nature communications

Article

https://doi.org/10.1038/s41467-025-58341-0

Analysis of allohexaploid wheatgrass genome reveals its Y haplome origin in Triticeae and high-altitude adaptation

Received: 8 November 2024

Accepted: 19 March 2025

Published online: 01 April 2025

Check for updates

Yi Xiong[1,10], Shuai Yuan[2,10], Yanli Xiong[1,3,10], Lizuiyue Li[4,5,10], Jinghan Peng[1,10], Jin Zhang[2], Xing Fan[6], Chengzhi Jiang[7], Li-na Sha[1], Zhaoting Wang[1], Xue Peng[1], Zecheng Zhang[1], Qingqing Yu[3], Xiong Lei[3], Zhixiao Dong[1], Yingjie Liu[1], Junming Zhao[1], Guangrong Li[7], Zujun Yang[7], Shangang Jia[8], Daxu Li[3], Ming Sun[9], Shiqie Bai[9] ✉, Jianquan Liu[2] ✉, Yongzhi Yang[2] ✉ & Xiao Ma[1] ✉

Phylogenetic origin of the Y haplome present in allopolyploid Triticeae species remains unknown. Here, we report the 10.47 Gb chromosome-scale genome of allohexaploid *Elymus nutans* (StStYYHH). Phylogenomic analyses reveal that the Y haplome is sister to the clade comprising V and Jv haplomes from *Dasypyrum* and *Thinopyum*. In addition, H haplome from the *Hordeum*-like ancestor, St haplome from the *Pseudoroegneria*-like ancestor and Y haplome are placed in the successively diverged clades. Resequencing data reveal the allopolyploid origins with St, Y, and H haplome combinations in *Elymus*. Population genomic analyses indicate that *E. nutans* has expanded from medium to high/low-altitude regions. Phenotype/environmental association analyses identify *MAPKKK18* promoter mutations reducing its expression, aiding UV-B adaptation in high-altitude populations. These findings enhance understanding of allopolyploid evolution and aid in breeding forage and cereal crops through intergeneric hybridization within Triticeae.

Elymus (wheatgrass or wild rye), the largest genus in the wheat tribe (Triticeae Dum.) of the grass family (Poaceae), encompasses a multitude of perennial grasses that are closely related to staple cereal crops, such as wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.) and rye (*Secale cereal* L.)[1]. According to the widely accepted taxonomic definition[2], the broadly defined genus *Elymus* sensu lato (s.l.) includes >150 miscellaneous species. These wild *Elymus* species have been widely used for ecological restoration and forage and cereal improvement[3] as a valuable tertiary gene pool[4,5]. This genus is composed entirely of allopolyploid species, originating from allopolyploidization events involving five basic haplomes or diploid genomes: St (*Pseudoroegneria*), Y (unknown), H (*Hordeum*), P (*Agropyron*), and W (*Australopyrum*)[6]. The polyploidization process gives rise to a diversity of genomic combinations, all sharing the St-genome, and spans a range of polyploidy levels

[1]College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu, Sichuan 611130, China. [2]State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystem, College of Ecology, Lanzhou University, Lanzhou 730000, China. [3]Sichuan Academy of Grassland Sciences, Chengdu, Sichuan 611700, China. [4]National Plateau Wetlands Research Center, Southwest Forestry University, Kunming 650224, China. [5]Yunnan Key Laboratory of Plateau Wetland Conservation Restoration and Ecological Services, Southwest Forestry University, Kunming 650224, China. [6]Triticeae Research Institute, Sichuan Agricultural University, Chengdu, Sichuan 611130, China. [7]School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China. [8]College of Grassland Science and Technology, China Agricultural University, Beijing 100193, China. [9]School of Life Science and Engineering, Southwest University of Science and Technology, Mianyang, Sichuan 621010, China. [10]These authors contributed equally: Yi Xiong, Shuai Yuan, Yanli Xiong, Lizuiyue Li, Jinghan Peng. ✉e-mail: baishiqie@swust.edu.cn; Liujq@nwipb.ac.cn; yangyz@lzu.edu.cn; maroar@126.com

