



Developing a High-Throughput SNP-Based Marker System to Facilitate the Introgression of Traits From *Aegilops* Species Into Bread Wheat (*Triticum aestivum*)

Alexandra M. Przewieslik-Allen^{1*}, Amanda J. Burridge¹, Paul A. Wilkinson¹, Mark O. Winfield¹, Daniel S. Shaw¹, Lorna McAusland², Julie King², Ian P. King², Keith J. Edwards¹ and Gary L. A. Barker¹

¹ Life Sciences, University of Bristol, Bristol, United Kingdom, ² Plant Sciences, Sutton Bonington Campus, Leicestershire, United Kingdom

OPEN ACCESS

Edited by:

István Molnár,
Centre for Agricultural Research
(MTA), Hungary

Reviewed by:

Parveen Chhuneja,
Punjab Agricultural University, India
Kentaro Yoshida,
Kobe University, Japan
Pilar Hernandez,
Instituto de Agricultura Sostenible
(IAS), Spain

*Correspondence:

Alexandra M. Przewieslik-Allen
a.allen@bristol.ac.uk

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 11 October 2018

Accepted: 21 December 2018

Published: 24 January 2019

Citation:

Przewieslik-Allen AM, Burridge AJ,
Wilkinson PA, Winfield MO, Shaw DS,
McAusland L, King J, King IP,
Edwards KJ and Barker GLA (2019)
Developing a High-Throughput
SNP-Based Marker System to
Facilitate the Introgression of Traits
From *Aegilops* Species Into Bread
Wheat (*Triticum aestivum*).
Front. Plant Sci. 9:1993.
doi: 10.3389/fpls.2018.01993

The genus *Aegilops* contains a diverse collection of wild species exhibiting variation in geographical distribution, ecological adaptation, ploidy and genome organization. *Aegilops* is the most closely related genus to *Triticum* which includes cultivated wheat, a globally important crop that has a limited gene pool for modern breeding. *Aegilops* species are a potential future resource for wheat breeding for traits, such as adaptation to different ecological conditions and pest and disease resistance. This study describes the development and application of the first high-throughput genotyping platform specifically designed for screening wheat relative species. The platform was used to screen multiple accessions representing all species in the genus *Aegilops*. Firstly, the data was demonstrated to be useful for screening diversity and examining relationships within and between *Aegilops* species. Secondly, markers able to characterize and track introgressions from *Aegilops* species in hexaploid wheat were identified and validated using two different approaches.

Keywords: *Aegilops*, wheat, genotyping array, single nucleotide polymorphism (SNP), introgression, wheat relative

INTRODUCTION

Aegilops is a genus of Eurasian annual grasses in the Poaceae known as the goatgrasses. There are 23 species within *Aegilops*; these species represent six different genome types (D, S, U, C, N, and M) and three different ploidy levels (diploid, tetraploid, and hexaploid) (The Plant List, 2013; Molnár et al., 2016; **Figure 1**). The genus *Aegilops* is the most closely related to the genus *Triticum*, which contains *Triticum aestivum* (bread wheat) and other domesticated wheats. Researchers have suggested that *Aegilops* and *Triticum* should be combined into a single evolutionary complex or even the same genus (Yamane and Kawahara, 2005). The close genetic relationship is evidenced by the numerous hybridisations that occur between members of both genera and by the presence of *Aegilops* in the evolutionary history of many *Triticum* species. Where geographic distributions are similar, gene flow has occurred between species; some species, such as *Aegilops cylindrica* have spread with wheat and have become uncontrollable weeds in wheat (Donal and Ogg, 1991). If treated separately, *Aegilops* appears to be basal to *Triticum*, with evidence indicating the genus' *Triticum* and *Aegilops* diverged an estimated 4 million years ago (Huang et al., 2002).

Aegilops has been divided into six sections based on morphological and genetic analysis. These are *Sitopsis* (Jaubert and Spach, 1850–1853) Zhuk., *Amblyopyrum* (Jaubert and Spach, 1850–1853) Eig, *Polyeides* Zhuk., *Cylindropyrum* (Jaubert and Spach, 1850–1853) Zhuk., *Comopyrum* (Jaubert and Spach, 1850–1853) Zhuk., *Vertebrata* Zhuk. (**Table 1**; Zhukovsky, 1928; Eig, 1929; Yamane and Kawahara, 2005; Schneider et al., 2008; Wang et al., 2013). *Aegilops mutica* (*syn. Amblyopyrum muticum*) has been separated by some researchers and placed into a monospecific genus called *Amblyopyrum* (Van Slageren, 1994) but for the purposes of this study has been included with other members of *Aegilops* for analysis. *Ae. speltoides* is thought to be the closest relative to the wheat B-genome and is also the donor of the G-genome of *Triticum timopheevii* (Dvorak et al., 1998a; Feldman, 2001). *Ae. tauschii* is the progenitor of the wheat D-genome, hybridizing with the AB-genome progenitor ~10,000 years ago to produce hexaploid bread wheat (McFadden and Sears, 1946). This rare hybridization event is thought to have only occurred once or a small number of times resulting in a severe genetic bottleneck (Charmet, 2011). Further inbreeding and domestication pressures have resulted in a narrow gene pool for modern bread wheat breeding.

The genus *Aegilops* promises to be an important resource for wheat breeding as it harbors a high level of genetic diversity, particularly with relation to adaptation to different ecological conditions and pest and disease resistance. All *Aegilops* species are undomesticated and have wide geographic distributions and natural variation (Ostrowski et al., 2016). *Aegilops* contains species belonging to the secondary gene pool of wheat, meaning they have a genome homologous with wheat and conventional crossing may be used to transfer genes to wheat (*Ae. tauschii* and *Ae. speltoides*). Other more distantly related members of the genus belong to the tertiary gene pool of wheat and may need specific breeding techniques for gene transfers to wheat, although crosses between the two genera have been reported to occur naturally (Popova, 1923; Leighty and Taylor, 1927; Schneider et al., 2008). Interspecific hybridization between bread wheat and members of the genus *Aegilops* has been used historically in wheat breeding to confer beneficial traits from *Aegilops* into bread wheat. These include resistance to rusts, powdery mildew, eyespot, nematodes, hessian fly and wheat aphid (see Schneider et al., 2008 for a full review). The genus *Aegilops* is a potential source of further genes conferring agronomically valuable traits, such as drought tolerance, salt tolerance, heat tolerance, tolerance to toxicity and nutritional and bread-making quality traits of potential use in plant breeding (Molnár et al., 2004; Colmer et al., 2006; Schneider et al., 2008; Kilian et al., 2011).

Advances in genome sequencing over the last decade have had huge impacts on our knowledge of the large and complex hexaploid wheat genome and our ability to develop molecular markers (Uauy, 2017; The International Wheat Genome Sequencing Consortium (IWGSC), 2018). The knock-on effects of these developments have been seen in the breeding lines developed, the widespread adoption of molecular markers in breeding programmes and the development of new breeding techniques, such as genomic selection (Schneider et al., 2008; Bassi et al., 2016). At the same time there has been recognition of

the importance of pre-breeding programmes specifically targeted at introducing genetic diversity from exotic sources, such as landraces and wheat relatives (Moore, 2015). The introduction of such diverse material has necessitated the development of specific molecular markers that are able to identify and characterize wheat relative DNA in the wheat genome (Winfield et al., 2016).

With the development of genomic tools and technologies enabling precise and efficient breeding *Aegilops* promises to be an increasingly important resource of genetic diversity in future wheat breeding. A potential drawback of utilizing wide crosses to introduce diversity in this way is the inclusion of large non-recombining blocks from a relative into the wheat genome. However, with the development of genomic technologies, improved crossing techniques and gene editing technologies it is becoming possible to target genomic regions with increased precision. To enable these techniques to be employed successfully there is a requirement for increasingly dense and precise molecular markers, which can be utilized in a high-throughput manner. A key challenge is to develop markers to track introgressed DNA in the wheat genetic background. This study describes the identification, validation and use of markers systems for facilitating the introgression of *Aegilops* species into hexaploid wheat. The wide range of species used in the study represent the three different ploidy levels and six genome types found within the genus. The markers developed have enabled the detection of *Aegilops* introgressions in newly developed lines with examples of how these markers have been deployed in different introgression projects.

