Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Intraspecific lineage divergence in Aegilops umbellulata Zhuk.: its association with variation in morphological and reproductive traits

In Son

koubo014971@gmail.com

Kobe University

Nozomi Kasazumi

Kobe University

Shigeo Takumi

Kobe University

Kentaro Yoshida

Kyoto University

Yoshihiro Matsuoka

Kobe University

Article

Keywords: wild wheats, genetic diversity, lineage structure, phenotypic diversity, phylogeography, genome-wide association study

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11376523/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Intraspecific lineage divergence in *Aegilops umbellulata* Zhuk.: its association with variation in morphological and reproductive traits

In Son[1*], Nozomi Kasazumi[1], Shigeo Takumi[1†], Kentaro Yoshida[2], Yoshihiro Matsuoka[1]

[1]Graduate School of Agricultural Science, Kobe University, Japan

[2]Graduate School of Agriculture, Kyoto University, Japan

[†]Deceased.

*Corresponding author: In Son

Graduate School of Agricultural Science, Kobe University, Kobe, Japan

Email: koubo014971@gmail.com

Abstract

Aegilops umbellulata Zhuk. is a wild diploid wheat species (UU genome) native to the Eastern Mediterranean and Middle East. To characterize the natural variation, 202 accessions were analyzed across their distribution ranges at the molecular and phenotypic levels. Population structure analysis revealed two major genetic lineages, UmbL1 and UmbL2. UmbL1 was further divided into western (UmbL1w) and eastern (UmbL1e) sublineages, whereas UmbL2 was restricted to the far west. Two additional groups, Intermediate-w and Intermediate-e, were genetically and geographically intermediate between UmbL1w and UmbL2 and between UmbL1w and UmbL1e, respectively. Among the 24 morphological and phenological traits, flowering time, awn length, trichome level, and seed-related traits showed clear lineage-associated variations. The UmbL1w accessions tended to flower later and had longer awns, higher trichome levels, and smaller seeds, with more spikelets per spike, whereas the UmbL1e accessions tended to flower earlier and produced larger seeds with fewer spikelets. The UmbL2 accessions showed intermediate flowering times and broad seeds with few spikelets. Genome-wide association analysis identified loci potentially associated with the trichome level, and comparative analysis highlighted a candidate gene on chromosome 3U. These findings provide insights into the diversity of *Ae. umbellulata* and its potential use in bread wheat breeding.

Keywords

wild wheats, genetic diversity, lineage structure, phenotypic diversity, phylogeography, genome-wide association study

Introduction

Aegilops umbellulata Zhuk. (2n=2x=14, UU genome) is an annual wild wheat species occurring across the Aegean Islands, Anatolia, the Middle East, and Transcaucasia. It typically occupies fallow lands, grasslands, roadsides, field margins, forest edges, and seaside sand dunes. This species belongs to section Aegilops, one of the five sections of the genus *Aegilops* (Aegilops, Cylindropyrum, Comopyrum, Cylindricum, and Vertebrata), among which section Aegilops is the most species-rich [1, 2]. *Ae. umbellulata* played a pivotal role in shaping the genetic composition of this genus. Through interspecific hybridization followed by allopolyploidization, the U genome was transmitted to several polyploid species, including tetraploid species such as *Aegilops geniculata* Roth (UUMM genome) and *Aegilops triuncialis* (L.) Á. Löve (UUCC genome) as well as the hexaploid species *Aegilops neglecta* Req. ex Bertol. (UUMMNN genome) within section Aegilops [3, 4]. In addition, *Ae. umbellulata* donates plasmons (i.e., cytoplasmic genomes) to several species within section Aegilops, including *Aegilops biuncialis* Vis., *Aegilops columnaris* Zhuk., *Ae. neglecta*, and *Ae. triuncialis* [5, 6]. Why do so many wild species share the U genome with *Ae. umbellulata*? Intraspecific diversity patterns in *Ae. umbellulata* may provide key insights into the mechanisms underlying the evolution not only of *Ae. umbellulata* but also of the genus *Aegilops*.

Previous studies have revealed various aspects of the species' phenotypic and genetic diversity. Accessions of this species display considerable morphological variations, prompting taxonomists to propose an intraspecific classification scheme (Fig. 1). Based on their glume-awn morphology, they can be classified into two varieties: *typica* and *pilosa* [7]. Hammer [8] further recognized two subspecies, *umbellulata* and *transcaucasica*, and two varieties, *umbellulata* and *pilosa*, within the subspecies *umbellulata*. However, the

relationship between these taxonomic classifications and the species' genetic structure remains unclear.

More recently, C-banding analysis of ten accessions of different geographic origins revealed no major structural chromosome rearrangements [9]. Hosseini et al. [10] reported intraspecific variations in DNA content and karyotypes through analyses of 96 Iranian accessions using flow cytometry and chromosome observations. Fluorescence *in situ* hybridization (FISH) karyotyping of 47 accessions identified six variants in chromosome 7, whereas only two or three variants were detected in each of the remaining six chromosomes [11]. Furthermore, morphological and isozyme analyses of approximately 200 accessions representing the entire species range showed that accessions sampled from the Greek Islands tended to have short spikes with small numbers of spikelets and awns and that they were genetically distantly related to continental accessions from Iran, Iraq, and Turkey [12].

RNA sequencing-based studies of nucleotide polymorphisms have provided a basis for understanding genetic patterns of diversity in *Ae. umbellulata*. Despite its relatively narrow geographic range, *Ae. umbellulata* exhibits a high level of nucleotide polymorphisms relative to *Aegilops tauschii* Coss., a species widely distributed in central Eurasia [13]. Notably, *Ae. umbellulata* comprises two divergent lineages, UmbL1 and UmbL2, identified based on nucleotide polymorphisms [14]. Consistent with Kawahara [12], accessions from the Greek Islands (mostly belonging to UmbL2) are genetically and morphologically distinct from those from continental regions (mostly belonging to UmbL1). UmbL1 occurs predominantly in continental regions, including Turkey, Iraq, Iran, and Cyprus, whereas UmbL2 is restricted to the Greek Islands, with a few exceptions. UmbL1 is further subdivided into two sublineages, UmbL1e (predominantly eastern continental) and UmbL1w (predominantly western continental), and accessions belonging to these sublineages differ in flowering time and other

89 morphological traits. Variation in flowering time exhibits a clear latitudinal cline and is
90 strongly correlated with solar radiation during the winter season.

91 Studies on wild wheat relatives are becoming increasingly important for global food security
92 because they provide fundamental genetic resources for wheat improvement. Bread wheat
93 (*Triticum aestivum* L. subsp. *aestivum*), which belongs to the genus capable of intercrossing
94 with *Aegilops*, is a major global crop used for bread, noodles, and a wide range of baked
95 products essential for human nutrition. However, the growth rate of global wheat production
96 has slowed down in recent years [15]. Concurrently, world population continues to increase,
97 and climate change-related stresses are expected to reduce global wheat production if
98 effective countermeasures are not implemented [16–18]. To enhance wheat resilience under
99 increasingly challenging production conditions, several wild relatives have already been
100 incorporated into breeding programs and have played critical roles in improving modern
101 bread wheat varieties. For example, *Aegilops ventricosa* Tausch serves as a source of
102 resistance to multiple diseases [19, 20] and *Leymus racemosus* (Lam.) Tzvelev improves
103 productivity under reduced nitrogen fertilizer input [21].

104 *Ae. umbellulata* possesses many agronomically important genes, including those conferring
105 resistance to leaf rust, stripe rust, stem rust, powdery mildew, and Hessian flies [22, 23] as
106 well as genes encoding low-molecular-weight glutenin subunits associated with flour quality
107 [24], and genes related to thermotolerance [25]. In addition, synthetic hexaploid lines
108 (AABBUU genome) exhibit broad variation in flowering time and morphological traits [26].
109 These synthetic lines can be used as bridging materials to introduce U-genome segments
110 carrying agronomically useful genes into bread wheat. Consequently, *Ae. umbellulata*
111 represents a useful resource for developing high-yielding, high-quality, and climate-resilient
112 wheat varieties as well as for countering the genetic uniformity of modern bread wheat
113 cultivars.

This study aimed to clarify the genetic structure and phenotypic variation in *Ae. umbellulata* across its species range and to identify the mechanisms underlying its diversity. To achieve these objectives, we analyzed a total of 202 accessions representing entire species range using genome-wide single nucleotide polymorphism (SNP) data and multi-year surveys of morphological and reproductive traits. High-resolution SNP data were generated by aligning next-generation sequencing reads obtained using the Genotyping by Random Amplicon Sequencing-Direct (GRAS-Di) method [27–30] to a recently published chromosome-scale genome assembly of *Ae. umbellulata* [31]. Genetic structure was characterized based on these SNP data, phenotypic variation was evaluated among genetic groups, and a genome-wide association study (GWAS) was conducted to identify genomic regions potentially associated with phenotypic variation. These findings provide novel insights into the evolutionary processes shaping diversity in *Ae. umbellulata* and offer avenues for its use in bread wheat breeding.

Results

Presence/absence and SNP genotypings using GRAS-Di amplicon reads

A total of 4,956,372–9,075,806 (mean 6,498,074) raw GRAS-Di amplicon reads were obtained for each of the 199 *Ae. umbellulata* accession (Fig. S1; Table S1 and S2). Based on these reads, the presence/absence and SNP genotypes were generated. For the presence/absence genotypes, a total of 196,349 markers were identified, including 139,019 markers of Quality A, 11,507 of Quality B, 17,921 of Quality C, and 27,902 of Quality D (Table S3). In some cases, multiple markers exhibited identical presence/absence patterns across the 199 accessions. The number of unique presence/absence patterns was 19,235 for

Quality A markers, 8,922 for Quality B markers, 13,881 for Quality C markers, and 21,872 for Quality D markers (Table S3). In total, 37,301 unique patterns were identified among the Quality A, B, and C markers. Based on these results, a total of 168,447 Quality A, B, and C markers were retained for subsequent analyses. No accession-wise missing data were observed for any marker. For SNP genotypes, 12,240 pruned biallelic sites were identified: 1,308 on chromosome 1U, 1,820 on chromosome 2U, 1,764 on chromosome 3U, 1,683 on chromosome 4U, 1,928 on chromosome 5U, 1,687 on chromosome 6U, and 2,050 on chromosome 7U. These SNPs were distributed across the genome, although their distribution was not uniform (Fig. S2).

