we attempted to elucidate phylogenetic relationship among Sitopsis species by performing RNA sequencing of the coding regions of each chromosome. Thus, genome-wide polymorphisms were extensively analyzed in 19 accessions of the Sitopsis species in reference to the tetraploid and hexaploid wheat B genome sequences and consequently were efficiently anchored to the B-genome chromosomes. The results of our genome-wide exon sequencing and resultant phylogenetic analysis indicate that *Ae. speltoides* is likely to be the direct donor of all chromosomes of the wheat B genome. Our results also indicate that the genome differentiation during wheat allopolyploidization from S to B proceeds at different speeds over the chromosomes rather than at constant rate and recombination could be a factor determining the speed. This observation is potentially generalized to genome differentiation during plant allopolyploid evolution.

Key words: chromosomal synteny, genome-wide polymorphisms, genome differentiation, RNA sequencing, wheat

1. Introduction

Common wheat (*Triticum aestivum* L., genome constitution AABBDD), a major food crop, is an allohexaploid species derived via allopolyploid speciation through interspecific crossing between

cultivated tetraploid wheat *Triticum turgidum* L. (AABB) and its diploid relative, *Aegilops tauschii* Coss. (DD).^{1–4} The cultivated tetraploid form was domesticated from the wild tetraploid wheat *T. turgidum* subspecies *dicoccoides* (AABB), which was thought to

be derived through interspecific hybridization between wild diploid progenitors of the A and B genomes. The A genome donor was the wild diploid wheat *T. urartu*,^{5,6} and the B genome could have been contributed by *Ae. speltoides* Tausch (SS).^{7–10} However, the origin of the B genome remains unclear, despite extensive research over the past few decades. The cytoplasmic genomes of allopolyploid wheat species were almost certainly transmitted from *Ae. speltoides*,^{11,12} indicating that at least *Ae. speltoides* contributed to establishment of the nuclear genome of allopolyploid wheat.

The indefinite origin of the wheat B genome is due to failure of homoeologous chromosome pairing between the B genome of allopolyploid wheat and the S genome of *Ae. speltoides* during meiosis in the respective interspecific hybrids.^{13,14} In addition, the section Sitopsis of *Aegilops* includes four wild diploid species, *Ae. bicornis* Jaub. et Spach. (S^bS^b), *Ae. searsii* Feldman et Kislev ex Hammer (S^SS), *Ae. sharonensis* Eig (S^Sl), and *Ae. longissima* Schweinf. & Muschl. (S^Sl), except *Ae. speltoides*. Of the five Sitopsis species that share the S genome,¹⁵ only *Ae. speltoides* of the subsection Truncata is cross-pollinating, whereas the other four subsection Emarginata species are self-pollinating. Two subspecies of *Ae. speltoides* (*ligustica* and *speltoides*; syn. *Ae. aucheri* Boiss.) have been defined to date,^{16,17} and they can be distinguished at least in part by a single locus, *Lig*, on chromosome 3S, which controls spike morphology.¹⁸ The two Emarginata species, *Ae. longissima* and *Ae. sharonensis*, are quite closely related and recognized as forming one complex.^{19–22} The F₁ hybrid plants among Sitopsis species show incomplete homoeologous pairing during meiosis,²³ suggesting that differentiation to the modified S genome occurred during diversification of the Sitopsis species. Some chromosomal rearrangements in the S genome, including translocations, have been reported in Sitopsis species.^{20,24} In addition, the B and S genomes can exhibit differences in the pattern of transposable element insertion.^{25,26} Structural differences in intergenic regions have therefore increased the phylogenetic distance between the B genome of polyploid wheat and the S genome of *Ae. speltoides*. Constitutive heterochromatic regions detected by chromosome staining are much more abundant in B-genome chromosomes than those of the A and D genomes,^{27,28} and the staining patterns of B-genome chromosomes appear to differ from those of S-genome chromosomes of *Ae. speltoides*.²⁷ The difference in heterochromatin bands results in reduced pairing between B and S homoeologous chromosomes.^{28,29} These structural modifications and distinct heterochromatin distribution have made it difficult to elucidate the origin of the B genome and assess the relationship between the Sitopsis genomes and B genome.

Molecular phylogenetic studies based on nuclear DNA polymorphisms (e.g. restriction fragment length polymorphisms [RFLPs] and amplified fragment length polymorphisms [AFLPs]) have revealed that the two subsections of Sitopsis are extensively differentiated^{7,21,30} and that the wheat B genome is much more closely related to the S genome of *Ae. speltoides* than to the other modified S genomes of subsection Emarginata species.^{7,21} Analyses of nucleotide sequence polymorphisms in single-copy genes also supported the hypothesis that *Ae. speltoides* is the donor of the B genome in allopolyploid wheat.^{31,32} In contrast, a few reports have suggested a polyphyletic origin of the wheat B genome via the introgression of several parental Sitopsis species.^{33–35} For example, a low copy number, non-coding sequence located in the region comprising 19% of the distal portion of the long arm of chromosome 3B exists only in *Ae. searsii* among all Sitopsis species.³³ Moreover, nucleotide sequence analyses have revealed increased divergence in the B genome of modern common wheat compared with *Ae. speltoides*, and this

divergence is thought to be a result of polyploidization events affecting B-genome evolution.³⁴ Thus, the phylogenetic relationship between the B and S genomes of section Sitopsis should be reconsidered based on the polymorphisms of each limited chromosomal region as well as those covering the entire chromosomal regions of the B and S genomes.

RNA sequencing is an effective approach for surveying large number of genome-wide polymorphisms derived only from the exon sequences in *Aegilops* species.^{36–39} In RNA sequencing of *Aegilops* species, polymorphisms identified without any reference genome information can be efficiently anchored to the homoeologous chromosomes of related species, such as common wheat and barley, based on conserved chromosomal synteny.⁴⁰ Here, we conducted RNA sequencing analyses of leaf transcripts from section Sitopsis species to avoid the intergenic and repetitive sequences of wheat chromosomes. The objectives of the present study were to (i) identify genome-wide polymorphisms in the Sitopsis genomes, (ii) elucidate the phylogenetic relationship among Sitopsis species, and (iii) determine the wheat B-genome origin based on genome-wide polymorphisms anchored putatively to each chromosome of the B genome.

2. Materials and methods

2.1. Plant materials

Three accessions of *Ae. speltoides* ssp. *ligustica* (SS genome), four accessions of *Ae. speltoides* ssp. *speltoides* (SS genome), two accessions of *Ae. bicornis* (S^bS^b genome), three accessions of *Ae. longissima* (S^Sl genome), three accessions of *Ae. sharonensis* (S^Sl genome), and four accessions of *Ae. searsii* (S^SS genome) were chosen as representatives of each species from the collection of the section Sitopsis at the National Bio Resource Project–Wheat, Japan (Table 1). These accessions of Sitopsis species were originally collected in the Middle East (Supplementary Fig. S1). A tetraploid wheat (*T. turgidum*) cultivar Langdon (AABB genome) was also used in this study. *Triticum urartu* KU-199-5 (AA genome), *Ae. umbellulata* KU-4017 (UU genome), and *Ae. tauschii* KU-2075 (DD genome) were used as outgroup species.

2.2. RNA sequencing

Total RNA was extracted using Sepasol-RNA I Super G (Nacalai Tesque, Kyoto, Japan) from leaves of 2- to 3-month-old plants grown in a glass house. The extracted RNA was treated with DNase I at 37 °C for 20 min, after which paired-end libraries for RNA sequencing 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 a previously reported protocol⁴¹ and then sequenced with 300-bp paired-end reads on an Illumina MiSeq sequencer. The obtained reads were deposited in the DDBJ Sequence Read Archive under accession number DRA007097. RNA sequencing data for the outgroup species (300-bp paired-end reads) were obtained from the DDBJ Sequence Read Archives: BioProject PRJDB4683 for *Ae. tauschii* KU-2075 and DRA006404 for *Ae. umbellulata* KU-4017.

