Genetic studies on intraspecific variations of *Aegilops tauschii* to enhance bread wheat diversity

(パンコムギの多様性を拡大するためのタルホコムギの種内変異に関する遺伝学的研究)

Mazin Mahjoob Mohamed Mahjoob

2021

Genetic studies on intraspecific variations of *Aegilops tauschii* to enhance bread wheat diversity

(パンコムギの多様性を拡大するためのタルホコムギの種内変異に関する遺伝学的研究)

A thesis Submitted to the United Graduate School of Agricultural Sciences, Tottori University, in partial fulfillment of requirements for the doctoral degree (Ph.D.) in Agricultural Sciences, Plant Molecular Breeding

By

Mazin Mahjoob Mohamed Mahjoob

Approved by:

Prof., Dr. ……..

Dean, United Graduate School of Agricultural Sciences, Tottori University

Prof., Dr. Hisashi Tsujimoto …….

Chairman of Supervisory Committee

The United Graduate School of Agricultural Sciences

Tottori University

2021

Table and contents

List of Figures

List of Tables

List of abbreviation

Flag leaf length	FLL
Flag leaf width	FLW
Spike length	SPL
Spike width	SPW
Seed number/spike	SN/SP
Spike weight	SPWg
Days to heading	DH
Biomass weight	Bio
Normalized Difference Vegetation Index	NDVI
Canopy temperature	CT
Chlorophyll content	SPAD
Lineage 1	TauL1
Lineage 2	TauL2
Lineage 3	TauL3
Subspecies	Ssp.
Marker traits assoication	MTA
Genome wide assoication study	GWAS

Acknowledgement

I would like to express my deepest gratitude to Prof. H. Tsujimoto for his kind instruction, plentiful academic advice, stimulus suggestion, discussion providing many insights, critical review of manuscripts, and helpful advices for research life.

*I also thank Drs. Yasir Serag Alnor Gor*afi *and Nasrein Mohamed Kamal, whose guidance, encouragement, suggestions, and very constructive criticism have contributed immensely to the evolution of my ideas on this work.*

I am very much thankful to Prof. H. Iwata and Mr. Tai-Shen Chen of Tokyo University, for their kind collaborations, valuable guidance, and meaningful contributions towards this work.

I am grateful to the head and staff of Arid Land Research Center for their support and providing all the facilities and helping me to conduct this research with more advice and suggestions.

I acknowledge with thanks the kind of patronage, loving inspiration, and timely guidance which I have received from Drs. Izzat Sidahmed Ali Tahir, Amani Mohamed Idris and Hala Mostafa Elamein of the Wheat Research Program, Agriculture Research Corporation, Sudan.

I would also like to express my gratitude to the members of Molecular Breeding for their support during the making of this research.

I would like to thank the Ministry of Education, Culture, Sports, Science and Technology of Japan for financial supports.

At last, and not least, I always thank my family for their kind support.

Finally, I would like to acknowledge with gratitude, the support and love of my family; my parents Saida Mohamed and Mahjoob Mohamed, my wife Radwa Ismail Abdelrahman Mohamed, my brothers, Kamal, Serag, Shams Aldeen, Imad and Amar. They all kept me going and this project would not have been possible without them.

Mazin Mahjoob Mohamed Mahjoob

General introduction

To improve crop varieties in the future, genetic diversity is a key factor in plant breeding. The artificial selection practices have resulted in erosion of diversity in elite germplasm and thus constraints wheat evolution. Wheat wild relatives, *Aegilops tauschii*, D-genome progenitor of wheat, are a storehouse for unexploited genetic diversity that can be used for improving wheat for yield, quality, and tolerance to biotic and abiotic stresses.

Aegilops tauschii Coss. is a wild diploid wheat relative with a DD genome and a broad distribution range in central Eurasia, ranging from northern Syria and southeastern Turkey to western China (van Slagern 1994; Matsuoka *et al.* 2008b). Around 8,000 years ago, the D genome of *Ae. tauschii* was introduced into common wheat (*Triticum aestivum* L. BBAADD genome) by interspecific hybridization with domesticated emmer wheat (T. *turgidum* L. subspecies *dicoccum* Schübl. with the BBAA genome) and subsequent amphidiploidization. Crossing cultivated tetraploid wheat with *Ae. tauschii* yields synthetic hexaploid wheat, which reproduces this evolutionary path (Kihara and Lilienfeld 1949; Matsuoka and Nasuda 2004; Takumi *et al.* 2009a; Kajimura *et al.* 2011). The wide phenotypic variance of *Ae. tauschii* can be brought into common wheat using synthetic hexaploid lines as intermediates or directly through synthetic octoploid, making *Ae. tauschii* a valuable genetic resource for wheat breeding (Zohary *et al.* 1969; Mujeeb-Kazi *et al.* 1996; Jones *et al.* 2013; Jafarzadeh *et*

al. 2016). *Ae. tauschii* Coss. subspecies *tauschii* has cylindrical spikelets, while *Ae. tauschii* Coss. subspecies *strangulata* (Eig) Tzvel. has quadrate spikelets (Eig 1929; Hammer 1980). Subspecies *tauschii* has a broad distribution throughout the species' geographical ranges, while subspecies *strangulata* is only found in a small area of the Transcaucasus and the Caspian Sea's southern shore (Eig 1929). The presence of continuous morphological intermediates between the two subspecies is also common (Dudnikov 1998). Subspecies *strangulata* is a monotypic subspecies, while subspecies *tauschii* has three varieties: *typica*, anathera, and meyeri. Variety *meyeri* has rosette form, small spikes, and less spikelets per spikelet, whereas variety *anathera* is judged by the awnless phenotype. However, the genetic basis for intraspecific classification and the phylogenetic relationships between these subspecies and varieties are poorly understood.

The *Ae. tauschii* population is divided into 18 haplogroups and four major haplogroup lineages, HGL7, HGL9, HGL16, and HGL17, based on chloroplast DNA variations (Matsuoka *et al.* 2008b). HGL7, the most common haplogroup lineage, is found in the center of the genetic network and in the species' range, while HGL9 and HGL16 are found in the western and eastern regions, respectively, suggesting that HGL9 and HGL16 diverged from HGL7. Following that, studies using genome-wide marker systems and Bayesian population structure analyses revealed that *Ae. tauschii* accessions are classified into three groups; two main lineages TauL1 and TauL2, and a

minor lineage TauL3 (Mizuno *et al.* 2010; Matsuoka *et al.* 2013; Wang *et al.* 2013a). TauL1 includes all HGL16 and some HGL7 accessions and is found throughout the species' range from the Transcaucasus to Pakistan and Afghanistan, while TauL2 includes all HGL9 and other HGL7 accessions and is restricted to the western region, and TauL3 like HGL17, is found only in Georgia. Although all three lineages coexist in the western part of the species' distribution, there are relatively few genetic intermediates, and these lineages tend to be reproductively isolated (Wang *et al.* 2013a). Together, these genealogical studies of chloroplast and nuclear genomes tell a story of intraspecific diversification (Mizuno *et al.* 2010); HGL7 split from HGL17 (TauL3) and other extinct lineages first, then TauL1 and TauL2 emerged from HGL7, followed by the appearance of HGL9 in TauL1 and HGL16 in TauL2. TauL1 and TauL2 are further subdivided into ‘a' and ‘b' sub-lineages respectively. The eastward migration of the sub-lineage TauL1b is thought to be responsible for TauL1's broad spread (Matsuoka *et al.* 2015). While their ranges overlap, the TauL2a and TauL2b accessions are primarily found in the western and eastern regions of the TauL2 geographical ranges, respectively. Subspecies *tauschii* can be found in all three lineages, but subspecies *strangulata* is only found in TauL2 (Mizuno *et al.* 2010).

Previous isozyme and DNA polymorphism studies agreed that typical wheat speciation existed in the Transcaucasus to southwestern Caspian Iran area (Tsunewaki 1966; Nakai 1978; Dvorak *et al.* 1998). Furthermore, several studies have assumed that

the D genome was donated to common wheat by the subspecies *strangulata* (Nishikawa *et al.* 1980; Jaaska 1981; Lagudah and Halloran 1988; Dvorak *et al.* 2012; Dudnikov 2017). According to the population structure studies revealed by chloroplast DNA, TauL2 and TauL3 were more closely related to the wheat D genome than TauL1 (Matsuoka *et al.* 2013), and most of the accessions genetically similar to the typical wheat D genome were TauL2b subspecies *tauschii*, not subspecies *strangulata* (Wang *et al.* 2013a). TauL2 and the wheat D genome diverged around 0.5 million years ago, considerably earlier than the common wheat speciation (Matsuoka *et al.* 2013; Marcussen *et al.* 2014), suggesting that TauL2 is not the wheat D genome's closest sister. Two indices for reproductive isolation, anther length (an index for outcrossing potential or pre-pollination) and cross ability with a tetraploid wheat cultivar (for post-pollination), were measured in *Ae. tauschii* accessions in a recent study and the expected reproductive barrier was stronger in TauL1 than TauL2, empirically supporting the suggestion that the wheat D genome was derived from the wheat D genome of TauL2 (Matsuoka and Takumi 2017).

In triploid plants, germination failure and hybrid sterility have also been found (Nishikawa 1960; Matsuoka *et al.* 2007, 2013). The polyploidy of emmer and bread wheats, as well as the large and repetitive genomes of *Triticum* and *Aegilops* species, have hindered the establishment of their reference genomes. *Ae. tauschii*, tetraploid, and bread wheat genome sizes were estimated to be 4.36 Gb, 12 Gb, and 17 Gb,

respectively (Jia *et al.* 2013; Avni *et al.* 2017). The intergenic regions are densely packed with transposable elements that are usually longer than Illumina reads. Furthermore, the coding regions of the three homoeologous genomes of common wheat share 97% identity (Krasileva *et al.* 2013). Since 2012, three studies on common wheat genome sequencing have been published, with total assemblies of 5.42 Gb (Brenchley *et al.* 2012), 10.2 Gb (IWGSC 2014), and 12.7 Gb (Clavijo *et al.* 2017), all of which were highly fragmented. (Ling *et al.* 2013; Jia *et al.* 2013) released draft genomes of *Ae. tauschii* and *Triticum urartu*, the diploid progenitor of the A genome, with 4.23-Gb and 3.92-Gb assemblies, respectively, but many scaffolds and contigs were not anchored to chromosomes. A physical map of *Ae. tauschii*, covering 4 Gb of the genome, was also released, though access to the sequences was limited (Luo *et al.* 2013). However, recent technological advances in sequencing platforms and assembly algorithms, such as the Illumina HiSeq X Ten, PacBio RS II and Sequel system, 10x Genomics Chromium System, Bionano optical genome mapping, and NRGene DeNovoMAGIC, have made it possible to develop reference-level genome sequences of wheat and its relatives. In 2017, a near-complete assembly of common wheat was reported by (Zimin *et al.* 2017a), with a total assembly length of 15.3 Gb and a N50 of 232,659 bases. The pre-publication data of IWGSC are now available on https://wheaturgi.versailles.inra.fr/Seq-Repository/Assemblies. For wild emmer wheat, a 10.1-Gb assembly has been reported (Avni *et al.* 2017). In addition, three genome

assemblies for *Ae. tauschii* have been announced independently (Luo *et al.* 2017; Zhao *et al.* 2017; Zimin *et al.* 2017b). These virtually complete reference genome sequences are allowing genetic and genomic studies of wheat and its relatives on an unprecedented scale. *Ae. tauschii* natural population exhibits wide variation in morphological and physiological traits. Early flowering accessions are more common in the eastern region (Matsuoka *et al.* 2008b, 2015), and genetic variations in early flowering loci such as *Vrn-D1*, *Ppd-D1*, and *VRN2* have been identified (Takumi *et al.* 2011; Huang *et al.* 2012; Kippes *et al.* 2016; Koyama *et al.* 2018). Spikelet morphological trait has regional clines; spikelets appear to be small in eastern and southern range (Takumi *et al.* 2009b; Matsuoka *et al.* 2009). Eastern accessions also have high seed productivity (Matsuoka *et al.* 2015) and salt resistance during germination and seedling development (Saisho *et al.* 2016), suggesting that these characteristics, as well as the early flowering phenotype, which is strongly acquired in subspecies TauL1b, were the driving force behind the species' eastward expansion. Thus, the morphological variations might underlie the genealogical diversification, and detailed analyses of their genetic bases would provide us new insights on the evolutionary path of *Ae. tauschii*. Notably, the TauL1b accessions of these agronomically significant traits may seem not to have been involved in the common wheat speciation case (Matsuoka *et al.* 2013; Wang *et al.* 2013b), implying that the TauL1 phenotypic and genetic variation would be useful for wheat breeding.

Because of the significant importance of *Ae. tauschii* as storehouse of genetic diversity necessary for wheat improvement, this dissertation aimed to (1) clarify the phylogeny of *Ae. tauschii* and identify morpho-physiological traits that discriminate between the two main lineages (TauL1 and TauL2), ssp. *tauschii* belonging to TauL1 or TauL2, and the two subspecies (ssp. *tauschii* and ssp. *strangulata*) in Chapter 1, and, (2) to identifying markers or genes associated with morpho-physiological traits in *Ae. tauschii*, and at understanding the difference in genetic diversity between the two main lineages in Chapter 2. Overall objective of this dissertation is to study each lineage independently and study the difference between in morpho-physiological variation.

Chapter 1

Traits to Differentiate Lineages and Subspecies of *Aegilops tauschii*, the D Genome Progenitor Species of Bread Wheat

1.1. Abstract

Aegilops tauschii Coss., the D genome donor of hexaploid wheat (*Triticum aestivum* L.), is the most promising resource used to broaden the genetic diversity of wheat. Taxonomical studies have classified *Ae. tauschii* into two subspecies, ssp. *tauschii* and ssp. *strangulata*. However, molecular analysis revealed three distantly related lineages, TauL1, TauL2, and TauL3. TauL1 and TauL3 includes the only ssp. *tauschii*, whereas TauL2 includes both subspecies. This study aimed to clarify the phylogeny of *Ae. tauschii* and to find the traits that can differentiate between TauL1, TauL2 and TauL3, or between ssp. *tauschii* and ssp. *strangulata*. I studied the genetic and morpho-physiological diversity in 293 accessions of *Ae. tauschii*, covering the entire range of the species. A total of 5,880 high-quality SNPs derived from DArTseq were used for phylogenetic cluster analyses. As a result, I observed wide morpho-physiological variation in each lineage and subspecies. Despite this variation, no key traits can discriminate lineages or subspecies though some traits were significantly different. Of 124 accessions previously lacking the passport data, 66 were allocated to TauL1, 57 to TauL2, and one to TauL3.

1.2. Introduction

Wild relatives attract increasing attention because they can provide characters related to adaptation (Hu *et al.* 2012). The genus *Aegilops* L. (Poaceae) has been intensively studied because of its close relationship with cultivated wheats. The phylogenetic relationship between genera *Aegilops* and *Triticum* L. is widely reported (Kimber and Zhao 1983; Kellogg *et al.* 1996; Petersen *et al.* 2006; Alnaddaf *et al.* 2012)**,** and on a world scale, the genus *Aegilops* includes 23 wild annual species, of which 11 are diploids and 12 are allopolyploids (Hammer 1980; Kilian *et al.* 2011). The revision of the genus *Aegilops* with regards to its genome and taxonomy results in a total of 27 specific and intraspecific taxa (Van Slageren 1994). *Aegilops tauschii* Coss. (syn. *Ae. squarrosa* auct. non L.), a wild diploid self-pollinating species (2n = 2x = 14, DD), is the D genome donor of the hexaploid bread wheat (*Triticum aestivum* L.; 2n = 6x = 42, AABBDD). This wild species is found mainly at the edges of wheat fields in eastern Turkey, Iraq, Iran, Pakistan, India, China, Afghanistan, Central Asia, Transcaucasia (South Caucasus), and the Caucasus region (Feldman M (2001)). About 8,000 to 10,000 years ago, the ancestor of the current bread wheat appeared as a result of natural hybridization between cultivated wheat (*Triticum turgidum* L., 2n = 4x = 28, AABB) and *Ae. tauschii* (Feldman M (2001); Kihara 1944; McFadden and Sears 1944). Inside this last species, two subspecies were first described by Eig (1929) (Eig 1929) as *Ae. squarrosa* ssp. *eusquarrosa* and ssp. *strangulata,* and their nomenclature was revised

by Hammer (1980) as *Ae. tauschii* ssp. *tauschii* and ssp. *strangulata*. *Ae. tauschii* is genetically and morphologically diverse (Eig 1929), and the ssp. *tauschii* has elongated cylindrical spikelets, whereas ssp. *strangulata* has quadrate spikelets and empty glumes (Eig 1929; Hammer 1980). The ssp. *tauschii* has a wide distribution throughout the species range, whereas ssp. *strangulata* is limited to the south-eastern Caspian coastal region and the Caucasus (Matsuoka *et al.* 2009) Some of the molecular studies supported the subspecies division (Gill *et al.* 1991; Dvorak *et al.* 1998b; Pestsova *et al.* 2000), whereas others did not (Lelley *et al.* 2000; Saeidi *et al.* 2006).

The genetic diversity in *Ae. tauschii* has been studied at the molecular level by using isozymes (Dudnikov and Kawahara 2006), random amplified polymorphic DNA (RAPD) (Okuno *et al.* 1998), chloroplast DNA(Matsuoka *et al.* 2005, 2009) amplified fragment length polymorphisms (AFLPs) (Mizuno *et al.* 2010)[23], simple sequence repeats (SSRs) (Naghavi and Mardi 2010), and DArT-array markers (Sohail *et al.* 2012). Most of these studies classified *Ae. tauschii* into three lineages: TauL1 including only ssp. *tauschii*, TauL2 including both ssp. *tauschii* and ssp. *strangulata*, and TauL3 with intermediate forms. However, Arora et al. (Arora *et al.* 2017, 2019b) reported that TauL1 is mainly associated with ssp. *tauschii* and TauL2 with ssp. *strangulata*. Therefore, this study aims to clarify the phylogeny of *Ae. tauschii* and to identify morpho-physiological traits that discriminate between the two main lineages (TauL1

and TauL2), ssp. *tauschii* belonging to TauL1 or TauL2, and the two subspecies (ssp. *tauschii* and ssp. *strangulata*).

1.3. Materials and Methods

1.3.1. Plant Materials

I used 293 *Ae. tauschii* accessions collected from the entire range of the natural distribution of this species (Table 1-1, Fig. 1-1). Of these accessions, 201 have full passport data, including geographical coordinates, lineages and subspecies classification (Matsuoka *et al.* 2009) (Fig. 1). Five of the 201 accessions (AT 55, AT 60, AT 76, PI 499262, and PI 508262) represent adventive populations in the Shaanxi and Henan provinces of China. Among the 201 accessions, 132 belong to TauL1, 64 to TauL2, and 5 to TauL3 (Matsuoka *et al.* 2009). Based on *sensu stricto* criteria for subspecies classification, only accessions with distinctly moniliform spikes were classified to *Ae. tauschii* ssp. *strangulata*. In contrast, accessions having mildly moniliform and cylindrical spikes were classified to *Ae. tauschii* ssp. *tauschii* (Matsuoka *et al.* 2009). Of 293 accessions used in this study, 169 were previously studied by Matsuoka et al. (2009) (Matsuoka *et al.* 2009)who classified 110, 55, and 4 to TauL1, TauL2, and TauL3, respectively.

1.3.2. Genomic Analysis and Statistical Analysis of Molecular Data

Genomic DNA was extracted using the CTAB method (Saghai-Maroof *et al.* 1984). The DNA samples (30 µl; 50–100 ng μl^{-1}) were sent to Diversity Arrays Technology

Pty. Ltd, Australia (http://www.diversityarrays.com) for a whole-genome scan using the DArTseq platform. Sequencing-based DArT genotyping applies two complexity-reduction methods optimized for several plant species i.e., *Pst*I/*Hpa*II and *Pst*I/*Hha*I were used to select a subset of the corresponding fragments (Sansaloni *et al.* 2011). At the DArT facility, the DArT soft marker extraction pipeline was used to filter and identify the informative markers. We performed the hierarchical clustering analysis in the statistical software R with the pvclust package (Suzuki and Shimodaira 2006). The DArTseq SNPs data of 5,880 markers without any missing data for 293 accessions of *Ae. tauschii* from 16 countries (some accessions are from unknown origin) were used for the analysis. Pvclust package computes the AU (approximately unbiased) *P*-value and BP (bootstrap probability) value via multiscale bootstrap resampling. These values can show how strong the clustering result is supported by the data. The dendrogram was generated by using the Euclidean distance matrix and complete method.