Property comparison of the presence/absence and SNP genotype datasets

To compare the shapes of *Ae. umbellulata* genetic structure inferred from the presence/absence and SNP datasets, the phylogenetic trees were first examined. In the maximum likelihood (ML) tree based on the presence/absence genotypes, a distinct cluster comprising 25 accessions was identified with strong bootstrap support (1.00) (Fig. S3). A similarly well-supported cluster (bootstrap support = 1.00) containing these 25 accessions and an additional accession, KU-12701, was recovered from the ML tree based on the SNP genotypes (Fig. 2a). The 25 accessions were collected from the far western part of the species range, including the Greek Islands and western Turkey, apart from accession KU-5901 from Syria. The additional accession, KU-12701, originated from Cyprus.

Next, population structure patterns for each genotype dataset were examined. For each dataset, the most plausible number of ancestral populations (K) was estimated based on cross-entropy scores. Specifically, sNMF was run for K values from 2 to 10, and $K = 3$ was selected because it exhibited a sharp decrease in the cross-entropy score, followed by only marginal decreases with higher K values (Fig. S4). At $K = 3$, the 25 accessions that formed a distinct

cluster in the phylogenetic trees exhibited high q-values for a single ancestral population in both genotype datasets (q-value > 0.76 in the presence/absence dataset and > 0.67 in the SNP dataset) (Table S4). In addition, the q-values of individual accessions were highly consistent between the two datasets, with a Pearson's correlation coefficient of 0.98 (Table S4). These results indicated that the presence/absence and SNP datasets captured essentially the same patterns of genetic variation in the species. Because the SNP dataset also provided physical map information, it was the focus in the subsequent procedures.

Lineage structure in the 199 *Ae. umbellulata* accessions

The analyses described above identified two major groups of accessions: the 26 accessions originating from the far western part of the species' range, including the Greek islands, Cyprus, and western Turkey, and the remaining 173 accessions originating from Turkey, northern Iraq, northwestern Iran, Syria, and Azerbaijan. The same grouping was evident in the principal component analysis (PCA) based on the SNP dataset (Fig. 3). Based on these findings, the group of 173 and 26 accessions were designated as lineages UmbL1 and UmbL2, respectively.

Population structure analysis ($K = 3$) provided further insights into lineage structure (Fig. 2a). Within UmbL1, two distinct genetic groups of accessions were identified: one comprising accessions (53 accessions) from central Turkey with q-values greater than 0.7 for ancestral population 2 (P2), and the other comprising accessions (78 accessions) from far-eastern Turkey, northern Iraq, and northwestern Iran with q-values greater than 0.7 for ancestral population 3 (P3). The differentiation between these groups was relatively weak, as several accessions were genetically and geographically intermediate, with q-values below 0.7 for both P2 and P3. Based on these findings, the former group was designated as sublineage UmbL1w and the latter group as sublineage UmbL1e (Table S1). The genetically and

geographically intermediate accessions between UmbL1w and UmbL1e were further designated as Intermediate-e (23 accessions). Intermediate-e accessions occurred in the eastern region (longitude ≥ 36.0°), except for KU-2779 (longitude = 35.3°). In addition, a second group of genetically and geographically intermediate accessions was identified between UmbL1w and UmbL2, which was designated as Intermediate-w (18 accessions). Accessions of Intermediate-w occurred in the western region (longitude < 36.0°). This intra-lineage structure of UmbL1 was consistent with the pattern observed in the PCA plot (Fig. 3). Notably, one accession, KU-12189, originated from a region where UmbL2 was predominant; however, it exhibited a genotype intermediate between UmbL1e and UmbL2.

Phenotypic variation patterns and their association with lineage divergence and environmental factors

The 199 *Ae. umbellulata* accessions exhibited a wide range of phenotypic variation across the 24 phenological and morphological traits examined (Table 1, Fig. 4, S5, and S6). Although the measured values substantially overlapped among the lineages/sublineages across the traits examined, phenological differences were particularly pronounced in heading and flowering time; the UmbL1w accessions tended to flower late, whereas the UmbL1e and UmbL2 accessions tended to flower early. Similarly, the UmbL1w accessions exhibited particularly higher trichome levels and notably longer awns of the first and third spikelets than the other lineages/sublineages (Fig. 4). Furthermore, the first internode length, flag leaf length, spike length, number of spikelets per spike, spikelet length, spikelet width, and seed shape differed among the lineages/sublineages, although the magnitude of these differences was relatively small. The first internodes and flag leaves were longer in the UmbL1e and UmbL2 accessions than in the UmbL1w accessions. The UmbL2 accessions tended to have shorter spikes, fewer spikelets, and wider spikelets, indicating more compact spikes (Fig. S6). For most of the

examined traits, the Intermediate-w and -e accessions were phenotypically intermediate between UmbL2 and UmbL1w, and UmbL1w and UmbL1e, respectively (Fig. S7).

Next, the relationship between ecologically important reproductive traits, namely flowering time and seed-related traits, and intraspecific lineage divergence was examined. To simplify the analysis, PCA was conducted using four seed-related traits (seed length, seed width, 100-grain weight, and number of spikelets per spike), and the first principal component (PC1) was used as a composite trait. PC1 explained 64.40% of the total variance, with factor loadings of 0.61 for 100-grain weight, 0.58 for seed width, 0.54 for seed length, and −0.10 for number of spikelets per spike. Therefore, PC1 was defined as the seed-related PC. Higher values of seed-related PC indicated greater seed length, seed width, and 100-grain weight, but fewer spikelets per spike. A plot of standardized flowering time against standardized seed-related PC values for the 199 *Ae. umbellulata* accessions showed that 45 UmbL1e accessions (57.69% of all UmbL1e accessions) were located in the second quadrant. This distribution indicated that the UmbL1e accessions tended to flower earlier and produced larger, heavier seeds with fewer spikelets per spike. In contrast, 34 UmbL1w accessions (64.15% of all UmbL1w accessions) were located in the fourth quadrant, indicating a tendency toward later flowering, smaller and lighter seeds, and more spikelets per spike (Fig. 5, Table S5). The UmbL2 accessions tended to have intermediate flowering times and wider seeds with fewer spikelets per spike. Many Intermediate-w and Intermediate-e accessions also exhibited intermediate flowering times. Intermediate-w accessions tended to have slightly larger, heavier seeds and fewer spikelets per spike, whereas Intermediate-e accessions tended to have slightly smaller, lighter seeds and more spikelets per spike (Fig. 5, Table S5).

To elucidate the environmental factors associated with the patterns of phenotypic divergence, a PCA of environmental variables obtained from WorldClim version 2.1 [32] was performed for the collection sites of the 199 *Ae. umbellulata* accessions (Fig. 6a). In the principal

component plot, the environmental conditions of the UmbL1w and UmbL2 accessions were largely separated along PC1, which explained 55.03% of the total variance. Based on the PCA loadings, the major environmental variables contributing to this separation were temperatures and water vapor pressures during both winter and spring and elevation. UmbL2 accessions were distributed in regions with higher temperatures and water vapor pressures during both winter and spring and lower elevations than those of the UmbL1w accessions. Additionally, the environmental conditions of the UmbL1e and UmbL1w accessions were largely separated along PC2, which explained 26.19% of the total variance. PCA loadings indicated that the major environmental variables contributing to this separation were solar radiation during winter and precipitation during both winter and spring. The UmbL1e accessions were distributed in regions with higher winter solar radiation and higher precipitation during both winter and spring than the UmbL1w accessions. Furthermore, many accessions in the Intermediate-e and -w groups tended to be distributed in regions with intermediate environmental conditions between those of the UmbL1e and UmbL1w accessions and between those of the UmbL1w and UmbL2 accessions. Negative correlations were observed between flowering time and both PC1 and PC2 scores, indicating that accessions from warmer, lower-elevation environments and those with higher solar radiation and precipitation tended to flower earlier (Fig. 6b, c).

Genome-wide association studies on the 24 *Ae. umbellulata* traits

Among the 24 analyzed traits, significant GWAS peaks were detected only for the trichome level, with associations identified on chromosomes 3U and 5U (Fig. S8). The most prominent peaks on chromosome 3U were located at 489,982,974 and 488,998,099 bp (Fig. 7a). To explore potential candidate gene regions underlying the trichome-associated region identified in the GWAS, we compared this region with previously reported trichome-associated loci in

bread wheat. In a previous study, four closely linked QTLs for trichome density have been identified on chromosome 3A in 'Chinese Spring' (CS) wheat [33]. This QTL region contains three genes that are highly expressed in leaves: *TraesCS3A02G406000*, predicted to encode an NAC domain-containing protein; *TraesCS3A02G414900*, annotated as a cysteine-rich receptor kinase; and *TraesCS3A02G440900*, predicted to encode a Sec14p-like phosphatidylinositol transfer family protein.

BLASTN searches against the *Ae. umbellulata* reference genome sequence (TA1851), using the three CS wheat gene sequences as queries, identified putative homologous regions on chromosome 3U for *TraesCS3A02G406000* and *TraesCS3A02G440900*. The corresponding *Ae. umbellulata* loci were AeUmb.TA1851.r1.3UG0029890 (475,567,604–475,570,057 bp) and AeUmb.TA1851.r1.3UG0033290 (509,865,136–509,902,297 bp), respectively. Both were located near the GWAS peak SNP sites and contained genes with exon-intron structures similar to the corresponding wheat genes (Fig. 7a, b). For *TraesCS3A02G406000* and AeUmb.TA1851.r1.3UG0029890, both genes contained three exons, and their CDS sequences showed 98.1% sequence identity, with only a single nucleotide substitution detected in the second exon. For *TraesCS3A02G440900* and AeUmb.TA1851.r1.3UG0033290, the wheat query sequence approximately aligned with three gene-like regions within the AeUmb.TA1851.r1.3UG0033290 locus. These results suggest that this locus may contain three gene copies. Among these regions, one (509,865,136–509,870,361 bp) showed the highest CDS sequence identity (97.2%) with the wheat query sequence and corresponded to its four-exon structure. The longest predicted amino acid sequences of AeUmb.TA1851.r1.3UG0029890 and AeUmb.TA1851.r1.3UG0033290 were 299 and 248 aa, respectively. The former was annotated as an NAC domain-containing protein 48 based on sequence similarity to NAC048

from *Oryza sativa* L. subsp. *japonica*, whereas the latter was annotated as a CRAL-TRIO domain-containing protein.