2.3. Quality control, alignment of paired-end reads, de novo transcriptome assembly, and single-nucleotide polymorphism (SNP)/insertion-deletion (indel) calling
FASTQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (6 July 2017, date last accessed)) was used to evaluate

Table 1. List of the 19 accessions in the section Sitopsis used in RNA-seq analyses

Species	Accession number	Origins	Mating systems
<i>Aegilops speltoides</i> ssp. <i>speltoides</i>	KU-2208A	Turkey	Outcrossing
	KU-14601	Israel	Outcrossing
	KU-14605	Israel	Outcrossing
<i>Ae. speltoides</i> ssp. <i>ligustica</i>	KU-12963a	Syria	Outcrossing
	KU-2236	Turkey	Outcrossing
	KU-7716	Iraq	Outcrossing
<i>Ae. bicornis</i>	KU-7848	Iraq	Outcrossing
	KU-5784	Egypt	Self-pollinating
	KU-14613	Israel	Self-pollinating
<i>Ae. longissima</i>	KU-5752	Jordan	Self-pollinating
	KU-14624	Israel	Self-pollinating
	KU-14635	Israel	Self-pollinating
<i>Ae. searsii</i>	KU-5755	Syria	Self-pollinating
	KU-6142	Jordan	Self-pollinating
	KU-6143	Jordan	Self-pollinating
<i>Ae. sharonensis</i>	KU-14651	Israel	Self-pollinating
	KU-14661	Israel	Self-pollinating
	KU-14663	Egypt	Self-pollinating
	KU-14668	Israel	Self-pollinating

sequencing quality of the reads from each of the sequenced samples. Adapter sequences, low-quality bases with an average quality score per 4 bp of <30, and reads of less than 100 bp were removed using Trimmomatic software, version 0.33,⁴² and only filtered paired reads were retained for subsequent analyses. The filtered reads were aligned to the reference B genome sequences of *T. aestivum* cv. Chinese Spring⁴³ using HISAT2 software, version 2.1.0.⁴⁴ To select uniquely mapped reads for SNP and indel calling, reads with a mapping quality of <40 were filtered out using SAMtools.⁴⁵ SNPs and indels were called using Coval⁴⁶ under the same criterion reported by Nishijima et al.³⁸; the depth of read coverage was ≥ 10 , and >95% of the mapped reads designated different nucleotide sequences from the reference sequences. To obtain non-redundant SNPs, we selected the positions of SNPs at which the read depth was ≥ 10 and there were no ambiguous nucleotides in any of the samples. We prepared two sets of non-redundant SNPs for construction of phylogenetic trees for intra- and interspecific comparisons of nucleotide variations. One was estimated in all of the samples, including A, D, U genome species, and the other was in the section of Sitopsis species and the wheat B genome. The distribution of SNPs/indels was visualized on the physical map of the B genome using CIRCOS⁴⁷ and R statistical software. *De novo* transcriptome assembly of *Ae. speltoides* ssp. *ligustica* KU-7716 was performed using Bridger.⁴⁸ Unmapped reads of *Ae. speltoides* ssp. *ligustica* KU-7716 were aligned to its assembled transcripts using Bowtie.⁴⁹ If the breadth coverage of the unmapped reads over a transcript is 100%, its corresponding annotated transcript was searched in the published transcripts of *T. aestivum* cv. Chinese Spring (iwgsc_refseqv1.0)⁴³ using BLASTN.⁵⁰

2.4. Construction of phylogenetic trees

Neighbour-joining (NJ) and maximum-likelihood (ML) phylogenetic trees were constructed using Molecular Evolutionary Genetics Analysis (MEGA) software, version 5.05.⁵¹ A Kimura 2-parameter model was used as the substitution model for tree construction. To assess node reliability in the trees, bootstrap probability was calculated from 1,000 bootstrap replicates. To construct phylogenetic trees for each chromosomal segment, the chromosomal regions of

the B genome were divided into ranges of 60 Mbp each, generating a total of 86 segments that included 124–931 non-redundant SNPs and 79–442 informative polymorphic sites (Supplementary Table S1). NJ trees for each subset of non-redundant SNPs were constructed using MEGA software. Phylogenetic network trees were constructed using SplitsTree4.⁵²

2.5. Nucleotide divergence between Sitopsis species/subspecies

To estimate nucleotide divergence between Sitopsis species/subspecies, two distance parameters were calculated: fixed substitutions between species⁵³ and average number of nucleotide differences between species.⁵⁴ To clarify positional changes in nucleotide divergence between Sitopsis species and the B genome of *T. aestivum*, the chromosomal regions of the B genome were divided into ranges of 20 Mbp each. A total of 262 segments of subsets of non-redundant SNPs in Sitopsis species and the B genome of *T. aestivum* cv. Chinese Spring were obtained. The fixed nucleotide differences and average number of nucleotide differences between Sitopsis species and the B genome were estimated for each subset of non-redundant SNPs. Non-synonymous fixed nucleotide differences were estimated using the variant annotation and effect prediction tool SnpEff.⁵⁵ The number of genes per 20 Mbp was counted based on the Chinese Spring annotation.⁴³

3. Results and discussion

3.1. RNA sequencing detected numerous SNPs in Sitopsis species and the common wheat B genome

To identify genome-wide SNPs and indels in Sitopsis species and the wheat B genome, RNA sequencing of 19 representative accessions of the five Sitopsis species was performed, generating 2–3 million filtered paired reads for each species (Table 2). Of these short reads, 67–81% were uniquely aligned to the B genome sequences of Chinese Spring, and 32,836–130,687 SNPs and 323–1,890 indels were identified. The fewest SNPs and indels were found in *Ae.*

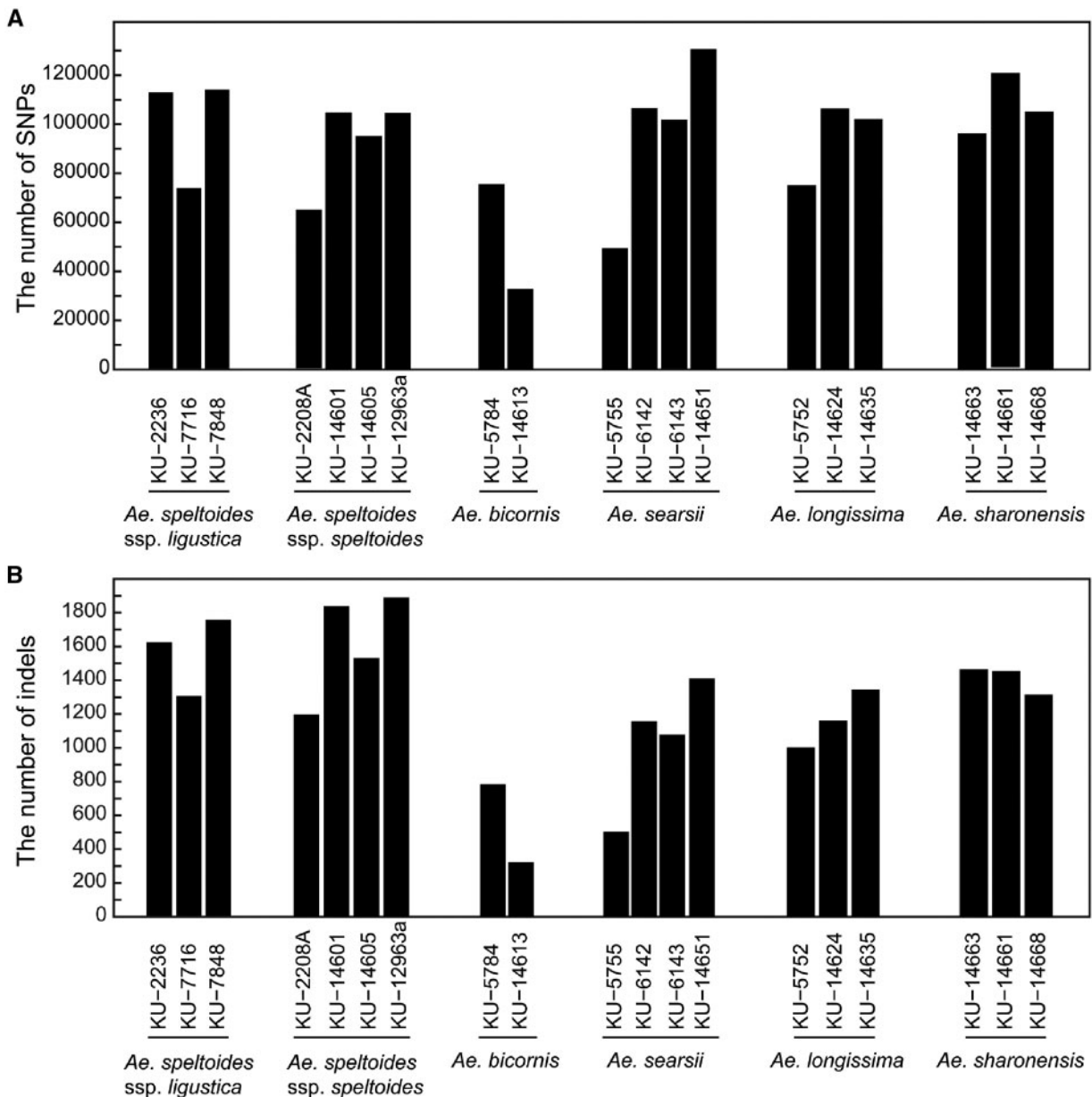


Figure 1. Bar charts illustrating the number of SNPs (A) and indels (B) between the B genome of *T. aestivum* cv. Chinese Spring and each of the 19 accessions of section Sitopsis species.