1.3.3. Morpho-Physiological Evaluation

The morphological and physiological traits of all the accessions were measured at the research field of the Arid Land Research Center, Tottori University (Tottori, Japan; 35°32′N, 134°13′E) during the winter and spring seasons of 2016/17 and 2017/18 by using an augmented complete block design with three randomly selected accessions as checks (GE12-14-O-1, GE12-28-O-2, and KU-20-2), and five plants were grown per accession. To estimate the phenotypic variation, we measured two leaf parameters (flag

leaf length, FLL; flag leaf width, FLW), four spike parameters (spike length, SPL; spike width, SPW; seed number per spike, SN/SP; spike weight, SPWg), days to heading (DH), biomass weight (Bio), and three physiological traits (Normalized Difference Vegetative Index, NDVI; canopy temperature, CT; and chlorophyll content, SPAD). To measure SPWg, I covered the spikes with a transparent envelope before physiological maturity to avoid shattering. The measurement methods are summarized in Table 1-2.

1.3.4. Statistical Analysis of Morpho-Physiological Data

Analyses of the phenotypic data, including mean, standard deviation, range distribution, and analysis of variance (*F* and *P*-values in one-way ANOVA) for the morpho-physiological variations were calculated using Plant Breeding Tools (PBTools) version 1.4 (International Rice Research Institute, http://bbi.irri.org/products). Because of significant genotype × season interaction, best linear unbiased predictions (BLUPs) were estimated for each trait.

1.4. Results

1.4.1. Phylogenetical Allocation of Uncertain Accessions by Molecular Markers

Following (Matsuoka *et al*. 2009), I carefully observed the key morphological traits of the 124 accessions that lacked taxonomical information and identified 7 accessions as ssp. *strangulata* and the remaining 117 as ssp. *tauschii*. Among the seven accessions identified as ssp. *strangulata*, AE 525 was collected from Iran, AE 692 from Uzbekistan, and AE 426, AE 428, AE 429, AE 430 and AE 434 from unknown regions.

To know the lineages (TauL1, TauL2 or TauL3) of all 124 accessions, I conducted cluster analysis using 5,880 DArTseq markers. As a result, 66, 57 and 1 were clustered in TauL1, TauL2 and TauL3, respectively (Fig. 1-2, Fig. 1-S1). All the accessions in TauL1 were ssp. *tauschii*, whereas in TauL2, 50 were ssp. *tauschii* and 7 ssp. *strangulata*. The accessions in the TauL3 were ssp. *tauschii*. These findings supported previous results that ssp. strangulata is present only in TauL2.

Previously, Matsuoka *et al.* (2009) classified *Ae. tauschii* accessions into TauL1, TauL2 and TauL3 based on the chloroplast DNA. To confirm their result, I analyzed the 169 accessions used in Matsuoka *et al.* (2009) using DArTseq markers. Most of the accessions were clustered as expected with 5 exceptions: KU-2109 and KU-2158 were in TauL1, whereas PI 486274, IG 127015, and IG 120735 were in TauL2.

From these studies, I found that all 293 accessions of *Ae. tauschii* were classified as 175 TauL1, 113 TauL2, and 5 TauL3. In TauL2, 15 accessions were ssp. *strangulata* and others including accessions in TauL1 and TauL3 were ssp. *tauschii*. The TauL1 cluster contained accessions from Syria, Turkey, Georgia, Armenia, Azerbaijan, Dagestan, Iran, Turkmenistan, Afghanistan, Pakistan, Tajikistan, Uzbekistan, Kyrgyzstan, Kazakhstan, China, and unknown countries. The TauL2 cluster contained accessions from Syria, Turkey, Georgia, Armenia, Azerbaijan, Dagestan, Iran, Turkmenistan, Uzbekistan, and unknown countries (Table 1-1, Fig. 1-2, Fig. 1-S1). The

ssp. *strangulata* accessions were clustered in one clade in TauL2, and most of the accessions were from Iran.

1.4.2. Morpho-Physiological Differences between TauL1 and TauL2

A large variation was observed for all the morpho-physiological traits in TauL1 and TauL2 (Table 1-3). Statistical analyses showed a significant difference between these two lineages in SPW, SPWg, DH, and Bio. The means in these traits were larger in TauL2 than in TauL1, indicating that the accessions in TauL2 tend to be higher than TauL1. On the other hand, the means of the physiological traits (NDVI, CT, and SPAD), and leaf traits (FLL and FLW) were not significantly different between them. The ranges of these traits overlapped between the two lineages, and thus I cannot discriminate the two groups with these traits (Table 1-3).

1.4.3. Morpho-Physiological Variation between ssp. **tauschii** *Belonging to TauL1 and TauL2*

I designated ssp. *tauschii* in TauL1 and TauL2 as 'TauL1T' and 'TauL2T', respectively, and compared accessions in these groups. A large variation was observed for all the morpho-physiological traits in TauL1T and TauL2T (Table 1-4). Statistical analyses showed significant differences between the two groups in FLL, DH, and Bio. The mean of FLL was higher in TauL1T, whereas those of DH and Bio were higher in TauL2T. On the other hand, the means of the physiological traits (NDVI, CT, and SPAD), and spike traits (SPL, SPW, SN/SP and SPWg) were not significantly different

between them. The ranges of these traits overlapped between TauL1T and TauL2T, and thus we cannot discriminate the two groups with these traits (Table 1-4).

1.4.4. Morpho-Physiological Variation between ssp. tauschii *and ssp.* strangulata

A large variation was observed for all the morpho-physiological traits in ssp. *tauschii* and ssp. *strangulata* (Table 1-5). Statistical analyses showed significant difference between these two subspecies in SPL, SN/SP, SPWg, and DH. The means of SPL and SN/SP were higher in ssp. *tauschii* than in ssp. *strangulata*, whereas those of SPWg and DH were higher in ssp. *strangulata* than in ssp. *tauschii*. On the other hand, the means of the leaf traits (FLL and FLW), SPW, and physiological traits (NDVI, CT, and SPAD) were not significantly different between them. The ranges of these traits overlapped between the two subspecies (Table 1-5).

1.4.5. Morpho-physiological traits of accessions in TauL3

In this study, only five accessions (AE 454, AE 929, AE 929a, KU-2829A and KU-2832) belong to TauL3. All the accessions originated from Georgia and showed a similar plant morphology to ssp. *tauschii* with an intermediate spike shape between TauL1 and TauL2. Genomic analysis revealed that these accessions are clearly differentiated from both TauL1 and TauL2.

1.5. Discussion

1.5.1. Geographical Clines of Morphological Variation in Subspecies and Lineage Classification

The main putative area of origin of *Ae. tauschii* is the Transcaucasus, from which it has spread to the east and south (Feldman M (2001)) (Fig. 1-1). While ssp. *tauschii* has cylindrical spike forms and ssp. *strangulata* moniliform spike forms, some *Ae. tauschii* accessions have mildly moniliform spike forms (TauL3) which suggest a hybrid origin. Overall, spikelet morphology is the main trait not only for discriminating the two subspecies but also for intraspecific diversification in *Ae. tauschii*, even though the genetic basis of spikelet morphology divergence has not yet been studied. Nishijima *et al*. (2017) divided *Ae. tauschii* into two main lineages TauL1 and TauL2, and a minor lineage (TauL3) by Bayesian population structure analysis with genome-wide marker genotyping. Using DArTseq genotyping of a large number of accessions, I confirmed their results (Fig. 1-2). The TauL1 accessions are spread from the western geographical range (Transcaucasus, northern regions of Iran) to the eastern geographical range (Pakistan and Afghanistan), whereas TauL2 is limited only to the western range, and ssp. *strangulata* is included only in TauL2.

This result is consistent with Mizuno *et al*. (2010b) using AFLPs. Thus, the differentiation of the ssp. *strangulata* is believed to have occurred in TauL2. Also, I found that the most probable origin of ssp. *strangulata* is Iran and that this subspecies clusters in one clade within TauL2 (Fig. 1-2). This finding strongly indicates that speciation had occurred in the ssp. *tauschii* included in TauL2, resulting in appearance of ssp. *strangulata*-type spike morphology. The D genome of ssp. *strangulata* is

involved in the D genome of bread wheat. This was revealed by sequencing (Ling *et al.* 2018), single nucleotide polymorphisms (Wang *et al.* 2013c), variation in the *AP2* homoeologs, the genes underlying lodicule development (Ning *et al.* 2013), SSR markers (Naghavi *et al.* 2009), NADP-dependent aromatic alcohol dehydrogenase (Jaaska 1978), and aspartate aminotransferase and alcohol dehydrogenase isoenzymes (Jaaska 1981). Overall, using the DArTseq genotyping platform, I have allocated 124 accessions with no previous lineage description into TauL1, TauL2 or TauL3. Also, based on this data, I have reclassified 5 accessions: 2 accessions from Iran (KU-2109 and KU-2158) formerly classified in TauL2 by chloroplast DNA (Matsuoka *et al.* 2009) were now placed in TauL1, and 3 accessions (PI 486274 from Turkey, IG 127015 from Armenia, and IG 120735 from Turkmenistan) formerly classified in TauL1 were now placed in TauL2. The inconsistency of the nucleus and cytoplasmic genomes may be attributable to the cytoplasmic substitution origin by hybrids between the two lineages and the backcrossing in the evolution of these accessions. Furthermore, previous studies reported that accessions in TauL2 were distributed in the regions near the Caspian Sea. However, here I found that five accessions (AE 192, AE 213, AE 250, CGN10733 and IG 120735) which originated from Turkmenistan and AE 692 from Uzbekistan were clustered in TauL2 (Table 1-1). These accessions may have been transferred to the regions naturally or by human activity.

1.5.2. Potential for Adaptive Convergence in Ae. tauschii *Evolution*

Molecular evolutionary studies have explained the origin of crops more clearly than before (Londo *et al.* 2006; Doebley *et al.* 2006; Purugganan and Fuller 2009), especially for the main crops that were domesticated without ploidy modification. Phylogeographic analyses based on nuclear and chloroplast DNA sequences have shown multiple evolutionary origins of cultivated rice in East Asia (Londo *et al.* 2006) and barley in the Fertile Crescent and Central Asia (Saisho and Purugganan 2007; Morrell and Clegg 2007), whereas phylogenetic analysis based on multilocus microsatellite genotyping has shown a single domestication event for maize ca. 9,000 years ago (Matsuoka *et al.* 2002). One of the fundamental problems in understanding the evolution of *Ae. tauschii* is the relationship between the different lineages and subspecies. In the current study, although some traits examined differed significantly between the lineages and subspecies, the range of the diversity was overlapped (Tables 1-3 -1-5). The phenotypes convergence may have originated through either divergent genetic solutions (Wittkopp *et al.* 2003; Pascoal *et al.* 2014) or the same pathways, genes, or even nucleotide positions in independent lineages (Zhen *et al.* 2012; Martin and Orgogozo 2013). Convergence at the genetic level can in turn result from (i) mutations arising independently in separate populations or organisms (parallel genetic evolution); (ii) evolution of a polymorphic allele in a common ancestral population or species (trans-specific polymorphism); and (iii) evolution of an allele introduced by hybridization (introgression) from one population to another (e.g., TauL1 and TauL2).

Another possibility that can explain the phenotypic similarities between the different *Ae. tauschii* lineages is the occurrence of genetic differentiation after the geographical isolation under similar environmental condition without morphological or physiological differentiation. Local standing genetic diversity combined with spatial population structure restricting dispersal in an ecologically patchy area promotes rapid convergence (Ralph and Coop 2015).

4.3. Implications of Ae. tauschii *Diversity in Wheat Breeding*

Among the species in genus *Aegilops*, only *Ae. tauschii* can be used efficiently for wheat improvement owing to the mostly regular pairing of its chromosomes with the D genome chromosomes of bread wheat (Kishii 2019). It is believed that *Ae. tauschii* is an excellent source to widen the narrow genetic base of bread wheat. Currently, with the new advances in plant science and the rapid development of sequencing and genome-editing tools, identification, and characterization of genes of interest in wheat are in progress and can be expected to become easier and more straightforward in the coming decades. Once the gene in question is identified and characterized, it is easy to transfer and utilize the gene in breeding programs. This will pave the way to utilize the genes from *Ae. tauschii* as it will help to overcome the limitations related to the irregular chromosome pairing.

Tables and Figures of chapter 1

Table 1-1. *Aegilops tauschii* accessions used in this study

Origin	TauL1					TauL2					TauL3	
Syria	*AE 1069*	IG 47259				IG 46623						
Turkey	KU-2131	KU-2132	KU-2133	KU-2136	KU-2137	PI 486267	**PI 486274**					
	KU-2138	KU-2140	KU-2141	PI 486270	PI 486277							
	PI 554319											
Georgia	*AE 254*	*AE 461*	*GE12-28-O-2*	*KU-20-2*	KU-2826	AE 1037	*GE12-14-O-1*	KU-2827	KU-2835B		AE 929	AE 454
	KU-2828	KU-2834									KU-2829A	KU-2832
											AE 929a	
Armenia	*AE 245*	*AE 253*	*AE 476*	*AE 721*	CGN 10734	*AE 229*	*AE 231*	*AE 940*	*AE 941*	IG 126991		
	IG 126273	IG 126280	IG 126293	IG 126353	IG 48748	**IG 127015**	KU-2811					
	IG 48758	KU-2809	KU-2810	KU-2814	KU-2816							
	KU-2821	KU-2822A	KU-2823	KU-2824								
Azerbaijan	*AE 143*	*AE 220*	*AE 251*	*AE 723*	*AE 724*	*AE 144*	*AE 191*	*AE 194*	*AE 195*	*AE 197*		
	AE 725	*AE 1055*	IG 47196			*AE 198*	*AE 199*	*AE 200*	*AE 202*	*AE 203*		
						AE 204	*AE 205*	*AE 206*	*AE 207*	*AE 210*		
						AE 211	*AE 216*	*AE 217*	*AE 218*	*AE 219*		
						AE 221	*AE 222*	*AE 223*	*AE 224*	*AE 226*		
						AE 230	*AE 255*	*AE 260*	*AE 261*	*AE 262*		
						AE 263	*AE 264*	*AE 267*	*AE 270*	*AE 272*		
						AE 273	*AK 228*	IG 47182	IG 47186	IG 47188		
						IG 47193	IG 47199	IG 47202	IG 47203	KU-2801		
						KU-2806						
Dagestan	*AE 234*					*AE 498*	IG 120863	IG 120866	IG 48274	KU-20-1		
Iran	*AE 183*	*AE 184*	*AE 541*	IG 49095	KU-2082	*AE 525**	*AE 526*	KU-20-8	KU-20-9*	KU-20-10		
	KU-2109	KU-2113	KU-2115	KU-2116	KU-2120	KU-2069	KU-2075*	KU-2079*	KU-2080*	KU-2083		
	KU-2121	KU-2142	KU-2143	KU-2144	KU-2148	KU-2086	KU-2088*	KU-2090*	KU-2092*	KU-2093*		
	KU-2152	KU-2153	KU-2154	KU-2157	**KU-2158**	KU-2096	KU-2097	KU-2098	KU-2100	KU-2101		
						KU-2102	KU-2103	KU-2104	KU-2105	KU-2106		
						KU-2110	KU-2111	KU-2112	KU-2118	KU-2124		
						KU-2126	KU-2155	KU-2156	KU-2159	KU-2160		
Turkmenistan	*AE 141*	*AE 146*	*AE 242*	*AE 248*	*AE 249*	*AE 192*	*AE 213*	*AE 250*	*CGN 10733*	**IG 120735**		
	AE 291	*AE 398*	*AE 472*	*AE 473*	*AE 499*							
	AE 637	*AE 964*	IG 126387	IG 126489	IG 48508							
	IG 48518											
Afghanistan	*AE 193*	*AE 275*	*AE 276*	*AE 277*	*AE 279*							
	AE 280	*AE 281*	*AE 1087*	KU-2010	KU-2012							
	KU-2016	KU-2018	KU-2022	KU-2025	KU-2027							
	KU-2035	KU-2039	KU-2042	KU-2043	KU-2044							
	KU-2050	KU-2051	KU-2056	KU-2059	KU-2061							
	KU-2063	KU-2066	*KU-2616*	KU-2617	KU-2619							
	KU-2621	KU-2624	KU-2630	KU-2632	KU-2633							
	KU-2635	KU-2636	KU-2638	KU-2639	PI 476874							
Pakistan	CGN 10767	CGN 10768	CGN 10769	CGN 10771	IG 108561							
	IG 46663	IG 46666	KU-2003	KU-2006	KU-2008							
Tajikistan	*AE 189*	*AE 233*	*AE 647*	*AE 817*	*AE 858*							
	AE 955	*AE 956*	AE 1038	*AE 1039*	*AE 1040*							
	IG 48554	IG 48559	IG 48564									
Uzbekistan	*AE 3*	*AE 239*	*AE 469*	*AE 560*	IG 120736	*AE 692**						
	IG 123910	IG 48539	IG 48565	IG 48567								
Kyrgyzstan	*AE 256*	*AE 257*	*AE 1180*	IG 131606								
Kazakhstan	AE 1090											
China	AT 55	AT 60	AT 76	PI 499262	PI 508262							
Unknown location site	*AE 26*	*AE 32*	*AE 67*	*AE 147*	*AE 150*	*AE 426**	*AE 428**	*AE 429**	*AE 430**	*AE 431*		
	AE 422	*AE 427*	*AE 433*	*AE 594*		*AE 432*	*AE 434**					

Roman accessions are known from Matsuoka et al. (2009) (Matsuoka *et al.* 2009). Italic accessions are classified in this study into TauL1, TauL2 or TauL3. Bold accessions have different taxonomy based on chloroplast DNA. AE accessions were received from the Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Germany; AT accessions from the Faculty of Agriculture, Okayama University, Japan; CGN accessions from the Instituut Voor Planten Veredeling, Landbouwhoge

School, Wageningen, the Netherlands; IG accessions from the International Center for Agricultural Research in the Dry Areas (ICARDA), Syria; KU accessions from the Germplasm Institute, Faculty of Agriculture, Kyoto University, Japan; and PI accessions from the US Department of Agriculture. * Ssp. *strangulata*. The subspecies classified morphologically following Matsouka et al. (2009) (Matsuoka *et al.* 2009) and confirmed by cluster analysis in this study (Supplementary Fig. 1-1).

Table 1-2. Phenotypic traits analyzed.

Trait	Abbreviation (unit)	Measurement/Definition
Flag leaf length	FLL (cm)	Measured from three tillers per accession.
Flag leaf width	FLW (mm)	Measured from three tillers per accession.
Spike length	SPL (cm)	Measured from the middle five spikes after maturity stage.
Spike width	SPW (cm)	Measured from the middle of five spikes after maturity stage.
Seed number/spike	SN/SP	Counted from five spikes at harvesting.
Spike weight	SPWg (g)	Weighed from five spikes (one per tiller) using a sensitive scale.
Days to heading	DH	Recorded when the whole spike above the flag leaf fully emerged on the earliest tiller in each plant of each accession.
Biomass weight	Bio (g)	Weighed after harvesting and drying of five plants in a glasshouse.
Normalized Difference Vegetation Index	NDVI	A vegetative index that compares reflectance in the red and near-infrared regions. Measured during flowering using a handheld optical sensor unit (Green Seeker), NTech Industries, Inc., Ukiah, CA, USA.
Canopy temperature	CT (°C)	Measured during flowering using an infrared thermometer AD-5611A.
Chlorophyll content	SPAD	Measured at the flowering stage from the middle of the flag leaf of three tillers using a Minolta brand chlorophyll meter (Model SPAD-502; Spectrum Technologies Inc., Plainfield, IL, USA).

Table 1-3. Morpho-physiological variation in two *Aegilops tauschii* lineages, TauL1 (175 accessions) and TauL2 (113 accessions).