Discussion

Based on the presence/absence and SNP genotypes, the present study identified two major divergent lineages, UmbL1 (173 accessions) and UmbL2 (26 accessions). UmbL1 occurs widely across the species range, from Turkey and Iran to Iraq and Cyprus, whereas UmbL2 occurs mostly on the Greek Islands. UmbL1 was subdivided into UmbL1w (53 accessions) and UmbL1e (78 accessions). This lineage structure was broadly consistent with those reported in previous studies [12, 14]. In addition, two intermediate groups, Intermediate-e (between UmbL1w and UmbL1e, 23 accessions) and Intermediate-w (between UmbL1w and UmbL2, 18 accessions), were identified within UmbL1. This relatively complex lineage structure markedly contrasts with that of *Ae. tauschii*, in which three highly distinct lineages have been identified [34]. A possible explanation for this observation is that *Ae. umbellulata* may have occasionally outcrossed despite its predominantly selfing reproductive system. Therefore, accessions of the intermediate groups may represent descendants of possible inter-sublineage or inter-lineage hybridization. The possibility of occasional outcrossing is consistent with the fact that the U genome is shared with several tetraploid *Aegilops* species that arose through natural hybridization involving *Ae. umbellulata.*

Among the 24 traits evaluated, relatively clear phenotypic divergence among the lineages/sublineages was observed for heading time, flowering time, awn length of the first and third spikelets, and trichome level, whereas the remaining traits showed a relatively weak lineage structure in their patterns of phenotypic variation (Fig. 3, Fig. S6). The pronounced lineage structure during heading and flowering time is consistent with the potential adaptive significance of flowering phenology. UmbL1w accessions tended to exhibit later heading and

312 flowering and were predominantly distributed in habitats characterized by lower temperatures
313 and solar radiation, whereas UmbL1e and UmbL2 accessions tended to exhibit earlier
314 heading and flowering and were associated with habitats characterized by higher solar
315 radiation and temperatures, respectively (Fig. 5). These patterns raise the possibility that
316 variations in flowering phenology contribute to adaptations to contrasting environmental
317 conditions among the lineages/sublineages. Similar geographic variations in flowering time
318 have been reported in another wild wheat relative, *Ae. tauschii*, where longitudinal and
319 latitudinal clines were observed. The longitudinal cline is associated with winter temperature,
320 whereas the latitudinal cline is influenced by environmental factors such as photoperiod
321 length [35]. Together, these findings suggest that flowering phenology is an important
322 component of environmental differentiation in wild wheat relatives and highlight their
323 potential value as a genetic resource for improving flowering-time variation in wheat.

324 The awn lengths of the first and third spikelets also showed a lineage structure, with longer
325 awns generally observed in UmbL1w and UmbL2 and shorter awns in UmbL1e (Fig. 3).
326 Awns have diverse potential functions, including photosynthesis [39] and seed dispersal [40].
327 Awns increase the adhesion of seeds to animal fur, potentially facilitating epizoochorous
328 dispersal [41]. Whether the lineage-associated variation in awn length observed in the present
329 study reflects adaptation remains unclear. Notably, UmbL1w had relatively long awns and a
330 broad geographic distribution across the Anatolian Peninsula (Fig. 1b). Given the occurrence
331 of large herbivores in Anatolia, long awns may facilitate seed dispersal through attachment to
332 animals; however, this possibility remains to be tested.

333 The analysis of reproductive traits showed that within UmbL1, early-flowering accessions,
334 mostly belonging to UmbL1e, tended to produce larger seeds, whereas late-flowering
335 accessions, mostly belonging to UmbL1w, tended to produce smaller seeds (Fig. 4). In
336 contrast, in *Ae. tauschii*, early- and late-flowering accessions tend to produce smaller and

larger seeds, respectively [42]. This contrasting pattern between *Ae. umbellulata* and *Ae. tauschii* may reflect the differences in resource allocation associated with flowering phenology. Early-flowering accessions with a shorter vegetative phase may allocate a greater proportion of resources to individual seeds, whereas late-flowering accessions with a longer vegetative phase may allocate more resources to the development of other organs, such as awns, potentially resulting in smaller seeds. Alternatively, under our field conditions, the period available for grain filling in late-flowering accessions may have been compressed by the onset of terminal heat, despite later flowering, which could have further contributed to their smaller seed size. The contrasting relationship between flowering time and seed size in these two wild wheat species warrants further investigation into the relationship between phenology and resource allocation during reproductive development.

The strong lineage structure observed for the trichome level also suggests that this trait may be associated with environmental differentiation among the lineages/sublineages. Trichomes have diverse functions, including capturing atmospheric moisture and defending against insect herbivores [36–38]. The UmbL1w accessions generally exhibited higher trichome levels and occurred in habitats with lower precipitation, suggesting a possible association between trichome levels and adaptation to relatively dry environments.

Trichome traits are potentially important in wheat breeding because variations in trichome density, length, and morphology can influence plant interactions under both biotic and abiotic stress. Leaf and glume trichomes can provide physical barriers against insect herbivores [36, 37] and may contribute to adaptation to drought, high temperatures, and increased solar radiation by modifying the leaf microenvironment and protecting photosynthetic tissues from excessive radiation [43, 44]. Therefore, variation in trichome characteristics among wild wheat relatives may represent a source of useful alleles for improving stress resistance/tolerance and environmental adaptation in cultivated wheat. Characterizing the

genetic basis of trichome variation in wild relatives could facilitate the identification and utilization of novel alleles in wheat breeding programs.

In the present study, a GWAS of the trichome level, followed by comparative analysis, identified a region on chromosome 3U corresponding to the QTL region on chromosome 3A in wheat. The corresponding 3A region contains three genes that are highly expressed in leaves and have been previously proposed as candidate genes for trichome density [33]. Of these three genes, putative homoeologous genes corresponding to *TraesCS3A02G406000* (predicted to encode an NAC domain-containing protein) and *TraesCS3A02G440900* (predicted to encode a Sec14p-like phosphatidylinositol transfer family protein) were identified within the 3U region. Although the region corresponding to *TraesCS3A02G414900*, annotated as a cysteine-rich receptor kinase, was detected on chromosome 3U by BLAST, no genes were annotated in the corresponding 3U region. Identification of the putative corresponding genes in *Ae. umbellulata* raises the possibility that *TraesCS3A02G406000* and AeUmb.TA1851.r1.3UG0029890, as well as *TraesCS3A02G440900* and AeUmb.TA1851.r1.3UG0033290, may be involved in the regulation of trichome traits. In particular, *TraesCS3A02G406000* and AeUmb.TA1851.r1.3UG0029890 were both annotated as encoding NAC domain-containing proteins, suggesting that they may have similar functions. Therefore, these genes represent potential targets for further investigation of the genetic basis of stress adaptation and trichome traits in wheat. Collectively, these findings highlight the potential of comparative analyses with wild relatives to improve our understanding of the genetic mechanisms underlying agronomically important wheat traits.

Methods

Plant materials

A total of 202 accessions of *Ae. umbellulata* were obtained from the National BioResource Project-Wheat, Kyoto University, Japan. Of these, 201 accessions were collected from Greece (23 accessions), Turkey (104), Syria (2), Iraq (63), Iran (7), Azerbaijan (1), and Cyprus (1) (Table S1, Fig. S1). Locality information was unavailable for one accession (KU-8-3). This set of accessions, representing the entire species range, was the same as that used by Kawahara (2002). For each accession, the latitude and longitude of the collection site were estimated from the locality information using Google Maps (https://maps.google.com/). Since their introduction, all accessions have been maintained through the self-fertilization of a single plant. All accessions were used for trait measurements, whereas all but KU-4065, KU-4085, and KU-4090 were used for genotyping (Table S1). One plant per accession was grown in a pot containing soil under unregulated temperature conditions in a vinyl house at Kobe University, Nada-Ku, Kobe, Japan (34.7° N, 135.2° E). Seeds were sown from mid-November to mid-December, and plants were grown until early June in each of the three growing seasons: 2019–2020, 2020–2021, and 2021–2022.

Genotyping

Total DNA was extracted from the young leaf tissue of a single healthy plant from each of the 199 accessions using the CTAB (cetyltrimethylammonium bromide) method [45]. For each accession, 100-bp paired-end sequencing was performed using the GRAS-Di method, an amplicon sequencing approach that employs a mixture of the total DNA and 12 random polymerase chain reaction primers (Illumina HiSeq 4000, Eurofins Genomics Inc., Tokyo, Japan) [27, 28]. The raw amplicon reads were used for genotyping based on two complementary approaches: ‘presence/absence’ and ‘SNP’.

For the presence/absence approach, the GRAS-Di software version 1.0.4 (TOYOTA, Aichi, Japan) was used to detect the presence or absence of amplicon reads as polymorphic markers.

Marker quality was assessed by Eurofins Genomics Inc., and the markers were classified into five quality categories (A–E, from high to low quality) based on their reliability. Markers classified as Quality E, which exhibited relatively high levels of missing data, were excluded from subsequent analyses.