bicornis, whereas the other four species had similar numbers (Fig. 1). High within-species variance in the number of SNPs and indels was detected but considered a potential artifact because the number of filtered reads differed among the tested accessions (Table 2). To examine this possibility, the correlation between the number of filtered reads and SNPs/indels was determined. However, no correlation was observed between the number of SNPs and indels and the number of filtered reads, suggesting that the high variance is a genetic characteristic of section Sitopsis (Supplementary Fig. S2).

To confirm that RNA sequencing could identify genome-wide SNPs and indels, the chromosomal distribution of SNPs and indels was examined (Fig. 2). SNPs and indels identified in all of the tested accessions of Sitopsis species and the B genome entirely covered all of the chromosomes, with no clear difference in the distribution of SNPs and indels among Sitopsis species. Regions with scant or

abundant SNPs on the chromosomes were quite consistent between species. For each chromosome, the number of SNPs ranged from 4,194 to 21,175, and the number of indels ranged from 34 to 317, with the high variance reflecting differences in SNP and indels numbers within species but not between species (Supplementary Fig. S3).

3.2. Phylogenetic relationship between Sitopsis species and B genome of bread wheat

To clarify the phylogenetic relationship and nucleotide divergence between species, we estimated sets of non-redundant SNPs anchored to each chromosome of the B genome in the 19 tested accessions of Sitopsis species and the B genome of Chinese Spring, with/without three outgroup species: *T. urartu*, *Ae. umbellulata*, and *Ae. tauschii*. Without the outgroup species, 30,589 non-redundant SNPs were

Table 2. Summary of RNA sequencing data for the 19 accessions in the section Sitopsis

Species	Accession number	Read pairs	Filtered read pairs	Alignment rate (%) ^a
<i>Aegilops speltoides</i> ssp. <i>speltoides</i>	KU-2208A	4,317,201	2,721,855 (63.05%)	74.77
	KU-14601	5,279,430	3,088,916 (58.51%)	78.29
	KU-14605	4,005,919	2,480,301 (61.92%)	77.17
	KU-12963a	4,779,168	3,041,573 (63.64%)	70.94
<i>Ae. speltoides</i> ssp. <i>ligustica</i>	KU-2236	4,378,445	2,313,014 (52.83%)	74.40
	KU-7716	4,036,660	2,553,962 (63.27%)	67.92
	KU-7848	4,962,422	2,843,436 (57.30%)	77.83
<i>Ae. bicornis</i>	KU-5784	4,470,520	2,713,658 (60.70%)	80.71
	KU-14613	4,567,698	2,881,072 (63.07%)	81.04
<i>Ae. longissima</i>	KU-5752	4,371,287	2,490,621 (56.98%)	78.48
	KU-14624	5,012,064	2,865,373 (57.17%)	73.05
	KU-14635	4,623,166	2,660,230 (57.54%)	75.52
<i>Ae. searsii</i>	KU-5755	5,240,442	3,150,732 (60.12%)	81.23
	KU-6142	5,024,725	3,108,286 (61.86%)	77.75
	KU-6143	4,852,741	2,776,399 (57.21%)	78.16
	KU-14651	5,345,759	3,178,031 (59.45%)	77.87
<i>Ae. sharonensis</i>	KU-14661	6,500,270	3,241,926 (49.87%)	75.55
	KU-14663	8,544,460	4,157,728 (48.66%)	78.27
	KU-14668	10,079,695	4,891,456 (48.53%)	75.25

^aThe alignment rate against the B genome of *T. aestivum* cv. Chinese Spring was calculated with HISAT2 version 2.1.0.

obtained. When the outgroup species were included, 39,148 non-redundant SNPs were identified. These sets of non-redundant SNPs covered all of the chromosome of the B genome (Supplementary Fig. S4), allowing evolutionary analyses of section Sitopsis based on genome-wide polymorphisms.

NJ and ML phylogenetic trees and a phylogenetic network tree were constructed based on the set of non-redundant SNPs with the outgroup species (Fig. 3 and Supplementary Fig. S5). The Sitopsis species were clearly divided into two clades. One clade included *Ae. speltoides* ssp. *speltoides* and *Ae. speltoides* ssp. *ligustica*, and the other clade included *Ae. longissima*, *Ae. sharonensis*, *Ae. bicornis*, and *Ae. searsii*; the two clades corresponded to subsections Truncata and Emarginata, respectively. This result was consistent with the results of previous studies based on RFLPs and AFLPs.^{7,21,30} The Emarginata clade was more closely related to *Ae. tauschii* and *Ae. umbellulata* than the Truncata clade. The B genomes of *T. aestivum* and *T. turgidum* were closely related to *Ae. speltoides* ssp. *speltoides* and *Ae. speltoides* ssp. *ligustica* in the Truncata clade. The average number of nucleotide differences and fixed nucleotide differences between *Ae. speltoides* ssp. and the B genome were the lowest in pairwise comparisons between species of section Sitopsis and the B genome (Table 3). These results supported the previous hypothesis that the B genome originated from the S genome of *Ae. speltoides*. Considering that the wheat B genome was not nested within the Truncata clade, the most recent common ancestor of *Ae. speltoides* ssp. *speltoides* and *Ae. speltoides* ssp. *ligustica* is likely the direct donor of the wheat B genome.

The two subspecies of *Ae. speltoides* were not clearly divided in the Truncata clade (Fig. 3 and Supplementary Fig. S5). The Truncata clade had longer external branches than the Emarginata clade. This observation could be explained by differences in the mating systems of the two clades: species of the Truncata clade are outcrossing, whereas species of the Emarginata clade are self-pollinating. The mating system of *Ae. speltoides* is highly outcrossing.^{16,21} RFLP analyses indicated that *Ae. speltoides* contained a higher proportion of heterozygous loci compared with other self-pollinating species of

Sitopsis.⁵⁶ Natural populations of *Ae. speltoides* would harbour nucleotide variations as heterozygous states. The tested accessions of *Ae. speltoides* had been maintained by self-pollinating for several decades in the Japanese gene bank, which could have led to fixation of one of the alleles in heterozygous sites, increasing the number of singletons within the *Ae. speltoides* accessions. This resulted in detection of a relatively large number of SNPs in the *Ae. speltoides* accessions (Fig. 1). In the Emarginata clade, *Ae. searsii* was monophyletic and separated from the other three species, whereas *Ae. bicornis* nested within *Ae. sharonensis* and *Ae. longissima*. *Aegilops longissima* diverged from *Ae. sharonensis*. The close relationship between *Ae. sharonensis* and *Ae. longissima* was consistent with previous reports^{7,21} and not inconsistent with a recent proposal that *Ae. sharonensis* is a subspecies of *Ae. longissima*.⁵⁷

3.3. Identification of SNPs distinguishing subspecies *ligustica/speltoides* and *Ae. sharonensis/Ae. longissima*

The phylogenetic tree did not discriminate well between subspecies of *Ae. speltoides* ssp. *speltoides* and *Ae. speltoides* ssp. *ligustica*, whereas the subspecies of *Ae. longissima* and *Ae. sharonensis* were distinguished. These subspecies are morphologically classified. Notably, differences in spike morphology between *Ae. speltoides* ssp. *speltoides* and *Ae. speltoides* ssp. *ligustica* can be explained by a single locus, *Lig*, located on chromosome 3S.¹⁸ If there are SNPs that would distinguish these subspecies (fixed nucleotide differences between subspecies), they could be related to the morphologic differences between the subspecies. Between *Ae. speltoides* ssp. *speltoides* and *Ae. speltoides* ssp. *ligustica*, 99 fixed nucleotide differences were detected (Table 3). The positions of these fixed nucleotide differences were scattered over the chromosomes (Fig. 4A). Ten of the fixed nucleotide differences caused amino acid substitutions between the two subspecies (Supplementary Table S2).