Trait	TauL1				TauL2				*P*-value (TauL1 versus TauL2)
	Min	Max	Mean	STD	Min	Max	Mean	STD	
FLL	5.35	20.65	13.74	2.48	5.77	20.32	12.96	2.84	0.052
FLW	4.80	11.00	8.10	1.20	4.20	10.90	7.80	1.10	0.145
SPL	9.08	17.55	12.61	1.50	8.80	17.27	12.03	1.55	0.325
SPW	0.40	0.71	0.53	0.06	0.40	0.75	0.58	0.07	0.011
SN/SP	15.82	29.67	22.00	2.32	15.42	29.93	19.51	2.05	0.081
SPWg	0.35	0.67	0.50	0.06	0.34	0.71	0.54	0.07	0.005
DH	150.78	184.03	169.19	5.78	159.77	191.45	174.39	4.04	0.000
Bio	60.53	189.78	99.24	23.61	73.90	227.09	134.50	37.11	0.000
NDVI	0.60	0.63	0.62	0.01	0.60	0.64	0.62	0.01	0.389
CT	15.11	25.14	18.34	1.91	14.49	24.50	17.91	1.84	0.303
SPAD	40.92	45.37	43.50	0.73	42.06	45.46	43.69	0.71	0.413

Table 1-4. Morpho-physiological variation in ssp. *tauschii* in TauL1 (TauL1T, 175 accessions) and TauL2 (TauL2T, 98 accessions).

Trait	TauL1T				TauL2T				*P*-value (TauL1T versus TauL2T)
	Min	Max	Mean	STD	Min	Max	Mean	STD	
FLL	5.35	20.65	13.74	2.48	5.77	20.32	12.78	2.89	0.040
FLW	4.80	11.00	8.10	1.20	4.20	1.90	7.80	1.20	0.239
SPL	9.08	17.55	12.61	1.50	8.80	16.75	12.19	1.41	0.271
SPW	0.40	0.71	0.53	0.06	0.40	0.72	0.57	0.07	0.145
SN/SP	15.82	29.67	22.00	2.32	16.13	29.93	19.72	2.07	0.106
SPWg	0.35	0.67	0.50	0.06	0.37	0.67	0.53	0.06	0.091
DH	150.78	184.03	169.19	5.78	159.77	191.45	174.46	4.28	0.001
Bio	60.53	189.78	99.24	23.61	73.90	227.09	135.39	37.28	0.000
NDVI	0.60	0.63	0.62	0.01	0.60	0.64	0.62	0.01	0.327
CT	15.11	25.14	18.34	1.91	14.49	24.50	17.84	1.87	0.377
SPAD	40.92	45.37	43.50	0.73	42.31	45.46	43.67	0.69	0.278

Table 1-5. Morpho-physiological variation in ssp. *tauschii* (273 accessions) and spp. s*trangulata* (15 accessions) of *Aegilops tauschii.*

Trait	Ssp. *tauschii*				Ssp. *strangulata*				***P*-value (*tauschii* versus *strangulata*)**
	Min	**Max**	**Mean**	**STD**	**Min**	**Max**	**Mean**	**STD**	
FLL	5.35	20.65	13.40	2.68	11.04	17.97	14.15	2.17	0.228
FLW	4.20	11.00	8.00	1.20	6.40	9.50	7.90	0.90	0.123
SPL	8.80	17.55	12.46	1.48	8.82	17.27	11.00	1.97	0.027
SPW	0.40	0.72	0.54	0.07	0.58	0.75	0.66	0.06	0.432
SN/SP	15.82	29.93	21.18	2.48	15.42	20.42	18.13	1.30	0.006
SPWg	0.35	0.67	0.51	0.06	0.34	0.71	0.58	0.09	0.004
DH	150.78	191.45	171.08	5.86	170.37	178.51	173.94	1.72	0.000
Bio	60.53	227.09	112.21	34.01	89.70	223.05	128.67	35.40	0.294
NDVI	0.60	0.64	0.62	0.01	0.60	0.63	0.62	0.01	0.088
CT	14.49	25.14	18.16	1.91	16.11	21.55	18.37	1.55	0.280
SPAD	40.92	45.46	43.56	0.72	42.06	44.83	43.82	0.84	0.151

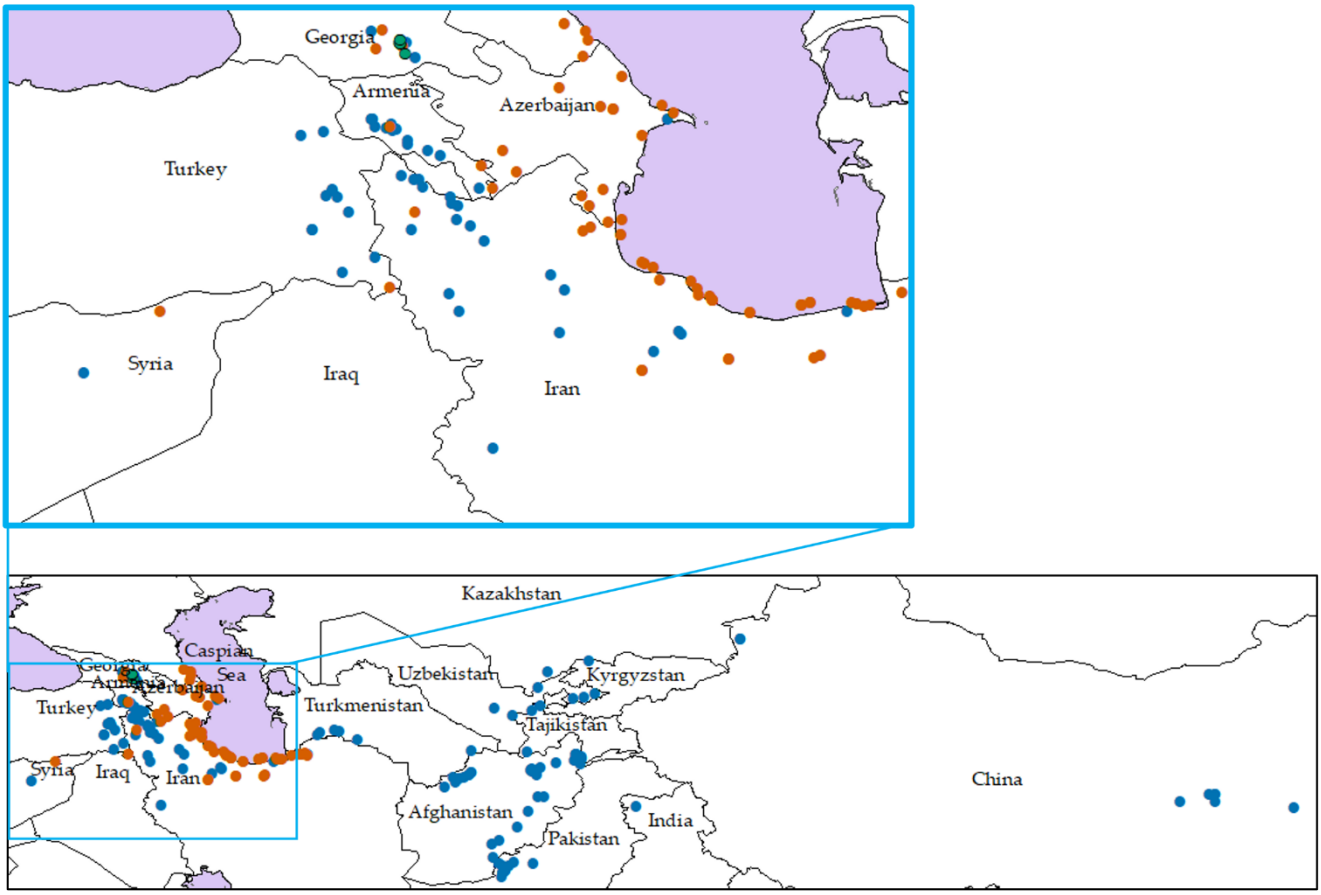

Figure 1-1. Geographical distribution of 293 *Aegilops tauschii* accessions. Blue circles, lineage 1 accessions (TauL1); red circles, lineage 2 accessions (TauL2); and green circle, lineage 3 accessions (TauL3). Western range is enlarged.

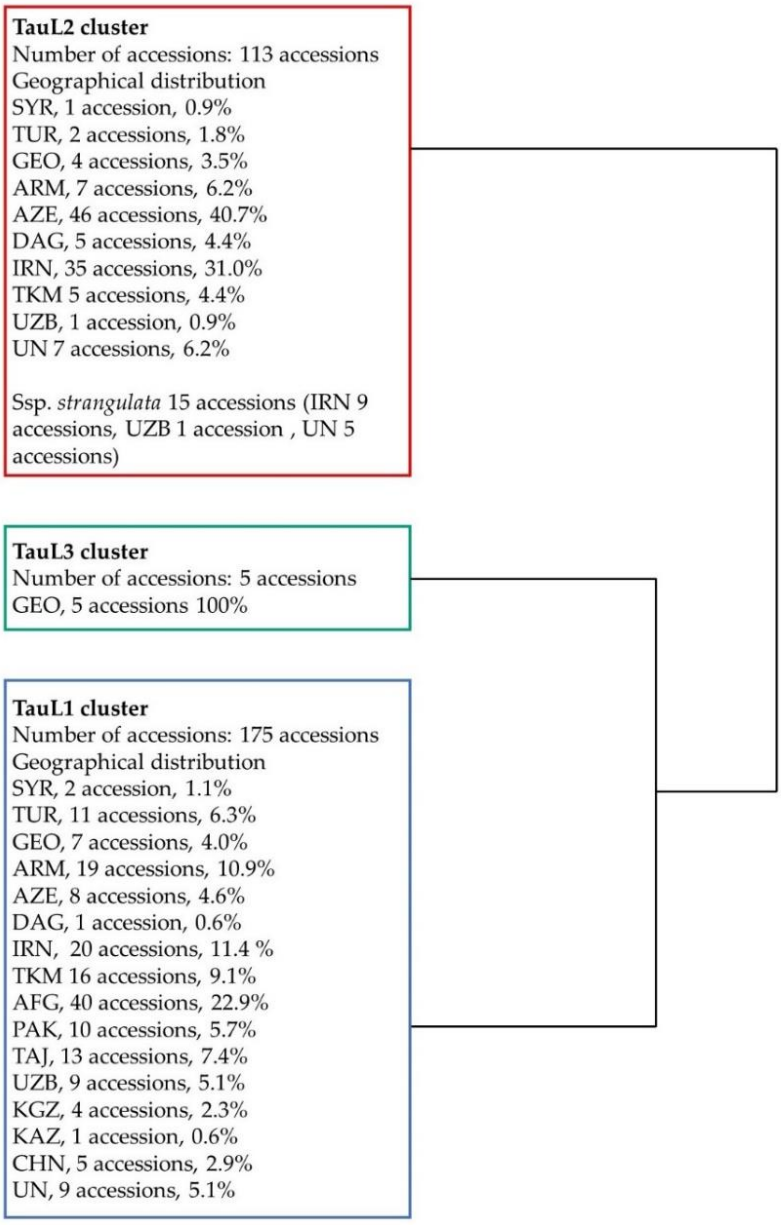

Figure 1-2. Schematic form of hierarchical clustering of 293 *Ae. tauschii* accessions showing the classification of TauL1, TauL2, and TauL3 based on high-quality SNPs derived from 5,880 DArTseq markers. Origin of accessions: SYR, Syria; TUR, Turkey; GEO, Georgia; ARM, Armenia; AZE, Azerbaijan; DAG, Dagestan; IRN, Iran; TKM, Turkmenistan; AFG, Afghanistan; PAK, Pakistan; TAJ, Tajikistan; UZB, Uzbekistan; KGZ, Kyrgyzstan; KAS, Kazakhstan and CHN, China, and UN, unknown country.

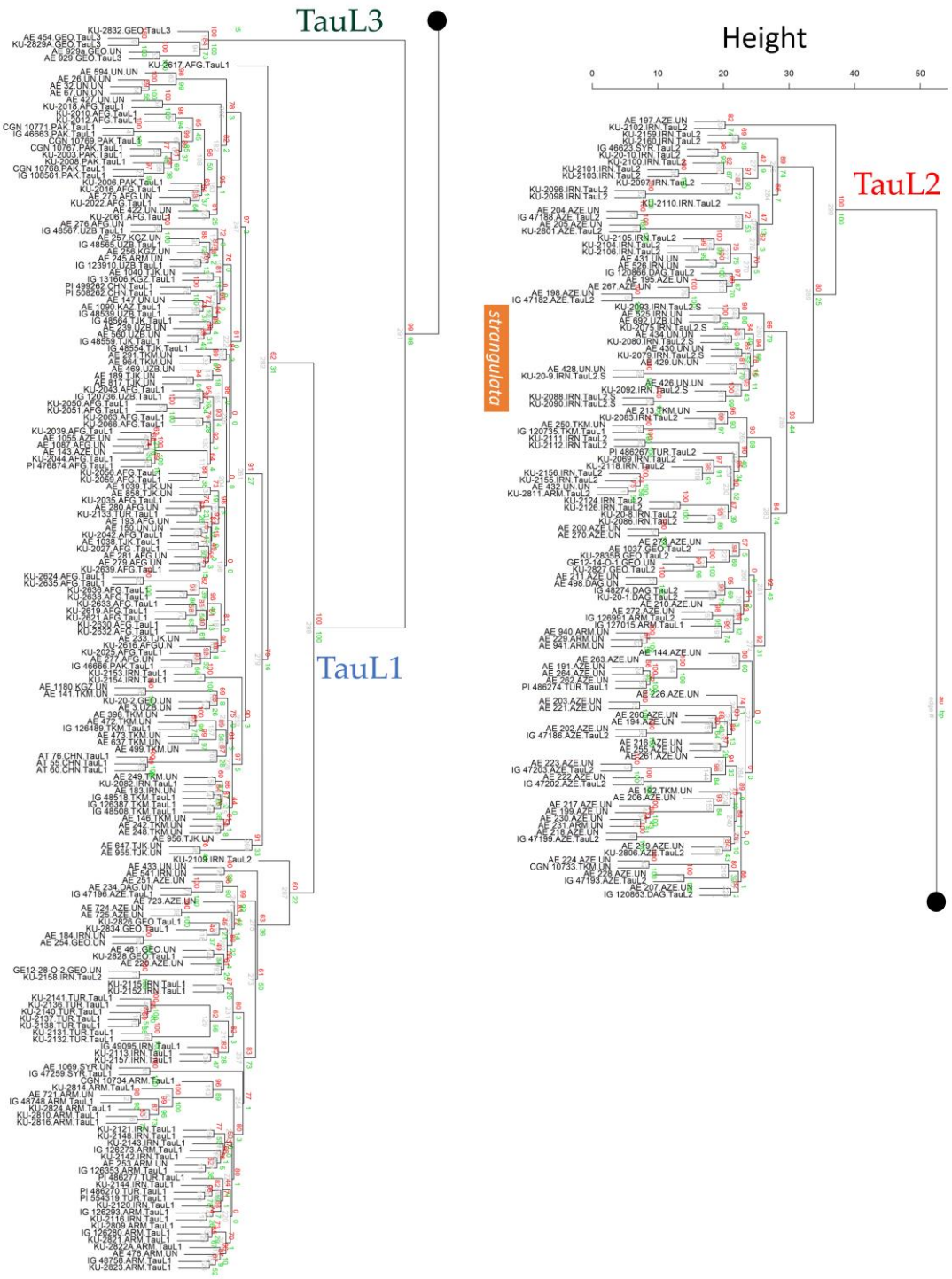

Figure 1-S1. Hierarchical clustering of 293 *Ae. tauschii* accessions showing the classification of TauL1, TauL2, and TauL3 based on high-quality SNP markers derived from 5,880 DArTseq markers.

Values at branches are AU values (upper, red), BP values (down, blue), and cluster labels (medium, gray). Ssp. *strangulata* is indicated, and others belongs to ssp. *tauschii*. UN, unknown lineages, or country. Origin of accessions: SYR, Syria; TUR, Turkey; GEO, Georgia; ARM, Armenia; AZE, Azerbaijan; DAG, Dagestan; IRN, Iran; TKM, Turkmenistan; AFG, Afghanistan; PAK, Pakistan; TAJ, Tajikistan; UZB, Uzbekistan; KGZ, Kyrgyzstan; KAS, Kazakhstan and CHN, China. The two black circles indicate where these two trees are connected.

Chapter 2

Genome-wide association study of morpho-physiological traits in *Aegilops tauschii* to broaden wheat genetic diversity

2.1. Abstract

Aegilops tauschii, the D-genome donor of bread wheat, is a storehouse of genetic diversity that can be used for wheat improvement. This species consists of two main lineages (TauL1 and TauL2) and one minor lineage (TauL3). Its morpho-physiological diversity is large, with adaptations to a wide ecological range. Identification of allelic diversity in *Ae. tauschii* is of utmost importance for efficient breeding and widening of the genetic base of wheat. This study aimed at identifying markers or genes associated with morpho-physiological traits in *Ae. tauschii*, and at understanding the difference in genetic diversity between the two main lineages. I performed genome-wide association studies of 11 morpho-physiological traits of 293 *Ae. tauschii* accessions representing the entire range of habitats using 34, 920 DArTseq markers. I observed a wide range of morpho-physiological variation among all accessions. I identified 79 marker–trait associations (MTAs) in all accessions, 14 specifics to TauL1 and 17 specific to TauL2, suggesting independent evolution in each lineage. Some of the MTAs could be novel and have not been reported in bread wheat. The markers or genes identified in this study

will help reveal the genes controlling the morpho-physiological traits in *Ae. tauschii*, and hence in bread wheat even if the plant morphology is different.

2.1. Introduction

Aegilops tauschii Coss. (syn. *Ae. squarrosa* auct. non L.), a wild diploid self-pollinating species ($2n = 2x = 14$, DD), is the D-genome donor of hexaploid bread wheat (*Triticum aestivum* L.; $2n = 6x = 42$, AABBDD). It is native to Central Asia throughout the Caspian Sea region and China. About $10,000$ years ago, natural hybridization between tetraploid wheat and *Ae. tauschii* (Renfrew 1973; Gill and Raupp 1987; Lubbers *et al.* 1991) led to the formation of hexaploid wheat (Kihara 1944; Mcfadden and Sears 1946)**.** Only a few *Ae. tauschii* lines from a limited area were involved in this hybridization (Lagudah *et al.* 1991). This has resulted in a narrow genetic base of the wheat D-genome during the evolution of bread wheat. This fact has been confirmed by various studies, and indicates that the D-genome of wheat has low genetic diversity compared with the A and B genomes (Kam-Morgan *et al.* 1989; Lubbers *et al.* 1991; Akhundov and Nevzorov 2010)**.** However, much greater genetic diversity is present in the wild D-genome donor (Naghavi *et al.* 2009). It is believed that *Ae. tauschii* is an excellent source of genes to widen the narrow genetic base of bread wheat, such as for drought and heat-stress tolerance (Elbashir *et al.* 2017b; Itam *et al.* 2020)**.** To use the genetic diversity in *Ae. tauschii* effectively, a precise genomic and morpho-physiological analysis is needed.

Genome-wide association study (GWAS) is a leading approach to the dissection of complex traits and the detection of novel and superior alleles for crop breeding. GWAS has been used to untangle the genetic architecture of numerous traits in different crops (Suwarno *et al.* 2015; Sun *et al.* 2017). Many studies have focused on understanding the genetic and morphological diversity of *Ae. tauschii* germplasm (Dudnikov and Kawahara 2006; Matsuoka *et al.* 2008a, 2009, 2015; Naghavi *et al.* 2009; Mizuno *et al.* 2010; Sohail *et al.* 2012; Nishijima *et al.* 2017). However, only a few studies in *Ae. tauschii* have used GWAS, focusing on cadmium stress (Qin *et al.* 2015), phosphorus deficiency (Liu *et al.* 2015a), grain architecture (Arora *et al.* 2017), grain micronutrient concentrations(Arora *et al.* 2019a), or other morphological traits (Liu *et al.* 2015b). Here I investigated marker–trait associations (MTAs) of morpho-physiological traits that could contribute greatly to improving yield and stress adaptation in bread wheat through GWAS, and sought specific MTAs to define the sources of evolution in two of its three lineages, TauL1 and TauL2.