MATERIALS AND METHODS

Germplasm

The accessions grown for DNA extraction (listed in **Supplementary File 1**) were grown in peat-based soil in pots and maintained in a glasshouse at 15–25°C with 14-h light, 8-h dark. Leaf tissue was harvested 4 weeks after germination, frozen in liquid nitrogen and stored at –20°C prior to nucleic acid extraction. Genomic DNA was prepared using a phenol–chloroform extraction method (BurrIDGE et al., 2017), treated with RNase-A (QIAGEN Ltd., Manchester, UK) according to the manufacturer's instructions and purified using the QiaQuick PCR purification kit (QIAGEN Ltd).

Genotyping

The original SNP collection consisted of 819,571 SNPs obtained from genic sequences derived via targeted capture re-sequencing of numerous wheat lines and validated on the Axiom[®] HD Wheat Genotyping Array (Winfield et al., 2016; Affymetrix UK Ltd, High Wycombe, UK; EVA accession PRJEB29561). The most informative 36,711 SNPs were selected for inclusion on The Axiom[®] Wheat-Relative Genotyping Array, based on data from screening ten wild relative species (*Ae. mutica*, *Ae. speltoides*, *Aegilops. tauschii*, *Triticum timopheevii*, *T. urartu*, *Secale cereale*, *Thinopyrum bessarabicum*, *Th. elongatum*, *Th. Intermedium*, and *Th. Ponticum*, **Supplementary File 2**; http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/axiom_download.php). Markers were selected to maximize polymorphism between

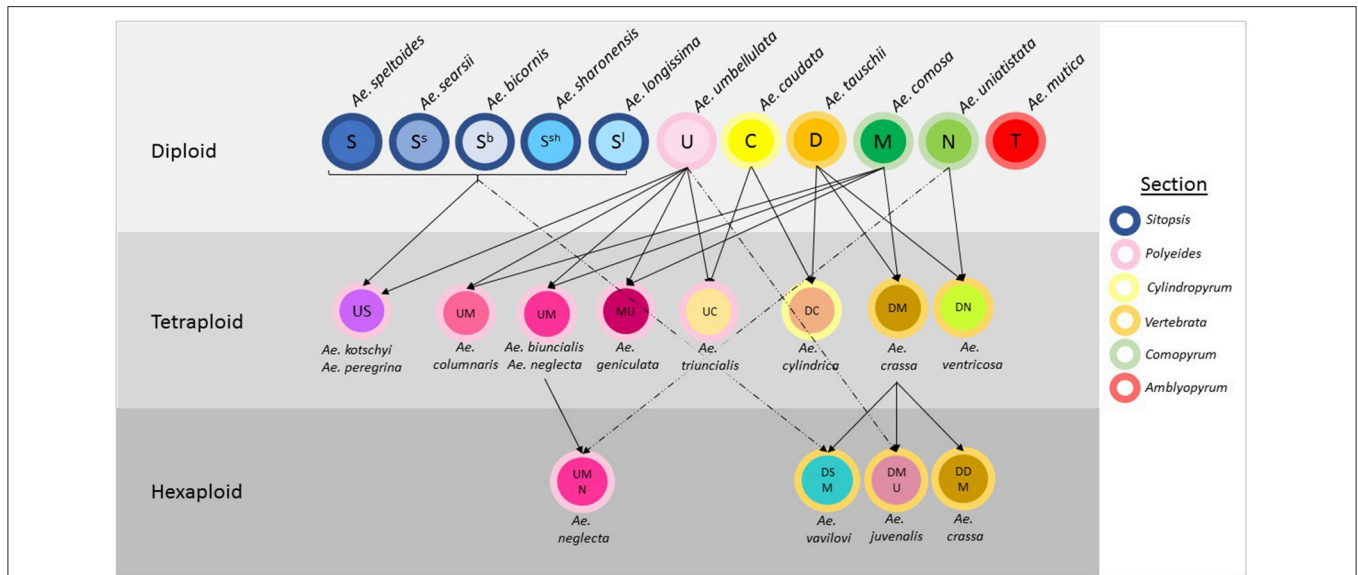


FIGURE 1 | Relationships between species in the genus *Aegilops*. The genome classification of each species is indicated within circles representing the species and arrows designate hybridisations between species. The color of the outline of each circle represents the section the species is allocated to.

the relative species' and wheat. Markers chosen behaved in a co-dominant manner making them potentially useful for identifying heterozygous calls in wheat-relative crosses (Allen et al., 2012). The Axiom[®] Wheat-Relative Genotyping Array was used to screen 278 *Aegilops* accessions using the Affymetrix GeneTitan[™] system according to the procedure described by Affymetrix (Axiom[®] 2.0 Assay Manual Workflow User Guide Rev3). Allele calling was carried out using the Affymetrix proprietary software packages Axiom Analysis Suite[™]. A variant Dish QC threshold of 0.75 was used instead of the default value (0.8) to account for the lower call rates typically obtained from hybridizing wheat relatives and progenitors to the array (Winfield et al., 2016).

The probes on the array are biallelic; for each locus there is a maximum of three calls possible (AA, AB, BB or 0, 1, 2) possible. The clustering pattern for each locus will depend upon the other lines that have been screened. For diploid species the clustering and genotype calling is straightforward, however screening polyploid lines is more complex. Although the intended application of the array is for specific intraspecific crosses where discrete populations are analyzed separately it is also possible to use the array to screen more diverse collections. For more complex populations or collections, a recommended approach would be to focus on particular loci of interest that produce clear clustering patterns. If necessary, it would be possible to examine the behavior of this subset of probes on the array and relate these to known genotypes. An important factor to note is whether the probe is "co-dominant" and interacts with just the genome of interest or if homeologous genomes also hybridize which will complicate the clustering pattern (Allen et al., 2012). Generally, probes that preferentially or specifically hybridize to a single genome will give higher quality clustering patterns even when screening diverse

lines and users could preferentially choose these for further analysis.

For example, the accessions of DDMMUU genome have the genotype ABAAAA and BBAAAA and an accession of the DD genome has the genotype AB. For a dominant probe you would get the following calling pattern: 0 = ABAAAA, 1 = BBAAAA, 2 = AB (assuming the interaction of the MM and UU genomes also). For a co-dominant, D-genome specific probe you would get 0 = BB, 1 = AB. Introducing more different lines with different genome compositions could further complicate the clustering pattern.

Assignment of a physical map position to the SNP markers was achieved by BLAST searching the probe sequences to the International Wheat Genome Sequencing Consortium (IWGSC) whole genome assembly v1.0. For further analysis (see below) data were screened for quality using the monomorphic and call rate filtering of Axiom Analysis Suite. High quality probes taken forward for PCoA and phylogenetic analysis had a call rate of 80% or higher.

Dimensionality Reduction

A distance matrix was generated from the genotype scores using R package SNPRelate (Zheng et al., 2012). The proportion of variance for the first six eigenvalues was as follows: 26.45, 15.93, 7.21, 4.60, 4.42, 3.97. The first two eigenvalues with over 42% of the variance were plotted as a PCA plot.

Phylogenetic Analysis

Evolutionary relationships between *Aegilops* varieties were investigated using the genotype calls from the wheat relative genotyping array. The SNPhylo pipeline (version 20160204; Lee et al., 2014) was used to construct phylogenetic trees based on a haplotype map of 278 *Aegilops* varieties that was derived from

TABLE 1 | Details of germplasm used in study.