from tetraploids with StH, StY or StP combinations to hexaploids and octoploids (Supplementary Fig. 1)[7–9]. The Y haplome is present in >75% of the known Asian tetraploids and almost all known Asian hexaploids within *Elymus*[10]. Based on cytogenetic and molecular phylogenetic results, species with a St and Y genomic combination have been separated and recognized as an independent genus, *Roegneria*[6]. Due to the lack of discovery of diploid donor species of the Y haplome in nature to date, these tetraploid *Roegneria* species are considered to be the ancestral species of the StY genome for three distinct hexaploid genera, *Campeiostachys* (StYH), *Kengyilia* (StYP) and *Anthosachne* (StYW), which have also been separated from *Elymus* s.l.[6]. The removal of these allopolyploid genera resulted in the restriction of *Elymus* to the StH haplome combination, *Elymus* s.s.[9]. It is noteworthy that some Asian *Campeiostachys* species with StYH, for example, *C. nutans* = *E. nutans*, are prone to being confused with co-occurring *Elymus* s.s. species with StH genomes, for example, *E. sibiricus*, in their highly similar morphological characters, which may arise from the potential recessive effects of the Y haplome[9]. This was supported by phylogenetic studies of Triticeae species based on FISH paintings and ITS sequences that the Y and St haplomes may have originated from a recent ancestor[11,12]. In contrast, evidence from single-copy nuclear genes such as *Acc1*, *Pgk1* and *EF-G* has challenged this notion and suggested that the Y haplome evolved independently, distinct from the St haplome[6,13]. Therefore, it is crucial to identify which basic diploid genome of Triticeae has the closest evolutionary relationship to the Y haplome.

The Qinghai-Tibet Plateau (QTP), with an average altitude of over 4000 meters and intense UV-B radiation[14], serves as a critical center for species biodiversity of Asian *Elymus* species, with almost half of them thriving there[2,10,15,16]. These species show diverse karyotypes and phenotypic diversity with complex speciation processes and distributional dynamics[8,10,13,17–23]. Furthermore, native *Elymus* species in the QTP region exhibit both short-term phenotype plasticity and long-term genetic adaptation to the challenges of extreme environmental conditions[22,23]. Except for genome-wide association studies (GWAS) using recombinant inbred lines from crossing contrasting phenotypes, genotype-environment association (GEA) has been developed to identify key genes and allelic variations for adaptive local phenotypes with utilization values[24,25]. By integrating GEA and GWAS with key adaptive traits, we can obtain a more comprehensive map of the genomic controls underlying local adaptation. What's more, the annotation of candidate adaptive genes in representative native species of the QTP can provide insights into the traits and signaling pathways involved in adaptation in non-model high altitude organisms.

E. nutans (i.e. *Campeiostachys nutans*), commonly known as drooping wheatgrass, is a perennial, self-compatible and long-awned allohexaploid (2n = 6x = 42, StStYYHH) bunchgrass occurring widely in the QTP[18]. It not only possesses desirable forage characteristics, but is also characterized by its robust tolerance to diverse stress, including drought, coldness and UV-B radiation that are generally lacking in cereal crops[7,21]. These traits are essential for the maintenance and renewal of the soil seed bank in alpine grasslands and sustainability of local ecosystems[26,27]. However, research progress on *E. nutans* has been constrained by the unavailability of a high-quality reference genome, particularly in areas including environmental adaptation mechanisms, genetic resource characterization, molecular breeding development, and interspecific genetic improvement.

In this work, we present a chromosome-scale de novo reference genome assembly for *E. nutans*. We resequence 225 accessions of *E. nutans* and other *Elymus* s.l. species consisting of diverse St, H and Y haplome combinations. By detecting genome-wide single nucleotide polymorphisms (SNPs), our findings determine the evolutionary origin of the Y haplome through detailed comparisons with haplome-known and genome-sequenced representative species in the Triticeae. We also examine multiple reticulate allopolyploidization events in *Elymus* s.l. species. Finally, we infer the demographic dynamics and explore the potential key genes and allelic variations related to high-altitude UV-B adaptation of *E. nutans*.