2.2. Materials and Methods

2.2.1. Plant materials

I used 293 *Ae. tauschii* accessions representing the entire range of natural habitats (Supplementary Table 2-1). These comprised AE accessions from the Institut für Pflanzengenetik und Kulturpflanzenforschung, Germany; AT accessions from the Faculty of Agriculture, Okayama University, Japan; CGN accessions from the Instituut

Voor Planten Veredeling, Landbouwhoge School, Wageningen, the Netherlands; IG accessions from the International Center for Agricultural Research in the Dry Areas, Syria; KU accessions from the Germplasm Institute, Faculty of Agriculture, Kyoto University, Japan; and PI accessions from the US Department of Agriculture. Within the panel, 175 accessions belong to TauL1, 133 to TauL2, and 5 to TauL3 (Supplementary Table 2-1).

2.2.2. Morpho-physiological evaluation

Details of the morpho-physiological evaluations and data collection are summarized in (Table 2-8). Spike length and width were measured using ruler as shown in (Fig. 2-1). All accessions were characterized in the research field of the Arid Land Research Center, Tottori University (Tottori, Japan; 35°32′N, 134°13′E), during the winter–spring seasons of 2016–17 (S1) and 2017–18 (S2), in an augmented complete block design with three checks selected randomly. I measured 11 morpho-physiological traits: flag leaf length (FLL), flag leaf width (FLW), spike length (SPL), spike width (SPW), seed number per spike (SN/SP), spike weight (SPWg), days to heading (DH), biomass (Bio), normalized difference vegetative index (NDVI), canopy temperature (CT), and chlorophyll content (SPAD).

2.2.3. Statistical analysis of agronomic traits

ANOVA was conducted in Plant Breeding Tools (PBTools) v. 1.4 software (International Rice Research Institute, http://bbi.irri.org/products). Using genetic

variance (V_g) and environmental variance (V_e), I calculated broad-sense heritability [$H^2 = V_g/(V_g + V_e)$] of each trait (Comstock and Robinson 1952). Because genotype × season interactions were significant, we estimated best linear unbiased predictions (BLUPs) for each trait. I used BLUP data for trait correlation analysis in TauL1, TauL2, and all accessions in SPSS v. 25 software (Gouda 2015).

2.2.4. Genotyping and marker–trait association (MTA) analysis

Genomic DNA was extracted from young leaves by using the CTAB method (Saghai-Maroof *et al.* 1984). The DNA samples (30 µL; 50–100 ng μL^{-1}) were sent to Diversity Arrays Technology Pty Ltd, Australia (http://www.diversityarrays.com), for a whole-genome scan on the DArTseq platform (DArT P/L, Canberra, Australia)). DArTseq is a genotyping-by-sequencing method which utilizes Next-Generation-Sequencing approach to sequence the most informative representations of genomic DNA samples to aid marker discovery. In total, DArTseq generates 59,193 silico and 55,390 SNP markers. I selected the markers with a call rate of 90% (10% missing data) and obtained 3,117 SNP and 47,072 Silico markers, The Fisher exact test was applied to determine if the two alleles were independent SNP markers. Single nucleotide polymorphism (SNP) or Silico DArT markers with a minor allele frequency of <5% were removed from the analysis. The remaining 34, 920 SNPs and Silico DArT markers were used for genomic analysis.

I performed GWAS with BLUP values for each phenotype using a Mixed Linear Model (MLM) in TASSEL v. 5 software (Bradbury *et al.* 2007). For all traits, as the Bonferroni-Holm correction for multiple testing ($\alpha = 0.05$) was too stringent markers with an adjusted -log10 (P-value) ≥ 4.0 were regarded as significant. To search for candidate genes, I performed a BLAST search of the sequence of each significant marker against the Chinese Spring RefSeq v. 1.0 wheat reference genome (IWGSC 2020). The position where the tag hit the best match was extended by 0.5 Mb in both directions, and that sequence was then used in a BLAST search of the Ensembl *T. aestivum* database (http://plants.ensembl.org/*Triticum_aestivum*/Info/Index) to find predicted genes or proteins within this region. To study the validate the usefulness of the MTAs revealed in *Ae. tauschii* to wheat breeding I compare it with previous revealed in bread wheat using GWAS.

2.3. Results

2.3.1. *Morpho-physiological variation*

I studied eight morphological traits (FLL, FLW, SPL, SPW, SN/SP, SPWg, DH, and Bio) and three physiological traits (NDVI, SPAD, and CT). Spike length and width measurement methodology shown in (Fig. 2-1). ANOVA revealed high genetic variation among all accessions in all traits (Table 2-1; Fig. 2-2).

The effect of seasonal difference (S) was significant ($P < 0.05$) for all traits except for FLW and DH. The effect of genotype × seasonal difference interaction (G × S) was

significant for DH, Bio, NDVI, SPAD, and CT. Morpho-physiological variations among accessions in each trait were confirmed by range, mean, standard deviation, and coefficient of variation. The coefficient of variation ranged from 4.6% to 35.5% in S1 and from 4.4% to 57.9% in S2. Heritability values were higher in morphological traits (>0.90; FLL, FLW, SPL, and SPW) than in physiological traits (<0.60; NDVI, SPAD, and CT; Table 2-1).

2.3.2. Correlation of morpho-physiological traits in TauL1, TauL2, and all accessions

In TauL1 and TauL2, I analyzed correlations among morpho-physiological traits (Tables 2-2, 2-3). Both lineages had significant positive correlations between SPWg and SPW (r = 0.781 in TauL1, r = 0.907 in TauL2), DH and Bio (r = 0.631 and 0.574), and SPL and SN/SP (r = 0.497 and 0.564). Both had negative correlations between CT and NDVI (r = −0.439 and −0.324), and CT and Bio (r = −0.427 and −0.163) (Tables 2-2, 2-3).

The correlations between spike-related traits (SPL, SPW, SN/SP, and SPWg) were slightly higher in TauL2 accessions than in TauL1 accessions. I also analyzed correlations in all accessions combined (TauL1, TauL2, and TauL3) (Table 2-4). I found positive correlations between SPWg and SPW (r = 0.843), DH and Bio (r = 0.594), SPL and SN/SP (r = 0.536), FLL and FLW (r = 0.483), and NDVI and Bio (r =

0.457). I found negative correlations between CT and NDVI ($r = -0.388$), and CT and Bio ($r = -0.304$).

2.3.3. GWAS in TauL1 and TauL2 to reveal allelic diversity in each lineage

GWAS revealed 14 MTAs in TauL1 and 17 in TauL2 (Figs. 2-3, 2-4; Table 2-5). TauL1 had one MTA for each SPL and SPW, 4 for Bio, 3 for DH, 2 each for SN/SP and SPAD, and 1 for each SN/SP, SPWg, and NDVI (Fig. 2-3; Table 2-5). R^2 values ranged from 0.10 to 0.19, and were higher than those of the significant markers in all accessions combined (0.05–0.09; Table 2-6). TauL2 had one MTA for each of FLW, SPW, SN/SP, SPWg, and SPAD, 7 MTAs for SPL and 5 MTAs for DH. R^2 ranged from 0.10 to 0.23 (Fig. 2-4; Table 2-5).

Among the MTAs detected for DH in all accessions combined, marker 32782144, 32765508, 32756332 on chromosomes 5D, 2D and 7D, was detected in TauL1 also, where it had pleiotropic effects on DH and Bio (Tables 2-5, 2-6). All other significant MTAs differed between all accessions combined, TauL1 and TauL2. Marker 32740588, detected in TauL2, had a pleiotropic effect on SPW and SPWg. An MTA for CT was detected only in TauL2 (Fig. 2-4; Table 2-5). TauL1 and TauL2 had no MTAs in common. TauL2 had fewer MTAs than TauL1.

2.3.4. GWAS in all accessions of **Aegilops tauschii**

GWAS in all 293 accessions identified 79 MTAs: one each for FLL and SPW, 7 for FLW, 16 for SPL; 13 for SN/SP; 5 for SPWg; 11 for DH; 13 for Bio; 6 for NDVI;

2 for SPAD; and 4 for CT (Fig. 2-5, Table 2-6). R^2 values ranged from 0.05 to 0.09. Most of these MTAs were different from those in TauL1 and TauL2. The exception markers 32785848 for FLL; 32717768 for SPL; 32749747 and 32749753 for SN/SP; 32782144, 32765508 and 32756332 for DH, appeared also in TauL1 controlling same traits. Most of the MTAs contributed less to variability (R^2) than those in TauL1 and TauL2.

2.3.5. Candidate gene identification

I searched for candidate genes for the MTAs in TauL1 and TauL2, and identified the possible functions. The functions show that the MTAs found here play an important role in plant adaptation and survival.

2.4. Discussion

2.4.1. *Morpho-physiological variation in* Aegilops tauschii

Among the wild species in the tribe Triticeae, *Ae. tauschii* is considered the most suitable for the genetic enhancement of wheat. The diversity of D-genome of *Ae. tauschii* is much larger than that of hexaploid wheat's D genome. The *Ae. tauschii* genome contains many useful genes for resistance to biotic and abiotic stresses and for seed storage proteins (Gill *et al.* 1991; Pestsova *et al.* 2000; Assefa and Fehrmann 2004; Naghavi and Mardi 2010)**.** The 293 *Ae. tauschii* accessions analyzed showed significant variation in most traits studied. Spike and leaf traits had higher heritabilities than physiological traits (CT, SPAD, and NDVI) (Table 2-1), indicating that environmental

factors greatly influence physiological traits. As spike and leaf traits are genetically determined, they are less influenced by the environment (Table 2-1). Selection of highly heritable traits will be effective for widening the genetic base of wheat diversity (Maniee *et al.* 2009)**.** Highly correlated traits are likely to be inherited together, widening the genetic base. A positive correlation between SPW and SPWg (r = 0.781 in TauL1, r = 0.907 in TauL2, r = 0.843 in all accessions; Tables 2–4) indicates that an increase in SPW increases SPWg. SPW had a greater effect on grain weight than SPL. On average, grains in TauL2 were heavier and larger. Moderate to strong correlations between grain weight and size in wheat have been reported (Rasheed *et al.* 2014). A mutation in *TaGW2-A1* increased both grain width and length in tetraploid and hexaploid wheat, which increased 1000-grain weight (Simmonds *et al.* 2016). The correlation between SPW and SPWg was highest in TauL2 (r = 0.907; Table 2-3), indicating that TauL2 is a more suitable source for improving grain weight. A positive correlation between SPL and SN/SP indicates that an increase in SPL increases SN/SP. SPL thus affects kernel number per spike and plays an essential role in improving wheat yield (Guo *et al.* 2017)**.** Moreover, the number of grains per m^2 and grain weight are the most important traits for determining grain yield (Arora *et al.* 2017).

Among physiological traits, a significant positive correlation of NDVI with Bio indicates that an increase in NDVI enhances Bio production and subsequently plant production and adaptation. The negative correlation between CT and Bio indicates that

a decrease in CT increases Bio. In other words, plants with better cooling capacity will maintain better Bio. A positive correlation of DH with Bio indicates that a longer vegetative period is preferable for a higher Bio, if the environment is favorable (Tables 2–4).

2.4.2. GWAS of morpho-physiological traits in TauL1 and TauL2

GWAS revealed that MTAs of morpho-physiological traits differed in both chromosome name and location between TauL1 and TauL2 (Table 2-5). These findings indicate that the traits have evolved independently in each lineage. TauL1 had more MTAs for SPAD, SN/SP, and Bio than TauL2 (Figs. 2-3, 2-4), indicating higher variation in these traits in TauL1. As most of the accessions in TauL2 originated from Northern Iran, which has a warm and mild environment, I can speculate that these two traits contribute to the adaptation of these accessions to their habitats. Conversely, NDVI was found only in TauL1. TauL1 could be a source for NDVI gene mining, whereas TauL2 could be a source for CT and SPAD gene mining.

Mahjoob et al., unpublished study found that spike traits are potentially useful for differentiating between TauL1 and TauL2: SPL, SPW, and SPWg all differed significantly. In TauL1, no significant MTA was detected for SPW, and the marker R^2 for SPWg was lower in TauL1 than in TauL2. These results support our conclusion that TauL2 has more diversity in SPW and SPWg than TauL1. Moreover, the SPW and SPWg candidate genes *TraesCS5D02G042200* and *TraesCS5D02G041500*, identified

in TauL2, are orthologous to *Arabidopsis thaliana AT2G03590*, which encodes a transmembrane transporter that increases nitrogen fixation and promotes seed development (Carter and Tegeder 2016)**.** Thus, TauL2 could be an essential source of genes related to these two traits.

2.4.3. GWAS of morpho-physiological traits in all accessions

The phenotypic contribution of markers revealed by GWAS was lower in all accessions than in TauL1 and TauL2 (Table 2-6). These may relate to the difference in population structures, these what reduced the contribution of markers to phenotypic variation (R^2).

2.4.4. Candidate genes revealed by GWAS in **Aegilops tauschii**

I found several MTAs and candidate genes associated with specific functions that play an important role in plant growth and survival. This study is the first study to use GWAS analysis of many morphological and physiological traits in *Ae. tauschii* of important agronomic value to wheat breeding though Liu et al. (Liu *et al.* 2015b) conducted GWAS in *Ae. tauschii* in which traits, SPL, FLL, and FLW are common. Liu et al. (Liu *et al.* 2015b) identified 18 MTAs for only 10 of the 29 traits studied. Our study identified more MTAs, with higher R^2 values (0.5 – 0.23) than most of those, because I used GWAS for two lineages independently with more molecular markers.

2.4.5. ***Marker traits revealed in wheat from*** **Aegilops tauschii**

To study the usefulness of the markers revealed in *Ae. tauschii* and their appearance in wheat, I reviewed previous GWAS studies of wheat (Table 2-7). Li et al. (2019), Ward et al. (2019), Jamil et al. (2019) (Jamil *et al.* 2019; Li *et al.* 2019; Ward *et al.* 2019) reported several MTAs for DH, FLL, SN/SP, and SPL on different chromosomes. I found MTAs for DH on chromosomes 1D, 2D, 3D, 4D, 5D, and 7D also found by Lie *et al.* (2019). I identified novel MTAs on chromosomes 3D and 4D for DH; on 7D for FLL; on 2D, 3D, 4D and 6D for FLW; on 1D, 2D, 3D, 5D, and 6D for SN/SP; and on 1D, 2D, 3D, 4D, and 6D for SPL.

In TauL1, I found novel MTAs on 5D for SN/SP; and on 6D for SPL. In TauL2 (which supplied the D-genome of hexaploid wheat (Matsuoka *et al.* 2013), I identified 6 novel MTAs: one each on 6D associated with DH and SN/SP; 5 MTAs on 1D, 2D, 3D, 5D and 6D associated with SPL. Those MTAs can be easily transferred to the D-genome of wheat where they would be expected to increase yield. Markers on 7D associated with DH can be transferred to improve early flowering in later-flowering variants, especially in drylands.

2.5. Conclusions

I conducted GWAS analysis of morpho-physiological traits in a diverse panel of *Ae. tauschii* accessions and identified several MTAs and corresponding candidate genes. Some of the candidate genes had exact functions related to the trait studied. Morphological traits are more stable and less affected by environmental factors than

physiological traits. GWAS analysis revealed that morphological traits had higher number of MTAs compared to physiological traits (Tables 2-5, 2-6). This facilitates the use of morphological trait selection in wheat breeding through marker-assisted selection. Comparing our findings with other studies in wheat suggested that some of the MTAs and genes identified here are not present in bread wheat. Our results reveal some of the hidden diversity in *Ae. tauschii* and provide a basis for its use in wheat breeding through direct and indirect crossing (Kishii 2019)**.** The information presented here could also help explain the mechanisms controlling the morpho-physiological traits in *Ae. tauschii*, which will pave the way to a better understanding of the mechanisms in bread wheat. Multiple-synthetic-derivative wheat lines incorporate a wide range of genetic diversity of *Ae. tauschii* including both lineages. The developing these materials from both lineages resulted to obtain heat and drought-resistant lines (Elbashir *et al.* 2017a; Gorafi *et al.* 2018; Itam *et al.* 2020). These facts support the indispensable role of the D-genome of *Ae. tauschii* in wheat breeding for high productivity and stress adaptation.

Tables and Figures of chapter 2

Table 2-1. Analysis of variance (ANOVA) of 11 morpho-physiological traits measured in 293 *Aegilops tauschii* accessions grown under field conditions during seasons 2016–17 (S1) and 2017–18 (S2).

Trait	Season	Accession range	Mean	*P*-value (G)	*P*-value (S)	*P*-value (G × S)	SED ± (G)	H^2	CV (%)
FLL (cm)	S1	5.11–22.72	14.98	0.001			3.3292	0.96	21.2
	S2	2.78–21.66	11.89	0.1394			3.539		27.6
	BLUP	4.29–21.42	13.44	<0.001	<0.001	1	1.2933		
FLW (cm)	S1	0.41–1.14	0.80	<0.001			0.1248	0.97	17.0
	S2	0.43–1.12	0.79	<0.001			0.1259		16.3
	BLUP	0.39–1.17	0.80	<0.001	0.9996	0.9975	0.0482		
SPL (cm)	S1	9.89–18.70	13.94	<0.001			1.021	0.98	10.3
	S2	6.92–17.03	10.66	<0.001			1.0216		15.4
	BLUP	8.63 -17.76	12.30	<0.001	<0.001	0.9998	0.4564		
SPW (cm)	S1	0.46–0.76	0.62	<0.001			0.0436	0.96	10.8
	S2	0.30–0.74	0.48	<0.001			0.0386		16.5
	BLUP	0.38–0.75	0.55	<0.001	<0.001	0.8922	0.028		
SN/SP	S1	11.83–32.89	20.42	0.0156			3.128	0.89	18.3
	S2	11.29–31.29	21.50	<0.001			2.3419		17.2
	BLUP	13.15–31.83	20.95	<0.001	<0.001	0.7002	2.0166		
SPWg	S1	0.30–0.77	0.56	0.1799			0.1075	0.90	15.2
	S2	0.29–0.74	0.47	0.0225			0.0849		17.5
	BLUP	0.27–0.76	0.52	<0.001	<0.001	0.9998	0.0496		
DH	S1	134–194	170.99	<0.001			1.4354	0.86	4.6
	S2	132–196	171.76	<0.001			2.6035		4.4
	BLUP	147–195	171.39	<0.001	0.1052	<0.001	3.89		
Bio	S1	50.30–260.90	140.34	<0.001			4.862	0.78	35.5
	S2	50.30–260.40	86.35	<0.001			2.1117		57.9
	BLUP	42.53–260.59	113.51	<0.001	<0.001	<0.001	31.766		
NDVI	S1	0.30–0.79	0.58	<0.001			0.0503	0.13	17.7
	S2	0.28–0.82	0.66	<0.001			0.0087		16.8
	BLUP	0.41- 0.78	0.62	0.1323	<0.001	<0.001	0.0979		
SPAD	S1	29.10–52.40	42.82	<0.001			2.3947	0.28	10.0
	S2	33.40–52.36	44.33	<0.001			0.4336		7.9
	BLUP	33.10–51.19	43.58	0.0047	<0.001	<0.001	3.5841		
CT (°C)	S1	10.62–34.48	18.96	<0.001			1.2967	0.55	21.9
	S2	9.40- 36.90	17.52	<0.001			0.7403		26.5
	BLUP	11.14–31.73	18.25	<0.001	<0.001	<0.001	3.4195		

CV: Coefficient of variation, SED: Significant error of a difference.

Table 2-2. Morpho-physiological correlation analysis in TauL1 performed using best linear unbiased predictions (BLUPs) of two consecutive seasons (2016–17 and 2017–18).