Species	Section	Genome	Ploidy	Number of accessions	Number of SNPs within species	Number of SNPs compared to <i>Triticum aestivum</i>
<i>Aegilops mutica</i>	<i>Amblyopyrum</i>	TT	2	12	15,617	33,816
<i>Aegilops comosa</i>	<i>Comopyrum</i>	MM	2	14	16,103	33,305
<i>Aegilops uniaristata</i>	<i>Comopyrum</i>	NN	2	6	9,190	33,120
<i>Aegilops caudata</i>	<i>Cylindropyrum</i>	CC	2	7	11,668	33,433
<i>Aegilops cylindrica</i>	<i>Cylindropyrum</i>	DDCC	4	7	13,147	30,052
<i>Aegilops biuncialis</i>	<i>Polyeides</i>	UUMM	4	8	13,040	33,129
<i>Aegilops columnaris</i>	<i>Polyeides</i>	UUMM	4	7	13,953	33,419
<i>Aegilops geniculata</i>	<i>Polyeides</i>	MMUU	4	5	12,412	33,772
<i>Aegilops kotschyii</i>	<i>Polyeides</i>	SSUU	4	8	11,658	33,024
<i>Aegilops neglecta</i>	<i>Polyeides</i>	UUMM/UUMMNN	4 or 6	6	17,373	33,306
<i>Aegilops peregrina</i>	<i>Polyeides</i>	UUSS	4	9	16,428	33,127
<i>Aegilops triuncialis</i>	<i>Polyeides</i>	UUCG	4	17	17,265	33,378
<i>Aegilops umbellulata</i>	<i>Polyeides</i>	UU	2	15	15,254	33,365
<i>Aegilops bicornis</i>	<i>Sitopsis</i>	SS	2	5	17,817	34,085
<i>Aegilops longissima</i>	<i>Sitopsis</i>	SS	2	10	15,835	33,093
<i>Aegilops searsii</i>	<i>Sitopsis</i>	SS	2	13	10,575	32,804
<i>Aegilops sharonensis</i>	<i>Sitopsis</i>	SS	2	13	18,220	33,339
<i>Aegilops speltoides</i>	<i>Sitopsis</i>	SS	2	38	24,524	32,401
<i>Aegilops crassa</i>	<i>Vertebrata</i>	DDMM/DDMMMM	4 or 6	7	11,798	30,103
<i>Aegilops juvenalis</i>	<i>Vertebrata</i>	DDMMUU	6	5	12,990	31,120
<i>Aegilops tauschii</i>	<i>Vertebrata</i>	DD	2	22	21,867	31,212
<i>Aegilops vavilovii</i>	<i>Vertebrata</i>	DDSSMM	6	5	13,893	30,969
<i>Aegilops ventricosa</i>	<i>Vertebrata</i>	DDNN	4	11	14,631	30,211

the genotype calls. The SNPhylo pipeline removed 13,253 lines of low quality data from the hapmap file, using the software's default parameters. Low quality data was defined as monomorphic, a MAF score of <0.1 or where 10% or more of the varieties had missing data. *Aegilops mutica* was assigned as an outgroup to root the tree based upon PCA results and putative reclassification to a separate genus from *Aegilops*. The hapmap file was then submitted to the SNPhylo pipeline, and a maximum likelihood tree was generated with a bootstrapping value set to 10,000. The Newick strings generated by SNPhylo were imported into the R package ggtree (version 1.2.17; Yu et al., 2017) which was used to construct a circular dendrogram.

Introgression Detection

The identification of putative introgressions was performed by comparing the genotype calls of hexaploid lines to *Aegilops* accessions over a 10 SNP window and calculating a percentage match. Analysis of control introgression lines indicated that a match of 40% or higher within the 10 SNP window was indicative of an introgression in the wheat background. This threshold was chosen based on the screening of known introgressions, such as 1B/1RS.

RESULTS

Diversity Within the Genus *Aegilops*

Of the 36,711 SNPs on the wheat relative array 34,602 (94.3%) were polymorphic in the entire collection of *Aegilops* accessions

used in the study (**Supplementary File 2**). The SNPs fell into the following classification categories: Poly High Resolution, 18.8%; No Minor Homozygote, 10.3%; Off Target Variant, 19.3%; Mono High Resolution, 5.6%; Call Rate Below Threshold, 5.6%; Other, 40.3%. The number polymorphic within each species ranged from 9,190 (25.0%; *Ae. uniaristata*) to 24,524 (66.9%; *Ae. speltoides*) and was related to the number of accessions genotyped ($R^2 = 0.56$) (**Table 1**). This was explored further by selecting four random subsets of five accessions for each of *Ae. speltoides* and *Ae. tauschii*. The number of polymorphisms detected in the *Ae. speltoides* subsets ranged from 10,842 to 12,366 and for *Ae. tauschii* ranged from 6,555 to 11,866. These figures are comparable to the other species sampled and suggest that the number of accessions has a clear effect on the number of polymorphisms detected rather than the species in question. The number of SNPs polymorphic between each species and the wheat samples genotyped ranged from 30,052 (81.9%; *Ae. cylindrica*) to 34,085 (92.8%; *Ae. bicornis*) and there appeared to be no relationship ($R^2 = 0$) between the number of accessions genotyped and the number of polymorphisms detected when compared with wheat. There was no significant relationship between the number of polymorphic SNPs within a species compared to between the species and wheat ($R^2 = 0.0047$).

A principal component (PCoA) analysis was used to visualize the relationship between genotyped accessions (**Figure 2; Supplementary File 3**). In general species containing a D-genome (section *Vertebrata*) were distributed in discrete clusters along the PC2 axis. The diploid *Ae. tauschii* clusters

