Accessions of the genetically distinct TauL3 lineage of Aegilops tauschii Coss. from newly identified habitats in Armenia, Azerbaijan, and Iran

Koyama, Yoshihiro
Tanaka, Hiroyuki
Sato, Kazuhiro
Tsujimoto, Hisashi
Matsuoka, Yoshihiro

(Citation)
Genetic Resources and Crop Evolution

(Issue Date)
2025-05-13

(Resource Type)
journal article

(Version)
Version of Record

Genet Resour Crop Evol
https://doi.org/10.1007/s10722-025-02438-4

BRIEF REPORT

Accessions of the genetically distinct TauL3 lineage of *Aegilops tauschii* Coss. from newly identified habitats in Armenia, Azerbaijan, and Iran

Yoshihiro Koyama · Hiroyuki Tanaka · Kazuhiro Sato · Hisashi Tsujimoto · Yoshihiro Matsuoka

Received: 11 March 2025 / Accepted: 7 April 2025

Abstract *Aegilops tauschii* Coss. (DD genome) is a wild, annual species recognized as a progenitor of common wheat (*Triticum aestivum* L., AABBDD genome). As an important genetic resource for wheat breeding, many accessions of this species are now available from genebanks, providing opportunities for evolutionary and agronomic research. To date, three divergent lineages—TauL1, TauL2, and TauL3—have been identified in this species based on DNA-level genotypes. Among them, TauL3 may have played a critical role in shaping the D genome of common wheat, but its accessions are limited in both geographic origin (previously known only from Georgia) and availability. In this paper, we report TauL3 accessions sampled from previously undocumented habitats in Armenia, Azerbaijan, and Iran, discovered while assessing a collection of *Ae. tauschii* accessions. The newly identified TauL3 accessions suggest that this lineage has a broader distribution in the Transcaucasus and adjacent regions than previously thought. Together with Georgian TauL3 accessions, they provide valuable materials for research on the role of TauL3 in the evolution of common wheat, as well as for breeding practices utilizing *Ae. tauschii* germplasm. Nevertheless, TauL3 remains an exceptionally rare lineage. Conserving its natural habitats is an urgent priority.

Keywords Intraspecific lineages · The Transcaucasus · Wheat genetic resources · Wild crop relative conservation · Wild wheat biogeography

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10722-025-02438-4.

Y. Koyama · Y. Matsuoka (✉)
Graduate School of Agricultural Science, Kobe University, 1-1 Rokko, Nada, Kobe 657-8501, Japan
e-mail: pinehill@port.kobe-u.ac.jp

Y. Koyama
e-mail: me@yox5ro.net

H. Tanaka
Faculty of Agriculture, Tottori University, 4-101, Koyama-cho Minami, Tottori 680-8553, Japan
e-mail: htanaka@tottori-u.ac.jp

K. Sato
Faculty of Agriculture, Setsunan University, 45-1 Nagaotoge-cho, Hirakata, Osaka 573-0101, Japan
e-mail: kazuhiro.sato@setsunan.ac.jp

K. Sato
Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan

H. Tsujimoto
Laboratory of Molecular Breeding, Arid Land Research Center, Tottori University, 1390 Hamasaka, Tottori 680-0001, Japan
e-mail: tsujim@tottori-u.ac.jp

Published online: 13 May 2025

Introduction

Aegilops tauschii Coss. (DD genome) (formerly known as *Ae. squarrosa* L.) is a wild annual species known as the D genome progenitor of common wheat (*Triticum aestivum* L., AABBDD genome) (Kihara 1944; McFadden and Sears 1944). This species is thought to have hybridized with a cultivated form of *T. turgidum* L. (AABB genome), giving rise to common wheat in certain regions of northern Iran and/or the Transcaucasus no later than 8000 years ago (Hillman 1978; Nesbit 2001). The natural habitat of this species expands widely across Central Eurasia, from Syria, eastern Turkey, the Transcaucasus, and northern Iran, through Turkmenistan, Afghanistan, and Pakistan, to Central Asia and China (van Slageren 1994; Kilian et al. 2011). As an important genetic resource for wheat breeding, many accessions of this species—collected across its vast range through several botanical expeditions since the mid-twentieth century (Kihara et al. 1965)—are now available from genebanks, providing opportunities to study the species' role in the formation of common wheat (Matsuoka and Takumi 2017).