Between *Ae. longissima* and *Ae. sharonensis*, 121 fixed nucleotide differences were found (Table 3). All of the chromosomes had fixed nucleotide differences (Fig. 4B), and they were most densely located

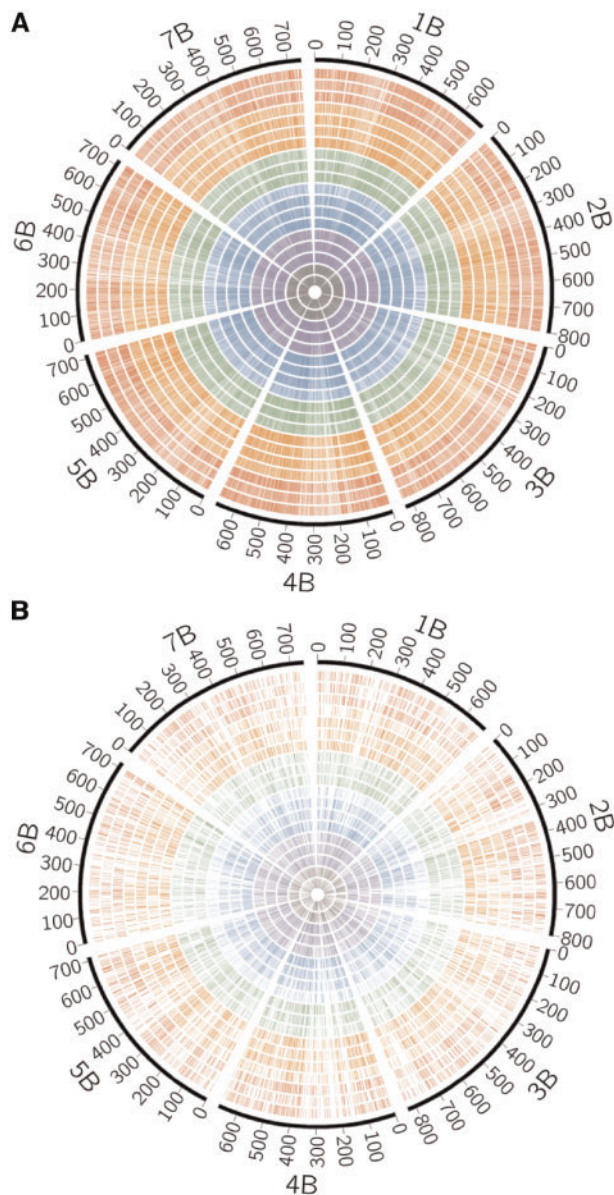


Figure 2. Distribution of SNPs (A) and indels (B) between the *T. aestivum* cv. Chinese Spring (CS) B genome and each of the 19 accessions of section Sitopsis species on the physical map of the CS B genome. From outer to inner circle, CS B genome (scale in Mb) and chromosome number (black); *Ae. speltoides* ssp. *ligustica* KU-2236, KU-7716, and KU-7848 (red); *Ae. speltoides* ssp. *speltoides* KU-2208A, KU-14601, KU-14605, and KU-12963a (orange); *Ae. longissima* KU-5752, KU-14624, and KU-14635 (green); *Ae. searsii* KU-5755, KU-6142, KU-6143, and KU-14651 (blue); *Ae. sharonensis* KU-14661, KU-14663, and KU-14668 (purple); and *Ae. bicornis* KU-5784 and KU-14613 (grey).

on the end of the long arm of chromosome 4B. Of all total fixed nucleotide differences, 19 caused amino acid substitutions between the two species. Genes with these fixed nucleotide differences encoded proteins involved in a variety of biological functions (Supplementary Table S3). Lateral awn elongation is a key character distinguishing *Ae. sharonensis* from *Ae. longissima*, as *Ae. longissima* lacks a lateral awn.²² The heading time and growth habitats of these two closely related species are also distinct.²² The distal region of chromosome 4S¹, in which many fixed nucleotide differences accumulated, might

control the morphologic and physiologic differences between *Ae. sharonensis* and *Ae. longissima*. Identification of the causal genes is a focus for future research.

3.4. Contrasting pattern of nucleotide divergence in the distal and proximal regions of the chromosomes

Some previous studies suggested the possibility of introgression from several parental Sitopsis species, supporting a polyphyletic origin of the wheat B genome.^{33–35} If introgression contributed to the origin of the wheat B genome, the phylogenetic relationship between species could possibly be verified based on chromosomal positions. To test this hypothesis, NJ trees were constructed for 60-Mbp regions on each chromosome (Fig. 5). Of a total of 86 phylogenetic trees, 83 exhibited a similar topology to that of trees based on entire chromosomes, in which the B genome of Chinese Spring was closely related to the Truncata clade including *Ae. speltoides* (Fig. 3). Two trees indicated that the B genome was closely related to the Emarginata clade (Fig. 5). These irregular trees were detected at the end of the long arm of chromosomes 1B and 3B. In the other tree, the B genome at the end of the short arm of chromosome 3B was located outside the Sitopsis species.

We estimated the average number of nucleotide differences between the B genome of Chinese Spring and each of the Truncata and Emarginata clades as a parameter of genetic divergence between the B genome and each clade (Fig. 6). In the distal regions of the chromosomes, genetic divergence between the B genome and Emarginata clade was slightly less than that between the B genome and Truncata clade, whereas in the proximal regions of the chromosomes, genetic divergence between the B genome and Emarginata clade was greater than that between the B genome and Truncata clade. The proximal chromosomal regions tended to exhibit conspicuous disparity in terms of genetic divergence ($(D_{E-B} - D_{T-B})/D_{T-B} \times 100 > 0$ in Fig. 6), but the range of this disparity differed among the chromosomes. Almost the entire region (~660 Mbp) of chromosome 3B exhibited clear disparity. In contrast, on chromosome 5B, the region exhibiting disparity was limited to within about 360 Mbp of the long arm. Except for chromosome 7B, the disparity appeared to be negatively correlated with the total number of non-redundant SNPs in Sitopsis species and the B genome. In chromosomal regions with larger number of non-redundant SNPs, genetic divergence between the B genome and Emarginata clade increased to the same extent as that between the B genome and Truncata clade. In addition, the disparity also reflected the number of genes; regions with fewer genes were found to exhibit greater disparity. Interestingly, this contrasting pattern of genetic divergence between the proximal and distal chromosomal regions corresponded to the gradient recombination rate along the chromosomes.⁵⁸

In common wheat, the recombination rate and chromosomal gene density increase as the centromeric region recedes.^{58,59} Multiple genes created by gene duplication are more frequently located in the distal regions of the chromosomes, where they potentially drive increases in gene density and the recombination rate.⁵⁵ This observation coincides with the observed positive correlation between the disparity of genetic divergence and number of genes (Fig. 6). Incomplete lineage sorting (ILS) is known to generate gene trees in which the topology is discordant with that of species trees and tends to more frequently occur in rapid successions of speciation events.^{60,61} If speciation events in section Sitopsis had occurred over a relatively short time, ILS in the ancestral population of Sitopsis represents a potential factor blurring the phylogenetic