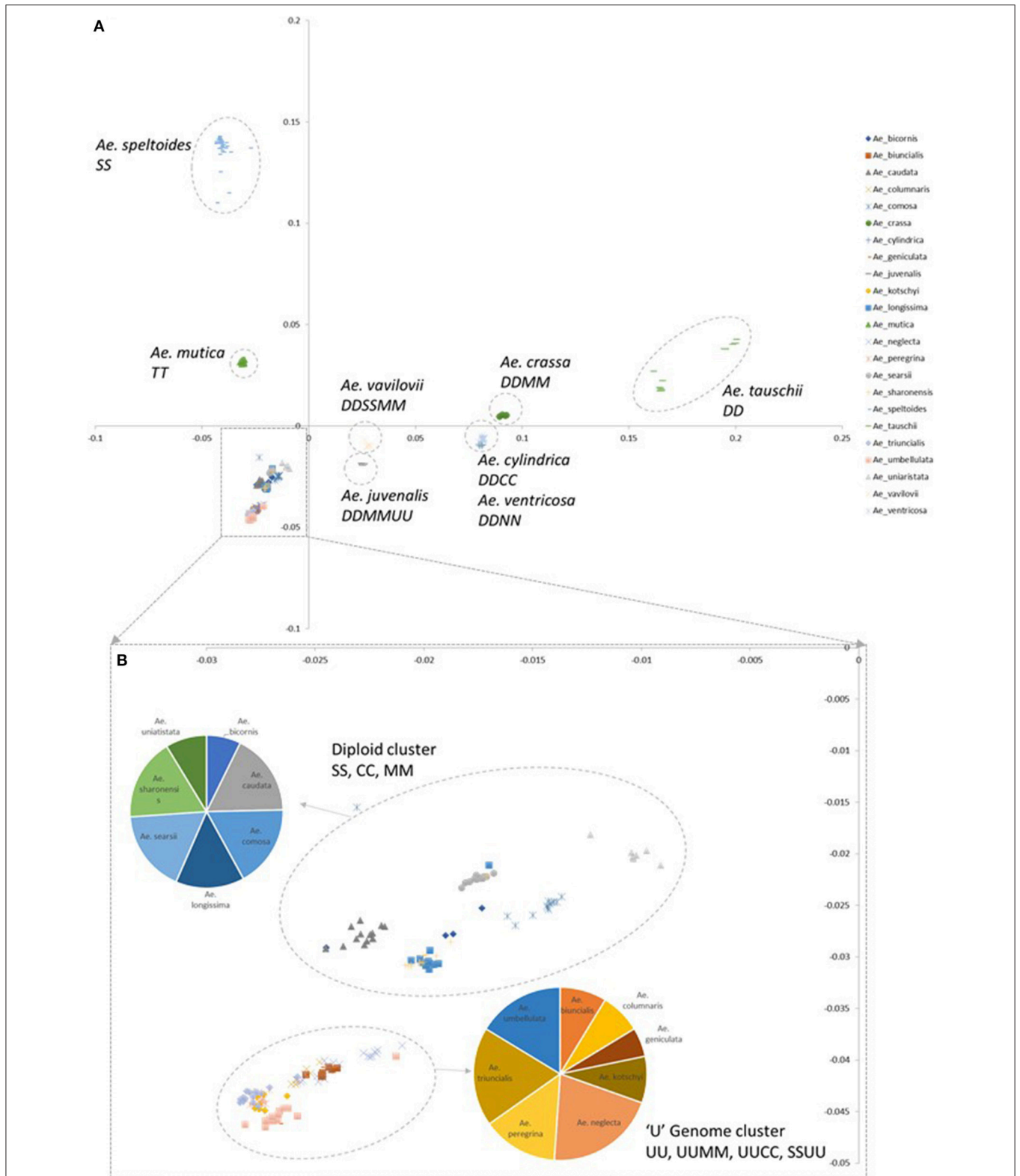


FIGURE 2 | Principal coordinate analysis (PCoA) plots colored by species. **(A)** Coordinate 1 is plotted along the y-axis, coordinate 2 is plotted along the x-axis. **(B)** Detail of clusters with negative PC1 and PC2 values. Pie charts indicate the numbers of each species belonging to designated clusters.

TABLE 2 | Genetic differentiation measured as Nei's standard genetic distance (above diagonal) and Fst measures (below diagonal) between species.

	<i>Ae ventricosa</i>	<i>Ae speltooides</i>	<i>Ae bicornis</i>	<i>Ae columnaris</i>	<i>Ae mutica</i>	<i>Ae neglecta</i>	<i>Ae biuncialis</i>	<i>Ae juvenalis</i>	<i>Ae triuncialis</i>	<i>Ae tauschii</i>	<i>Ae caudata</i>	<i>Ae umbellulata</i>	<i>Ae longissima</i>	<i>Ae geniculata</i>	<i>Ae searsii</i>	<i>Ae crassa</i>	<i>Ae uniaristata</i>	<i>Ae kotschy</i>	<i>Ae peregrina</i>	<i>Ae comosa</i>	<i>Ae cylindrica</i>	<i>Ae vavilovii</i>	<i>Ae sharonensis</i>
<i>Ae ventricosa</i>	0	0.33	0.22	0.22	0.28	0.19	0.2	0.16	0.22	0.13	0.25	0.25	0.25	0.21	0.27	0.12	0.18	0.23	0.23	0.2	0.12	0.17	0.25
<i>Ae speltooides</i>	0.58	0	0.2	0.22	0.17	0.2	0.22	0.25	0.21	0.48	0.23	0.24	0.22	0.22	0.24	0.35	0.25	0.22	0.21	0.23	0.34	0.25	0.22
<i>Ae bicornis</i>	0.5	0.44	0	0.12	0.17	0.11	0.13	0.16	0.1	0.36	0.11	0.15	0.04	0.13	0.07	0.25	0.17	0.07	0.07	0.15	0.21	0.13	0.04
<i>Ae columnaris</i>	0.54	0.5	0.38	0	0.16	0.02	0.06	0.13	0.03	0.38	0.1	0.05	0.16	0.05	0.18	0.26	0.17	0.08	0.07	0.14	0.21	0.18	0.15
<i>Ae mutica</i>	0.58	0.44	0.44	0.48	0	0.15	0.16	0.21	0.15	0.45	0.17	0.19	0.2	0.17	0.22	0.31	0.21	0.18	0.17	0.18	0.28	0.22	0.2
<i>Ae neglecta</i>	0.48	0.49	0.34	0.1	0.44	0	0.03	0.11	0.04	0.37	0.11	0.05	0.15	0.03	0.17	0.24	0.11	0.07	0.06	0.1	0.2	0.16	0.14
<i>Ae biuncialis</i>	0.52	0.51	0.4	0.27	0.49	0.14	0	0.13	0.05	0.38	0.12	0.06	0.17	0.02	0.18	0.26	0.13	0.08	0.07	0.07	0.22	0.18	0.16
<i>Ae juvenalis</i>	0.44	0.52	0.42	0.4	0.51	0.36	0.41	0	0.12	0.25	0.18	0.13	0.2	0.12	0.21	0.09	0.21	0.13	0.12	0.19	0.16	0.06	0.19
<i>Ae triuncialis</i>	0.54	0.51	0.35	0.16	0.47	0.18	0.25	0.4	0	0.39	0.07	0.04	0.15	0.05	0.17	0.26	0.16	0.06	0.05	0.14	0.18	0.17	0.14
<i>Ae tauschii</i>	0.41	0.67	0.63	0.67	0.7	0.65	0.67	0.56	0.67	0	0.4	0.4	0.39	0.38	0.4	0.11	0.4	0.39	0.39	0.38	0.11	0.25	0.38
<i>Ae caudata</i>	0.57	0.52	0.36	0.4	0.5	0.37	0.44	0.5	0.3	0.69	0	0.14	0.18	0.13	0.19	0.29	0.18	0.14	0.13	0.15	0.16	0.2	0.17
<i>Ae umbellulata</i>	0.58	0.54	0.46	0.24	0.54	0.21	0.27	0.44	0.21	0.7	0.49	0	0.19	0.05	0.21	0.29	0.19	0.08	0.07	0.16	0.24	0.21	0.18
<i>Ae longissima</i>	0.56	0.5	0.15	0.49	0.53	0.44	0.5	0.51	0.47	0.67	0.52	0.56	0	0.17	0.1	0.27	0.21	0.08	0.07	0.18	0.26	0.16	0.01
<i>Ae geniculata</i>	0.52	0.5	0.39	0.25	0.49	0.12	0.11	0.4	0.23	0.67	0.45	0.23	0.5	0	0.18	0.26	0.14	0.08	0.07	0.09	0.22	0.18	0.16
<i>Ae searsii</i>	0.6	0.54	0.3	0.56	0.59	0.49	0.56	0.56	0.52	0.71	0.58	0.61	0.41	0.57	0	0.29	0.22	0.14	0.13	0.2	0.27	0.13	0.1
<i>Ae crassa</i>	0.38	0.59	0.52	0.59	0.61	0.54	0.59	0.32	0.58	0.37	0.62	0.64	0.6	0.59	0.65	0	0.29	0.27	0.26	0.27	0.11	0.09	0.27
<i>Ae uniaristata</i>	0.49	0.54	0.49	0.53	0.56	0.39	0.47	0.54	0.5	0.7	0.56	0.58	0.58	0.5	0.64	0.65	0	0.18	0.18	0.12	0.26	0.22	0.2
<i>Ae kotschy</i>	0.56	0.51	0.28	0.33	0.52	0.28	0.35	0.43	0.28	0.69	0.49	0.35	0.33	0.34	0.51	0.61	0.57	0	0.02	0.16	0.23	0.16	0.07
<i>Ae peregrina</i>	0.54	0.5	0.26	0.28	0.49	0.25	0.3	0.4	0.23	0.67	0.44	0.29	0.29	0.29	0.47	0.58	0.52	0.09	0	0.15	0.22	0.15	0.06
<i>Ae comosa</i>	0.51	0.51	0.41	0.45	0.51	0.34	0.28	0.49	0.44	0.67	0.48	0.51	0.51	0.34	0.57	0.59	0.43	0.5	0.46	0	0.24	0.19	0.18
<i>Ae cylindrica</i>	0.37	0.58	0.48	0.52	0.58	0.49	0.54	0.43	0.5	0.38	0.48	0.58	0.57	0.54	0.62	0.36	0.6	0.57	0.54	0.55	0	0.17	0.25
<i>Ae vavilovii</i>	0.45	0.51	0.36	0.48	0.52	0.44	0.49	0.23	0.48	0.56	0.52	0.54	0.45	0.48	0.44	0.31	0.55	0.47	0.45	0.5	0.45	0	0.15
<i>Ae sharonensis</i>	0.56	0.5	0.14	0.48	0.53	0.43	0.49	0.5	0.45	0.67	0.52	0.54	0.04	0.49	0.41	0.59	0.57	0.31	0.27	0.51	0.57	0.45	0

are located furthest from other *Aegilops* species, the tetraploid D-genome species *Ae. crassa*, *Ae. cylindrica*, and *Ae. ventricosa* mid-way along the axis and the hexaploid *Ae. vavilovii* and *Ae. juvenalis* closest to other *Aegilops* species. *Aegilops tauschii* formed two discrete clusters reflecting the two separate gene pools sampled in the accessions (Dvorak et al., 1998b; Mizuno et al., 2010; Jones et al., 2013). Along the PC1 axis the bread wheat B-genome progenitor species *Ae. speltooides* clustered furthest from other *Aegilops* species and the other samples clustering with positive PC1 values belong to *Ae. mutica*. The remaining samples cluster more closely with negative PC1 and PC2 values. They are split into two locations; one consisting of diploid S, C, M, and N genome species; the other

of diploid and tetraploid U genome containing species (section *Polyeides*).