Results

Assembly of the giant hexaploid *E. nutans* genome

The genome of the hexaploid *E. nutans* wild accession FS20130 (2n = 6x = 42, StStYYHH) was sequenced and assembled. This sample, originating from the QTP in western China, represents the highest altitude recorded for this species, at 5012 meters (Supplementary Fig. 2). To ensure a high-quality assembly, we generated a total of 366.72 Gb (~35.03×) PacBio HiFi reads and 1.21 Tb (~115.62×) Hi-C reads (Supplementary Table 1). The initial contig-level assembly performed using hifiasm[28] results in a comprehensive assembly size of 10.47 Gb, accompanied by an exceptional contig N50 of 76.25 Mb, significantly surpassing the reported assembly for *E. sibiricus*[29]. Utilizing the Hi-C reads, we successfully scaffolded 95.05% of the contigs into 21 chromosomes (Supplementary Table 2; Supplementary Fig. 3a). A BUSCO completeness of 99.50% and an LAI value of 13.97 both indicated high accuracy, completeness, and continuity of our assembly (Supplementary Table 2). A total of 8.67 Gb of TE repetitive sequences were detected, accounting for 83.02% of the total genome (Supplementary Table 3). Among them, LTR repetitive sequences were the most abundant, occupying 62.56% of the genome (Supplementary Table 3), a proportion similar to that observed in maize[30] (62%), barley[31] (68.3%), and wheat[32] (66.9%). Using a combination of transcript evidence, ab initio prediction and homologous alignment, we predicted a total of 120,538 genes, with 91.58% protein-coding genes could be annotated in the gene function databases (Supplementary Table 4).

All chromosomes could be classified into three distinct subgenomes through SubPhaser[33] (Supplementary Fig. 3b; Supplementary Fig. 4). We then revealed the syntenic relationships between these three subgenomes and previously reported genomes containing similar subgenomes, namely barley (H) and *E. sibiricus* (St and H)[9] (Fig. 1a, b; Supplementary Fig. 5). Finally, we have successfully designated the assembled chromosomes as En-St01 to En-St07, En-H01 to En-H07 and En-Y01 to En-Y07 (Fig. 1b; Supplementary Fig. 5). The divergence of the synonymous mutation (Ks) between the three subgenomes indicates that the St and Y subgenomes have differentiated more recently (Supplementary Fig. 6), suggesting a closer genetic relationship between these two subgenomes. We found that the En-St subgenome exhibited the highest collinearity with the St subgenome of *E. sibiricus* and the H subgenome of barley, showing contiguous syntenic relationships without translocations, which may reflect an ancestral state (Fig. 1b; Supplementary Fig. 5). In contrast, the En-H and En-Y subgenomes contained more chromosomal rearrangements. We have detected a significant non-Robertsonian reciprocal translocation (non-RTA) event between the H and Y subgenomes based on both the subgenome-specific k-mers and syntenic analyses (Fig. 1; Supplementary Figs. 4, 6a, and 7). This non-RTA occurred at the long arm of H04 and the short arm of Y03, resulting in En-H04 being composed of the larger segment and centromere of H04 along with a short portion of Y03, while En-Y03 retained the remaining sequences and the centromere of Y03. This non-RTA was validated by our Hi-C connected mapping results (Supplementary Fig. 3a) and a non-denaturing fluorescence in situ hybridization (ND-FISH) assay utilizing probe combinations (Fig. 1c). Additionally, we detected a small non-RTA within Y subgenome when using the En-St genome as a reference, occurring at the ends of Y04 and Y05, resulting in En-Y04 and En-Y05 each containing a small portion of the other (Supplementary Fig. 6a). We also identified that the previously reported non-RTA event between H04 and H06 in *E. sibiricus* is a specific event[29], which is absent in *E. nutans* (Fig. 1b; Supplementary Fig. 5).

Polyploidization-triggered differences in the expression of subgenome/homoeologue genes may occur due to the sub-functionalization