Previous studies have shown that *Ae. tauschii* has a well-defined genetic structure. To date, three divergent lineages—TauL1, TauL2, and TauL3—have been identified based on patterns of polymorphisms observed in nuclear and chloroplast DNA (Matsuoka et al. 2013; Gaurav et al. 2022). TauL1 is distributed across the species' range and consists of two sublineages: TauL1a, occurring in the western part, including the Transcaucasus and adjacent regions, and TauL1b, occurring in the eastern part, including Afghanistan, Pakistan, Central Asia, and China. TauL2 is mostly distributed in the western part; however, reports on its occurrence in Central Asia are accumulating (Singh et al. 2019; Li et al. 2022). TauL2 consists of two sublineages: TauL2a, occurring mainly in the Transcaucasus and adjacent regions, and TauL2b, occurring mainly in Caspian Iran (Matsuoka et al. 2015). TauL3 has been reported only from Georgia.

The relationship of these lineages/sublineages with the D genome of common wheat is complex. In general, TauL2 and TauL3 are genetically closer to the D genome of common wheat than TauL1 (Matsuoka et al. 2013). Previous studies have revealed the composite nature of the common wheat D genome (Cavalet-Giorsa et al. 2024; Wang et al. 2024): it is primarily a patchwork of the genomes of the TauL2 sublineages, with interwoven components from the TauL1 and TauL3 genomes. The TauL3 component exhibits intriguing geographic distribution patterns: common wheat accessions with negligible to low proportions of the TauL3 component are spread across Eurasia and northern Africa, whereas those with substantial proportions are largely restricted to the Transcaucasus and adjacent regions (Cavalet-Giorsa et al. 2024). Additionally, the genomes of several TauL1 and TauL2 *Ae. tauschii* accessions derived from the Transcaucasus and northern Iran also contain notable proportions of a component derived from TauL3 (Koyama et al. 2025).

All these observations suggest that TauL3 may have played a critical role in shaping the genomes of common wheat and *Ae. tauschii*, highlighting the need for further studies on the origins and functions of the TauL3 component in these species. Nevertheless, TauL3 represents a lineage with a small population size, and its accessions are limited in both geographic origin and availability. In this paper, we report on TauL3 accessions sampled from previously undocumented habitats in Armenia, Azerbaijan, and Iran, discovered while assessing a collection (more than 570 accessions) of *Ae. tauschii* accessions, and discuss the implications of this finding.

Materials and methods

Plant materials

A total of five *Ae. tauschii* accessions, sampled from Armenia (accession IG 126999), Azerbaijan (accessions TN10 and TN11), and Iran (accessions IG 48883 and IG 49140), were used (Table 1).

Notes on the Azerbaijani accessions

The Azerbaijani accessions were originally collected from a roadside site at Kjuduly village, Sheki district, in 2008 (Fig. 1). When the collected samples were grown in an ex situ field, their spike colors segregated into black and white. Based on this observation, individuals with black spikes and those with white spikes

Table 1 *Ae. tauschii* accessions used

Accession [a]	Source [b]	Country	Province [c]	Locality [d]	Latitude [e]	Longitude [e]	Altitude [f]	Note [g]
IG 126999	ICARDA	Armenia	Sjunik	Megri distr.; near VV. Agarak and Karchewan	38.87	46.19	769 m	
TN10	Tottori University	Azerbaijan	NA	Kjuduly village, Sheki district	41.16	47.17	329 m	Black spike
TN11	Tottori University	Azerbaijan	NA	Kjuduly village, Sheki district	41.16	47.17	329 m	White spike
IG 48883	ICARDA	Iran	East Azerbaijan	Tabriz	38.08	46.30	NA	
IG 49140	ICARDA	Iran	East Azerbaijan	10 km Ahar to Meshkin Shahr	38.50	47.20	1110 m	

[a]The accession names "TN10" and "TN11" are tentative

[b]ICARDA stands for the International Center for Agricultural Research in the Dry Areas

[c]An 'NA' denotes 'not available'

[d]Presented as received from the source

[e]Decimal degrees

[f]An 'NA' denotes 'not available'

[g]Presented when available

Fig. 1 Photographs taken at the collection site of the Azerbaijani accessions: site view (**A**) and *Ae. tauschii* spikes (**B**)

were designated as separate accessions. The accession names "TN10" and "TN11" are tentative.