Trait	FLL	FLW	SPL	SPW	SN/SP	SPWg	DH	Bio	NDVI	SPAD	CT
FLL		0.530**	0.264**	0.086	0.178*	0.174*	-0.170*	0.026	0.151*	-0.085	-0.035
		0.000	0.000	0.250	0.016	0.019	0.022	0.730	0.042	0.253	0.641
FLW			0.196**	0.241**	0.067	0.292**	-0.315**	-0.088	0.092	-0.029	-0.043
			0.008	0.001	0.367	0.000	0.000	0.237	0.219	0.694	0.565
SPL				0.049	0.497**	-0.014	0.183*	0.134	0.271**	-0.060	-0.209**
				0.510	0.000	0.851	0.014	0.071	0.000	0.417	0.005
SPW					-0.264**	0.781**	-0.094	0.035	0.084	0.162*	-0.057
					0.000	0.000	0.208	0.637	0.260	0.029	0.442
SN/SP						-0.224**	0.239**	0.065	0.170*	-0.093	-0.152*
						0.002	0.001	0.381	0.022	0.210	0.040
SPWg							-0.177*	-0.007	0.093	0.213**	0.011
							0.017	0.930	0.210	0.004	0.882
DH								0.631**	0.240**	0.068	-0.286**
								0.000	0.001	0.364	0.000
Bio									0.460**	0.085	-0.427**
									0.000	0.256	0.000
NDVI										-0.050	-0.439**
										0.501	0.000
SPAD											-0.022
											0.772

Asterisks: Correlation is significant at *0.05 or **0.01 level. Upper values are correlation coefficients (R^2); lower values are probabilities (P).

Table 2-3. Morpho-physiological correlation analysis in TauL2 performed using best linear unbiased predictions (BLUPs) of two consecutive seasons (2016–17 and 2017–18).

Trait	FLL	FLW	SPL	SPW	SN/SP	SPWg	DH	Bio	NDVI	SPAD	CT
FLL		0.433**	0.254**	0.085	0.162*	0.124	0.005	0.181*	0.256**	-0.091	-0.084
		0.000	0.001	0.292	0.044	0.124	0.950	0.024	0.001	0.257	0.295
FLW			0.062	0.308**	-0.071	0.245**	-0.334**	-0.097	0.053	0.128	-0.108
			0.442	0.000	0.381	0.002	0.000	0.226	0.507	0.110	0.181
SPL				-0.051	0.564**	-0.108	0.137	0.151	0.180*	0.052	-0.197*
				0.525	0.000	0.181	0.088	0.060	0.025	0.515	0.014
SPW					-0.285**	0.907**	-0.161*	0.101	0.228**	0.019	-0.096
					0.000	0.000	0.044	0.208	0.004	0.818	0.231
SN/SP						-0.260**	0.189*	0.063	0.004	0.005	-0.167*
						0.001	0.018	0.434	0.963	0.946	0.037
SPWg							-0.106	0.083	0.222**	0.001	-0.063
							0.186	0.303	0.005	0.990	0.432
DH								0.574**	0.213**	0.046	-0.003
								0.000	0.008	0.566	0.970
Bio									0.457**	-0.003	-0.163*
									0.000	0.968	0.042
NDVI										-0.003	-0.324**
										0.974	0.000
SPAD											-0.116
											0.148

Asterisks: Correlation is significant at *0.05 or **0.01 level. Upper values are correlation coefficients; lower values are probabilities (*P*).

Table 2-4. Morpho-physiological correlation analysis in *Aegilops tauschii* performed using best linear unbiased predictions (BLUPs) of two consecutive seasons (2016–17 and 2017–18).

Trait	FLL	FLW	SPL	SPW	SN/SP	SPWg	DH	Bio	NDVI	SPAD	CT
FLL		0.483**	0.268**	0.088	0.176**	.155**	-0.101	0.093	0.192**	-0.092	-0.047
		0.000	0.000	0.105	0.001	0.004	0.061	0.085	0.000	0.088	0.390
FLW			.126*	0.269**	-0.001	.265**	-0.331**	-0.088	0.083	0.047	-0.074
			0.020	0.000	0.986	0.000	0.000	0.102	0.125	0.383	0.172
SPL				0.005	.536**	-0.050	.147**	0.140**	.219**	-0.022	-0.183**
				0.933	0.000	0.352	0.006	0.009	0.000	0.683	0.001
SPW					-.269**	.843**	-.129*	0.066	0.148**	0.092	-0.073
					0.000	0.000	0.017	0.223	0.006	0.088	0.179
SN/SP						-.236**	.206**	0.065	0.090	-0.055	-0.149**
						0.000	0.000	0.232	0.097	0.313	0.006
SPWg							-.144**	0.037	0.152**	0.107*	-0.022
							0.007	0.489	0.005	0.048	0.680
DH								0.594**	.215**	0.054	-0.156**
								0.000	0.000	0.321	0.004
Bio									0.457**	0.042	-0.304**
									0.000	0.435	0.000
NDVI										-0.025	-.388**
										0.651	0.000
SPAD											-0.068
											0.209

Asterisks: Correlation is significance at *0.05 or **0.01 level. Upper values are correlation coefficients; lower values are probabilities (*P*).

Table 2-5. Marker–trait associations in TauL1 and TauL2 revealed by DArTseq markers.

Lineage	Trait	Marker	Chromo-some	Marker (R^2)	SNPs
TauL1	FLL	32785848	7D	0.13	A/C
	SPL	32717768	6D	0.10	A/C
	SPW	32760139	7D	0.10	A/C
	SN/SP	32749747	5D	0.11	A/C
	SN/SP	32749753	5D	0.11	A/C
	DH	32782144	5D	0.13	A/C
	DH	32765508	2D	0.12	A/C
	DH	32756332	7D	0.13	A/C
	Bio	32736226|F|0-57	1D	0.19	C/T
	Bio	32785723	7D	0.15	A/C
	Bio	32726273	2D	0.11	A/C
	Bio	32772268	7D	0.13	A/C
	SPAD	32729785	4D	0.11	A/C
	SPAD	32730976	4D	0.11	A/C
TauL2	FLW	32784824|F|0-48	3D	0.18	C/G
	SPL	32759935|F|0-24	6D	0.20	C/G
	SPL	32784064|F|0-46	2D	0.20	C/G
	SPL	32779458|F|0-14	5D	0.20	G/A
	SPL	32787428|F|0-35	1D	0.19	A/C
	SPL	32743820|F|0-45	3D	0.20	C/A
	SPL	32738139|F|0-14	1D	0.19	A/G
	SPL	32765872|F|0-42	1D	0.15	T/C
	SPW	32784172	5D	0.21	A/C
	SN/SP	32713693	6D	0.18	A/C
	SPWg	32734854	5D	0.16	A/C
	DH	32784386|F|0-19	2D	0.21	A/G
	DH	32773864|F|0-53	7D	0.22	T/C
	DH	32778505|F|0-44	7D	0.10	G/A
	DH	32762941|F|0-10	6D	0.23	C/G
	DH	32749704	2D	0.18	A/C
	SPAD	32727677	6D	0.19	A/C

Table 2-6. Marker–trait associations in all accessions combined revealed by DArTseq markers.

Lineage	Trait	Marker	Chromo-some	Marker (R^2)	SNPs
All accessions combined	FLL	32785848	7D	0.08	A/C
	FLW	32718764	6D	0.07	A/C
	FLW	32759292|F|0-65	4D	0.08	T/G
	FLW	4308876|F|0-13	2D	0.07	A/C
	FLW	32786154|F|0-26	3D	0.07	G/C
	FLW	32744997|F|0-29	6D	0.07	G/T
	FLW	32741109|F|0-7	6D	0.07	G/A
	FLW	32744675|F|0-22	3D	0.07	C/A
	SPL	32765734|F|0-65	5D	0.09	A/T
	SPL	32783978|F|0-22	1D	0.09	T/C
	SPL	32762629|F|0-44	4D	0.09	C/T
	SPL	32761977|F|0-43	2D	0.09	A/C
	SPL	32785020|F|0-6	1D	0.08	T/G
	SPL	32783857|F|0-18	4D	0.08	T/C
	SPL	32784064|F|0-46	2D	0.08	C/G
	SPL	32759935|F|0-24	6D	0.08	C/G
	SPL	32771485|F|0-17	5D	0.08	G/C
	SPL	32779458|F|0-14	5D	0.08	G/A

Table 2-6. continue Marker–trait associations in all accessions combined revealed by DArTseq markers.

Lineage	Trait	Marker	Chromo-some	Marker (R^2)	SNPs
All accessions combined	SPL	32787428|F|0-35	1D	0.08	A/C
	SPL	32781608|F|0-12	6D	0.08	G/A
	SPL	32717768	6D	0.06	A/C
	SPL	32776612|F|0-42	4D	0.07	T/A
	SPL	32723745	6D	0.06	A/C
	SPL	4313687|F|0-14	4D	0.06	C/G
	SPW	32717545	6D	0.05	A/C
	SN/SP	32749747	5D	0.08	A/C
	SN/SP	32719710	2D	0.08	A/C
	SN/SP	32749753	5D	0.07	A/C
	SN/SP	32767889	5D	0.08	A/C
	SN/SP	32758509|F|0-48	1D	0.08	A/G
	SN/SP	32752366	2D	0.07	A/C
	SN/SP	32719225	2D	0.07	A/C
	SN/SP	4316286	6D	0.07	A/C
	SN/SP	32717034	1D	0.06	A/C
	SN/SP	32767716	3D	0.06	A/C
	SN/SP	32732120	2D	0.06	A/C
	SN/SP	32722401	1D	0.06	A/C
	SN/SP	32765944|F|0-5	1D	0.06	A/C

Table 2-6. continue Marker–trait associations in all accessions combined revealed by DArTseq markers.

Lineage	Trait	Marker	Chromo-some	Marker (R^2)	SNPs
All accessions combined	SPWg	32734854	5D	0.09	A/C
	SPWg	32740167	4D	0.06	A/C
	SPWg	32728690	4D	0.06	A/C
	SPWg	32777696	5D	0.07	A/C
	SPWg	32768696	5D	0.06	A/G
	DH	32782144	5D	0.09	A/C
	DH	32765508	2D	0.09	A/C
	DH	32756332	7D	0.09	A/C
	DH	32736226|F|0-57	1D	0.08	C/T
	DH	32732332	2D	0.06	A/C
	DH	32788932	7D	0.07	A/C
	DH	32784386|F|0-19	2D	0.07	A/G
	DH	32778000|F|0-41	3D	0.07	G/A
	DH	32748170	5D	0.06	A/C
	DH	32738692|F|0-21	5D	0.06	T/G
	DH	32743805|F|0-13	4D	0.07	C/T
	Bio	32736226|F|0-57	1D	0.09	C/T
	Bio	32729301	3D	0.07	A/C

Table 2-6. continue Marker–trait associations in all accessions combined revealed by DArTseq markers.

Lineage	Trait	Marker	Chromo-some	Marker (R^2)	SNPs
All accessions combined	Bio	32776881|F|0-6:T>G-6:T>G	7D	0.09	T/G
	Bio	32785173|F|0-32	7D	0.09	A/G
	Bio	4301634|F|0-37	4D	0.06	A/G
	Bio	32774257	7D	0.07	A/C
	Bio	32711185	6D	0.06	A/C
	Bio	32748250	5D	0.07	A/C
	Bio	32752563|F|0-27	3D	0.07	C/G
	Bio	32730781	3D	0.06	A/C
	Bio	32783241	3D	0.06	A/C
	Bio	32729873	2D	0.06	A/C
	Bio	4323592	3D	0.06	A/C
	NDVI	32785664	7D	0.07	A/C
	NDVI	4329000	6D	0.06	A/C
	NDVI	32764127	3D	0.06	A/C
	NDVI	32754805	2D	0.06	A/C
	NDVI	32751192	5D	0.06	A/C
	NDVI	32781729	1D	0.06	A/C
	SPAD	32727677	6D	0.08	A/C
	SPAD	32753001|F|0-8	5D	0.07	G/C
	CT	32729931	6D	0.08	A/C
	CT	32788658|F|0-19	2D	0.08	A/G
	CT	32784004|F|0-7	3D	0.08	T/C
	CT	32787808|F|0-25	7D	0.08	C/T

Table 2-7. Comparison of MTAs in bread wheat reported previously and those identified in this study in *Aegilops tauschii*

Reference	Species	Trait	Chromosome						
			1D	2D	3D	4D	5D	6D	7D
Li et al. (2019)	*T. aestivum*	DH							
Ward et al. (2019)	*T. aestivum*	DH		**x**					**x**
Jami et al. (2019)	*T. aestivum*	DH	**x**				**x**		**x**
Current study	TauL1	DH		x			x		x
Current study	TauL2	DH		x				x	x
Current study	All	DH	x	x	x	x	x		x
Li et al. (2019)	*T. aestivum*	FLL							**x**
Current study	TauL1	FLL							x
Current study	TauL2	FLL							
Current study	All	FLL							x
Li et al. (2019)	*T. aestivum*	FLW							
Current study	TauL1	FLW							
Current study	TauL2	FLW			x				
Current study	All	FLW		x	x	x		x	
Ward et al. (2019)	*T. aestivum*	SN/SP				**x**			
Current study	TauL1	SN/SP					x		
Current study	TauL2	SN/SP						x	
Current study	All	SN/SP	x	x	x		x	x	
Li et al. (2019)	*T. aestivum*	SPL							**x**
Current study	TauL1	SPL						x	
Current study	TauL2	SPL	x	x	x		x	x	
Current study	All	SPL	x	x	x	x	x	x	

Bold x: Marker identified in previous studies.

Table 2-8. Morpho-physiological traits measured, their abbreviations and definitions.

Trait	Abbreviation	Measurement/Definition
Flag leaf length	FLL (cm)	Measured from three tillers of each accession.
Flag leaf width	FLW (cm)	Measured from three tillers of each accession.
Spike length	SPL (cm)	Measured at the middle spike after maturity stage in five spikes.
Spike width	SPW (cm)	Measured at the middle of five spikes after maturity stage in five spikes.
Seed number/Spike	SN/SP	Counted from five spikes at harvesting.
Seed weight/Spike	SPWg (g)	Measured using five spikes one from each tiller using a sensitive scale.
Days to heading	DH	Recorded when the whole spike above the flag leaf position fully emerged on the earliest tiller in each plant of each accession.
Biomass weight	Bio (g)	Measured after harvesting and drying in a glasshouse from five plants were counted.
Normalized Difference Vegetation Index	NDVI	A vegetative index that compares reflectance in the red and near infrared regions. Measured during flowering using a handheld optical sensor unit (Green Seeker), 2012 NTech Industries, Inc., Ukiah, CA, USA.
Canopy temperature	CT (°C)	Measured during flowering using an inferred thermometer AD-5611A.
Chlorophyll content	SPAD	Measured at the flowering stage from the middle of the flag leaf of three tillers using A Minolta brand chlorophyll meter (Model SPAD-502; Spectrum Technologies Inc. Plainfield, IL).

Supplementary Table 2-1. Phenotypic traits analyzed.

Origin	TauL1					TauL2					TauL3	
Syria	*AE 1069*	IG 47259				IG 46623						
Turkey	KU-2131	KU-2132	KU-2133	KU-2136	KU-2137	PI 486267	**PI 486274**					
	KU-2138	KU-2140	KU-2141	PI 486270	PI 486277							
	PI 554319											
Georgia	*AE 254*	*AE 461*	*GE12-28-O-2*	*KU-20-2*	KU-2826	AE 1037	*GE12-14-O-1*	KU-2827	KU-2835B		AE 929	AE 454
	KU-2828	KU-2834									KU-2829A	KU-2832
											AE 929a	
Armenia	*AE 245*	*AE 253*	*AE 476*	*AE 721*	CGN 10734	*AE 229*	*AE 231*	*AE 940*	*AE 941*	IG 126991		
	IG 126273	IG 126280	IG 126293	IG 126353	IG 48748	**IG 127015**	KU-2811					
	IG 48758	KU-2809	KU-2810	KU-2814	KU-2816							
	KU-2821	KU-2822A	KU-2823	KU-2824								
Azerbaijan	*AE 143*	*AE 220*	*AE 251*	*AE 723*	*AE 724*	*AE 144*	*AE 191*	*AE 194*	*AE 195*	*AE 197*		
	AE 725	*AE 1055*	IG 47196			*AE 198*	*AE 199*	*AE 200*	*AE 202*	*AE 203*		
						AE 204	*AE 205*	*AE 206*	*AE 207*	*AE 210*		
						AE 211	*AE 216*	*AE 217*	*AE 218*	*AE 219*		
						AE 221	*AE 222*	*AE 223*	*AE 224*	*AE 226*		
						AE 230	*AE 255*	*AE 260*	*AE 261*	*AE 262*		
						AE 263	*AE 264*	*AE 267*	*AE 270*	*AE 272*		
						AE 273	*AK 228*	IG 47182	IG 47186	IG 47188		
						IG 47193	IG 47199	IG 47202	IG 47203	KU-2801		
						KU-2806						
Dagestan	*AE 234*					*AE 498*	IG 120863	IG 120866	IG 48274	KU-20-1		
Iran	*AE 183*	*AE 184*	*AE 541*	IG 49095	KU-2082	*AE 525**	*AE 526*	KU-20-8	KU-20-9*	KU-20-10		
	KU-2109	KU-2113	KU-2115	KU-2116	KU-2120	KU-2069	KU-2075*	KU-2079*	KU-2080*	KU-2083		
	KU-2121	KU-2142	KU-2143	KU-2144	KU-2148	KU-2086	KU-2088*	KU-2090*	KU-2092*	KU-2093*		
	KU-2152	KU-2153	KU-2154	KU-2157	**KU-2158**	KU-2096	KU-2097	KU-2098	KU-2100	KU-2101		
						KU-2102	KU-2103	KU-2104	KU-2105	KU-2106		
						KU-2110	KU-2111	KU-2112	KU-2118	KU-2124		
						KU-2126	KU-2155	KU-2156	KU-2159	KU-2160		
Turkmenistan	*AE 141*	*AE 146*	*AE 242*	*AE 248*	*AE 249*	*AE 192*	*AE 213*	*AE 250*	*CGN 10733*	**IG 120735**		
	AE 291	*AE 398*	*AE 472*	*AE 473*	*AE 499*							
	AE 637	*AE 964*	IG 126387	IG 126489	IG 48508							
	IG 48518											
Afghanistan	*AE 193*	*AE 275*	*AE 276*	*AE 277*	*AE 279*							
	AE 280	*AE 281*	*AE 1087*	KU-2010	KU-2012							
	KU-2016	KU-2018	KU-2022	KU-2025	KU-2027							
	KU-2035	KU-2039	KU-2042	KU-2043	KU-2044							
	KU-2050	KU-2051	KU-2056	KU-2059	KU-2061							
	KU-2063	KU-2066	*KU-2616*	KU-2617	KU-2619							
	KU-2621	KU-2624	KU-2630	KU-2632	KU-2633							
	KU-2635	KU-2636	KU-2638	KU-2639	PI 476874							
Pakistan	CGN 10767	CGN 10768	CGN 10769	CGN 10771	IG 108561							
	IG 46663	IG 46666	KU-2003	KU-2006	KU-2008							
Tajikistan	*AE 189*	*AE 233*	*AE 647*	*AE 817*	*AE 858*							
	AE 955	*AE 956*	AE 1038	*AE 1039*	*AE 1040*							
	IG 48554	IG 48559	IG 48564									
Uzbekistan	*AE 3*	*AE 239*	*AE 469*	*AE 560*	IG 120736	*AE 692**						
	IG 123910	IG 48539	IG 48565	IG 48567								
Kyrgyzstan	*AE 256*	*AE 257*	*AE 1180*	IG 131606								
Kazakhstan	AE 1090											
China	AT 55	AT 60	AT 76	PI 499262	PI 508262							
Unknown location site	*AE 26*	*AE 32*	*AE 67*	*AE 147*	*AE 150*	*AE 426**	*AE 428**	*AE 429**	*AE 430**	*AE 431*		
	AE 422	*AE 427*	*AE 433*	*AE 594*		*AE 432*	*AE 434**					

Roman accessions are known from Matsuoka et al. (2009) (Matsuoka *et al.* 2009). AE accessions were received from the Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Germany; AT accessions from the Faculty of Agriculture, Okayama University, Japan; CGN accessions from the Instituut Voor Planten Veredeling, Landbouwhoge School, Wageningen, the Netherlands; IG accessions from the International Center for Agricultural Research in the Dry Areas (ICARDA), Syria;

KU accessions from the Germplasm Institute, Faculty of Agriculture, Kyoto University, Japan; and PI accessions from the US Department of Agriculture. * Ssp. *strangulata*.