The SNP approach involved read alignment and SNP calling based on amplicon reads. Quality control was performed using Trimmomatic version 0.39 [46] to remove adapter sequences, trim reads when the average quality within a 4-bp sliding window fell below a Phred score of 30, and discard reads shorter than 50 bp. The filtered reads were aligned to the *Ae. umbelulata* reference genome sequence (TA1851, NCBI assembly: ASM3246443v1, https://www.ebi.ac.uk/ena/browser/view/GCA_032464435.1) [31] using BWA version 0.7.18 [47]. SAMtools version 1.20 [48] was used to remove aligned reads with library insert sizes shorter than 70 bp or longer than 600 bp as well as PCR duplicates. SNP calling was then performed using BCFtools version 1.20 [49] with the options: '-A' (include anomalous read pairs in variant calling), '-B' (disable probabilistic realignment for the computation of base alignment quality), and '-q 20' (minimum mapping quality). The options '-G' (apply the Hardy-Weinberg equilibrium assumption within but not across the populations) and '-m' (alternative model for multiallelic and rare-variant calling) were also used. SNPs were filtered using the criterion '-i DP > = 5 & MQ > = 40'. Subsequently, biallelic SNPs were selected using the view command with the option '-v snps -m 2 -M 2'. Sites with a minor allele frequency (MAF) below 0.05 and missing data frequency above 0.2 were further removed using VCFtools version 0.1.16 [50]. The remaining SNPs were pruned using PLINK version 1.90b7.2 [51] with the command '--indep-pairwise 20000 2000 0.5'.

Genetic analyses

ML phylogenetic analyses of the 199 *Ae. umbellulata* accessions were performed using the presence/absence and SNP datasets in Molecular Evolutionary Genetics Analysis (MEGA) version 10.2.6 [52]. Genetic distances between accessions were calculated using the Kimura 2-parameter model [53]. Bootstrap support values for the branches were calculated from 1,000 replicates.

Population structure analysis was performed using sNMF software version 1.2 [54]. The number of ancestral populations (*K*) was set from 1 to 10, and the most plausible *K* was assessed using the cross-entropy criterion, which measures the ability of the algorithm to correctly impute masked genotypes. Lower cross-entropy values generally indicate better performance. In this study, the *K* value at which the cross-entropy score showed a sharp decrease followed by only marginal decreases with higher *K* was selected as the optimal *K*. Ancestry coefficients were estimated for each accession at each *K*, representing the proportion of ancestry assigned to each inferred population; these coefficients are hereafter referred to as q-values. Composite figures combining phylogenetic trees and population structure bar plots were created using the Interactive Tree of Life (iTOL) tool version 7.6 [55].

Trait measurements

A total of 24 traits—heading time, flowering time, plant height, culm length, spike length (as the dispersal unit), length of the part above the spike neck (including spike length), number of spikelets per spike, flag leaf length, flag leaf width, first internode length, awn length of the first, third, and fifth spikelets, trichome level, width and length of the first, second, and third spikelets, 100-grain weight, seed length, seed width, and seed length-to-width ratio—were evaluated over three growing seasons. All traits, except 100-grain weight, seed length, seed width, and seed length-to-width ratio, were assessed using three spikes per plant. For the 100-

grain weight, the weight of 25 grains per plant was measured and multiplied by four. Seed length and width were measured for 25 grains per plant using *SmartGrain* software version 1.2 [56]. The seed length-to-width ratio was calculated by dividing the seed length by the seed width using measurements obtained with *SmartGrain* software. Heading and flowering times were defined as the number of days from sowing to heading and flowering, respectively. Measurements were conducted between 9:00 and 10:30 a.m. Trichome levels were assessed using a tactile test, in which the degree of trichome development on the surface of the flag leaf was rated on a three-level scale (1–3), with higher scores indicating greater trichome development.

To account for variation among accessions and years, linear mixed models (LMMs) were evaluated for each of the 23 traits, excluding the trichome level. Candidate models were compared based on the Akaike's information criterion (AIC) using ML (restricted maximum likelihood [REML] = FALSE), and the model with the lowest AIC was selected for each trait. For 22 traits, excluding 100-grain weight, the selected model was 'trait ~ accession + (1 | year) + (1 | year:accession)', where accession was included as a fixed effect, while year and the year × accession interaction were included as random effects. After model selection, the selected model for each trait was refitted using REML, and the estimated marginal means (EMMs) for each accession were calculated. These EMMs, which accounted for variation among years and accession-by-year combinations, were used as accession-level trait values in subsequent analyses.

For 100-grain weight, the model that included accession as a fixed effect had a higher AIC (1039.74) than the corresponding null model (1019.67) and was therefore not selected. Ordinal mixed models were evaluated for the trichome level, which was recorded as an ordinal categorical trait with three levels; however, these models could not be reliably estimated. In addition, trichome levels showed little variation among years. Therefore,

model-based adjustment was not applied to these two traits, and the arithmetic mean across the three years was calculated for each accession and used as the accession-level trait value in subsequent analyses.

All statistical analyses were performed in R version 4.6.1 [57] using the "lme4" package version 2.0.6 [58] for fitting LMMs and the "emmeans" package version 2.0.4 [59] for calculating EMMs. Data visualization was performed using the "ggplot2" package version 4.0.3 [60].

Environmental data

Environmental datasets for the collection sites of the 199 *Ae. umbellulata* accessions were obtained from the WorldClim version 2.1 database (1970–2000; https://www.worldclim.org) [32]. The dataset contained the monthly mean temperature, precipitation, solar radiation, and water vapor pressure. To represent the growing season of *Ae. umbellulata*, the mean values were calculated for winter (November to February) and spring (March to June). Elevation data were obtained from the Shuttle Radar Topography Mission elevation dataset provided by WorldClim version 2.1.

PCA

PCA of the SNP, trait, and environmental datasets were conducted using the R package "pcaMethods" version 2.4.0 [61]. The SNP dataset was mean-centered but not scaled and was analyzed using the probabilistic PCA (PPCA) method with the pca() function. The 23 trait variables, excluding the seed length-to-width ratio, and nine environmental variables were mean-centered and scaled to unit variance (UV scaling) and analyzed using the singular value decomposition method with the pca function. Four selected seed-related traits (seed length,

seed width, 100-grain weight, and number of spikelets per spike) were mean-centered, scaled to unit variance, and analyzed using the pca function. PC1 score was used as the seed-related PC.

GWAS

GWAS was conducted using the R package GAPIT version 3.4 [62]. A mixed linear model (MLM) was used, incorporating population structure as a fixed effect and kinship among accessions as a random effect. The kinship matrix accounted for the genetic relatedness among the accessions in the model. The data for all 24 traits were used as phenotypic data, and the SNP dataset was used as the marker dataset. The q-values obtained from the sNMF analysis ($K = 3$) were used as covariates to account for population structure. For candidate genes associated with trichome-related GWAS signals, sequence and annotation data were obtained from Ensembl Plants [63] and GrainGenes [64]. BLASTN searches [65] were conducted against *Ae. umbellulata* reference genome sequence (TA1851), using CS wheat gene sequences as queries to identify the corresponding regions on chromosome 3U. Exon-intron structures of the identified wheat and *Ae. umbellulata* genes were compared and visualized using SeqView [66].

Data availability

The phenotypic, GRAS-Di marker, and SNP genotype data generated in this study are deposited in Zenodo and available to the editors and reviewers via a private link (https://zenodo.org/records/23182934?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjQyZDNjZjBkLTM5YzItNDNkZS1hMzVjLTAwNGY0OWEwYWJkMSIsImRhdGEiOnt9LCJyYW5kb20iOiIwNDQ1ODg4OGU0NmE0MmM0ODFiMDk5NGI3YjgyMDBjNiJ9.nT_cpueoK8Y0tybViEEP_InBch_PkG3MREjWuDFDTVk8fbj9herw-

ry4xO_8Z8atKi8myjoNgNx1_Pdox5i94g). The datasets will be made publicly available upon publication. The raw GRAS-Di sequencing data are deposited in the DNA Data Bank of Japan (DDBJ) Sequence Read Archive under BioProject accession PRJDB45878.

References

1. van Slageren, M. W. *Wild Wheats: A Monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae).* **94-7**, 1–512 (ICARDA/Wageningen Agricultural University, Wageningen, 1994).
2. Kilian, B. et al. Aegilops. *in Wild Crop Relatives: Genomic and Breeding Resources* (ed. Kole, C.) 1–76 (Springer, Berlin, Heidelberg, 2011). https://doi.org/10.1007/978-3-642-14228-4_1
3. Kihara, H. Considerations on the Evolution and distribution of *Aegilops* species based on the analyser-method. *Cytologia* **19**:336-357.19.336 (1954). https://doi.org/10.1508/cytologia
4. Kimber, G. & Tsunewaki, K. Genome symbols and plasma types in the wheat group. *Proc. 7th International Wheat Genetics Symposium* 2, 1209–1210 (Cambridge, 1988).
5. Tsunewaki, K. Plasmon analysis in the *Triticum-Aegilops* complex. *Breed. Sci.* **59**:455-470 (2009).
6. Tsunewaki, K. Erratum to: “Plasmon analysis in the *Triticum-Aegilops* complex.”. *Breed. Sci.* **60**:177–178 (2010).
7. Eig, A. Monographisch-Kritische Übersicht der Gatteung *Aegilops*. *Repert. Spec. Nov. Regni Veg. Beih.* **55**: 1–228 (1929).
8. Hammer, K. Vorarbeiten zur monographischen Darstellung von Wildp anzensortimenten: *Aegilops* L. Die Kulturp anze 28: 33–80 (in German with English summary). (1980).

9. Friebe, B., Jiang, J., Tuleen, N. & Gill, B. S. Standard karyotype of *Triticum umbellulatum* and the characterization of derived chromosome addition and translocation lines in common wheat. *Theor. Appl. Genet.* **90**: 150–156 (1995). https://doi.org/10.1007/BF00221010

10. Hosseini, F., Aghaei, M. J., Vaezi, Sh. & Shahli, M. K. S. Karyotypic diversity in *Aegilops umbellulata* collection of Iran. *Iran. J. Rangelands For. Plant Breed. Genet. Res.* **21**(1): 140-148 (in Percian with English summary) (2013).