Single nucleotide polymorphism (SNP) genotyping

For each accession, total DNA was extracted from the young, healthy leaves of a single plant grown in a greenhouse using the CTAB (cetyltrimethylammonium bromide) method (Saghai-Maroof et al. 1984). The accessions were genotyped as described in Koyama et al. (2025) and outlined below. The genotyping-by-random-amplicon-sequencing-direct (GRAS-Di) method (Enoki and Takeuchi 2018) was used to generate 5–10 million raw reads (150 bp long, paired-end) per sample on a NovaSeq 6000 instrument (flow cell type S4) at Eurofins Genomics, Inc (Tokyo, Japan). To prepare the sequencing libraries, the final product of two sequential polymerase chain reaction steps performed for each accession, with the total DNA as the template and a mix of 12 random primers, were pooled. Raw reads were adaptor-trimmed and quality-filtered using Trimmomatic (Bolger et al. 2014) by setting the ILLUMINACLIP option to SLIDINGWINDOW:4:30 MINLEN:50.

BWA version 0.7.18 (r1243) (Li 2013) was used with the mem option to align high-quality reads to the *Ae. tauschii* AL8/78 reference genome sequence (Aet v5.0)

(Wang et al. 2021). In addition, previously obtained, quality-filtered GRAS-Di reads from 210 accessions (Table S1) representing the entire species range were aligned to the reference genome sequence. Reads that could not be mapped to the chromosomal sequences were removed from subsequent analyses. Aligned reads with library insert sizes shorter than 70 bp in length or longer than 600 bp in length were removed using SAMtools version 1.20 (Li et al. 2009) with the command 'samtools view -e '((pnext + 150) − pos) > = 70 && ((pnext + 150) − pos) < = 600'. Duplicated reads were removed using SAMtools.

SNPs were identified across 215 accessions relative to the reference genome sequence using BCFtools mpileup version 1.20 (Li 2011) with the '-q 30' setting (minimum mapping quality) and BCFtools call with the '-G' option (Hardy-Wineberg equilibrium not assumed). The filtering was performed using BCFtools with the '− i DP > = 5 & MQ > = 40' setting. Sites with a minor allele frequency below 5% and missing data frequency greater than 20% were additionally filtered out using VCFtools version 0.1.16 (Danecek et al. 2011). The resulting SNPs were pruned using PLINK version 1.90b7 with the '–indep-pairwise' option set to '20,000 2000 0.5' for subsequent analyses (Purcell et al. 2007).

Principal component analysis (PCA)

PCA was performed based on a covariance matrix generated from the SNP genotypes of the accessions, with the genotype at each site coded as '− 1' for homozygous reference, '0' for heterozygous, and '1' for homozygous alternate. The probabilistic PCA method available in the pcaMethods (version 1.94.0) package (Stacklies et al. 2007) was applied in R version 4.4.1 (R Core Team 2016). The principal component plot was generated using the scatterplot3d package for R (Liggest and Mächler 2003).

Pairwise F_{ST} estimation

The F_{ST} values between the five accessions and TauL3, and between the five accessions and each *Ae. tauschii* sublineage, were estimated based on SNPs using the method of Weir and Cockerham (1984), as implemented in the hierfstat package (version 0.5-11) for R (Goudet and Jombart 2022).

Geographic distribution maps

Maps illustrating the geographic distribution of the accessions were generated using a spatial dataset obtained from Natural Earth, a source of free vector and raster map data (naturalearthdata.com), with the rnaturalearth package for R (Massicotte and South 2024). Plotting was performed using the ggplot2 package for R ver. 4.4.1 (Wickham 2016).

Results

Across the 215 accessions, high-quality SNPs were widely mapped to the AL8/78 reference genome sequence at 17,893 sites, including 2,180 SNPs on chromosome 1D, 3,095 on 2D, 2,735 on 3D, 2,110 on 4D, 2,738 on 5D, 2,259 on 6D, and 2,776 on 7D (Fig. 2). A PCA performed on the SNP genotype scores revealed a clear association between the five accessions from Armenia, Azerbaijan, and Iran and the previously identified Georgian TauL3 accessions. In a plot based on the first three principal