Figure 2-1. Methodology of spike measurements in *Ae. tauschii*. (A) Spike length was measured from the base of the lowest spikelet to the top of the highest spikelet. (B) Spike width was measured from the widest part of the spikelet.

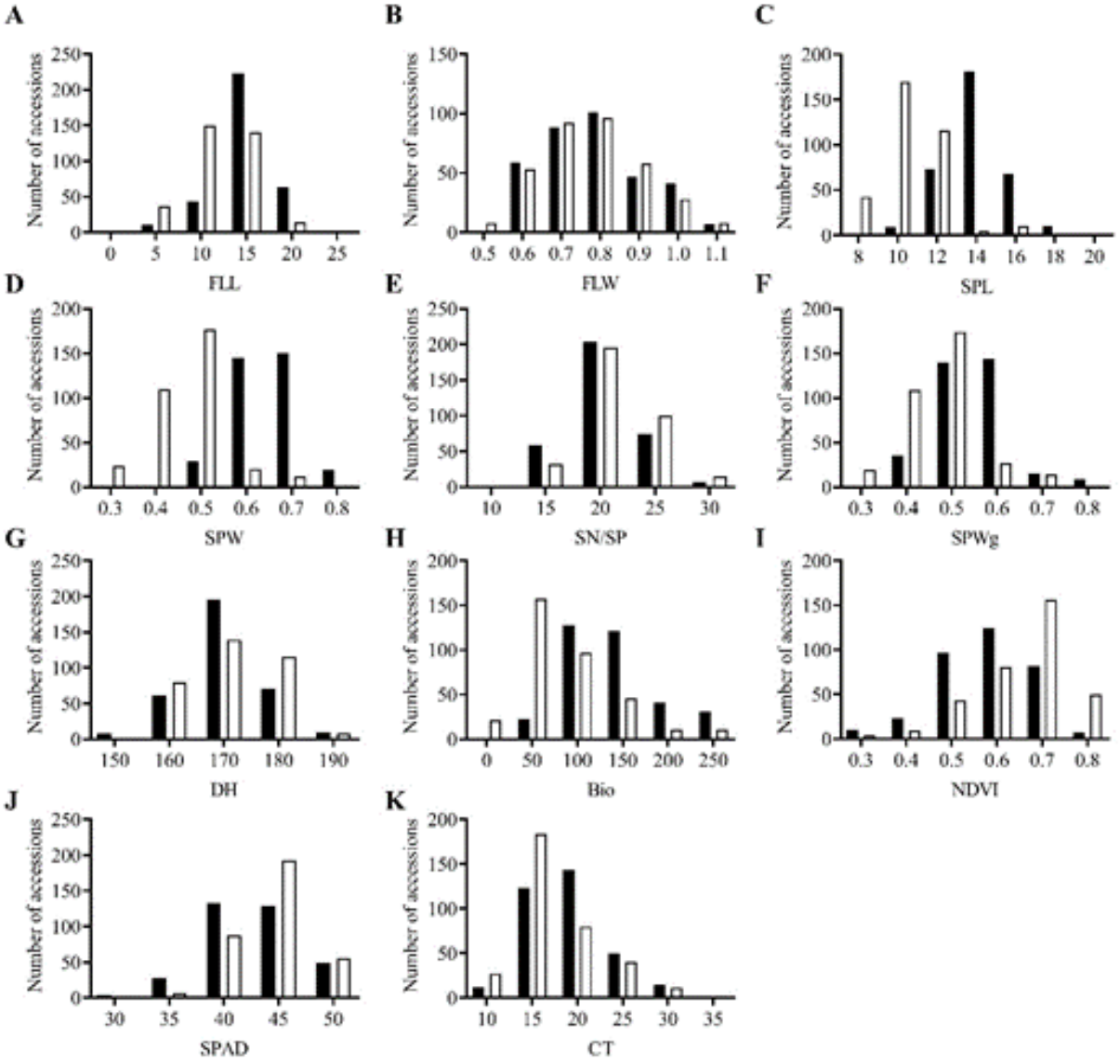

Figure 2-2. Morpho-physiological variation in *Aegilops tauschii* accessions in ■ season 1 and □ season 2. FLL, flag leaf length; FLW, flag leaf width; SPL, spike length; SPW, spike width; SN/SP, seed number per spike; SPWg, spike weight; DH, days to heading; Bio, biomass weight; NDVI, normalized difference vegetative index; CT, canopy temperature; SPAD, chlorophyll content.

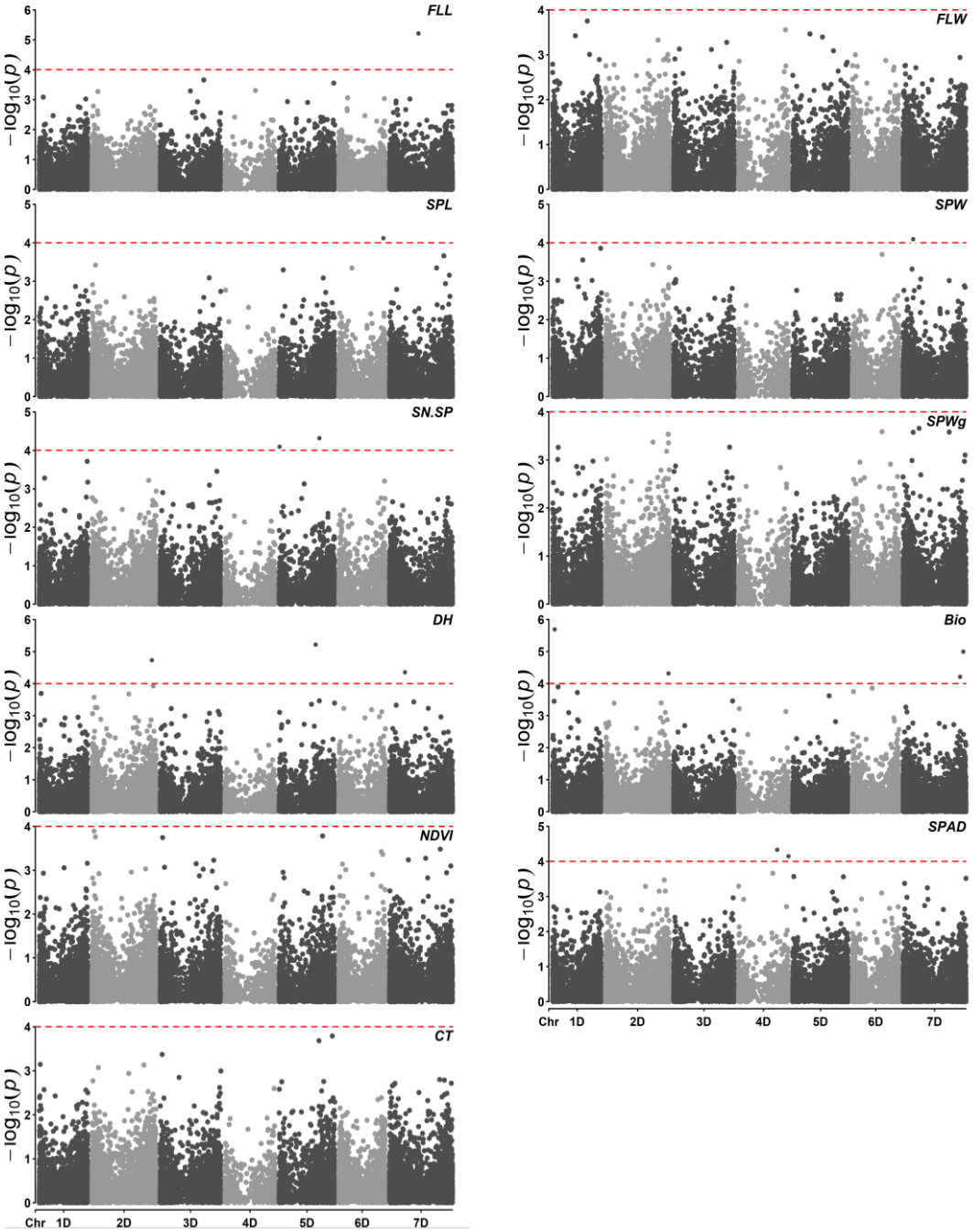
FLL
FLW
SPL
SPW
SN.SP
SPWg
DH
Bio
NDVI
SPAD
CT
−log10(p)
Chr 1D 2D 3D 4D 5D 6D 7D

Figure 2-3. Manhattan plots representing seven chromosomes carrying significant markers detected by Mixed Linear Model using BLUP values in TauL1. FLL, flag leaf length; FLW, flag leaf width; SPL, spike length; SPW, spike width; SN/SP, seed number per spike; SPWg, spike weight; DH, days to heading; Bio, biomass weight; NDVI, normalized difference vegetative index; CT, canopy temperature; SPAD, chlorophyll content. Genomic coordinates are displayed along the X-axis, with the negative logarithm of the association p-value for each single nucleotide polymorphism (SNP) displayed on the Y-axis, meaning that each dot on the Manhattan plot signifies a SNP. Black rules indicate the significance threshold.

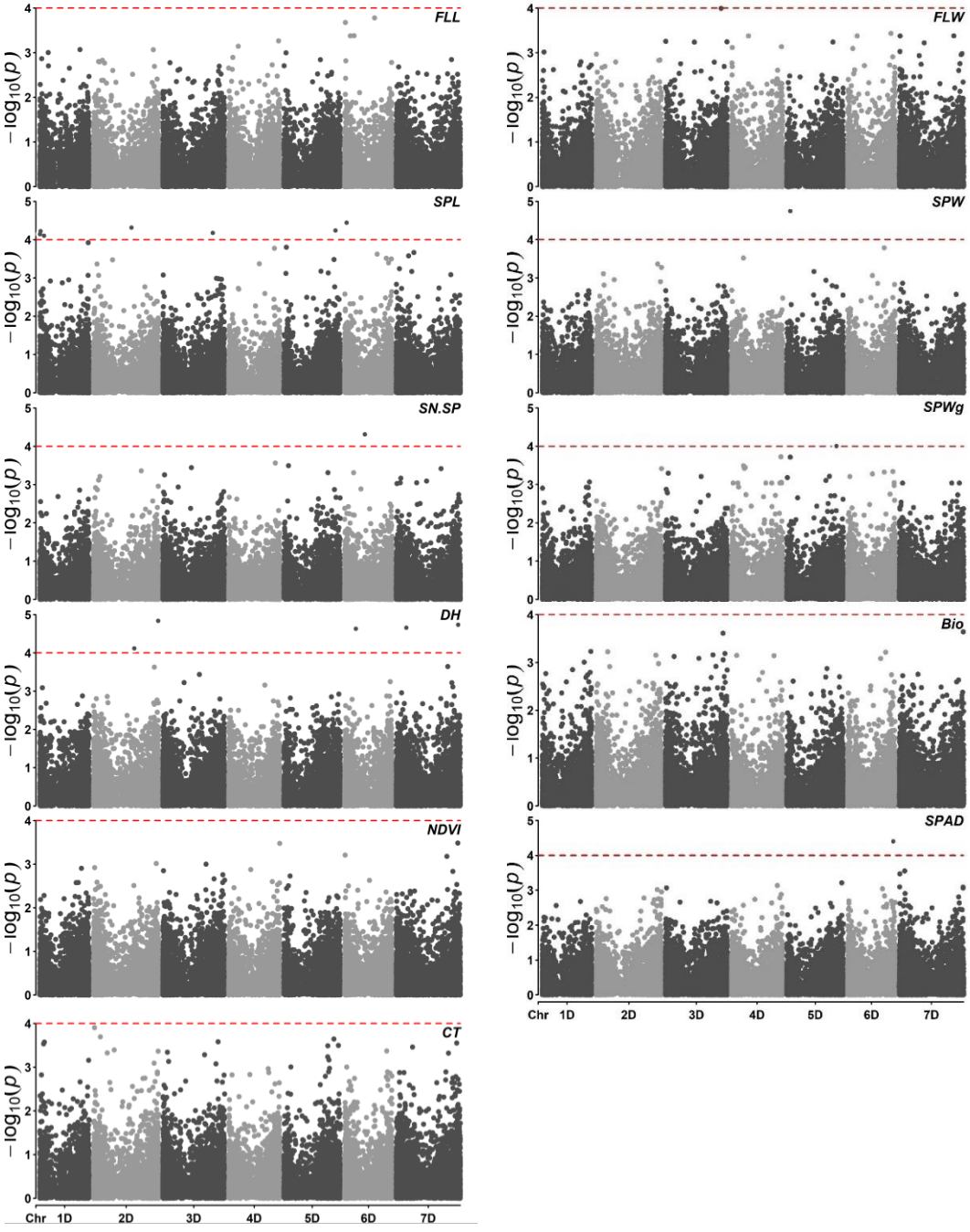

Figure 2-4. Manhattan plots representing seven chromosomes carrying significant markers detected by Mixed Linear Model using BLUP values in TauL2. FLL, flag leaf length; FLW, flag leaf width;

SPL, spike length; SPW, spike width; SN/SP, seed number per spike; SPWg, spike weight; DH, days to heading; Bio, biomass weight; NDVI, normalized difference vegetative index; CT, canopy temperature; SPAD, chlorophyll content. Genomic coordinates are displayed along the X-axis, with the negative logarithm of the association p-value for each single nucleotide polymorphism (SNP) displayed on the Y-axis, meaning that each dot on the Manhattan plot signifies a SNP. Black rules indicate the significance threshold.

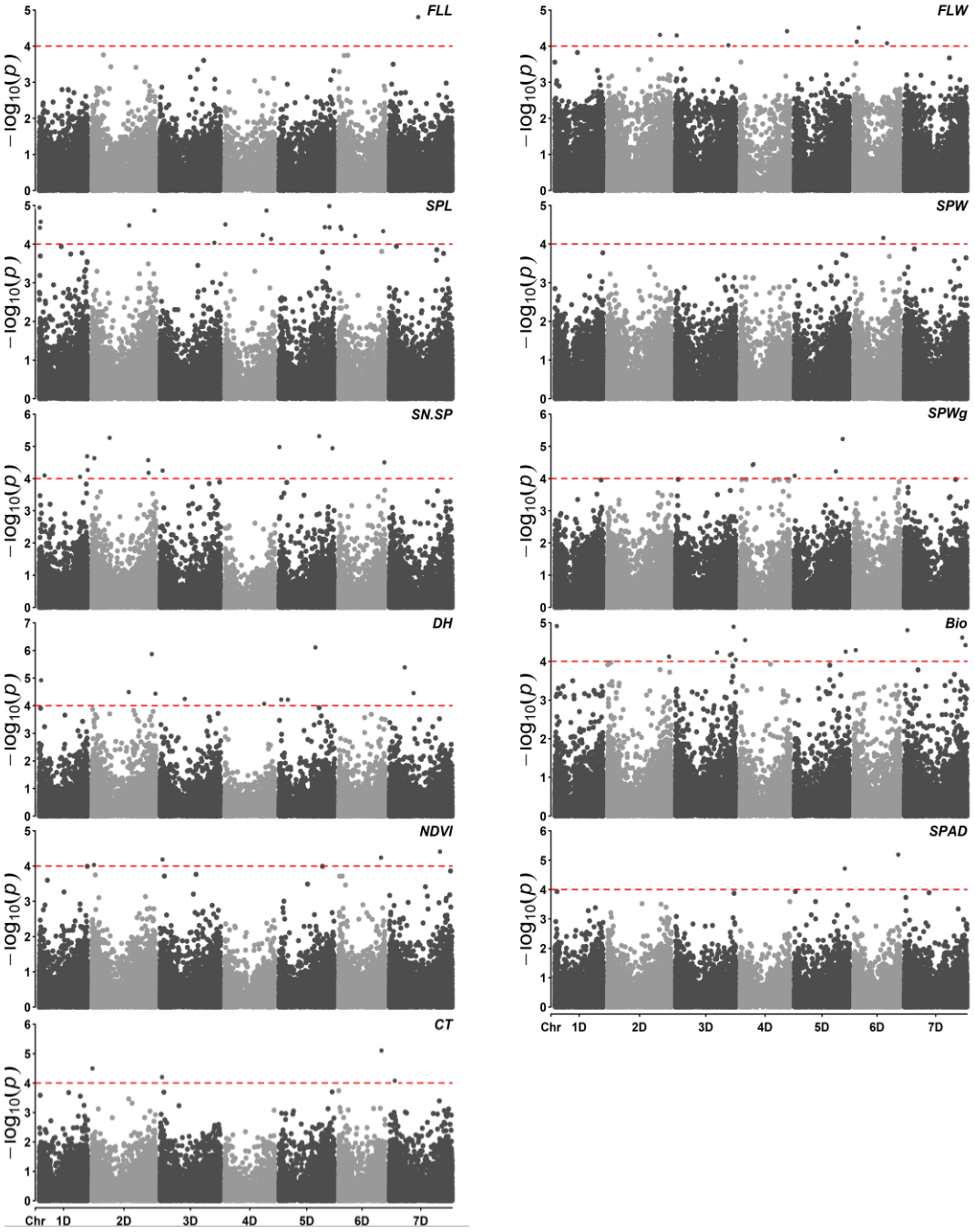

Figure 2-5. Manhattan plots representing seven chromosomes carrying the significant markers detected by Mixed Linear Model using BLUP values in all accessions. FLL, flag leaf length; FLW,

flag leaf width; SPL, spike length; SPW, spike width; SN/SP, seed number per spike; SPWg, spike weight; DH, days to heading; Bio, biomass weight; NDVI, normalized difference vegetative index; CT, canopy temperature; SPAD, chlorophyll content. Genomic coordinates are displayed along the X-axis, with the negative logarithm of the association p-value for each single nucleotide polymorphism (SNP) displayed on the Y-axis, meaning that each dot on the Manhattan plot signifies a SNP. Black rules indicate the significance threshold.

General discussion and conclusion

In this dissertation, phenotypic and genetic analyses provided new insight about genetic diversity in *Ae. tauschii*. Although it is one species, the TauL1 and TauL2 are different. Though wheat breeders should consider the diversity of each lineage independently for wheat breeding. In Chapter 1, I revealed that *Ae. tauschii* has wide range of morpho-physiological variation and spike traits significantly different between the two main lineages TauL1 and TauL2, and between ssp. *tauschii* and ssp. *strangulata* although the range of values overlapped between them. Our result indicated that there is high level of phenotypic convergency presented in *Ae. tauschii.* Genomic analysis showed that three independent lineages are existed; TauL1 and TauL3 include only ssp. *tauschii*, whereas TauL2 includes both ssp. *tauschii* and ssp. *strangulata*. This result is consisting with previous results (Matsuoka et al. 2009, Mizuo 2010). Using DArTseq platform I could allocate 124 accessions previously lacking the passport data, 66 were to TauL1, 57 to TauL2, and one to TauL3. This study identified the genomic and phenotypic diversity of three lineages and two subspecies. This will significantly improve the utilization of *Ae. tauschii* in wheat breeding and increase the outcome of breeding values for different breeding targets.

In chapter 2, I studied allelic diversity in *Ae. tauschii* for agronomically important traits to widen the genetic base of wheat. I revealed 79 marker–trait associations (MTAs) in all accessions, 14 specific to TauL1 and 17 specific to TauL2, suggesting

independent evolution in each lineage. Some of the MTAs could be novel and have not been reported in bread wheat. The markers or genes revealed in this study will help reveal the genes controlling the morpho-physiological traits in *Ae. tauschii*, and hence in bread wheat even if the plant morphology is different. In conclusion, I clarified the phylogenetic of *Ae. tauschii*, through DArTseq markers and studied the phenotypic variation of agronomically important traits for intraspecific species and lineages. In addition, genome-wide association revealed novel genetic loci for agronomically important traits. These analyses revealed some genetic loci are related to TauL1 and others are related to TauL2. These markers can contribute to improve bread wheat greatly for different breeding purposes.

From thesis studies, I revealed that, although the *Ae. tauschii* is one genome incorporate different lineages with high phenotypic convergency among them. These different lineages contribute to increase the genetic diversity in wheat independently. Thus, breeders should consider these lineages are different genome. In another study (not included in this dissertation), we revealed a similar result, where each lineage was contributing independently to control leaf hair density. Multiple-synthetic-derivative lines developed from both lineages showed a wide genetic diversity for heat, drought, and phosphorus use efficiency.