https://doi.org/10.1038/s41467-025-58341-0

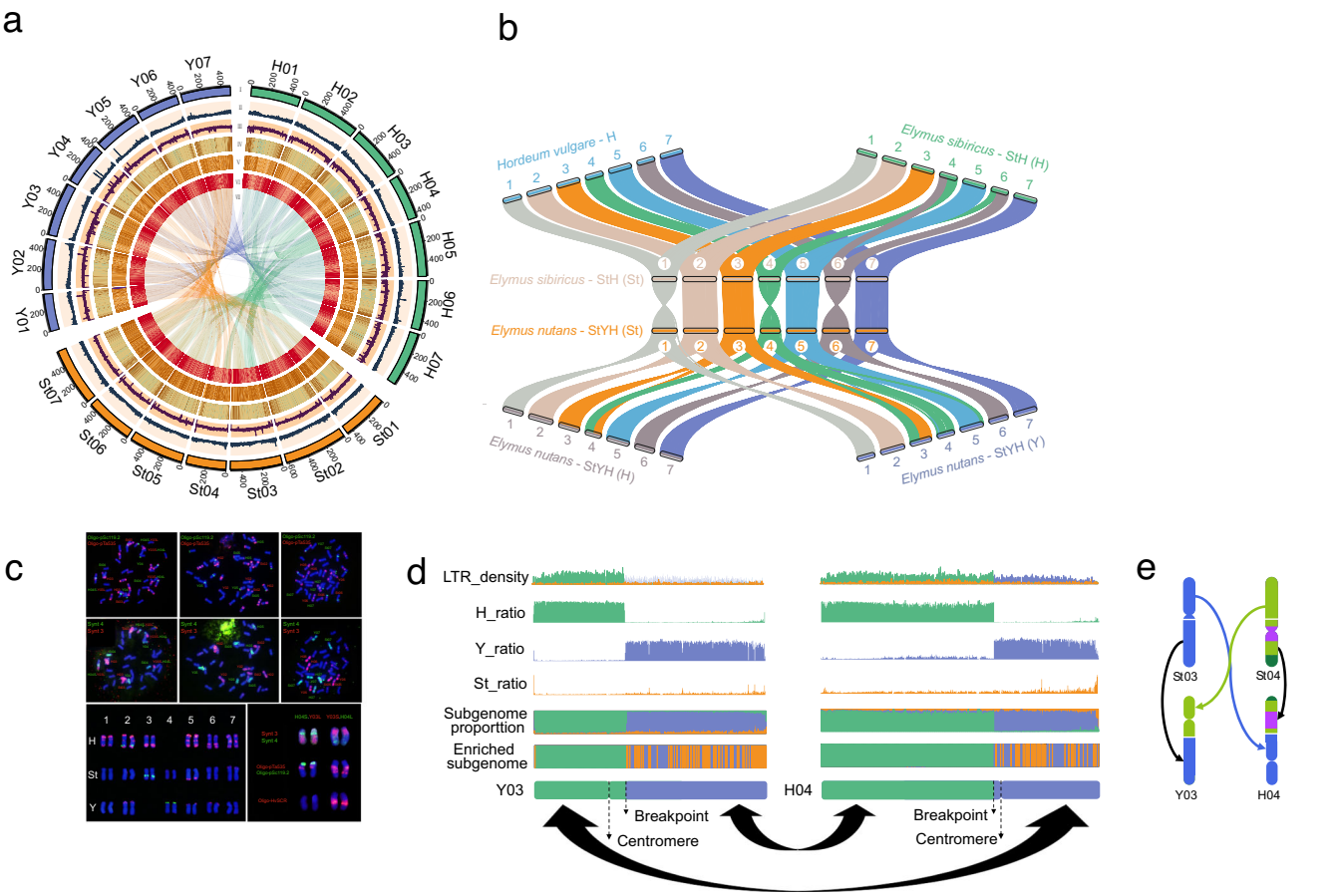


Fig. 1 | Comprehensive analysis of the allohexaploid *Elymus nutans* genome assembly. a The feature of the *E. nutans* genome. I, the 21 chromosomes; II, gene density within a 500 kb sliding window; III, the average GC content in each sliding window (500 kb); IV, the density of LTR retrotransposons; V, the density of Copia retrotransposons; VI, the density of Gypsy retrotransposons; VII, syntenic blocks among the three subgenomes, and chromosomal translocation events were highlighted with colors. **b** The syntenic analysis of the three subgenomes (St, Y, H) of *E. nutans* with barley (H) and *E. sibiricus* (St, H), which possess similar genomic structures. **c** The validation of the non-RTA event between En-Y03 and En-H04 using a FISH assay with three independent experiment. The red and green colors on the chromosomes correspond to the probe colors of the synthetic oligonucleotide libraries. **d**, **e**, A significant non-RTA event between En-Y03 and En-H04 was validated based on specific K-mers (**d**) and karyotype (**e**) analysis. Source data are provided as a Source Data file.