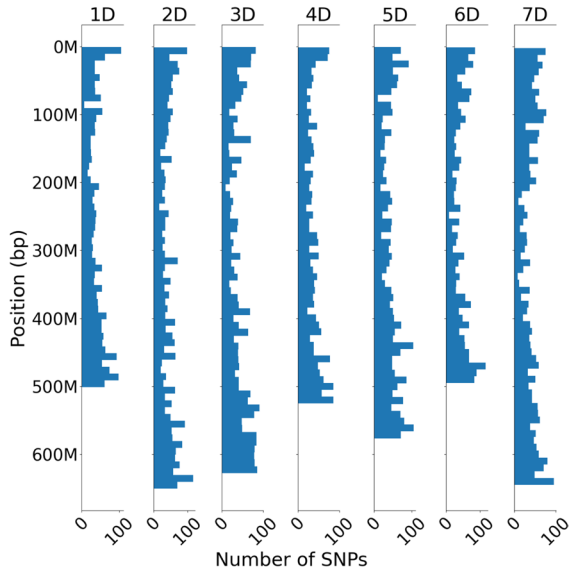

Fig. 2 Bar plots of the number of GRAS-Di-derived single nucleotide polymorphisms (SNPs) mapped to the *Ae. tauschii* AL8/78 reference genome sequence (displayed every 10 Mbp)

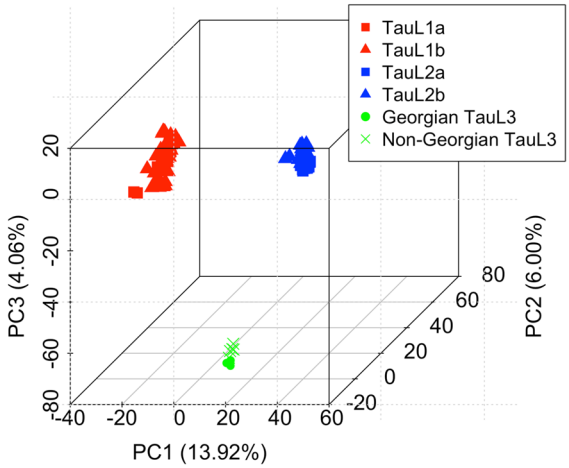

Fig. 3 Graph of the first three principal components from a principal componente analysis (PCA) based on the GRAS-Di-derived single nucleotide polymorphism (SNP) genotypes of 215 accessions. The first component (PC1) accounts for 13.92%, the second (PC2) for 6.00%, and the third (PC3) for 4.06% of the total variance. The accession groups are represented by color and shape for each accession according to the key

components, the five accessions closely aligned with the Georgian TauL3 accessions, whereas they were only distantly related to the TauL1 and TauL2 accessions (Fig. 3; Table S1). The proportions of variance explained by the first three principal components (i.e., PC1, PC2, and PC3) were 13.92%, 6.00%, and 4.06%, respectively.

The five accessions exhibited a low level of genetic differentiation relative to the Georgian TauL3 accessions (pairwise $F_{ST}=0.04$), whereas they showed a higher level of genetic differentiation relative to TauL1a (pairwise $F_{ST}=0.21$), TauL1b (pairwise $F_{ST}=0.21$), TauL2a (pairwise $F_{ST}=0.28$), and TauL2b (pairwise $F_{ST}=0.26$).

The spikes of these five accessions are mildly moniliform, similar in shape to those of the Georgian TauL3 accessions (Fig. 4).

Discussion

Based on the findings described above, we identify the five accessions from Armenia, Azerbaijan, and Iran as belonging to TauL3. The newly identified TauL3 accessions provide novel insights into the

Fig. 4 Green spikes of TauL3 *Ae. tauschii* accessions: IG 126999 (Armenia), TN10 (Azerbaijan), IG 48883 (Iran), IG 49140 (Iran), and KU-2832 (Georgia) (left to right). Scale bar: 1 cm

geographic distribution of TauL3, suggesting that this lineage has a broader distribution in the Transcaucasus and adjacent regions than previously thought (Fig. 5). Its actual range most likely includes the North Caucasus, as Dudnikov (1998, 2012) reported a Dagestan population genetically close to a Georgian TauL3 accession.