Summary (In English)

The germplasm of related wild species attracts increasing attention because they can provide characters related to adaptation to cultivated species by breeding. The genus *Aegilops* L. (Poaceae) has been intensively studied because of its close relationship with cultivated wheats. The phylogenetic relationship between genera *Aegilops* and *Triticum* L. is widely reported, and on a world scale, the genus *Aegilops* includes 23 wild annual species, of which 11 are diploids and 12 are allopolyploids.

About 8000 to 10,000 years ago, the ancestor of the current bread wheat appeared as a result of natural hybridization between cultivated tetraploid wheat (*Triticum turgidum* L., 2n = 4x = 28, AABB) and *Ae. tauschii*. Inside this last species, two subspecies were first described by Eig (1929) as *Ae. squarrosa* ssp. *eusquarrosa* and ssp. *strangulata* and their nomenclature was revised by Hammer (1980) as *Ae. tauschii* ssp. *tauschii* and ssp. *strangulata*. *Ae. tauschii* is genetically and morphologically diverse, and the ssp. *tauschii* has elongated cylindrical spikelets, whereas ssp. *strangulata* has quadrate spikelets and empty glumes. The ssp. *tauschii* has a wide distribution throughout the species range, whereas ssp. *strangulata* is limited to the south-eastern Caspian coastal region and the Caucasus.

The genetic diversity in *Ae. tauschii* has been studied at the molecular level including isozymes, random amplified polymorphic DNA (RAPD), chloroplast DNA,

amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs) and DArT array markers. Most of these studies classified *Ae. tauschii* into three lineages, TauL1, TauL2 and TauL3: TauL1 including only ssp. *tauschii*, TauL2 including both ssp. *tauschii* and ssp. *strangulata* and TauL3 with intermediate forms.

Ae. tauschii is the easiest species in this genus to utilize in wheat breeding, because there is little to no inhibition to meiotic chromosome pairing with the D genome chromosomes of bread wheat. There is a few research proposed that TauL2 is closer to the D genome of bread wheat compare to TauL1. Furthermore, a few studies have assumed that the D genome of bread wheat was donated to common wheat by the ssp. *strangulata*. This was revealed by several isozyme polymorphism such as NADP-dependent aromatic alcohol dehydrogenase and alcohol dehydrogenase isoenzymes.

It has been confirmed that *Ae. tauschii* have useful traits widely used over the past 60 years for wheat breeding for biotic and abiotic stresses tolerance. It has been utilized via synthetic hexaploid wheat as bridge crossing and direct crossing however, both of these methods have limitation. To use the genetic diversity in *Ae. tauschii* effectively in wheat breeding, a precise genomic and morpho-physiological analysis is needed.

In the first part of this dissertation, I clarified the phylogeny of *Ae. tauschii* using 5,880 high-quality SNPs derived from DArTseq and further measured the traits that significantly different between TauL1, TauL2 and TauL3, or between ssp. *tauschii* and ssp. *strangulata*. Genetic and 11 morpho-physiological diversity was examined in 293

accessions covering the entire range of *Ae. tauschii*, including lines that previously lacked passport data. As a result, we were able to allocate 175, 113 and 5 to TauL1, TauL2 and TauL3, respectively. Of 124 lines lacking passport data were assigned 66 to TauL1, 57 to TauL2 and 1 to TauL3. To study the morpho-physiological variation, I measured two leaf parameters (flag leaf length; flag leaf width), four spike parameters (spike length; spike width; seed number per spike; spike weight), days to heading biomass weight and three physiological traits (Normalized Difference Vegetative Index; canopy temperature; and chlorophyll content).

As a result, I observed wide morpho-physiological variation in each lineage and subspecies. Although some of the spike related traits examined differed significantly between the lineages and subspecies, the range of the variation was overlapped. These similarities may resulted due to adaptive convergence which possibly have originated between the different lineages after the geographical isolation under similar environmental condition

In the second part of this dissertation, I identified allelic diversity in *Ae. tauschii* which is of utmost importance for efficient breeding and widening of the genetic base of wheat. Here I identified markers or genes associated with morpho-physiological traits in *Ae. tauschii*, and understood the difference in genetic diversity between the two main lineages. I performed genome-wide association studies of the same 11 morpho-physiological traits used the first part of this dissertation for 293 *Ae. tauschii* accessions

representing the entire range of natural species range to cover most genetic diversity that existed in *Ae. tauschii* using 34,920 DArTseq markers. I observed a wide range of morpho-physiological variation among all accessions. I identified 79 marker-trait associations (MTAs) in all accessions, 14 specifics to TauL1 and 17 specifics to TauL2, suggesting independent evolution in each lineage. Some of the MTAs are novel and have not been reported in bread wheat. The MTAs identified in each lineage are different from each other, which mean both lineages have highly adopted different genes. This should be considered when we use *Ae. tauschii* in wheat breeding. The markers or genes identified in this study will help to reveal the genes controlling the morpho-physiological traits in *Ae. tauschii*, and thus in bread wheat even if the plant morphology is different.

From the phylogenetic study, I have revealed the most traits to discriminate between and subspecies are spike-related traits (spike shape). Also, it is difficult to discriminate between lineage with plant phenology, and the easiest and accurate method is genomic analysis. Also, I have confirmed that genotyping by the DArTseq platform is an accurate platform to study genomic analysis of plant species. Using this platform, I could give an accurate taxonomy for 124 accessions lacked data on their lineages and subspecies. Furthermore, I could propose the geographical origin of these accessions. These findings will largely facilitate the utilization of *Ae. tauschii* in wheat breeding.

From and GWAS analysis, I have revealed that there is intraspecific lineages

variation excited in both lineages. This variation must be utilized efficiently to transfer most genetic variation to bread wheat. Furthermore, I have discovered that each lineage has contributed differently to enhance specific traits. This mean breeder must utilize the specific lineage according to the breeding targets. In another word, developing new germplasm from both lineages could transfer a large variation of D-genome diversity compare with using one lineage. A number of studies on different traits and in different environments are needed to gain a better understanding of the genetic diversity present in *Ae. tauschii*. It is also necessary to employ advanced genome analysis so that the large and complex genomes can be easily analyzed and a large number of genetic markers can be generated.

Summary (In Japanese)

近縁野生種は、適応性を栽培種に提供することができるため、育種において、ますます注目を集めている。エギロプス属（*Aegilops* L.、イネ科）は、栽培コムギと密接な関係があるため、集中的に研究されてきた。エギロプス属とコムギ属（*Triticum* L.）との間の系統関係は詳細に報告されており、世界的に見ても、エギロプス属には23種の野生一年生種が含まれ、そのうち11種が2倍体であり、12種が異質倍数体である。

約8000～1万年前、現在のパンコムギの祖先は、栽培4倍体コムギ(*Triticum turgidum* L., 2n = 4x = 28, AABB)とタルホコムギとの間の自然交雑の結果として現れた。タルホコムギは、Eig (1929)によって*Ae. squarrosa*と命名され、その中に、*eusquarrosa*亜種と*strangulata*亜種の2つの亜種が存在することが最初に記載された。後に、Hammer (1980)は、タルホコムギを、*Ae. tauschii*と改名し、それに伴って、それらは *tauschii*亜種と*strangulata*亜種という名称に改訂された。タルホコムギは遺伝的にも形態的にも多様性が大きく，*tauschii*亜種

は細長い円筒形の小穂を持つのに対し，*strangulata*亜種は四角形の小穂および外穎をもつ。*Tauschii*亜種は、種の分布域全体に分布しているのに対し、*strangulata*亜種はカスピ海沿岸南東部とコーカサス地域に限られている。

タルホコムギの遺伝的多様性はアイソザイム、RAPD、葉緑体DNA、AFLP、SSRおよびDArTarrayマーカーなど分子レベルで研究されている。これらの研究の多くは本種を3つの系統群、TauL1系統群、TauL2系統群およびTauL3系統群に分類している。TauL1系統群は*tauschii*亜種のみを含み、TauL2系統群は*tauschii*亜種と*strangulata*亜種の両方を含み、TauL3系統群は中間型を含む。

タルホコムギは、パンコムギの D ゲノム染色体との減数分裂染色体の対合がほとんど阻害されないため、エギロプス属の中ではコムギの育種に利用しやすい種である。TauL 系統群 1 と比較して TauL2 系統群がパンコムギの D ゲノムに近いという研究はいくつかある。さらに、パンコムギの D ゲノムが *strangulata* 亜種の供与親となったとする研究もある。このことは、NADP 依存

芳香族アルコール脱水素酵素やアルコール脱水素酵素のアイソザイムの多形成でも明らかとなっている。

タルホコムギは、過去60年間に渡り、生物的・非生物的ストレス耐性コムギ育種に広く利用されてきた有用な形質を持つことが確認されている。タルホコムギは、橋渡し交配や直接交配などの人工的な合成6倍体コムギを介して利用されてきたが、いずれの方法にも限界があった。タルホコムギの遺伝的多様性をパンコムギ育種に効果的に利用するためには、正確なゲノムおよび形態生理学的な解析が必要である。

本論文の前半では、DArTseqに由来する5,880個の高品質一塩基多型を用いて、タルホコムギの系統関係を明らかにし、さらに、TauL1系統群、TauL2系統群、TauL3系統群、または*tauschii*亜種と*strangulata*亜種の間で有意に異なる形質を見いだした。タルホコムギの全範囲をカバーする293の系統について、これまでパスポートデータを欠いていた系統を含めて、遺伝的および11の形態生理学的な多様性を調べた。その結果、175、113、5をそれぞれTauL1、

TauL2、TauL3に割り当てることができた。パスポートデータを持たない124の系統は、66をTauL1に、57をTauL2に、1をTauL3に割り当てることができた。形態生理学的な多様性を調べるために、2つの葉関連形質（止葉長、止葉幅）、4つの穂関連形質（穂長、穂幅、穂あたり種子数、穂重）、到穂日数、3つの生理学的形質（NDVI、葉面温度、葉緑素含量）を測定した。

その結果、各系統群および亜種において、形態生理学的に幅広い変異のあることが確認できた。また、調べた穂関連形質の中には、系統群や亜種間で有意に異なるものもあったが、その変異幅は重なっていた。この類似性は適応集中、地理的隔離の後に、同様な環境条件におかれることにより異なる系統群の形質が類似する、適応集中の結果であると考えられた。

本論文の後半では、効率的な育種やコムギの遺伝的基盤の拡大のために最も重要である、タルホコムギの対立遺伝子の多様性を明らかにした。ここでは、タルホコムギの形態生理学的形質に関連するマーカーや遺伝子を同定し、2つの主要系統群間の遺伝的多様性の違いを解明した。タルホコムギに存

在するほとんどの遺伝的多様性をカバーするために、自然分布の全範囲から代表する293系統のタルホコムギを対象に、本論文の前半部分で使用したのと同じ11の形態生理学的形質について、34,920のDArTseqマーカーを用いてゲノムワイド関連研究を行った。その結果、すべての系統において、幅広い形態生理学的な変異が観察された。その結果、79の形質相関マーカー（MTA）が全系統で確認され、そのうち14はTauL1系統群に特異的なもの、17はTauL2系統群に特異的なもので、それぞれの系統群で独立した進化を遂げていることが示唆された。MTAの中には、パンコムギでは報告されていない新規のものもあった。各系統群で同定されたMTAは互いに異なっており、これは両系統群において異なる遺伝子を高度に蓄積していることを意味する。このことは、タルホコムギをコムギの育種に利用する際に考慮すべきである。本研究で同定されたマーカーや遺伝子は、植物の形態が異なっていても、タルホコムギ、ひいてはパンコムギの形態生理学的形質を支配する遺伝子を明らかにするのに役立つと考えられる。

本研究の系統学的研究から、亜種との識別に最も必要な形質は、穂関連の形質（穂の形状）であることを明らかになった。また、植物の表現型で系統群を判別することは難しく、簡単で正確な方法はゲノムの解析であることが判明した。また、DArTseqプラットフォームによるジェノタイピングは、植物種のゲノム解析を研究するための正確なプラットフォームであることを確認した。このプラットフォームを使うことで、系統群や亜種のデータが不足していた124系統に対して、正確な分類を行うことが可能となった。さらに、これらの系統の起源地も提案することがで可能となった。これらの研究成果は、コムギの育種におけるタルホコムギの利用を大きく促進すると思われる。

私は、GWAS 分析から、両系統群には種内変異があることを明らかにした。パンコムギに遺伝的変異の大半を転移するためには、この変異を効率的に利用する必要がある。さらに、それぞれの系統群において、特定の形質が現れるために、異なる遺伝子が関与していることを発見した。つまり、育種家は育種目標に応じて特定の系統群を別個に利用する必要のある事が分かった。

言い換えれば、両方の系統群から新しい生殖質を開発すれば、1つの系統群を使用する場合と比較して、Dゲノムの多様性の大きな変異をコムギに移すことができる。タルホコムギに存在する遺伝的多様性をより深く理解するためには、異なる形質や異なる環境下での多くの研究が必要である。また、大規模で複雑なゲノムを容易に解析し、多数の遺伝子マーカーを作成することができるように、高度なゲノム解析を採用する必要がある。

References

Akhundov, I., and V.B. Nevzorov (2010) A simple characterization of Student's t3 distribution. Stat. Probab. Lett. 80: 293–295.

Alnaddaf, L.M., M.Y. Moualla, and N. Haider (2012) Resolving genetic relationships among *Aegilops* L . and *Triticum* L . Species using analysis of chloroplast DNA by cleaved amplified polymorphic sequence (CAPS). Asian J. Agric.Sci. 4: 270–279.

Arora, S., J. Cheema, J. Poland, C. Uauy, and P. Chhuneja (2019a) Genome-wide association mapping of grain micronutrients concentration in *Aegilops tauschii*. Front. Plant Sci. 10: 1–14.

Arora, S., N. Singh, S. Kaur, N.S. Bains, C. Uauy, J. Poland, and P. Chhuneja (2017) Genome-wide association study of grain architecture in wild wheat *Aegilops tauschii*. Front. Plant Sci. 8: 886.

Arora, S., B. Steuernagel, K. Gaurav, S. Chandramohan, Y. Long, O. Matny, R. Johnson, J. Enk, S. Periyannan, N. Singh, *et al.* (2019b) Resistance gene cloning from a wild crop relative by sequence capture and association genetics. Nat. Biotechnol. 37: 139–143.

Assefa, S., and H. Fehrmann (2004) Evaluation of *Aegilops tauschii* Coss. for resistance to wheat stem rust and inheritance of resistance genes in hexaploid wheat. Genet Resour. Crop Evol. 51: 663–669.

Avni, R., M. Nave, O. Barad, K. Baruch, S.O. Twardziok, H. Gundlach, I. Hale, M. Mascher, M. Spannagl, K. Wiebe, *et al.* (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357: 93–97.

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

Brenchley, R., M. Spannagl, M. Pfeifer, G.L.A. Barker, R. D'Amore, A.M. Allen, N. McKenzie, M. Kramer, A. Kerhornou, and D. Bolser (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491: 705–710.

Carter, A.M., and M. Tegeder (2016) Increasing nitrogen fixation and seed development in soybean requires complex adjustments of nodule nitrogen metabolism and partitioning processes. Curr. Biol. 26: 2044–2051.

Clavijo, B.J., L. Venturini, C. Schudoma, G.G. Accinelli, G. Kaithakottil, J. Wright, P. Borrill, G. Kettleborough, D. Heavens, and H. Chapman (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome Res. 27: 885–896.

Comstock, R.E., and H.F. Robinson (1952) Genetic parameters, their estimation and significance. In: Proceedings of the 6th international Grassland congress. Washington, DC, pp 248–291.

Doebley, J.F., B.S. Gaut, and B.D. Smith (2006) The molecular genetics of crop domestication. Cell 127: 1309–1321.

Dudnikov, A.J. (1998) Allozyme variation in Transcaucasian populations of *Aegilops squarrosa*. Heredity (Edinb) 80: 248–258.

Dudnikov, A.J. (2017) Polymorphism of Got2 DNA sequences sheds light on *Aegilops tauschii* Coss. intraspecies divergence and origin of *Triticum aestivum* L. Genet. Resour. Crop Evol. 64: 1623–1640.

Dudnikov, A.J., and T. Kawahara (2006) *Aegilops tauschii*: Genetic variation in Iran. Genet. Resour. Crop Evol. 53: 579–586.

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

Dvorak, J., M.C. Luo, Z.L. Yang, and H.B. Zhang (1998a) The structure of the *Aegilops tauschi*i genepool and the evolution of hexaploid wheat. Theor. Appl. Genet. 97:657–670.

Dvorak, J., M.C. Luo, Z.L. Yang, and H.B. Zhang (1998b) The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor. Appl. Genet. 97: 657–670.

Eig, A. (1929) Monographisch-kritische Ubersicht der Gattung *Aegilops*. Repert. Spec. Nov. Reg. Veget. 55: 1–228.

Elbashir, A.A.E., Y.S.A. Gorafi, I.S.A. Tahir, A.M.A. Elhashimi, M.G.A. Abdalla, and H. Tsujimoto (2017a) Genetic variation in heat tolerance-related traits in a population of wheat multiple synthetic derivatives. Breed. Sci. 67: 483–492.

Elbashir, A.A.E., Y.S.A. Gorafi, I.S.A. Tahir, J.S. Kim, and H. Tsujimoto (2017b) Wheat multiple synthetic derivatives: A new source for heat stress tolerance adaptive traits. Breed. Sci. 67: 248–256.

Feldman, M. (2001) Origin of cultivated wheat. In: Bonjean, A.P. and W.J. Angus (eds.) The World Wheat Book: A history of Wheat Breeding, Lavoisier Publishing, Paris, pp. 3–56.

Gill, B.S., and W.J. Raupp (1987) Direct genetic transfers from *Aegilops squarrosa* L. to Hexaploid Wheat 1. Crop Sci. 27: 445–450.

Gill, K.S., E.L. Lubbers, B.S. Gill, W.J. Raupp, and T.S. Cox (1991) A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD). Genome 34: 362–374.

Gorafi, Y.S.A., J.S. Kim, A.A.E. Elbashir, and H. Tsujimoto (2018) A population of wheat multiple synthetic derivatives: an effective platform to explore, harness and utilize genetic diversity of *Aegilops tauschii* for wheat improvement. Theor. Appl. Genet. 131: 1615-1626.

Gouda, M.A. (2015) Common pitfalls in reporting the use of SPSS software. Med. Princ. Pract. 24: 300.

Guo, Z., D. Chen, A.M. Alqudah, M.S. Röder, M.W. Ganal, and T. Schnurbusch (2017) Genome-wide association analyses of 54 traits identified multiple loci for the determination of floret fertility in wheat. New Phytol. 214: 257–270.

Hammer, K. (1980) Vorarbeiten zur monographischen Darstellung von Wildpflanzensortimenten: *Aegilops* L. Die. Kult. 28: 33–180.

Hu, X.-G., B.-H. Wu, Z.-H. Yan, S.-F. Dai, L.-Q. Zhang, D.-C. Liu, and Y.-L. Zheng (2012) Characteristics and polymorphism of NAM gene from *Aegilops* section sitopsis species. African J. Agric. Res. 7: 5252–5258.

Huang, L., Q. Wang, L.Q. Zhang, Z.W. Yuan, J.R. Wang, H.G. Zhang, Y.L. Zheng, and D.C. Liu (2012) Haplotype variations of gene Ppd-D1 in *Aegilops tauschii* and their implications on wheat origin. Genet. Resour. Crop Evol. 59: 1027–1032.

Itam, M., M. Abdelrahman, Y. Yamasaki, R. Mega, Y. Gorafi, K. Akashi, and H. Tsujimoto (2020) *Aegilops tauschii* introgressions improve physio-biochemical traits and metabolite plasticity in bread wheat under drought stress. Agronomy 10: 1–17.

Jaaska, V. (1981) Aspartate aminotransferase and alcohol dehydrogenase isoenzymes: intraspecific differentiation in *Aegilops tauschii* and the origin of the D genome polyploids in the wheat group. Plant Syst. Evol. 137: 259–273.