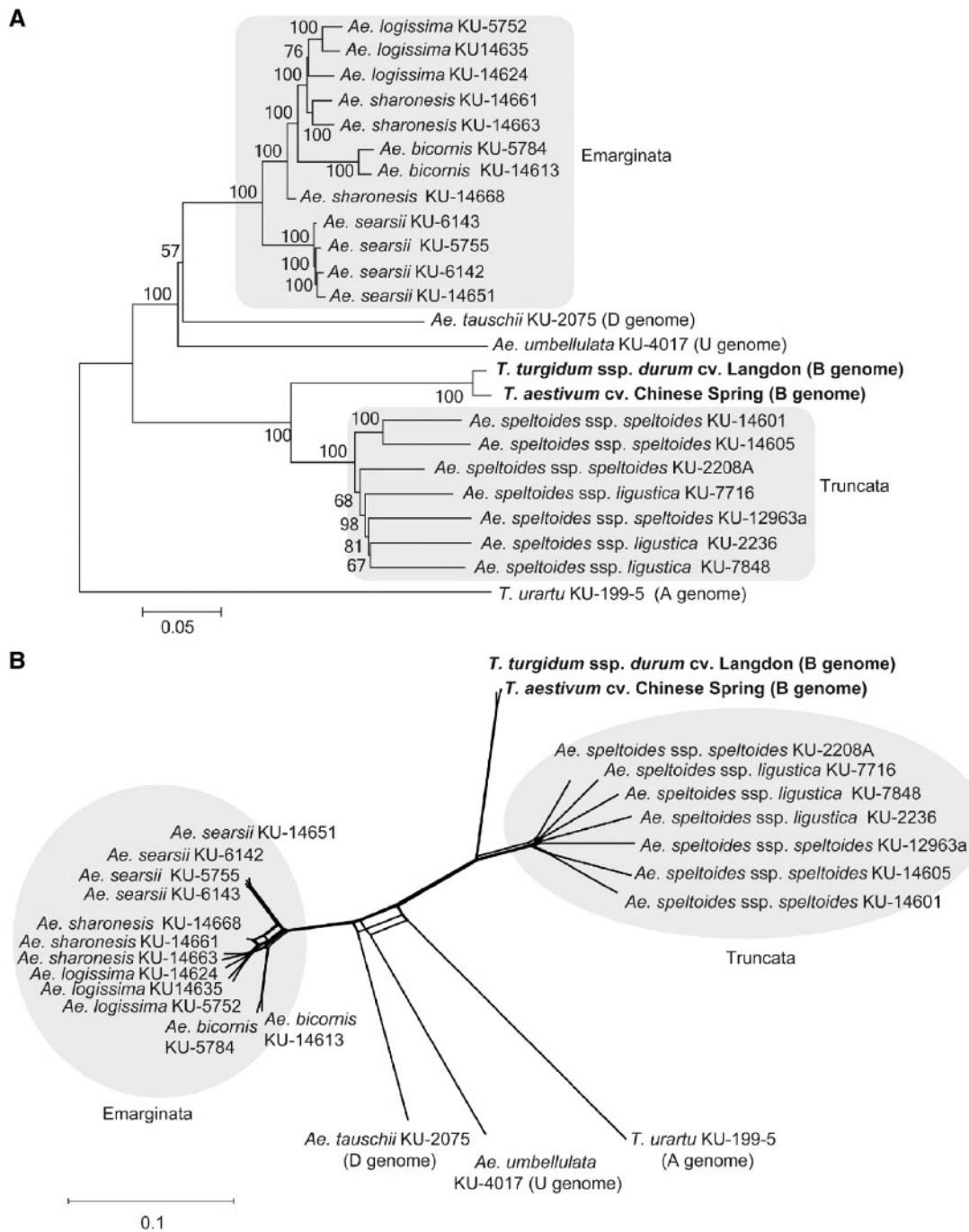


Figure 3. Phylogenetic relationship among the 19 accessions of section Sitopsis species (S genome), the B genomes of *T. aestivum* cv. Chinese Spring and *T. turgidum* ssp. *durum* cv. Langdon, *T. urartu* (A genome), *Ae. tauschii* (D genome), and *Ae. umbellulata* (U genome). NJ tree (A) and phylogenetic network (B) are shown. Bootstrap probabilities are shown on the branches (number of bootstrap replications = 1000). The scale bar is shown below each phylogenetic tree.

relationship between Sitopsis species and the B genome. ILS is positively correlated with recombination.^{61,62} Indeed, distal chromosomal regions with an irregular phylogenetic topology (Fig. 5) exhibited more complex reticulate structures in the network tree compared with whole-chromosomal regions (Fig. 3B and Supplementary Fig. S6). In the interspecies comparisons, the internal reticulate structures represent data conflicts caused by phenomena such as ILS.⁶³ Therefore, ILS could explain the irregular topology of phylogenetic trees in Sitopsis species (Fig. 5) and the

unclear disparity of genetic divergence (Fig. 6) that were prominent in the distal chromosomal regions with a higher recombination rate.

In the present study, a large number of SNPs and indels were discovered in the wheat B genome and five Sitopsis species based on RNA sequencing of leaf-derived transcripts. The polymorphic data would be useful for developing genome-wide markers on the S-genome chromosomes as performed for the wild diploid relatives, *Ae. tauschii* and *Ae. umbellulata*.^{38,39}

Table 3. Summary of pairwise comparisons of fixed divergence (upper) and average number of nucleotide differences (lower) between *Sitopsis* species

	<i>Ae. speltoides</i> ssp. <i>ligustica</i>	<i>Ae. speltoides</i> ssp. <i>speltoides</i>	<i>Ae. speltoides</i> ssp. <i>speltoides</i>	<i>Ae. bicornis</i>	<i>Ae. longissima</i>	<i>Ae. searsii</i>	<i>Ae. sharonensis</i>	B-genome Subspecies	
									<i>lig/spel</i> <i>lon/sha</i>
<i>Ae. speltoides</i> ssp. <i>ligustica</i>		99	9,032	8,153	8,431	7,854	6,270	0	7,693
<i>Ae. speltoides</i> ssp. <i>speltoides</i>	5,411.4		8,558	7,704	7,981	7,400	5,779	0	7,251
<i>Ae. bicornis</i>	12,785.5	12,666		1,884	3,653	1,771	13,093	7,935	1,557
<i>Ae. longissima</i>	12,412.1	12,313.5	3,239.7		2,530	121	12,175	7,110	0
<i>Ae. searsii</i>	12,052.8	11,971.9	4,318.8	3,782.8		1,976	12,498	7,356	1,841
<i>Ae. sharonensis</i>	12,070	11,965.5	3,048.5	1,720.9	3,244.9		11,811	6,821	0
B-genome	9,654	9,507.8	13,489.5	13,187	12,774.8	12,803.3		5,170	11,577
Sub-species	<i>lig/spel</i>	4,587.4	4,544.3	12,717.2	12,355.8	12,006.5	12,010.3	9,570.4	6,690
	<i>lon/sha</i>	12,241.1	12,139.5	3,144.1	1,358.7	3,513.8	1,331.2	12,995.2	12,183

lig/spel: comparisons between *Ae. speltoides* ssp. *ligustica* and *Ae. speltoides* ssp. *speltoides*.

lon/sha: comparisons between *Ae. longissima* and *Ae. sharonensis*.

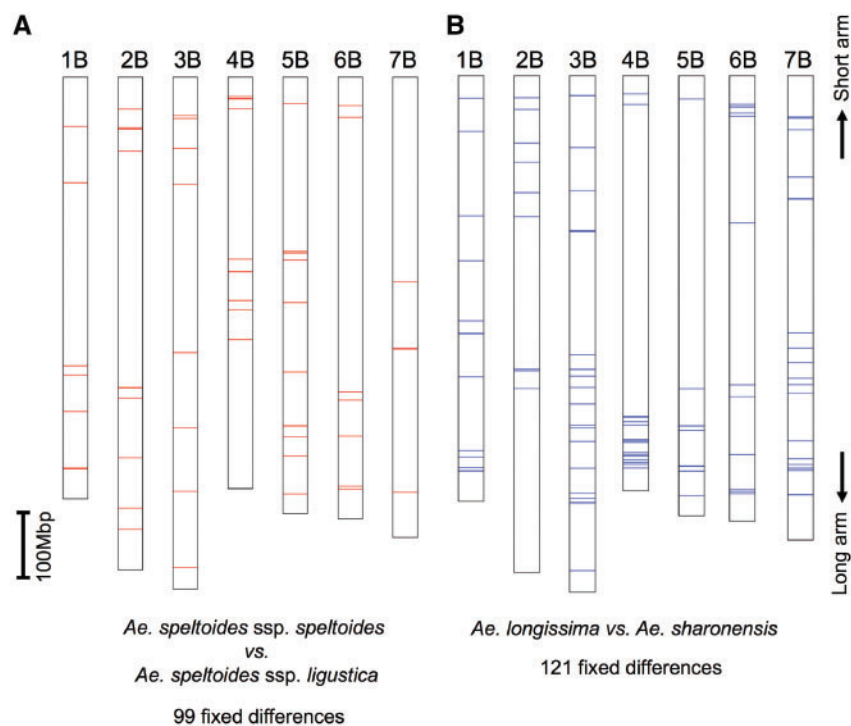


Figure 4. Distributions of the positions of fixed nucleotide differences between subspecies of *Aegilops speltoides* ssp. *ligustica* and *Ae. speltoides* ssp. *speltoides* (A) and between *Ae. sharonensis* and *Ae. longissima* (B) on the physical map of the B genome of *T. aestivum* cv. Chinese Spring from chromosomes 1B to 7B.