Nevertheless, TauL3 remains a lineage with a very small population size; so far, only 11 accessions have been found among more than 570 in our collection. Clearly, conserving this lineage's natural habitats is an urgent priority, necessitating further studies on *Ae. tauschii* in the Transcaucasus and adjacent regions. The TauL3 accessions that will be discovered in

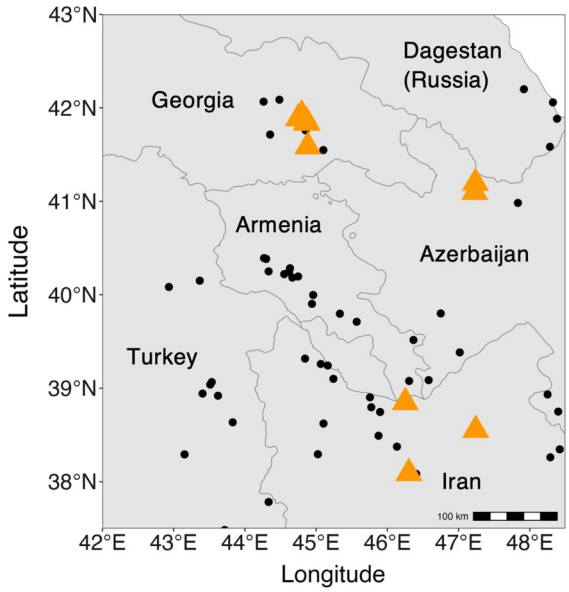

Fig. 5 Geographic distribution of *Ae. tauschii* accessions in the Caucasus region. TauL3 accessions are represented by orange triangles, while other accessions are represented by black points

future studies, together with the newly identified TauL3 accessions from Armenia, Azerbaijan, and Iran, as well as those from Georgia, will provide valuable materials for studies on the role of TauL3 in the evolution of common wheat and *Ae. tauschii*, and for breeding practices utilizing *Ae. tauschii* germplasm.

Acknowledgements We thank the late Tamara Smekalova (N.I. Vavilov Research Institute of Plant Industry), Evgenyi Zhuev (N.I. Vavilov Research Institute of Plant Industry), the late Naib Aminov (Azerbaijan National Academy of Sciences, Genetic Resources Institute), and Aybeniz Alieva (Azerbaijan National Academy of Sciences, Genetic Resources Institute) for their collaboration in collecting the Azerbaijani accessions.

Author contributions Yoshihiro Koyama and Yoshihiro Matsuoka contributed to the study conception and design. Material preparation was performed by Hiroyuki Tanaka, Kazuhiro Sato, Hisashi Tsujimoto, Yoshihiro Koyama, and Yoshihiro Matsuoka. Data collection and analysis were performed by Yoshihiro Koyama and Yoshihiro Matsuoka. The first draft of the manuscript was written by Yoshihiro Matsuoka and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided by Kobe University. This work was supported by the Japan Society for the Promotion of Science KAKENHI Grant (19H02935, 19KK0157, 20K20720, 23K23573, and 23K26927), Moonshot Agriculture, Forestry and Fisheries Research and Development Program (JPJ009237), Hyogo Science and Technology Association Academic Research Grant (number 4002, fiscal year 2022), and the Joint Research Program of Arid Land Research Center, Tottori University (04A2002 and 06B2007).

Data availability Datasets generated and/or analyzed during the current study are available in the National Library of Medicine repository under accession number PRJNA1229811 and DDBJ BioProject repository under accession number PRJDB19942.

Declarations

Conflict of interest The authors declare no competing interests.

References

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Cavalet-Giorsa E, González-Muñoz A, Athiyannan N et al (2024) Origin and evolution of the bread wheat D genome. Nature 633:848–855. https://doi.org/10.1038/s41586-024-07808-z

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dudnikov AJ (1998) Allozyme variation in Transcaucasian populations of *Aegilops squarrosa*. Heredity 80:248–258. https://doi.org/10.1038/sj.hdy.6882810

Dudnikov AJ (2012) Chloroplast DNA non-coding sequences variation in *Aegilops tauschii* Coss.: evolutionary history of the species. Genet Resour Crop Evol 59:683–699. https://doi.org/10.1007/s10722-011-9711-8

Enoki H, Takeuchi Y (2018) New genotyping technology, GRAS-Di, using next generation sequencer. Plant and

Animal Genome XXVI Conference, San Diego, CA, United States. https://pag.confex.com/pag/xxvi/meetingapp.cgi/Paper/29067