11. Song, Z. et al. Analysis of structural genomic diversity in *Aegilops umbellulata, Ae. markgrafii, Ae. comosa, and Ae. uniaristata* by fluorescence in situ hybridization karyotyping. *Front. Plant Sci.* **11**. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2020.00710 (2020). https://doi.org/10.3389/fpls.2020.00710

12. Kawahara, T. Morphological and isozyme variation in genebank accessionsof *Aegilops umbellulata* Zhuk., a wild relative of wheat. *Genet. Resour. Crop Evol.* **49**:89–94 (2002). https://doi.org/10.1023/A:1013895912971

13. Okada, M. et al. RNA-seq analysis reveals considerable genetic diversity and provides genetic markers saturating all chromosomes in the diploid wild wheat relative *Aegilops umbellulata*. *BMC Plant Biol.* **18**, 271 (2018). https://doi.org/10.1186/s12870-018-1498-8

14. Son, I., Kasazumi, N., Okada, M., Takumi, S. & Yoshida, K. Discrepancy of flowering time between genetically close sublineages of *Aegilops umbellulata* Zhuk. *Sci. Rep.* **14**, 7437 (2024). https://doi.org/10.1038/s41598-024-57935-w

15. FAO. FAOSTAT: Data: Production: Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL (2025).

16. United Nations. GLOBAL ISSUES: Population. https://www.un.org/en/global-issues/population (2025).

17. Li, C., Camac, J., Robinson, A. & Kompas, T. Predicting changes in agricultural yields under climate change scenarios and their implications for global food security. *Sci. Rep.* **15**, 2858 (2025). https://doi.org/10.1038/s41598-025-87047-y
18. Lu, M. Climate change and wheat production: A comprehensive review of challenges and adaptation strategies. *Resources Data Journal* **4**:183-186.4.0_183 (2025). https://doi.org/10.50908/rdj
19. Gao, L. et al. The *Aegilops ventricosa* 2NvS segment in bread wheat: cytology, genomics and breeding. *Theor. Appl. Genet.* **134**: 529–542 (2021). https://doi.org/10.1007/s00122-020-03712-y
20. Doussinault, G., Delibes, A., Sanchez-Monge, R. & Gracia-Olmedo. Transfer of a dominant gene for resistance to eyespot disease from a wild grass to hexaploid wheat. *Nature* **303**: 698–700 (1983). https://doi.org/10.1038/303698a0
21. Subbarao, G. V. et al. Enlisting wild grass genes to combat nitrification in wheat farming: A nature-based solution. *Proc. Natl Acad. Sci. USA* **118,** e2106595118 (2021). https://doi.org/10.1073/pnas.2106595118
22. Chhuneja, P. et al. Transfer of leaf rust and stripe rust resistance from *Aegilops umbellulata* Zhuk. to bread wheat (*Triticum aestivum* L.). *Genet. Resour. Crop Evol.* **55**:849–859 (2008). https://doi.org/10.1007/s10722-007-9289-3
23. Edae, E. A., Olivera, P. D., Jin, Y., Poland, J. A. & Rouse, M. N. Genotype-by-sequencing facilitates genetic mapping of a stem rust resistance locus in *Aegilops umbellulata,* a wild relative of cultivated wheat. *BMC Genomics* **17**, 1039 (2016). https://doi.org/10.1186/s12864-016-3370-2
24. Wang, J., Wang, C., Zhen, S., Li, X. & Yan, Y. Low-molecular-weight glutenin subunits from the 1U genome of *Aegilops umbellulata* confer superior dough rheological properties and improve breadmaking quality of bread wheat. *J. Sci. Food Agric.* **98**, 2156–2167 (2018). https://doi.org/10.1002/jsfa.8700

25. Rampino, P., Pascali, M. D., Perrotta, C. & Gullì, M. New gene functions are involved in the thermotolerance of the wild wheat relative *Aegilops umbellulata*. *Plant Physiol. Biochem.* **156**:115–124.plaphy.2020.09.005 (2020). https://doi.org/10.1016/j
26. Okada, M. et al. Phenotypic effects of the U-genome variation in nascent synthetic hexaploids derived from interspecific crosses between durum wheat and its diploid relative *Aegilops umbellulata*. *PLoS ONE* **15**(4):e0231129.pone.0231129 (2020). https://doi.org/10.1371/journal
27. Enoki, H. & Takeuchi, Y. New genotyping technology, GRAS-Di, using next generation sequencer. *Plant and Animal Genome Conference XXVI*, San Diego, CA, P0153 (2018). https://pag.confex.com/pag/xxvi/meetingapp.cgi/Paper/29067
28. Enoki, H. The construction of psedomolecules of a commercial strawberry by DeNovoMAGIC and new genotyping technology, GRAS-Di. *Plant and Animal Genome Conference XXVII*, San Diego, CA (2019). https://pag.confex.com/pag/xxvii/meetingapp.cgi/Paper/37002
29. Hosoya, S. et al. Random PCR-based genotyping by sequencing technology GRAS-Di (genotyping by random amplicon sequencing, direct) reveals genetic structure of mangrove fishes. *Mol. Ecol. Resour.* **19**:1153–1163 (2019). https://doi.org/10.1111/1755-0998
30. Miki, Y. et al. GRAS-Di system facilitates high-density genetic map construction and QTL identification in recombinant inbred lines of the wheat progenitor *Aegilops tauschii*. *Sci. Rep.* **10**, 21455 (2020). https://doi.org/10.1038/s41598-020-78589-4
31. Abrouk, M. et al. Chromosome-scale assembly of the wild wheat relative *Aegilops umbellulata*. *Sci. Data* **10,** 739. (2023). https://doi.org/10.1038/s41597-023-02658-2
32. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017). https://doi.org/10.1002/joc.5086

637 33. Fan, H. et al. QTL mapping of trichome traits and analysis of candidate genes in leaves
638 of wheat (*Triticum aestivum* L.). *Genes* **15,** 42 (2024).
639 https://doi.org/10.3390/genes15010042
640 34. Matsuoka, Y. et al. Genetic basis for spontaneous hybrid genome doubling during
641 allopolyploid speciation of common wheat shown by natural variation analyses of the
642 paternal species. *PLoS ONE* **8**(8), 1–17 (2013).
643 https://doi.org/10.1371/journal.pone.0068310
644 35. Matsuoka, Y., Takumi, S. & Kawahara, T. Flowering time diversification and dispersal
645 in Central Eurasian wild wheat *Aegilops tauschii* coss.: Genealogical and ecological
646 framework. *PLoS ONE* **3**(9) (2008). https://doi.org/10.1371/journal.pone.0003138
647 36. Correa, L. D. J. et al. A comprehensive analysis of wheat resistance to *Rhopalosiphum*
648 *padi* (Hemiptera: Aphididae) in Brazilian wheat cultivars. *J. Econ. Entomol.* **113**(3),
649 1493–1503 (2020). https://doi.org/10.1093/jee/toaa059
650 37. Gebretsadik, K. G., Zhang, Y. & Chen, J. Screening and evaluation for antixenosis
651 resistance in wheat accessions and varieties to grain aphid, *Sitobion miscanthi*
652 (Takahashi) (Hemiptera: Aphididae). *Plants* **11,** 1094 (2022).
653 https://doi.org/10.3390/plants11081094
654 38. Hakeem, S., Ali, Z., Saddique, M. A. B., Habib-ur-Rahman, M. & Trethowan, R. Leaf
655 prickle hairs and longitudinal grooves help wheat plants capture air moisture as a water-
656 smart strategy for a changing climate. *Planta* **254**(1), 18 (2021).
657 https://doi.org/10.1007/s00425-021-03645-w
658 39. Li, X., Tang, Y., Zhou, C., & Lv, J. Contributions of glume and awn to photosynthesis,
659 ^{14}C assimilates and grain weight in wheat ears under drought stress. *Heliyon*, **9**(10)
660 (2023). https://doi.org/10.1016/j.heliyon.2023.e21136

40. Ntakirutimana, F. et al. Potential effects of awn length variation on seed yield and components, seed dispersal and germination performance in siberian wildrye (*Elymus sibiricus* L.). *Plants* **8,** 561 (2019). https://doi.org/10.3390/plants8120561
41. Couvreur, M., Couvreur, M., Vandenberghe, B., Verheyen, K. & Hermy, M. An experimental assessment of seed adhesivity on animal furs. *Seed Science Research* **14**(2):147-159 (2004). https://doi.org/10.1079/SSR2004164
42. Matsuoka, Y., Takumi, S. & Kawahara, T. 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 (2015). https://doi.org/10.1186/s12862-015-0496-9
43. Pshenichnikova, T. A. et al. Quantitative characteristics of pubescence in wheat (*Triticum aestivum* L.) are associated with photosynthetic parameters under conditions of normal and limited water supply. *Planta* **249**, 839–847 (2019). https://doi.org/10.1007/s00425-018-3049-9
44. He, J. et al. Technological advances in imaging and modelling of leaf structural traits: a review of heat stress in wheat. *Journal of Experimental Botany* **77**(9): 2626–2644 (2026). https://doi.org/10.1093/jxb/eraf070
45. Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A. & Allard, R. W. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **81**(24): 8014-8018.81.24.8014 (1984). https://doi.org/10.1073/pnas.81.24.8014
46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15):2114-20 (2014). https://doi.org/10.1093/bioinformatics/btu170

47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14):1754-60 (2009). https://doi.org/10.1093/bioinformatics/btp324
48. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078–2079 (2009). https://doi.org/10.1093/bioinformatics/btp352
49. Li, H. A statistical framework for SNP calling mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**(21): 2987–2993 (2011). https://doi.org/10.1093/bioinformatics/btr509
50. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**(15): 2156–2158 (2011). https://doi.org/10.1093/bioinformatics/btr330
51. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**(3): 559-575 (2007). https://doi.org/10.1086/519795
52. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary. *genetics* analysis across computing platforms. *Molecular Biology and Evolution* **35**(6): 1547–1549 (2018). https://doi.org/10.1093/molbev/msy096
53. Kimura, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**(2): 111-120 (1980). https://doi.org/10.1007/BF01731581
54. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196** (4): 973-983.113.160572 (2014). https://doi.org/10.1534/genetics.113.160572
55. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52** (W1): W78-W82 (2024). https://doi.org/10.1093/nar/gkae268