and neo-functionalization process[34–38]. A total of 15,875 homoeologue gene triads that exhibited a 1:1:1 correspondence across the three *E. nutans* subgenomes were detected in the hexaploid *E. nutans* (Supplementary Data 1). We have identified 19,594, 19,802 and 18,040 1:1 homoeologous gene diplets between En-St and En-Y, En-St and En-H, En-H and En-Y, respectively (Supplementary Data 1). We further revealed that missed genes in the third subgenomes of these diplets were mainly experienced gene pseudogenization, in which nearly 60.18%, 51.90% and 51.52% missing genes in St, H and Y subgenomes were pseudogenes, respectively (Supplementary Fig. 8; Supplementary Data 2). We then evaluated the expression biases of each subgenome with these gene triads, in which seven homologous expression patterns were categorized under three abiotic stress conditions (drought, UV-B, and cold). The results showed that the majority of the gene triads (54.34% ~ 62.69%) displayed balanced expression (Supplementary Figs. 9–14; Supplementary Data 3–8), while gene triads with expression dominance in a single subgenome were the least frequent (2.84% ~ 6.47%). The number of gene triads across the seven homologous expression patterns demonstrated a negligible difference (degrees of freedom (df) = 6, $P > 0.05$, chi-square test) between the three stress conditions and their corresponding CK (Supplementary Fig. 13a; Supplementary Fig. 14). Furthermore, no significant difference was detected in the number of up- and down-regulated genes for each subgenome across all abiotic stress conditions (Supplementary Fig. 13b). The GO enrichment results also showed that these differentially expressed genes (DEGs) in different subgenomes had similar functions, indicating no significant function differentiation of the three subgenomes (Supplementary Fig. 15). We further detected the differences in the co-expression networks among the subgenomes and showed that genes from the Y-subgenome were more frequently found in divergent expression modules than their homoeologues from the H- and St-subgenomes (df = 3, $P < 0.05$, chi-square test) (Supplementary Fig. 16; Supplementary Data 9). These results suggest that the expression profiles of the three subgenomes in *E. nutans* are relatively balanced, but the divergent expression modules may change under different conditions, which is similar to hexaploid oats[36]. We also assessed the average Ka/Ks values for these gene triads, which revealed no significant differences in Ka/Ks values between the *E. nutans* subgenomes (Supplementary Fig. 17; Supplementary Data 10). However, for gene pairs with Ka/Ks > 1, the genes in St vs H, St vs Y, and Y vs H showed different gene function (Supplementary Fig. 18; Supplementary Data 11). The divergent expression pattern (sub-functionalization) and the "homologous gene function change (neo-functionalization)" (Supplementary Data 12) of the classical star DEGs (with 1:1:1 gene triads) under each stress condition were also found. What's more, less than half of the DEGs were included in the gene triads (Supplementary Table 5), indicating that most of the DEGs underlying stress tolerance are located in distinct syntenic regions. All these results suggested that most of the genes may have been conserved with less change and there are also some genes that experienced