A comparison of genetic differentiation and Fst between species (Table 2) revealed relationships between species with *Ae. tauschii* in particular showing a high degree of differentiation to most other species (except the related *Ae. crassa*, *Ae. ventricosa*, and *Ae. cylindrica*). Diploid species of different genome classifications showed a high degree of differentiation to each other but more similarity to polyploid species from the same section (e.g., *Ae. umbellulata* and polyploid species in section *Polyeides*) or containing a common genome (e.g., U-genome containing tetraploids). Other relationships are also revealed; the *Sitopsis* species showing greatest similarity to the

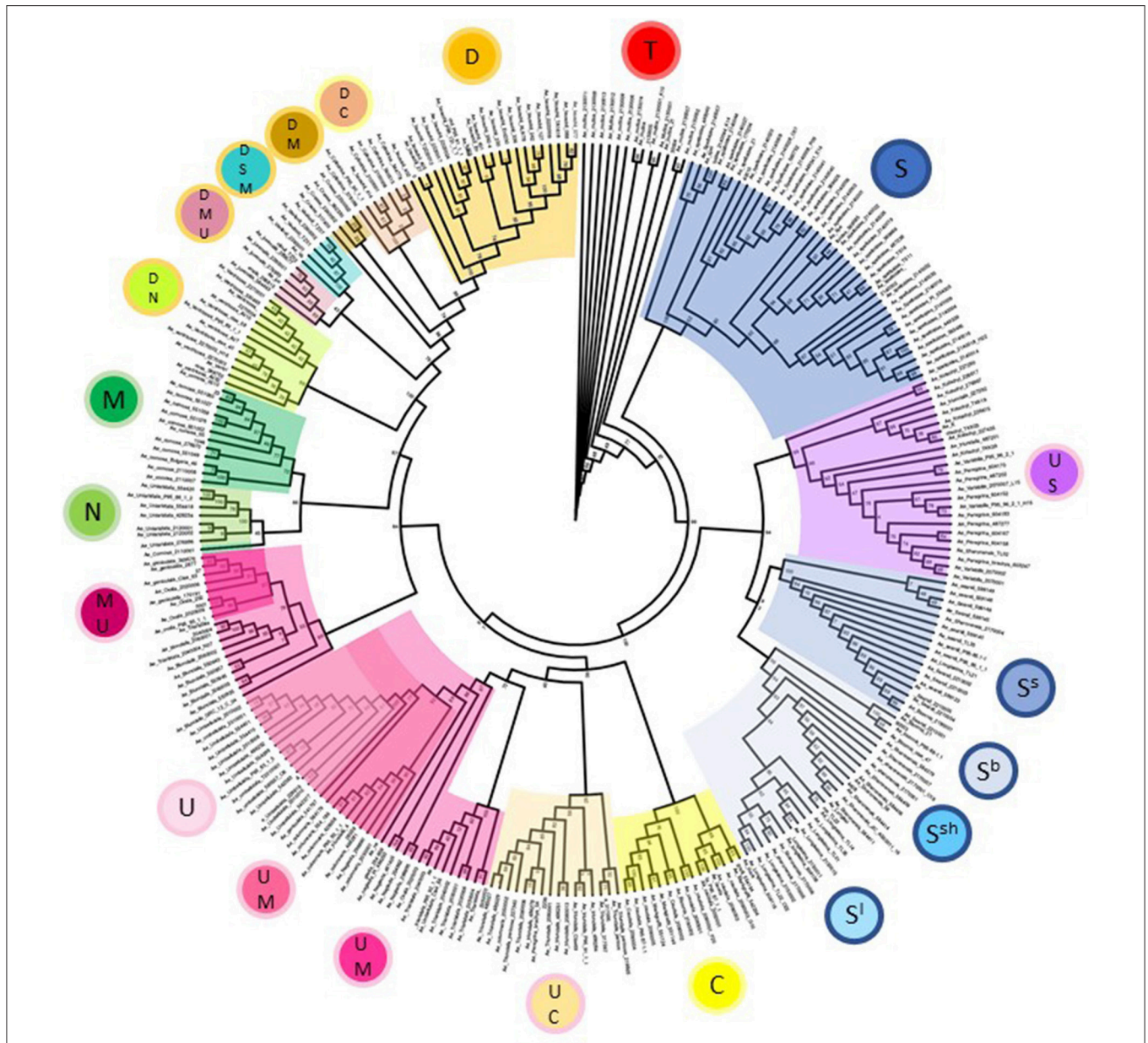


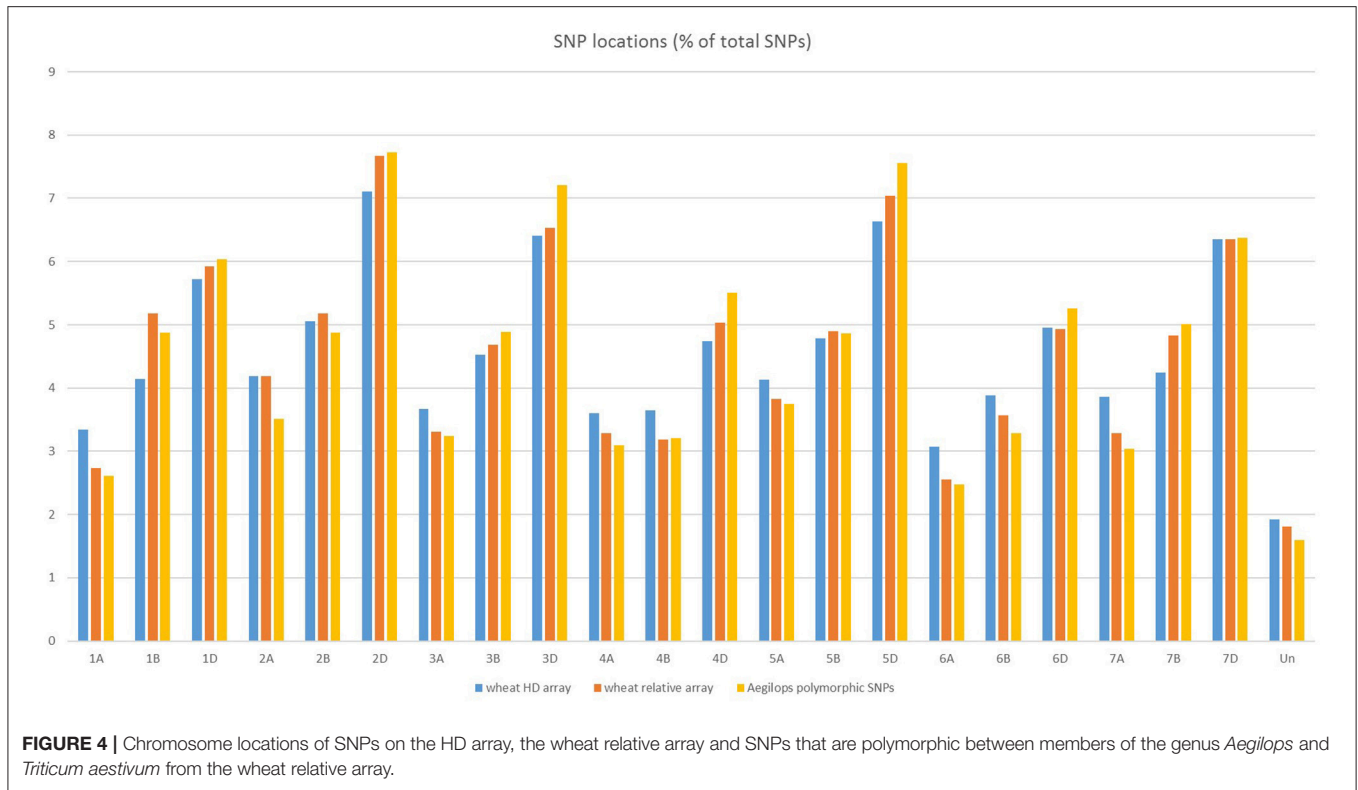
FIGURE 3 | Phylogeny of *Aegilops* accessions used in the study based upon a maximum likelihood tree generated using all genotype data. Bootstrap support values are given at branch points and are based upon 10,000 replicates. Clades are colored according to genome designations given in the inner circles in **Figure 1**.