56. Tanabata, T., Shibaya, T., Hori, K., Ebana, K. & Yano, M. SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiology* **160,** (4): 1871-1880 (2012). doi: 10.1104/pp.112.205120
57. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2026).
58. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**(1): 1-48 (2015). https://doi.org/10.18637/jss.v067.i01
59. Lenth, R. V. & Piaskowski, J. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 2.0.4 (2026).
60. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. ISBN 978-3-319-24277-4 (2016).
61. Stacklies, W., Redestig, H., Scholz, M., Walther, D. & Selbig, J. pcaMethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23** (9): 1164–1167 (2007). https://doi.org/10.1093/bioinformatics/btm069
62. Wang, J. & Zhang, Z. GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics* **19,** 629–640 (2021). https://doi.org/10.1016/j.gpb.2021.08.005
63. Yates, A. D. et al. Ensembl 2026. *Nucleic Acids Res.* **54,** D1053–D1060 (2026). https://doi.org/10.1093/nar/gkaf1239
64. Yao, E. et al. GrainGenes: genetics, genomes, and pangenomes. *Genetics*, **232** (4), iyaf270 (2026). https://doi.org/10.1093/genetics/iyaf270
65. Altschul, S. F. et al. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990). https://doi.org/10.1016/S0022-2836(05)80360-2

66. Ohtsubo, Y., Maruyama, F., Mitsui, H., Nagata, Y. & Tsuda, M. Complete genome sequence of *Acidovorax* sp. strain KKS102, a polychlorinated-biphenyl degrader. *J. Bacteriol.* **194,** 6970–6971 (2012). https://doi.org/10.1128/jb.01848-12

Acknowledgements

Seeds of *Aegilops umbellulata* accessions used in this study were provided by the National BioResource Project – Wheat (https://shigen.nig.ac.jp/wheat/komugi), Kyoto University, Japan. This work was supported by JSPS KAKENHI Grant Number 25KJ1812. We thank Moeko Okada for assistance in verifying the geographic information and valuable advice and discussion.

Author contributions

IS, NK, and ST generated the GRAS-Di sequencing data. IS evaluated the phenotypes and performed the genetic analyses. NK and KY provided advice on data analysis and contributed to the interpretation of the results. YM provided advice on analytical methods and contributed substantially to the interpretation of the results and revision of the manuscript. All authors, except ST, who passed away, contributed to the preparation and revision of the manuscript and read and approved the final version.

Competing interests

The authors declare no competing interests.

Figure Legends

Fig. 1 Spikes of *Ae. umbellulata*

Two accessions were selected to represent each lineage of *Ae. umbellulata* (see "Lineage structure in the 199 *Ae. umbellulata* accessions"), and the accession names and assigned lineage names are shown below each photograph. According to the seed supplier, the National BioResource Project – Wheat, KU-4010, KU-4075, and KU-5920 are classified as the variety *typica*, whereas KU-4061, KU-4078, and KU-5947 are classified as the variety *pilosa*.

Fig. 2 Genetic and geographic structure of the 199 *Ae. umbellulata* accessions based on the SNP dataset

(a) Maximum likelihood (ML) phylogenetic tree and population structure bar plot of the 199 *Aegilops umbellulata* accessions inferred from the SNP genotypes. Numbers on the branches indicate bootstrap support values ≥ 0.8. The arrow indicates the branch leading to a distinct cluster of 26 accessions with strong bootstrap support (1.00). The bar plot shows q-values for the three putative ancestral populations (P1, P2, and P3) at $K = 3$ estimated using sNMF. (b) Geographic distribution of the 199 accessions. Accessions are indicated by different shapes according to their lineages (UmbL1 and UmbL2) and are color-coded according to their sublineages and intermediate groups. Accession KU-12189 exhibited an intermediate genetic composition that did not correspond to either Intermediate-e or Intermediate-w and is therefore indicated by a separate color. The dotted line encloses the main distribution area of UmbL2 accessions, and the map on the left shows an enlarged view of this area.

Fig. 3 PCA plot of the SNP genotypes of the 199 accessions

Points indicate the scores of each accession based on the first (PC1) and second (PC2) principal components. Accessions are indicated by different shapes according to their

lineages (UmbL1 and UmbL2) and are color-coded according to their sublineages and intermediate groups. The percentages of variance explained by each principal component are shown in parentheses on the corresponding axes.

Fig. 4 Phenotypic variation in five traits showing remarkable intraspecific lineage divergence

Each plot shows heading time (a), flowering time (b), trichome level (c), awn length of the first spikelet (d), and awn length of the third spikelet (e). Individual points represent accession-level trait values, grouped by lineage on the x-axis. For each lineage, the mean is indicated by a larger point, with a 95% confidence interval indicated by an error bar. Accessions are indicated by different shapes according to their lineages (UmbL1 and UmbL2) and are color-coded according to their sublineages and intermediate groups.

Fig. 5 Relationship between flowering time and the seed-related PC in 199 *Ae. umbellulata* accessions

The plot shows the standardized flowering time values against the standardized seed-related PC1 values. The seed-related PC was derived from four seed-related traits: seed length, seed width, 100-grain weight, and number of spikelets per spike. Each point represents an individual accession. Accessions are indicated by different shapes according to their lineages (UmbL1 and UmbL2) and are color-coded according to their sublineages and intermediate groups. Pearson's correlation coefficient (r) is indicated in the upper-left corner.

Fig. 6 PCA biplot of environmental variables at the collection sites of the 199 accessions and relationships between flowering time and environmental PC scores

(a) Points represent the scores of individual accessions on the first (PC1) and second (PC2) principal components, and arrows represent the factor loadings of the environmental variables. The analysis included four climatic variables based on WorldClim version 2.1 (Temp, mean temperature; Prec, mean precipitation; Srad, solar radiation; Vapr, water vapor pressure) and elevation (Elev). Climatic variables were calculated for two seasons: winter (November to February, indicated by "_1") and spring (March to June, indicated by "_2"). The plot shows flowering time against the environmental PC1 (b) and PC2 (c) scores. Accessions are represented by different shapes according to their lineages (UmbL1 and UmbL2) and are color-coded according to their sublineages and intermediate groups. The percentage of variance explained by each principal component is shown in parentheses on the corresponding axis labels. Pearson's correlation coefficients (*r*) are indicated in (b) and (c).

Fig. 7 Genomic locations and exon–intron structures of candidate regions associated with trichome level on chromosome 3U

(a) Schematic representation of the locations of the GWAS peak SNP sites and the corresponding regions identified by BLASTN searches using *TraesCS3A02G406000* and *TraesCS3A02G440900* sequences as queries. The two GWAS peak SNP sites are indicated by red vertical lines. The corresponding *Ae. umbellulata* regions, AeUmb.TA1851.r1.3UG0029890 and AeUmb.TA1851.r1.3UG0033290, are shown below the chromosome 3U overview as blue blocks. (b) Comparison of the exon–intron structures of the wheat query genes and the corresponding *Ae. umbellulata* regions. Blue and red boxes indicate exons of the wheat query genes and *Ae. umbellulata* sequences, respectively. The three gene-like regions identified within AeUmb.TA1851.r1.3UG0033290 are shown separately, and the region showing the highest sequence identity to the wheat query sequence

is indicated as the top hit. Yellow lines indicate sequence alignments between the corresponding regions.

834 **Table 1. Summary of phenotypic traits evaluated in the 202 accessions across three**

835 **growing seasons**

836 Values are shown as the number of years, number of replicates per accession per year, total

837 observations (missing %), mean ± SD, and range.

Trait	Year	Repetition	Total observations (Missing %)	Mean±SD	Range
Heading time (days)	3	3	1806 (0.66%)	158.33 ± 12.2	115–186
Flowering time (days)	3	3	1798 (1.1%)	166.27 ± 11.14	120–190
Plant height (cm)	3	3	1808 (0.55%)	62.64 ± 11.42	22.3–94.92
Culm length (cm)	3	3	1811 (0.39%)	58.92 ± 11.39	10.6–91
Spike length (cm)	3	3	1809 (0.5%)	3.26 ± 0.48	2–5.3
Length of the part above spike neck (cm)	3	3	1809 (0.5%)	3.66 ± 0.52	2.3–5.8
Number of spikelet per spike (number)	3	3	1809 (0.5%)	5.01 ± 0.67	3–7
Length of flag leaf (cm)	3	3	1812 (0.33%)	6.08 ± 1.48	2.31–13.03
Width of flag leaf (mm)	3	3	1812 (0.33%)	4.82 ± 0.77	1.57–7.31
Length of the first internode (cm)	3	3	1812 (0.33%)	17.95 ± 4.24	2.2–50.5
Awn length of the first spikelet (cm)	3	3	1809 (0.5%)	2.68 ± 0.64	0.27–5.1
Awn length of the third spikelet (cm)	3	3	1809 (0.5%)	3.63 ± 0.69	1.81–6
Awn length of the fifth spikelet (cm)	3	3	1809 (0.5%)	3.36 ± 0.62	1.6–5.6
Trichome level	3	3	1812 (0.33%)	1.75 ± 0.84	1–3
Width of the first spikelet (mm)	3	3	1806 (0.66%)	5.01 ± 0.59	3.11–6.83
Length of the first spikelet (mm)	3	3	1806 (0.66%)	12.05 ± 1.26	7.73–16.42
Width of the second spikelet (mm)	3	3	1806 (0.66%)	5.26 ± 0.61	3.15–7.47
Length of the second spikelet (mm)	3	3	1806 (0.66%)	12.26 ± 1.23	6.31–16.8
Width of the third spikelet (mm)	3	3	1805 (0.72%)	4.15 ± 0.95	1.17–6.95
Length of the third spikelet (mm)	3	3	1805 (0.72%)	9.86 ± 1.73	2.26–14.92
100-grain weight (g)	3	1	595 (1.82%)	1.49 ± 0.69	0.06–3.28
Seed length (mm)	3	25	14368 (1.88%)	6.41 ± 1.07	1.97–9.71
Seed width (mm)	3	25	14368 (1.88%)	2.22 ± 0.51	0.71–3.72
Length-to-width ratio of seed	3	25	14368 (1.88%)	2.97 ± 0.47	1.72–5.26