In conclusion, the present phylogenetic analyses based on genome-wide polymorphisms suggest that the B genome of common wheat was derived from the S genome of *Ae. speltoides*. A few chromosomal regions demonstrated the clearly exceptional relationship between the B genome and *Sitopsis* species, whereas the irregular topology observed could be explained by higher recombination rates in the distal regions of wheat chromosomes. Therefore, based on genome-wide polymorphisms identified from the RNA sequencing data, all of the chromosomal regions of the wheat B genome could have originated from the S genome of *Ae. speltoides*. Moreover, the failure of pairing between homoeologous chromosomes between the

B and S genomes during meiosis could be due to factors associated with highly evolved regions or intergenic regions. The alignment rate to B genome of *Ae. speltoides* was not as high as those of *Ae. bicornis* and *Ae. searsii* (Table 2). Of the transcripts that were entirely covered with unaligned RNA sequencing reads, 18% encoded F-box proteins and disease resistance proteins such as NBS-LRR (Supplementary data S1). Positive selection is known to act on these protein genes and increase nucleotide substitutions between species.^{64–66} After separation of the B genome from the S genome, the different selective pressure between B and S genomes may contribute to enhancing their genetic differentiation. In addition, distinct

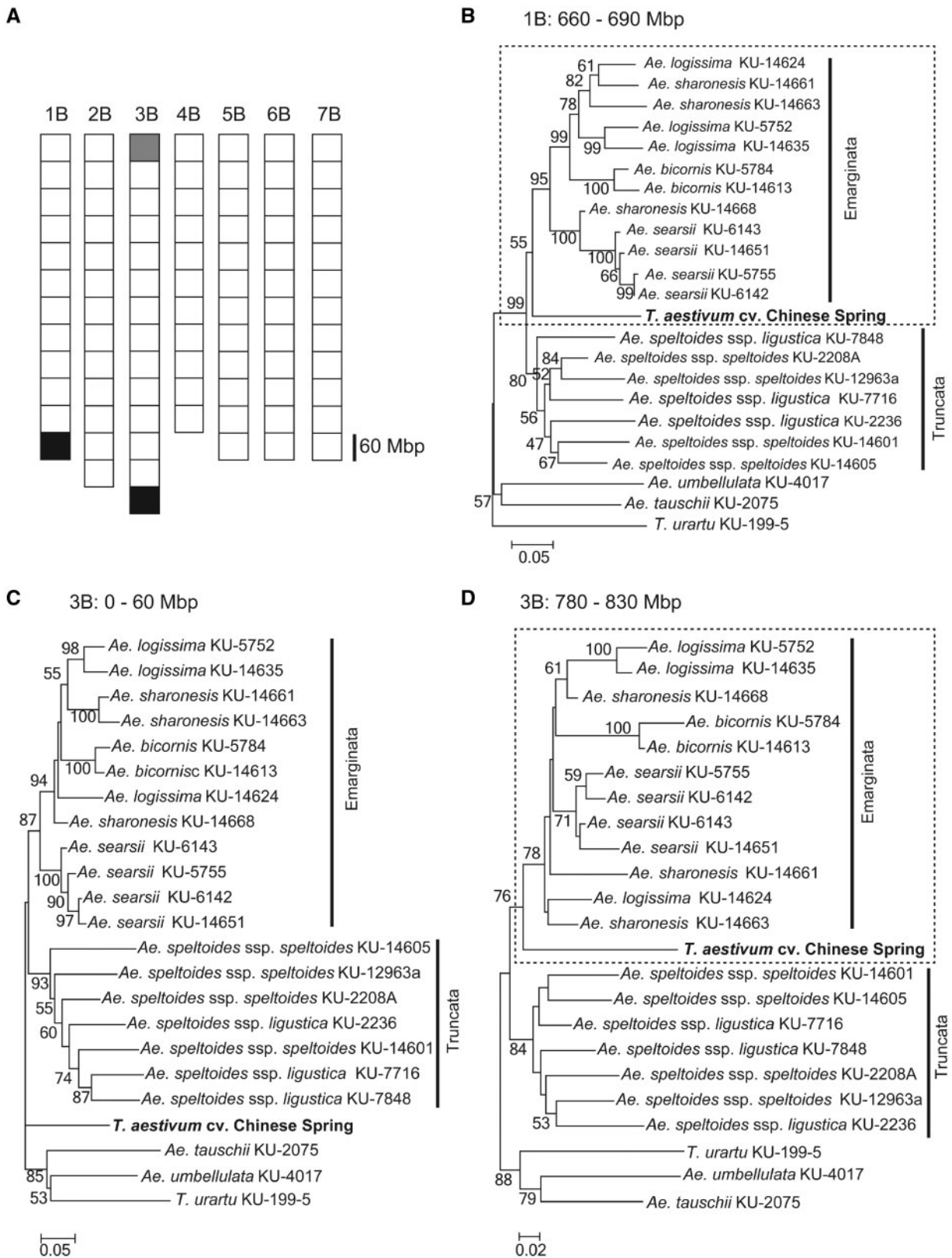


Figure 5. Irregular topologies of the phylogenetic trees in the distal chromosomal regions. The chromosomal regions were divided into 86 segments of 60 Mbp each. NJ trees were constructed based on non-redundant SNPs located in each segment. Squares on the chromosomes in panel (A) correspond to the 86 segments. Phylogenetic trees of the white squares showed that the B genome of *T. aestivum* cv. Chinese Spring was the most closely related to *Ae. speltooides* in the section Sitopsis. This observation was consistent with those for trees constructed based on all non-redundant SNPs (Fig. 5). Phylogenetic trees of the black squares showed that the B genome was closely related to Emarginata clades. A phylogenetic tree of the grey square showed that the B genome was located outside of Sitopsis species. Trees with irregular topologies at the end of the long arm of chromosome 1B (B) and the short arm (C) and long arm (D) of chromosome 3B are shown. Bootstrap probabilities with over 50% and scale bars are shown for each tree.