Jaaska, V. (1978) NADP-dependent aromatic alcohol dehydrogenase in polyploid wheats and their diploid relatives. On the origin and phylogeny of polyploid wheats. Theor. Appl. Genet. 53: 209–217.

Jafarzadeh, J., D. Bonnett, J.-L. Jannink, D. Akdemir, S. Dreisigacker, and M.E. Sorrells (2016) Breeding value of primary synthetic wheat genotypes for grain yield. PLoS One 11:e0162860.

Jamil, M., A. Ali, A. Gul, A. Ghafoor, A.A. Napar, A.M.H. Ibrahim, N.H. Naveed, N.A. Yasin, and A. Mujeeb-Kazi (2019) Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. BMC Plant Biol. 19: 1–18.

Jia, J., S. Zhao, X. Kong, Y. Li, G. Zhao, W. He, R. Appels, M. Pfeifer, Y. Tao, X. Zhang, *et al.* (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496: 91–95.

Jones, H., N. Gosman, R. Horsnell, G.A. Rose, L.A. Everest, A.R. Bentley, S. Tha, C. Uauy, A. Kowalski, and D. Novoselovic (2013) Strategy for exploiting exotic germplasm using genetic, morphological, and environmental diversity: the *Aegilops tauschii* Coss. example. Theor. Appl. Genet. 126: 1793–1808.

Kajimura, T., K. Murai, and S. Takumi (2011) Distinct genetic regulation of flowering time and grain-filling period based on empirical study of D-genome diversity in synthetic hexaploid wheat lines. Breed. Sci. 61: 130–141.

Kam-Morgan, L.N.W., B.S. Gill, and S. Muthukrishnan (1989) DNA restriction fragment length polymorphisms: a strategy for genetic mapping of D genome of wheat. Genome 32:724–732.

Kellogg, E.A., R. Appels, and R.J. Mason-Gamer (1996) When genes tell different stories: the diploid genera of Triticeae (Gramineae). Syst. Bot. 321–347.

Kihara, H. (1944) Discovery of the DD-analyzer, one of the ancestors of *Triticum vulgare*. Agric. Hortic. 19: 13–14.

Kihara, H., and F. Lilienfeld (1949) A new synthesized 6x-wheat. Hereditas 35: 307–319.

Kilian, B., K. Mammen, E. Millet, R. Sharma, A. Graner, F. Salamini, K. Hammer, and H. Ozkan (2011) *Aegilops*, wild crop relatives, genomic and breeding resources. Cereal (Ed Kole, C) 1–76.

Kimber, G., and Y.H. Zhao (1983) The D genome of the Triticeae. Can. J. Genet. Cytol. 25: 581–589.

Kippes, N., A. Chen, X. Zhang, A.J. Lukaszewski, and J. Dubcovsky (2016) Development and characterization of a spring hexaploid wheat line with no functional *VRN2* genes. Theor. Appl. Genet. 129: 1417–1428.

Kishii, M. (2019) An update of recent use of *Aegilops* species in wheat breeding. *Front.* Plant Sci. 10: 585.

Koyama, K., Y. Okumura, E. Okamoto, R. Nishijima, and S. Takumi (2018) Natural variation in photoperiodic flowering pathway and identification of photoperiod-insensitive accessions in wild wheat, *Aegilops tauschii*. Euphytica 214:1–12.

Krasileva, K. V, V. Buffalo, P. Bailey, S. Pearce, S. Ayling, F. Tabbita, M. Soria, S. Wang, E. Akhunov, and C. Uauy (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. Genome Biol. 14: 1–19.

Lagudah, E.S., R. Appels, and D. McNeil (1991) The *Nor-D3* locus of *Triticum tauschii* : natural variation and genetic linkage to markers in chromosome 5. Genome 34: 387–395.

Lagudah, E.S., and G.M. Halloran (1988) Phylogenetic relationships of *Triticum tauschii* the D genome donor to hexaploid wheat - 1. Variation in HMW subunits of glutenin and gliadins. Theor. Appl. Genet. 75: 592–598.

Lelley, T., M. Stachel, H. Grausgruber, and J. Vollmann (2000) Analysis of relationships between *Aegilops tauschii* and the D genome of wheat utilizing microsatellites. Genome 43: 661–668.

Li, F., W. Wen, J. Liu, Y. Zhang, S. Cao, Z. He, A. Rasheed, H. Jin, C. Zhang, J. Yan, *et al.* (2019) Genetic architecture of grain yield in bread wheat based on genome-wide association studies. BMC Plant Biol. 19: 1–19.

Ling, H.-Q., S. Zhao, D. Liu, J. Wang, H. Sun, C. Zhang, H. Fan, D. Li, L. Dong, and Y. Tao (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496: 87–90.

Ling, H.Q., B. Ma, X. Shi, H. Liu, L. Dong, H. Sun, Y. Cao, Q. Gao, S. Zheng, Y. Li, *et al.* (2018) Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. Nature 557: 424–428.

Liu, Y., L. Wang, M. Deng, Z. Li, Y. Lu, J. Wang, Y. Wei, and Y. Zheng (2015a) Genome-wide association study of phosphorus-deficiency-tolerance traits in *Aegilops tauschii*. Theor. Appl. Genet. 128: 2203–2212.

Liu, Y., L. Wang, S. Mao, K. Liu, Y. Lu, J. Wang, Y. Wei, and Y. Zheng (2015b) Genome-wide association study of 29 morphological traits in *Aegilops tauschii*. Sci. Rep. 5:15562.

Londo, J.P., Y.C. Chiang, K.H. Hung, T.Y. Chiang, and B.A. Schaal (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc. Natl. Acad. Sci. U S A 103: 9578–9583.

Lubbers, E.L., K.S. Gill, T.S. Cox, and B.S. Gill (1991) Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. Genome 34: 354–361.

Luo, M.C., Y.Q. Gu, D. Puiu, H. Wang, S.O. Twardziok, K.R. Deal, N. Huo, T. Zhu, L. Wang, Y. Wang, *et al.* (2017) Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature 551: 498–502.

Luo, M.C., Y.Q. Gu, F.M. You, K.R. Deal, Y. Ma, Y. Hu, N. Huo, Y. Wang, J. Wang, S. Chen, *et al.* (2013) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. Proc. Natl. Acad. Sci. U S A 110: 7940–7945.

Maniee, M., D. Kahrizi, and R. Mohammadi (2009) Genetic variability of some morpho-physiological traits in durum wheat (*Triticum turgidum* var. durum). J. Appl. Sci. 9: 1383–1387.

Marcussen, T., S.R. Sandve, L. Heier, M. Spannagl, M. Pfeifer, K.S. Jakobsen, B.B.H. Wulff, B. Steuernagel, K.F.X. Mayer, O.A. Olsen, *et al.* (2014) Ancient hybridizations among the ancestral genomes of bread wheat. Science 345: 1250092–1250092.

Martin, A., and V. Orgogozo (2013) The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. Evolution (N Y) 67:1235–1250.

Matsuoka, Y., M. Jaffar Aghaei, M.R. Abbasi, A. Totiaei, J. Mozafari, and S. Ohta (2008a) Durum wheat cultivation associated with *Aegilops tauschii* in northern Iran. Genet. Resour. Crop Evol. 55: 861–868.

Matsuoka, Y., N. Mori, and T. Kawahara (2005) Genealogical use of chloroplast DNA variation for intraspecific studies of *Aegilops tauschii* Coss. Theor. Appl. Genet. 111: 265–271.

Matsuoka, Y., and S. Nasuda (2004) Durum wheat as a candidate for the unknown female progenitor of bread wheat: An empirical study with a highly fertile F1 hybrid with *Aegilops tauschii* Coss. Theor. Appl. Genet. 109: 1710–1717.

Matsuoka, Y., S. Nasuda, Y. Ashida, M. Nitta, H. Tsujimoto, S. Takumi, and T. Kawahara (2013) Genetic basis for spontaneous hybrid genome doubling during allopolyploid speciation of common wheat shown by natural variation analyses of the paternal species. PLoS One 8: 1–17.

Matsuoka, Y., E. Nishioka, T. Kawahara, and S. Takumi (2009) Genealogical analysis of subspecies divergence and spikelet-shape diversification in central Eurasian wild wheat *Aegilops tauschii* Coss. Plant Syst. Evol. 279: 233–244.

Matsuoka, Y., and S. Takumi (2017) The role of reproductive isolation in allopolyploid speciation patterns: Empirical insights from the progenitors of common wheat. Sci. Rep. 7: 16004.

Matsuoka, Y., S. Takumi, and T. Kawahara (2008b) Flowering time diversification and dispersal in central Eurasian wild wheat *Aegilops tauschii* Coss.: genealogical and ecological framework. PLoS One 3:e3138.

Matsuoka, Y., S. Takumi, and T. Kawahara (2015) Intraspecific lineage divergence and its association with reproductive trait change during species range expansion in central Eurasian wild wheat *Aegilops tauschii* Coss. (Poaceae) Phylogenetics and phylogeography. BMC Evol. Biol. 15: 213.

Matsuoka, Y., S. Takumi, and T. Kawahara (2007) Natural variation for fertile triploid F 1 hybrid formation in allohexaploid wheat speciation. Theor. Appl. Genet .115: 509–518.

Matsuoka, Y., Y. Vigouroux, M.M. Goodman, J.G. Sanchez, E. Buckler, and J. Doebley (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. U S A 99: 6080–6084.

Mcfadden, E.S., and E.R. Sears (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J. Hered. 37: 107–116.

McFadden, E.S., and E.R. Sears (1944) The artificial synthesis of *Triticum spelta*. In: Records of the Genetics Society of America. pp 26–27.

Mizuno, N., M. Yamasaki, Y. Matsuoka, T. Kawahara, and S. Takumi (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.

Morrell, P.L., and M.T. Clegg (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc. Natl. Acad. Sci. U S A 104: 3289–3294.

Mujeeb-Kazi, A., V. Rosas, and S. Roldan (1996) Conservation of the genetic variation of *Triticum tauschii* (Coss.) Schmalh. *(Aegilops squarrosa* auct. non L.) in synthetic hexaploid wheats (*T. turgidum* L. s.lat. × *T. tauschii*; 2n = 6x = 42, AABBDD) and its potential utilization for wheat improvement. Genet. Resour. Crop. Evol. 43: 129–134.

Naghavi, M.R., M.J. Aghaei, A.R. Taleei, M. Omidi, J. Mozafari, and M.E. Hassani (2009) Genetic diversity of the D-genome in *T. aestivum* and *Aegilops* species using SSR markers. Genet. Resour. Crop Evol. 56: 499–506.

Naghavi, M.R., and M. Mardi (2010) Characterization of genetic variation among accessions of *Aegilops tauschii*. Asia-Pacific J. Mol. Biol. Biotechnol. 18: 91–94.

Nakai, Y. (1978) Variation and geographical distribution of esterase zymograms in *Aegilops squarrosa*. Wheat Inf. Serv.

Ning, S., N. Wang, S. Sakuma, M. Pourkheirandish, T. Koba, and T. Komatsuda (2013) Variation in the wheat AP2 homoeologs, the genes underlying lodicule development. Breed. Sci. 63: 255–266.

Nishijima, R., Y. Okamoto, H. Hatano, and S. Takumi (2017) Quantitative trait locus analysis for spikelet shape-related traits in wild wheat progenitor *Aegilops tauschii*:

Implications for intraspecific diversification and subspecies differentiation. PLoS One 12: 3e.0173210.

Nishikawa, K. (1960) Hybrid lethality in crosses between emmer wheats and *Aegilops squarrosa*. I. Vitality of F1 hybrids between emmer wheats and *Ae. squarrosa* var. *typica*. Seiken Jiho= Rep. Kihara Inst. biol. Res. 21–28.

Nishikawa, K., Y. Furuta, and T. Wada (1980) Genetic studies on α- amylase isozymes in wheat. III intraspecific variation in *Ae. tauschii aquarrosa* and birthplace of hexaploid wheat. Japanese J. Genet. 55:325–336.

Okuno, A., H. Tamemoto, K. Tobe, K. Ueki, Y. Mori, K. Iwamoto, K. Umesono, Y. Akanuma, T. Fujiwara, H. Horikoshi, *et al.* (1998) Troglitazone increases the number of small adipocytes without the change of white adipose tissue mass in obese Zucker rats. J. Clin. Invest. 101: 1354–1361.

Pascoal, S., T. Cezard, A. Eik-Nes, K. Gharbi, J. Majewska, E. Payne, M.G. Ritchie, M. Zuk, and N.W. Bailey (2014) Rapid convergent evolution in wild crickets. Curr. Biol. 24: 1369–1374.

Pestsova, E., V. Korzun, N.P. Goncharov, K. Hammer, M.W. Ganal, and M.S. Röder (2000) Microsatellite analysis of *Aegilops tauschii* germplasm. Theor. Appl. Genet. 101: 100–106.

Petersen, G., O. Seberg, M. Yde, and K. Berthelsen (2006) Phylogenetic relationships of T*riticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). Mol. Phylogenet. Evol. 39: 70–82.

Purugganan, M.D., and D.Q. Fuller (2009) The nature of selection during plant domestication. Nature 457:843–848.

Qin, P., L. Wang, K. Liu, S. Mao, Z. Li, S. Gao, H. Shi, and Y. Liu (2015) Genome wide association study of *Aegilops tauschii* traits under seedling-stage cadmium stress. Crop J. 3:405–415.

Ralph, P.L., and G. Coop (2015) Convergent evolution during local adaptation to patchy landscapes. PLoS. Genet. 11: 1–31.

Rasheed, A., X. Xia, F. Ogbonnaya, T. Mahmood, Z. Zhang, A. Mujeeb-Kazi, and Z. He (2014) Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. BMC Plant Biol. 14:1–21.

Renfrew, J.M. (1973) Palaeoethnobotany. The prehistoric food plants of the near east and Europe.

Saeidi, H., M.R. Rahiminejad, S. Vallian, and J.S. Heslop-Harrison (2006) Biodiversity of diploid D-genome *Aegilops tauschii* Coss. in Iran measured using microsatellites. Genet. Resour. Crop Evol. 53: 1477–14848.

Saghai-Maroof, M.A., K.M. Soliman, R.A. Jorgensen, and R.W. Allard (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance,

chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U S A 81: 8014–8018.

Saisho, D., and M.D. Purugganan (2007) Molecular phylogeography of domesticated barley traces expansion of agriculture in the old world. Genetics 177: 1765–1776.

Saisho, D., S. Takumi, and Y. Matsuoka (2016) Salt tolerance during germination and seedling growth of wild wheat *Aegilops tauschii* and its impact on the species range expansion. Sci. Rep. 6: 1–12.

Sansaloni, C., C. Petroli, D. Jaccoud, J. Carling, F. Detering, D. Grattapaglia, and A. Kilian (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proc. 5: P54.

Simmonds, J., P. Scott, J. Brinton, T.C. Mestre, M. Bush, A. del Blanco, J. Dubcovsky, and C. Uauy (2016) A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. Theor. Appl. Genet. 129: 1099–1112.

Sohail, Q., T. Shehzad, A. Kilian, A.E. Eltayeb, H. Tanaka, and H. Tsujimoto (2012) Development of diversity array technology (DArT) markers for assessment of population structure and diversity in *Aegilops tauschii*. Breed. Sci. 62:38–45.

Sun, C., F. Zhang, X. Yan, X. Zhang, Z. Dong, D. Cui, and F. Chen (2017) Genome-wide association study for 13 agronomic traits reveals distribution of superior

alleles in bread wheat from the Yellow and Huai Valley of China. Plant Biotechnol. J. 15: 953–969.

Suwarno, W.B., K. V. Pixley, N. Palacios-Rojas, S.M. Kaeppler, and R. Babu (2015) Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theor Appl Genet 128:851–864.

Suzuki, R., and H. Shimodaira (2006) Pvclust: An R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540–1542.

Takumi, S., K. Koyama, K. Fujiwara, and F. Kobayashi (2011) Identification of a large deletion in the first intron of the *Vrn-D1* locus, associated with loss of vernalization requirement in wild wheat progenitor *Aegilops tauschii* Coss. Genes Genet. Syst. 86:183–195.

Takumi, S., Y. Naka, H. Morihiro, and Y. Matsuoka (2009a) Expression of morphological and flowering time variation through allopolyploidization: an empirical study with 27 wheat synthetics and their parental *Aegilops tauschii* accessions. Plant Breed. 128:585–590.

Takumi, S., E. Nishioka, H. Morihiro, T. Kawahara, and Y. Matsuoka (2009b) Natural variation of morphological traits in wild wheat progenitor *Aegilops tauschii* Coss. Breed. Sci. 59:579–588.

Tsunewaki, K. (1966) Comparative gene analysis of common wheat and its ancestral species. II. Waxiness, growth habit and awnedness. Jpn. J. Bot. 19: 175–229.

Van Slageren, M.W. (1994) Wild wheats: a monograph of *Aegilops* L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae). Agricultural University Wageningen, The Netherlands.

van Slagern, M.W. (1994) Wild wheats: a monograph of *Aegilops* L. and amblyopyrum (Jaub. & Spach) Eig (Poaceae). Wageningen, Wageningen Agric. Univ. Pap. Netherlands 94–97.

Wang, J., M.C. Luo, Z. Chen, F.M. You, Y. Wei, Y. Zheng, and J. Dvorak (2013a) *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytol. 198: 925–937.

Wang, J., M.C. Luo, Z. Chen, F.M. You, Y. Wei, Y. Zheng, and J. Dvorak (2013b) *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytol. 198: 925–937.

Wang, J., M.C. Luo, Z. Chen, F.M. You, Y. Wei, Y. Zheng, and J. Dvorak (2013c) *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytol. 198: 925–937.

Ward, B.P., G. Brown-Guedira, F.L. Kolb, D.A. Van Sanford, P. Tyagi, C.H. Sneller, C.A. Griffey, F. Li, W. Wen, J. Liu, *et al.* (2019) Genome-wide association studies

for yield-related traits in soft red winter wheat grown in Virginia. PLoS One 19: 2044–2051.

Wittkopp, P.J., B.L. Williams, J.E. Selegue, and S.B. Carroll (2003) Drosophila pigmentation evolution: Divergent genotypes underlying convergent phenotypes. Proc. Natl. Acad. Sci. U S A 100: 1808–1813.

Zhao, G., C. Zou, K. Li, K. Wang, T. Li, L. Gao, X. Zhang, H. Wang, Z. Yang, and X. Liu (2017) The *Aegilops tauschii* genome reveals multiple impacts of transposons. Nat. Plants 3: 946–955.

Zhen, Y., M.L. Aardema, E.M. Medina, M. Schumer, and P. Andolfatto (2012) Parallel molecular evolution in an herbivore community. Science 337: 1634–1637.

Zimin, A. V, D. Puiu, R. Hall, S. Kingan, B.J. Clavijo, and S.L. Salzberg (2017a) The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. Gigascience 6: 11gix097.

Zimin, A. V, D. Puiu, M.-C. Luo, T. Zhu, S. Koren, G. Marçais, J.A. Yorke, J. Dvořák, and S.L. Salzberg (2017b) Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res. 27: 787–792.

Zohary, D., J.R. Harlan, and A. Vardi (1969) The wild diploid progenitors of wheat and their breeding value. Euphytica 18: 58–65.

List of published papers

Chapter1:

Title: Traits to differentiate lineages and subspecies of *Aegilops tauschii*, the D genome progenitor species of bread wheat

Authors: Mahjoob, M. M. M., Chen. T., Gorafi, Y. S. A., Yamasaki, Y., Kamal, N. M., Abdelrahman, M., Iwata, H., Matsuoka, Y., Tahir, I. S. A. and Tsujimoto, H

Full Name of Scientific Journal with Volume, Number and Pages: Diversity 13 (5) 217. doi.org/10.3390/d13050217

Published date: May 19, 2021

Chapter 2:

Title: Genome-wide association study of morpho-physiological traits in *Aegilops tauschii* to broaden wheat genetic diversity

Authors: Mahjoob, M. M. M., Gorafi, Y. S. A., Kamal, N. M., Yamasaki, Y., Tahir, I. S. A., Matsuoka, Y. and
Tsujimoto, H.

Full Name of Scientific Journal with Volume, Number and Pages: Plants 10 (2), 211 doi. 10.3390/plants10020211

Published date: January 22, 2021