Gaurav K, Arora S, Silva P et al (2022) Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. Nat Biotech 40:422–431. https://doi.org/10.1038/s41587-021-01058-4

Goudet J, Jombart T (2022) hierfstat: estimation and tests of hierarchical f-statistics. R package version 0.5-11. https://CRAN.R-project.org/package=hierfstat

Hillman G (1978) On the origins of domestic rye—*Secale Cereale*: the finds from Aceramic Can Hasan III in Turkey. Anatol Stud 28:157–174. https://doi.org/10.2307/3642748

Kihara H (1944) Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agric Hort 19:889–890 **(in Japanese)**

Kihara H, Yamashita K, Tanaka M (1965) Morphological, physiological, genetical and cytological studies in *Aegilops* and *Triticum* collected from Pakistan, Afghanistan and Iran. In: Yamashita K (ed) Results of the Kyoto University Scientific Expedition to the Karakoram and Hindukush, 1955, Vol. I, Cultivated plants and their relatives. Kyoto University, Kyoto, pp 1–118

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, Özkan H (2011) *Aegilops*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, Cereals. Springer, Berlin Heidelberg, pp 1–76

Koyama Y, Nasu M, Matsuoka Y (2025) f_4-statistics-based ancestry profiling and convolutional neural network phenotyping shed new light on the structure of genetic and spike shape diversity in *Aegilops tauschii* Coss bioRxiv 2025.02.16.638466; https://doi.org/10.1101/2025.02.16.638466

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li H, Nie F, Zhu L et al (2022) New insights into the dispersion history and adaptive evolution of taxon *Aegilops tauschii* in China. J Genet Genomics 49:185–194. https://doi.org/10.1016/j.jgg.2021.11.004

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. https://doi.org/10.48550/arXiv.1303.3997

Ligges U, Mächler M (2003) Scatterplot3d - an R package for visualizing multivariate data. J Stat Softw 8:1–20

Matsuoka Y, Takumi S (2017) The role of reproductive isolation in allopolyploid speciation patterns: empirical insights from the progenitors of common wheat. Sci Rep 7:16004. https://doi.org/10.1038/s41598-017-15919-z

Matsuoka Y, Nasuda S, Ashida Y, Nitta M, Tsujimoto H, Takumi S, Kawahara T (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:e68310. https://doi.org/10.1371/journal.pone.0068310

Matsuoka Y, Takumi S, Kawahara T (2015) Intraspecific lineage divergence and its association with reproductive trait change during species range expansion in central Eurasian wild wheat *Aegilops tauschii* Coss. (Poaceae). BMC Evol Biol 15:213. https://doi.org/10.1186/s12862-015-0496-9

McFadden ES, Sears ER (1944) The artificial synthesis of *Triticum spelta*. Rec Genet Soc Am 13:26–27

Nesbitt M (2001) Wheat evolution: integrating archaeological and biological evidence. The Linnean 3:37–59

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Amer J Human Genet 81:559–575. https://doi.org/10.1086/519795

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018. https://doi.org/10.1073/pnas.81.24.8014

Singh N, Wu S, Tiwari V, Sehgal S, Raupp J, Wilson D, Abbasov M, Gill B, Poland J (2019) Genomic analysis confirms population structure and identifies inter-lineage hybrids in *Aegilops tauschii*. Front Plant Sci 10:9. https://doi.org/10.3389/fpls.2019.00009

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods—A bioconductor package providing PCA methods for incomplete data. Bioinformatics 23:1164–1167. https://doi.org/10.1093/bioinformatics/btm069

van Slageren MW (1994) Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub & Spach) Eig (Poaceae). Wageningen Agricultural University, Wageningen

Wang L, Zhu T, Rodriguez JC et al (2021) *Aegilops tauschii* genome assembly Aet v5.0 features greater sequence contiguity and improved annotation. G3 11:jkab325. https://doi.org/10.1093/g3journal/jkab325

Wang Z, Wang W, He Y, Xie X, Yang Z, Zhang X, Niu J, Peng H, Yao Y, Xie C, Xin M, Hu Z, Sun Q, Ni Z, Guo W (2024) On the evolution and genetic diversity of the bread wheat D genome. Mol Plant 17:1672–1686. https://doi.org/10.1016/j.molp.2024.09.007

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370. https://doi.org/10.2307/2408641

Wickham H (2016) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York