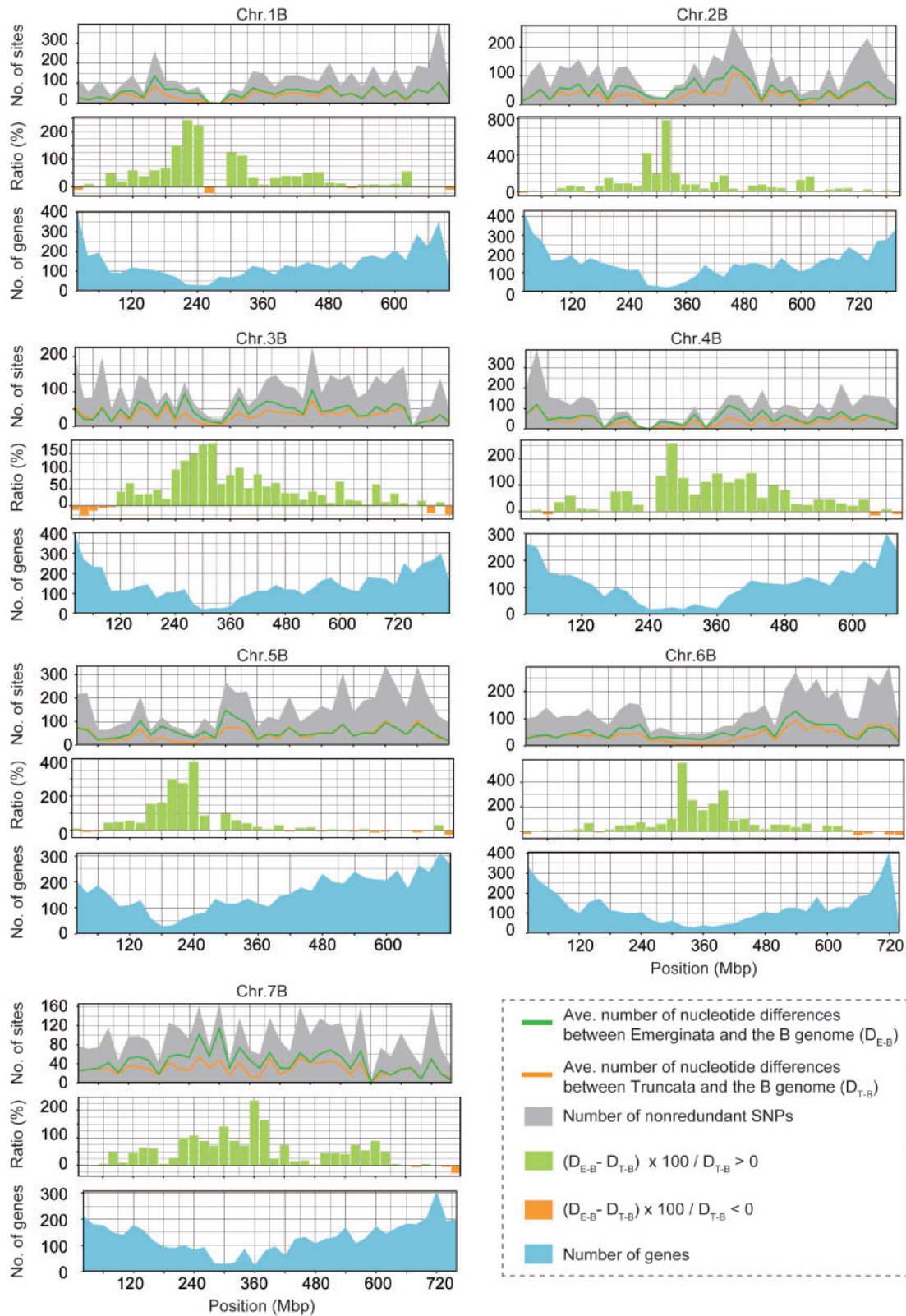


Figure 6. Contrasting pattern of nucleotide divergence in the distal and proximal regions of B-genome chromosomes. Average number of nucleotide differences per 20 Mbp between species in Truncata (= *Ae. speltoides* spp.) and the B genome of *T. aestivum* cv. Chinese Spring (D_{T-B}) and between Emerginata species and the B genome (D_{E-B}) are plotted on each chromosome. The distributions of these differences are shown by line graphs (orange and green) in the top panels. Area charts in grey colour in the top panels denote the distribution of the total number of non-redundant SNPs per 20 Mbp in Sitopsis species and the B genome. Middle panels show the distribution of the ratio expressing disparity between the two genetic divergences ($(D_{E-B} - D_{T-B}) \times 100 / D_{T-B}$) along each chromosome using bar charts. Area charts in blue colour indicate the distribution of the number of genes along each chromosome.

patterns of accumulation of repetitive sequences could have led to the differential distribution of heterochromatic regions between the B and S genomes. To elucidate the molecular nature of the differentiation of the B and S genomes, future studies should compare the repetitive sequences over all chromosomal regions.

Data availability

All read sequences were deposited into the DDBJ/GenBank/MMBL database with accession number DRA007097. RNA sequencing data for the outgroup species were used in the DDBJ/GenBank/MMBL database: BioProject PRJDB4683 for *Ae. tauschii* KU-2075 and DRA006404 for *Ae. umbellulata* KU-4017.

Acknowledgements

Seeds of section Sitopsis species used in this study were supplied by the National BioResource Project–Wheat, Japan (www.nbrp.jp). This work was supported by Grant-in-Aid for Scientific Research (B) No. 16H04862 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan to ST, by Grant-in-Aid for Scientific Research on Innovative Areas No. 17H05842 from MEXT to ST, 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. JPMJPR15QB).

Accession number

DRA007097

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at DNARES online.

References

- Kihara, H. 1944, Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*, *Agric. Hortic.*, **19**, 889–90 [in Japanese].
- McFadden, E.S. and Sears, E.R. 1946, The origin of *Triticum spelta* and its free-threshing hexaploid relatives, *J. Hered.*, **37**, 107–16.
- Matsuoka, Y. 2011, Evolution of polyploid *Triticum* wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification, *Plant Cell Physiol.*, **52**, 750–64.
- Dvorak, J., Deal, K.R., Luo, M.-C., You, F.M., von Borstel, K. and Dehghani, H. 2012, The origin of spelt and free-threshing hexaploid wheat, *J. Hered.*, **103**, 426–41.
- Dvořák, J., McGuire, P.E. and 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**, 680–9.
- Takumi, S., Nasuda, S., Liu, Y.G. and Tsunewaki, K. 1993, Wheat phylogeny determined by RFLP analysis of nuclear DNA. 1. Einkorn wheat, *Jpn. J. Genet.*, **68**, 73–9.
- Sasanuma, T., Miyashita, N.T. and Tsunewaki, K. 1996, Wheat phylogeny determined by RFLP analysis of nuclear DNA. 3. Intra- and interspecific variations of five *Aegilops Sitopsis* species, *Theor. Appl. Genetics*, **92**, 928–34.
- Daud, H.M. and Gustafson, J.P. 1996, Molecular evidence for *Triticum speltoides* as a B-genome progenitor of wheat (*Triticum aestivum*), *Genome*, **39**, 543–8.
- Maestra, B. and Naranjo, T. 1998, Homoeologous relationships of *Aegilops speltoides* chromosomes to bread wheat, *Theor. Appl. Genet.*, **97**, 181–6.
- Haider, N. 2013, The origin of the B-genome of bread wheat (*Triticum aestivum* L.), *Russ. J. Genet.*, **49**, 263–74.
- Terachi, T., Ogihara, Y. and Tsunewaki, K. 1990, The nuclear basis of genetic diversity among cytoplasts of *Triticum* and *Aegilops*. 7. Restriction endonuclease analysis of mitochondrial DNA from polyploidy wheats and their ancestral species, *Theoret. Appl. Genetics*, **80**, 366–73.
- Wang, G.-Z., Miyashita, N.T. and Tsunewaki, K. 1997, Plasmon analyses of *Triticum* (wheat) and *Aegilops*: PCR-single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs, *Proc. Natl. Acad. Sci. USA*, **94**, 14570–7.
- Jenkins, J.A. 1929, Chromosome homologies in wheat and *Aegilops*, *Am. J. Bot.*, **16**, 238–45.
- Sarkar, P. and Stebbins, G. L. 1956, Morphological evidence concerning the origin of the B genome, *Am. J. Bot.*, **43**, 297–304.
- van Slageren, M.W. 1994, *Wild Wheats: A Monograph of Aegilops L. and Amblyopyrum (Jaub. and Spach) Eig (Poaceae): A Revision of All Taxa Closely Related to Wheat, Excluding Wild Triticum Species, with Notes on Other Genera in the Tribe Triticeae, Especially Triticum*. Wageningen Agric. Univ., Wageningen.
- Eig, A. 1929, *Monographisch-kritische Übersicht der Gattung Aegilops, Repertorium Specierum Novarum Regni Vegetabilis*. Verlag des Repertoriums, Dahlem.
- Zohary, D. and Imber, D. 1963, Genetic dimorphism in fruit types in *Aegilops speltoides*, *Heredity*, **18**, 223–31.
- Luo, M.C., Deal, K.R., Yang, Z.L. and Dvořák, J. 2005, Comparative genetic maps reveal extreme crossover localization in the *Aegilops speltoides* chromosomes, *Theor. Appl. Genet.*, **111**, 1098–106.
- Hammer, K. 2008, Zur Taxonomie und Nomenklatur der Gattung *Aegilops* L., *Feddes Rep.*, **91**, 225–58.
- Zhang, H., Reader, S.M., Liu, X., Jia, J.Z., Gale, M.D. and Devos, K.M. 2001, Comparative genetic analysis of the *Aegilops longissima* and *Ae. sharonensis* genomes with common wheat, *Theor. Appl. Genet.*, **103**, 518–25.
- Kilian, B., Özkan, H., Deusch, O., et al. 2007, Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes, *Mol. Biol. Evol.*, **24**, 217–27.
- Ohta, A., Yamane, K. and Kawahara, T. 2017, Relationship between spike morphology and habitat of four *Aegilops* species of section Sitopsis, *Genet. Resour. Crop Evol.*, **64**, 889–99.
- Kihara, H. 1949, Genomanalyse bei *Triticum* und *Aegilops*. IX. Systematischer Aufbau der Gattung *Aegilops* auf genomanalytischer Grundlage, *Cytologia*, **14**, 135–44.
- Friebe, B., Tuleen, N., Jiang, J. and Gill, B.S. 1993, Standard karyotype of *Triticum longissimum* and its cytogenetic relationship with *T. aestivum*, *Genome*, **36**, 731–42.
- Salse, J., Chagué, V., Bolot, S., et al. 2008, New insights into the origin of the B genome of hexaploid wheat: evolutionary relationships at the SPA genomic regions with the S genome of the diploid relative *Aegilops speltoides*, *BMC Genomics*, **9**, 555.
- Salina, E.A., Sergeeva, E.M., Adonina, I.G., et al. 2011, The impact of Ty3-gypsy group LTR retrotransposons *Fatima* on B-genome specificity of polyploidy wheat, *BMC Plant Biol.*, **11**, 99.
- Gill, B.S. and Kimber, G. 1974, Giemsa C-banding and the evolution of wheat, *Proc. Natl. Acad. Sci. USA*, **71**, 4086–90.
- Endo, T.R. and Gill, B.S. 1984, Somatic karyotype, heterochromatin distribution, and nature of chromosome differentiation in common wheat, *Triticum aestivum* L. em Thell, *Chromosoma*, **89**, 361–9.
- Dvořák, J. and McGuire, P.E. 1981, Nonstructural chromosome differentiation among wheat cultivars, with special reference to differentiation of chromosomes in related species, *Genetics*, **97**, 391–414.