838



KU-4010 (UmbL1e) KU-4061 (UmbL1e)



KU-4075 (Intermediate-e) KU-4078 (Intermediate-e)



KU-5920 (UmbL1w) KU-5947 (UmbL1w)

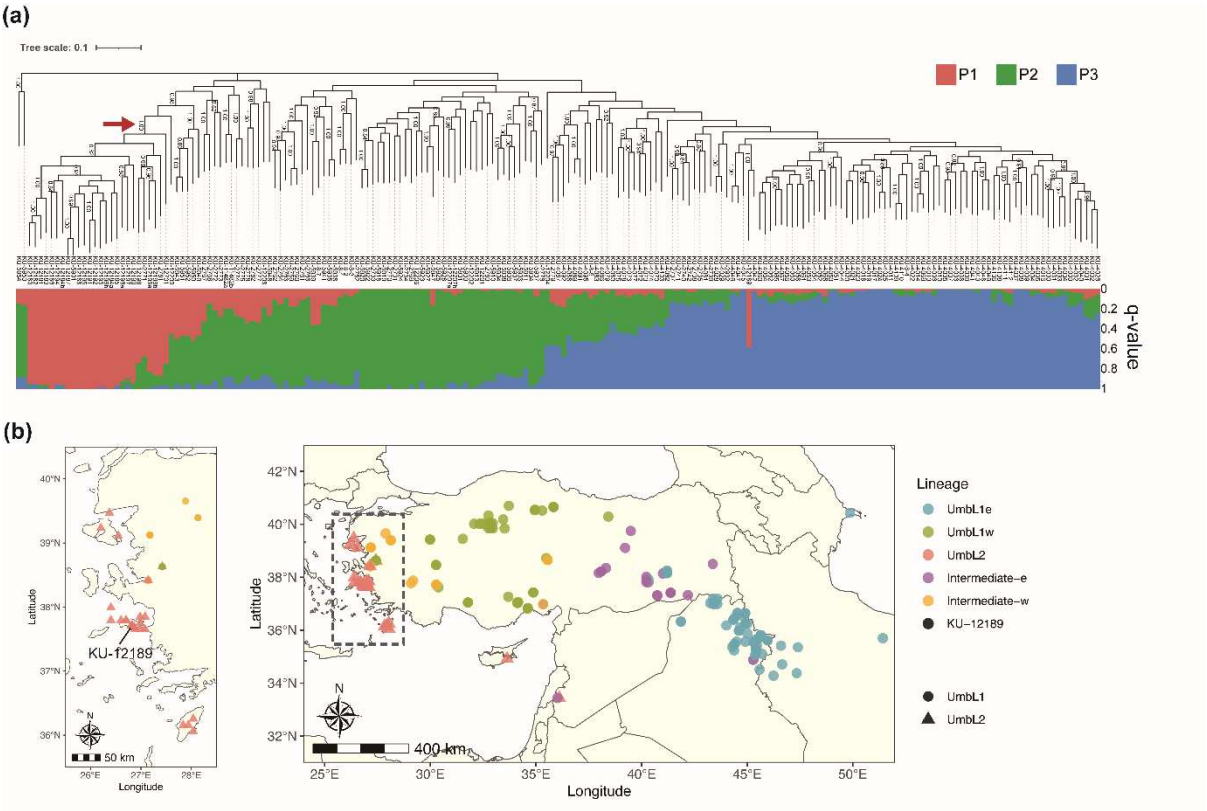


KU-5948 (Intermediate-w) KU-5939 (Intermediate-w)

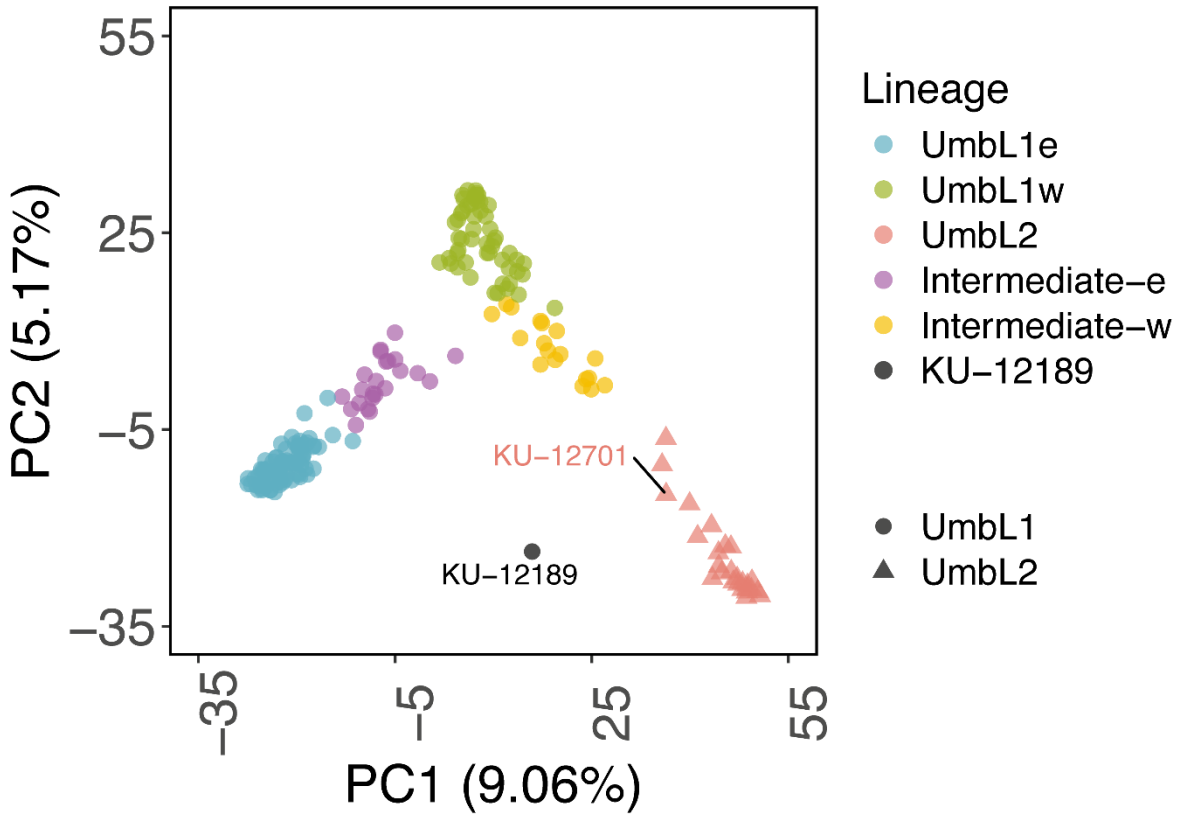


KU-12701 (UmbL2) KU-12185 (UmbL2)

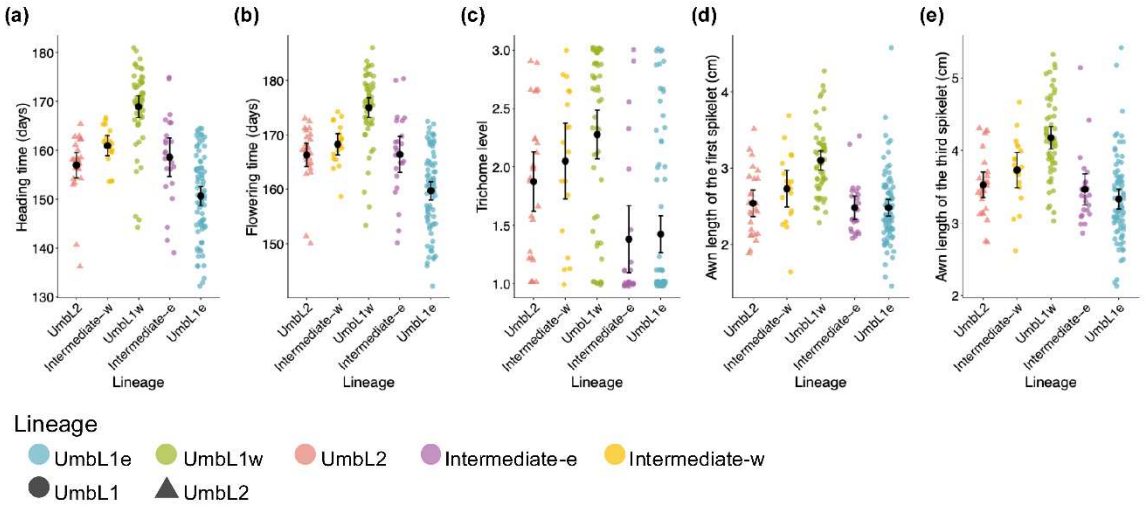
839


(a)
Tree scale: 0.1
P1
P2
P3
q-value
(b)
Lineage
UmbL1e
UmbL1w
UmbL2
Intermediate–e
Intermediate–w
KU–12189
UmbL1
UmbL2
KU–12189
50 km
400 km
Latitude
Longitude