S-genome of *Ae. kotschyi* and *Ae. peregrina* is *Ae. sharonensis*, while the UUMM tetraploid showing greatest similarity (*F_{st}*) to *Ae. juvenalis* (DDMMU) is *Ae. geniculata* suggesting evolutionary relationships between these species.

Phylogenetic Relationships Within the Genus *Aegilops*

Phylogenetic analysis of *Aegilops* species used in the study broadly reflects the picture seen in the principal component analysis but allowed us to examine relationships between species and accessions in greater detail (**Figure 3**). All major branches had good bootstrap support with values over 80%. *Aegilops*

speltoides accessions formed a separate clade to other S-genome species in the study at the base of the tree. Other S-genome containing species grouped together with tetraploids *Ae. kotschyi* and *Ae. peregrina* (UUSS) forming a separate clade to diploids *Ae. searsii*, *Ae. sharonensis*, *Ae. bicornis*, and *Ae. longissima*. The C-genome containing species *Ae. caudata* (diploid, CC) and *Ae. triuncialis* (tetraploid, UCCC) form two neighboring subclades of the clade also containing the U-genome containing diploid *Ae. umbellulata* (UU) and tetraploids *Ae. columnaris* and *Ae. neglecta* (tetraploid, UUMM). Interestingly, other UUMM tetraploids *Ae. geniculata*, *Ae. biuncialis*, and *Ae. ovata* (*syn. Ae. neglecta*, The Plant List) formed a separate clade to these,



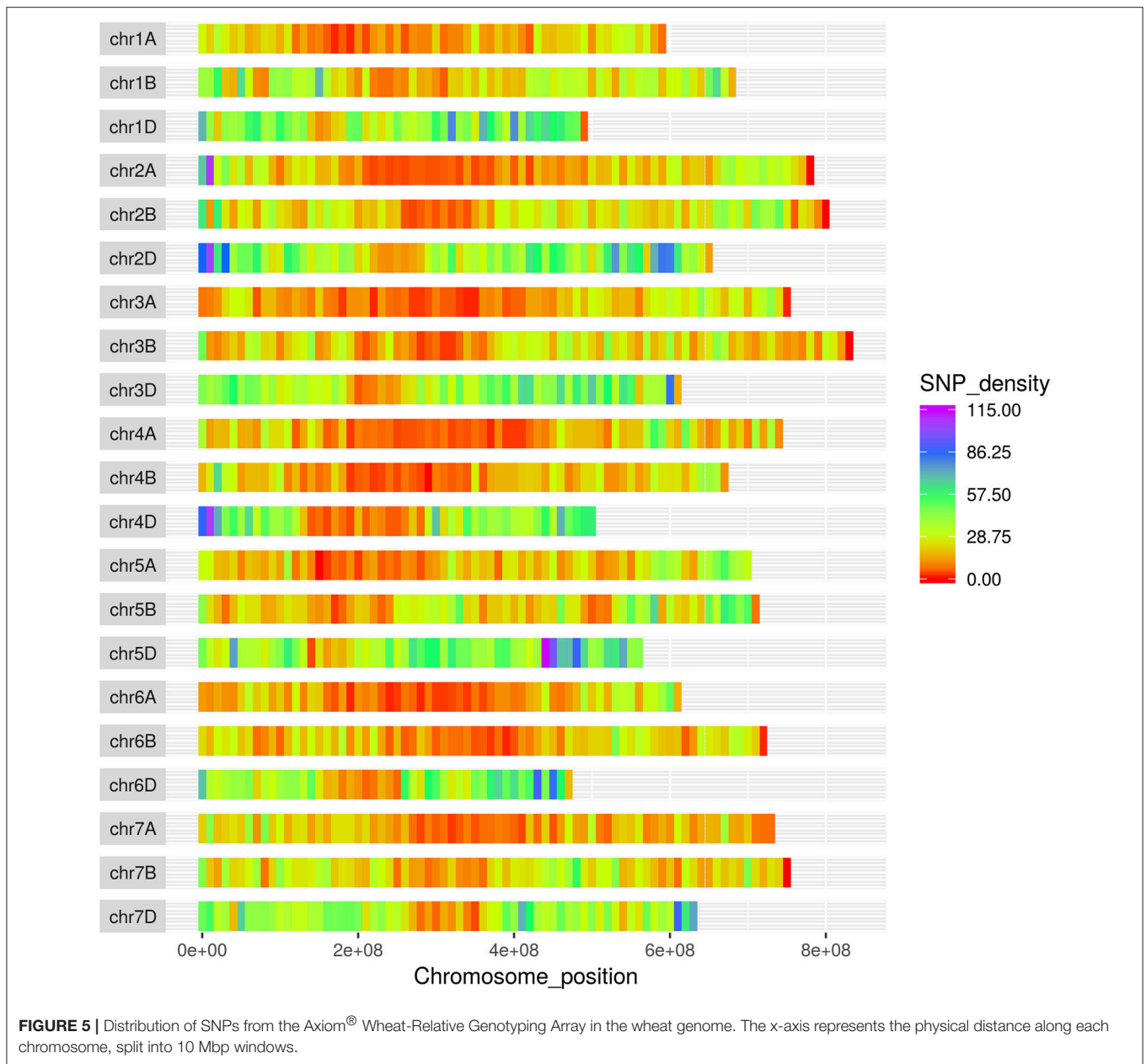
indicating diversity present in this group both within and between species. The diploid species *Ae. comosa* (MM) and *Ae. uniariastata* (NN) form neighboring subclades reflecting genetic similarity between these species. Finally, D-genome containing species were organized in a series of distinct subclades, including hexaploids *Ae. vavilovii* (DDSSMM) and *Ae. juvenalis* (DDMMUU); tetraploids *Ae. crassa* (DDMM), *Ae. cylindrica* (DDMM), and *Ae. ventricosa* (DDNN) and the diploid *Ae. tauschii* (DD). The *Ae. tauschii* subclade splits into the two distinct clusters also seen in the principal component analysis and relates to the geographic origin of the accessions selected (Dvorak et al., 1998b; Mizuno et al., 2010).

Detecting *Aegilops* Introgressions in Wheat

The SNPs on the wheat relative array were assigned a putative chromosomal map location in wheat based upon BLAST alignment of the probe sequence surrounding each SNP against the Chinese Spring v 1.0 genome assembly (Winfield et al., 2016; The International Wheat Genome Sequencing Consortium (IWGSC), 2018). Over 98% of SNPs (36,009) on the wheat relative array were assigned a chromosome location (**Supplementary File 4**). The proportion of chromosome assignments were compared to the larger wheat HD array and calculated separately for SNPs that were polymorphic between the *Aegilops* and wheat on the wheat array (**Figure 4**). All three datasets had the highest proportion of SNPs located in the D-genome, then B genome, with the lowest proportions assigned to the A-genome. This was seen most dramatically in the set which were polymorphic between *Aegilops* and wheat (A genome

21.7%; B genome 31.0%; D genome 45.7%). The distribution of SNPs along chromosomes (**Figure 5**) demonstrated the higher proportion of SNPs in the D genome and also revealed a bias of SNP distribution toward the telomeres, as has been previously reported for exome based SNPs.