30. Sasanuma, T., Chabane, K., Endo, T.R. and Valkoun, J. 2004, Characterization of genetic variation in and phylogenetic relationships among diploid *Aegilops* species by AFLP: incongruence of chloroplast and nuclear data, *Theor. Appl. Genet.*, **108**, 612–8.
31. Petersen, G., Seberg, O., Yde, M. and Berthelsen, K. 2006, Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat, *Mol. Phylogenet. Evol.*, **39**, 70–82.
32. Wang, J.-R., Zhang, L., Wei, Y.-M., Yan, Z.-H., Baum, B.R., Nevo, E. and Zheng, Y.-L. 2007, Sequence polymorphisms and relationships of dimeric α -amylase inhibitor genes in the B genomes of *Triticum* and S genomes of *Aegilops*, *Plant Sci.*, **173**, 1–11.
33. Liu, B., Segal, G., Rong, J.K. and Feldman, M. 2003, A chromosome-specific sequence common to the B genome of polyploid wheat and *Aegilops searsii*, *Plant Syst. Evol.*, **241**, 55–66.
34. El Baidouri, M., Murat, F., Veyssiere, M., et al. 2017, Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*), *New Phytol.*, **213**, 1477–86.
35. Zhang, W., Zhang, M., Zhu, X., et al. 2018, Molecular cytogenetic and genomic analyses reveal new insights into the origin of the wheat B genome, *Theor. Appl. Genet.*, **131**, 365–75.
36. Iehisa, J.C.M., Shimizu, A., Sato, K., Nasuda, S. and 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**, 487–97.
37. Iehisa, J.C.M., Shimizu, A. and Sato, K. 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**, 261–71.
38. Nishijima, R., Yoshida, K., Motoi, Y., Sato, K. and Takumi, S. 2016, Genome-wide identification of novel genetic markers from RNA sequencing assembly of diverse *Aegilops tauschii* accessions, *Mol. Genet. Genomics*, **291**, 1681–94.
39. Okada, M., Yoshida, K., Nishijima, R., et al. 2018, RNA-seq analysis reveals considerable genetic diversity and provides genetic markers saturating all chromosomes in the diploid wild wheat relative *Aegilops umbellata*, *BMC Plant Biol.* **18**, 271.
40. Mayer, K.F.X., Martis, M., Hedley, P.E., et al. 2011, Unlocking the barley genome by chromosomal and comparative genomics, *Plant Cell*, **23**, 1249–63.
41. Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y., Wu, J. and Itoh, T. 2016, Improvement of barley genome annotations by deciphering the Haruna Nijo genome, *DNA Res.*, **23**, 21–8.
42. Bolger, A.M., Lohse, M. and Usadel, B. 2014, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, **30**, 2114–20.
43. The International Wheat Genome Sequencing Consortium (IWGSC). 2018, The reference sequence of bread wheat, IWGSC RefSeq, v1.0, <https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies>.
44. Kim, D., Langmead, B. and Salzberg, S.L. 2015, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods.*, **12**, 357–60.
45. Li, H., Handsaker, B., Wysoker, A., et al. 2009, Subgroup 1000 Genome Project Data Processing. The sequence alignment/map format and SAMtools, *Bioinformatics*, **25**, 2078–9.
46. Kosugi, S., Natsume, S., Yoshida, K., et al. 2013, Coval: improving alignment quality and variant calling accuracy for next-generation sequencing data, *PLoS One*, **8**, e75402.
47. Krzywinski, M., Schein, J., Birol, I., et al. 2009, Circos: an information aesthetic for comparative genetics, *Genome Res.*, **19**, 1639–45.
48. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.*, **10**, R25.
49. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990, Basic local alignment search tool, *J. Mol. Biol.*, **215**, 403–10.
50. Chang, Z., Li, G., Liu, J., et al. 2015, Bridger: a new framework for *de novo* transcriptome assembly using RNA-seq data, *Genome Biol.*, **16**, 30.
51. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.*, **28**, 2731–9.
52. Huson, D.H. and Bryant, D. 2006, Application of phylogenetic networks in evolutionary studies, *Mol. Biol. Evol.*, **23**, 254–67.
53. McDonald, J.H. and Kreitman, M. 1991, Adaptive protein evolution at the *Adh* locus in *Drosophila*, *Nature*, **351**, 652–4.
54. Nei, M. 1987, *Molecular Evolutionary Genetics*: Columbia University Press, New York.
55. Cingolani, P., Platts, A., Wang, L., et al. 2012, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3, *Fly*, **6**, 80–92.
56. Dvorák, J., Luo, M.C. and Yang, Z.L. 1998, Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species, *Genetics*, **148**, 423–34.
57. Olivera, P.D. and Steffenson, B.J. 2009, *Aegilops sharonensis*: origin, genetics, diversity, and potential for wheat improvement, *Botany*, **87**, 740–56.
58. Akhunov, E.D., Goodyear, A.W., Geng, S., et al. 2003, The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms, *Genome Res.*, **13**, 753–63.
59. Choulet, F., Alberti, A., Theil, S., et al. 2014, Structural and functional partitioning of bread wheat chromosome 3B, *Science*, **345**, 1249721.
60. Degnan, J.H. and Rosenberg, N.A. 2009, Gene tree discordance, phylogenetic inference and the multispecies coalescent, *Trends Ecol. Evol.*, **24**, 332–40.
61. Pease, J.B. and Hahn, M.W. 2013, More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting, *Evolution*, **67**, 2376–84.
62. Hobolth, A., Duthel, J.Y., Hawks, J., Schierup, M.H. and Mailund, T. 2011, Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection, *Genome Res.*, **21**, 349–56.
63. Suh, A., Smeds, L. and Ellegren, H. 2015, The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds, *PLoS Biol.*, **13**, e1002224.
64. Clark, R.M., Schweikert, G., Toomajian, C., et al. 2007, Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*, *Science*, **317**, 338–42.
65. Chen, Q., Han, Z., Jiang, H., Tian, D. and Yang, S. 2010, Strong positive selection drives rapid diversification of R-genes in *Arabidopsis* relatives, *J. Mol. Evol.*, **70**, 137–48.
66. Schumann, N., Navarro-Quezada, A., Ullrich, K., Kuhl, C. and Quint, M. 2011, Molecular evolution and selection patterns of plant F-box proteins with C-terminal kelch repeats, *Plant Physiol.*, **155**, 835–50.