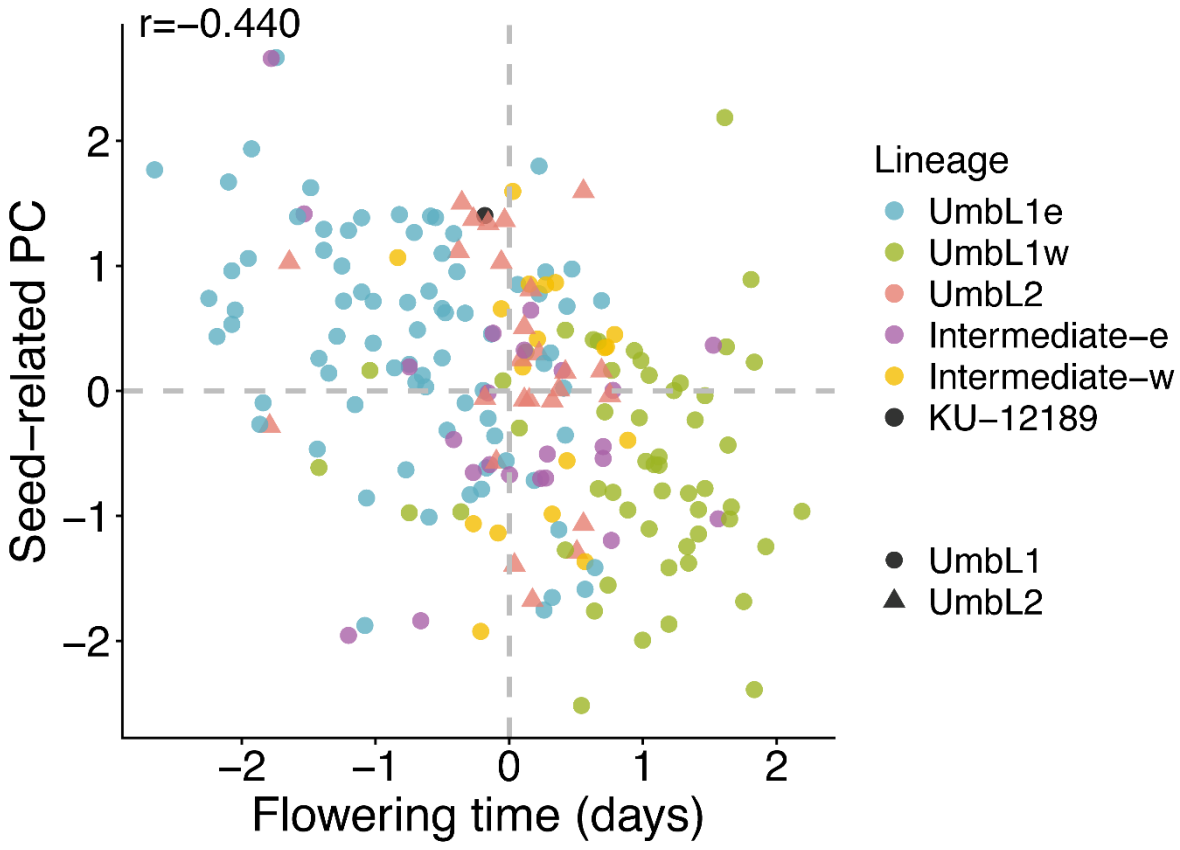
840


PC2 (5.17%)
PC1 (9.06%)
KU–12701
KU–12189
Lineage
UmbL1e
UmbL1w
UmbL2
Intermediate–e
Intermediate–w
KU–12189
UmbL1
UmbL2

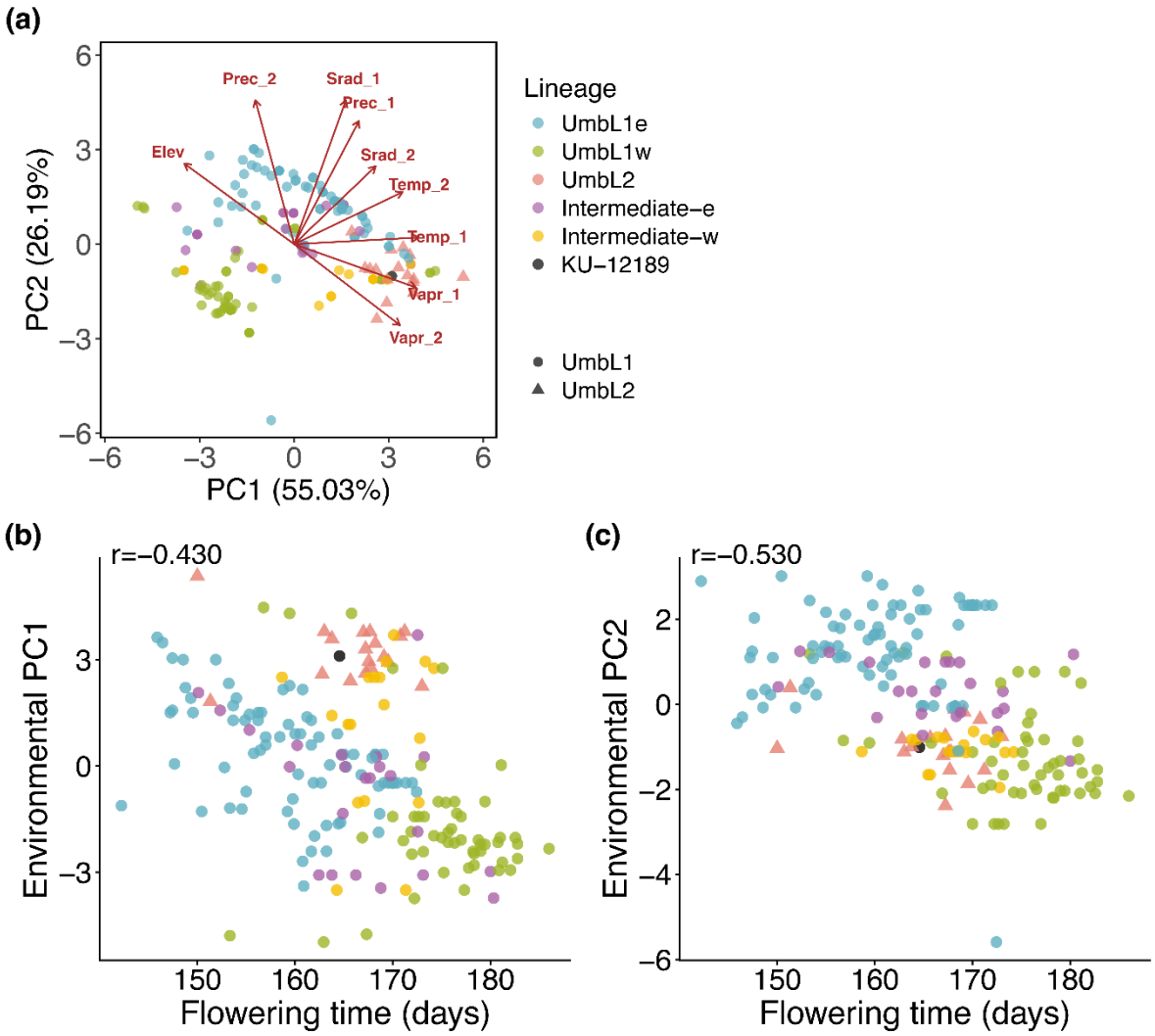
841


(a)
(b)
(c)
(d)
(e)
Heading time (days)
Flowering time (days)
Trichome level
Awn length of the first spikelet (cm)
Awn length of the third spikelet (cm)
UmbL2
Intermediate–w
UmbL1w
Intermediate–e
UmbL1e
Lineage
Lineage
UmbL1e
UmbL1w
UmbL2
Intermediate-e
Intermediate-w
UmbL1
UmbL2

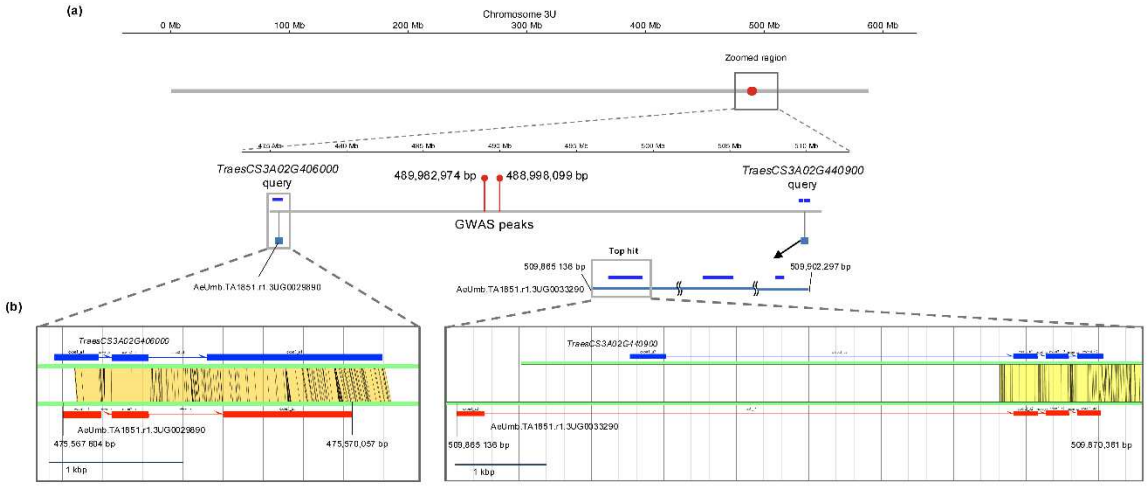
842


r=−0.440
Seed−related PC
Flowering time (days)
Lineage
UmbL1e
UmbL1w
UmbL2
Intermediate−e
Intermediate−w
KU−12189
UmbL1
UmbL2

843


(a)
Prec_2
Srad_1
Prec_1
Elev
Srad_2
Temp_2
Temp_1
Vapr_1
Vapr_2
PC2 (26.19%)
PC1 (55.03%)
Lineage
UmbL1e
UmbL1w
UmbL2
Intermediate−e
Intermediate−w
KU−12189
UmbL1
UmbL2
(b)
r=−0.430
Environmental PC1
Flowering time (days)
(c)
r=−0.530
Environmental PC2
Flowering time (days)

844


(a)
Chromosome 3U
Zoomed region
TraesCS3A02G406000
query
489,982,974 bp
488,998,099 bp
TraesCS3A02G440900
query
GWAS peaks
Top hit
(b)

845

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable.xlsx
- SupplementaryFig.pdf