The SNP collection on the wheat relative array was employed in two different projects to identify introgressed segments of *Aegilops sp.* in a hexaploid wheat background. These projects took two different approaches to introducing the introgressions and in identifying the markers to track the introgressions. The first study used a specific accession of *Ae. sharonensis* to produce recombinant plants resistant to African Stem Rust. This project used the Axiom wheat HD array to identify and track introgressed regions by comparing the recombinant line and the *Aegilops sharonensis* genotype over 10 SNP windows (Millet et al., 2017). In this study we have shown that this analysis can be repeated using the subset of SNPs on the wheat relative array (**Figure 6**). The second project introgressed *Ae. speltoides* into a hexaploid wheat background with the aim of generating a population where individuals contained specific introgressed segments, which together represent the majority of the *Ae. speltoides* genome (King et al., 2017a). The approach taken to achieve this was to generate a genetic map containing 22,258 polymorphic SNPs (60.6% of total SNPs on the array) and refined this to 544 high quality framework markers. This map was used to inform and identify introgressed segments across individuals from five backcrossed populations, which were then confirmed by genomic *in situ* hybridization (GISH). A high frequency of introgressions were identified and it was possible to track these



through the back-crossing process using markers from the wheat relative array.

DISCUSSION

This study describes the development and application of the first high-throughput genotyping platform specifically designed for screening wheat relative species. The Axiom® Wheat-Relative Genotyping Array contains a framework of over 36,000 SNP markers selected to be useful when screening a diverse range of species with a variety of genome structure and ploidy levels. The platform was used in this study to perform the largest screen of the genus *Aegilops* to date with multiple accessions representing all species in the genus. The array data was demonstrated to be

useful for screening diversity within and between *Aegilops* species and was able to be used for examining relationships within the genus. Furthermore, the data was used to identify and track introgressions from *Aegilops* species in hexaploid wheat using two different approaches.

Over 94% of the SNPs on the array detected a polymorphism between *Aegilops* species. The average number of polymorphic SNPs within a species was 14,231 (36.8%) although the data suggested that when higher numbers of accessions were screened the number of polymorphic SNPs also increased. In the cases of *Ae. tauschii* (22) and *Ae. speltoides* (38) the number of polymorphisms detected doubled compared to random sets of five accessions. This suggests there is a high level of genetic diversity present in these species and careful selection of

Aegilops where genomes in all related species were similar to each other and to that of the parental diploid species, whereas the second genomes were modified compared to the original. Our data suggest the pivotal genome in polyploid UUMM species' *Ae. columnaris*, *Ae. biuncialis*, *Ae. neglecta* and *Ae. geniculata* is the U-genome, while *Ae. triuncialis* is found to be similar to both of its parental species, suggesting the C-genome has not diverged significantly as has been previously observed (Badaeva et al., 2004). However, *Ae. kotschy* and *Ae. peregrina* cluster more closely with S-genome species suggesting that the U-genome in this instance is not pivotal.

Within the UUMM polyploid clade the four related tetraploid species are not organized into discrete sub-clades with and some mixing of species is seen within this clade and, where branching occurs, bootstrap support is not always high. This may suggest a polyphyletic origin of these species with different sources of parental genomes (one accession of *Ae. umbellulata* falls into a subclade of *Ae. neglecta*). Alternatively, this may indicate introgressive hybridization and gene flow between these genetically similar species. With multiple possible sources of the S-genome in *Ae. kotschy* and *Ae. peregrina* the genetic diversity scores were compared and revealed the closest species to both were *Ae. sharonensis* and *Ae. longissima* as has previously been observed (Badaeva et al., 2004). When observed in the tree this result is explained by an *Ae. sharonensis* accession falling into the *Ae. peregrina* clade, a result which requires further investigation. Map locations of the SNPs on the Wheat-Relative array revealed a greater proportion were located on the D-genome than the A- or B-genomes. This contrasts with the usual results obtained when mapping polymorphisms in bread wheat, which are usually lacking in D-genome markers and reflects untapped diversity in D-genome relatives (Allen et al., 2016). The high level of intraspecific diversity present within *Ae. tauschii* and between this species and all others in the genus (except close relatives *Ae. crassa* and *Ae. cylindrica*) was observed in the PCoA analysis and high genetic diversity measures obtained (Figure 2; Table 2).

All *Aegilops* species screened had a high proportion of polymorphic SNPs when compared to bread wheat (average 88.8%) indicating a high level of potentially useful markers for detecting introgressions in a hexaploid wheat background. The specific number of polymorphic SNPs for a cross would depend upon the accession chosen and the wheat cultivar used. Although the array has been demonstrated to be useful in detecting diversity and relationships between a large collection of species, the real power and intended application of this tool is for specific intraspecific crosses where discrete populations are analyzed separately. In two specific case studies we have shown that the array may be applied to detecting and tracking *Aegilops* introgressions by two different strategies; in both cases multiple

introgressions were detected. In the *Aegilops speltoides* project a high number of introgressions were detected and confirmed via GISH and additionally a gametocidal gene was detected (King et al., 2017a). In the *Ae. sharonensis* project recombinant chromosomes were identified and the gene conferring resistance to Ug99 group races was located on chromosome 1B, between 280 and 650 Mbp (Millet et al., 2017). The wheat-relative array has also been used in a third introgression project to screen wheat /*Am. muticum* (syn. *Ae. mutica*) recombinant chromosomes. The array enabled the identification and characterization of genome wide introgressions of various sizes (from large to very small) (King et al., 2017b).

The utility of the wheat relative genotyping array has been demonstrated to be effective in detecting both intra- and inter-specific diversity with insights into the structure and relationships within the genus *Aegilops*. The array data has been shown to be effective in identifying introgressed regions of *Aegilops* in a hexaploid bread wheat background. The application of the array extends beyond *Aegilops* and *Triticum*, having been designed for a wide range of species including those outside of these two genera. By linking with additional resources, such as the annotated genome sequence for wheat (The International Wheat Genome Sequencing Consortium (IWGSC), 2018) and with the development of technologies that enable targeted and precise introgression the potential gain from introducing selected beneficial genes from wild relatives of wheat is an exciting and increasingly feasible prospect.

AUTHOR CONTRIBUTIONS

AP-A analyzed the data and wrote the manuscript. AB, DS, and LM prepared samples. AB and DS performed genotyping. AB performed PCoA. PW performed phylogenetic analysis. IK, JK, KE, and GB developed the genotyping platform. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGMENTS

We are grateful to the Germplasm Resources Unit (GRU) for providing many of the accessions used in this paper and to the Biotechnology and Biological Sciences Research Council, UK for funding this work (award BB/N021061/1, BBS/E/J/000PR9781).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01993/full#supplementary-material>

REFERENCES

- Allen, A. M., Barker, G. L. A., Wilkinson, P., Burrridge, A., Winfield, M., Coghill, J., et al. (2012). Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 11, 279–295. doi: 10.1111/pbi.12009
- Allen, A. M., Winfield, M. O., Burrridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L., et al. (2016). Characterisation of a wheat breeders' array suitable for high throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnol. J.* 15, 390–401. doi: 10.1111/pbi.12635
- Badaeva, E. D., Amosova, A. V., Samatadze, T. E., Zoshchuk, S. A., Shostak, N. G., Chikida, N. N., et al. (2004). Genome differentiation in *aegilops*.

4. evolution of the U-genome cluster. *plant Syst. Evol.* 246, 45–76 doi: 10.1007/s00606-003-0072-4
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossae, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242, 23–26. doi: 10.1016/j.plantsci.2015.08.021
- Burridge, A. J., Wilkinson, P. A., Winfield, M. O., Barker, G. L. A., Allen, A. M., Coghill, J. A., et al. (2017). Conversion of array-based single nucleotide polymorphic markers for use in targeted genotyping by sequencing in hexaploid wheat (*Triticum aestivum*). *Plant Biotechnol. J.* 16, 867–876. doi: 10.1111/pbi.12834
- Charmet, G. (2011). Wheat domestication: lessons for the future. *C. R. Biol.* 334, 212–220. doi: 10.1016/j.crvi.2010.12.013
- Colmer, T. D., Flowers, T. J., and Munns, R. (2006). Use of wild relatives to improve salt tolerance in wheat. *J. Exp. Bot.* 57, 1059–1078. doi: 10.1093/jxb/erj124
- Donal, W. W., and Ogg, A. G. (1991). Biology and control of jointed goatgrass (*Aegilops cylindrica*), a review. *Weed Technol.* 5, 3–17. doi: 10.1017/S0890037X00033170
- Dvorak, J., Luo, M. C., and Yang, Z. L. (1998a). Restriction fragment length poly-morphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. *Genetics* 148, 423–434.
- Dvorak, J., Luo, M. C., and Yang, Z. L. (1998b). The structure of the aegilops tauschii gene pool and the evolution of hexaploid wheat. *Theor. Appl. Genet.* 97, 657–670. doi: 10.1007/s001220050942
- Eig, A. (1929). Monographisch-kritische ubersicht der gattung aegilops. *Reptorium Specierum Novarum Regni Vegetabilis* 55, 1–228.
- Feldman, M. (2001). “The origin of cultivated wheat,” in *The world wheat book. A history of wheat breeding*, eds A. P. Bonjean, and W. J. Angus (Paris: Lavoisier Tech & Doc), 3–56.
- Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., et al. (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8133–8138. doi: 10.1073/pnas.072223799
- Jaubert, H. F., and Spach, E. (1850–1853). *Illustrationes Plantarum Orientium*, Vol. 4. Paris: Apud Roret Bibliopolam.
- Jones, H., Gosman, N., Horsnell, R., Rose, G. A., Everest, L. A., Bentley, A. R., et al. (2013). Strategy for exploiting exotic germplasm using genetic, morphological, and environmental diversity: the *Aegilops tauschii* Coss. example. *Theor. Appl. Genet.* 126, 1793–1808. doi: 10.1007/s00122-013-2093-x
- Kihara, H. (1954). Considerations on the evolution and distribution of aegilops species based on the analyser-method. *Cytologia* 19, 336–357. doi: 10.1508/cytologia.19.336
- Kilian, B., Mammen, K., Millet, E., Sharma, R., and Graner, A. (2011). “Aegilops,” in *Wild Crop Relatives: Genomic and Breeding Resources*, ed C. Kole (New York, NY: Springer-Verlag), 1–76.
- Kimber, G., and Feldman, M. (1987). *Wild Wheat, An Introduction*. Columbia: College of Agriculture, University of Missouri, 142 pp.
- King, J., Grewal, S., Yang, C.-Y., Hubbart, S., Scholefield, D., Ashling, S., et al. (2017b). A step change in the transfer of interspecific variation into wheat from *Amblyopyrum Muticum*. *Plant Biotech. J.* 15, 217–226. doi: 10.1111/pbi.12606
- King, J., Grewal, S., Yang, C. Y., Hubbart Edwards, S., Scholefield, D., Ashling, S., et al. (2017a). Introgression of aegilops speltoides segments in triticum aestivum and the effect of the gametocidal genes. *Ann. Bot.* 121, 229–240. doi: 10.1093/aob/mcx149
- Lee, T. H., Guo, H., Wang, X., Kim, C., and Paterson, A. H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15:163. doi: 10.1186/1471-2164-15-162
- Leighty, G., and Taylor, J. W. (1927). Studies in natural hybridization of wheat. *J. Agric. Res.* 3, 865–887.
- McFadden, E. S., and Sears, E. R. (1946). The origin of *Triticum spelta* and its free-testing hexaploid relatives. *J. Heredity* 37, 107–116. doi: 10.1093/oxfordjournals.jhered.a105594
- Millet, E., Steffenson, B. J., Prins, R., Sela, H., Przewieslik-Allen, A. M., and Pretorius, Z. A. (2017). Genome targeted introgression of resistance to african stem rust from aegilops sharonensis into bread wheat. *Plant Genome* 10, 1–11. doi: 10.3835/plantgenome2017.07.0061
- Mizuno, N., Yamasaki, M., Matsuoka, Y., Kawahara, T., and Takumi, S. (2010). Population structure of wild wheat D-genome progenitor *Aegilops tauschii* Coss: implications for intraspecific lineage diversification and evolution of common wheat. *Mol. Ecol.* 19, 999–1013. doi: 10.1111/j.1365-294X.2010.04537.x
- Molnár, I., Gáspár, L., Sárvári, É., Dulai, S., Hoffmann, B., Molnár-Láng, M., et al. (2004). Physiological and morphological responses to water stress in *Aegilops biuncialis* and *Triticum aestivum* genotypes with differing tolerance to drought. *Funct. Plant Biol.* 31, 1149–1159. doi: 10.1071/FP03143
- Molnár, I., Vrána, J., Burešová, V., Čápal, P., Farkas, A., Darkó, É., et al. (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. *The Plant J.* 88, 452–467. doi: 10.1111/tpj.13266
- Moore, G. (2015). Strategic pre-breeding for wheat improvement. *Nat. Plants* 1:15018. doi: 10.1038/NPLANTS.2015.18
- Ostrowski, M.-F., Prosperi, J.-M., and David, J. (2016). Potential Implications of climate change on aegilops species distribution: sympatry of these crop wild relatives with the major european crop *triticum aestivum* and conservation issues. *PLoS ONE* 11:e0153974. doi: 10.1371/journal.pone.0153974
- Popova, G. (1923). Species of *Aegilops* and their mass hybridization with wheat in Turkestan. *Bull Appl. Bot.* 13, 461–482.
- Schneider, A., Molnár, I., and Molnár-Láng, M. (2008). Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat. *Euphytica* 163, 1–19. doi: 10.1007/s10681-007-9624-y
- The International Wheat Genome Sequencing Consortium (IWGSC) (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- The Plant List (2013). Version 1.1. Available online at: <http://www.theplantlist.org/>
- Uauy, C. (2017). Wheat genomics comes of age. *Curr. Opin. Plant Biol.* 36, 142–148. doi: 10.1016/j.pbi.2017.01.007
- Van Slageren, M. W. (1994). *Wild Wheats; A Monograph of Aegilops L. and Amblyopyrum (Jaub. and Spach) Eig (Poaceae)*. Aleppo: Agricultural University, Wageningen & International Centre for Agricultural Research in dry Areas.
- Wang, J., Zhang, W., Zhao, H., Li, F. R., Wang, Z. G., Ji, J., et al. (2013). Molecular cytogenetic characterization of the *Aegilops biuncialis* karyotype. *Genet. Mol. Res.* 12, 683–692. doi: 10.4238/2013.March.11.16
- Winfield, M., Allen, A. M., Burridge, A. J., Barker, G. L. A., Benbow, H. R., Wilkinson, P. A., et al. (2016). High density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 14, 1195–1206. doi: 10.1111/pbi.12485
- Yamane, K., and Kawahara, T. (2005). Intra- and interspecific phylogenetic relationships among diploid *Triticum-Aegilops* species (*Poaceae*) based in base-pair substitutions, indels and microsatellites in chloroplast noncoding sequences. *Am. J. Bot.* 92, 1887–1898. doi: 10.3732/ajb.92.11.1887
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628
- Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C., and Weir, B. (2012). A high-performance computing toolset for relatedness and principal component analysis of snp data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606
- Zhukovsky, P. M. (1928). A critical-systematical survey of the species of the genus *Aegilops* L. *Bull. Appl. Bot. Genet. Plant Breed.* 18, 417–609.
- Zohary, D., and Feldman, M. (1962). Hybridization between amphiploids and the evolution of polyploids in the wheat (*Aegilops-Triticum*) group. *Evolution* 16, 44–61. doi: 10.1111/j.1558-5646.1962.tb03197.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Przewieslik-Allen, Burridge, Wilkinson, Winfield, Shaw, McAusland, King, King, Edwards and Barker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.