https://doi.org/10.1038/s41467-025-58341-0

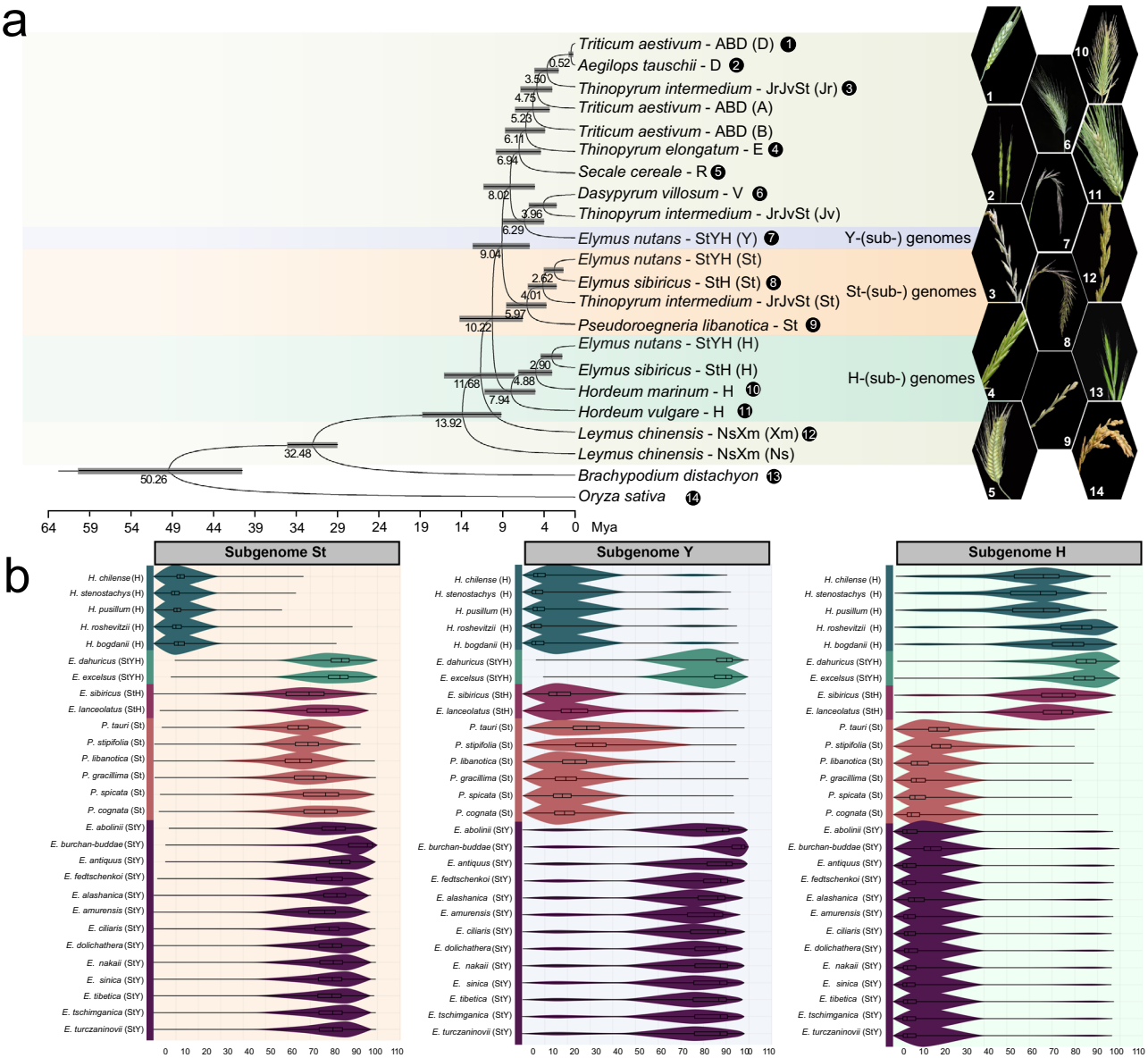


Fig. 2 | The evolutionary status of *E. nutans* among the tribe Triticeae and sequence similarity of the raw reads mapping to the three subgenomes (St/Y/H) of *E. nutans*. **a** The phylogenetic tree is constructed based on the 20 haplomes from 12 species belonging to Triticeae, using rice (*O. sativa*) and *B. distachyon* as outgroups. The gray rectangle bars indicated the timescale between nodes. The right image showcases their spikelets. **b** The sequence similarity analysis was conducted for raw reads of 28 species mapping with St/Y/H haplomes of *E. nutans*. The central box represents the interquartile range, bounded by the first (25th percentile) and third (75th percentile) quartiles. The horizontal line inside the box indicates the median (50th percentile). The whiskers denote the standard deviation of the data, while the shading outside the box plot illustrates the overall dispersion of the dataset. Source data are provided as a Source Data file.

the sub-functionalization and neo-functionalization process in the three *E. nutans* genomes, which may be important for the adaptation of *E. nutans*.

Evolutionary position of St, Y and H haplomes among Triticeae species

We collected 12 species representing 20 haplomes of Triticeae to construct their phylogenetic relationships (Fig. 2a). The results showed that Ns, Xm, H and St haplomes were successively sister to other haplomes, with divergence time of 13.92, 11.68, 10.22 and 9.04 million years ago (Mya), respectively. Notably, we observed that the divergences between H and St haplomes from *E. nutans* and *E. sibiricus* was estimated from 2.62 to 2.90 Mya, which may reflect distinct polyploidization trajectories among them. A key finding is that the En-Y haplome is closely related to the V (*Dasypyrum villosum*) and Jv (*Thinopyum intermedium*) haplomes, with a divergence time of 6.29 Mya. This suggests that the ancestral donor of the Y haplome in *Elymus* s.l. may be phylogenetically proximate to *Dasypyrum* and *Thinopyrum*, thereby challenging the prevailing hypothesis that the Y and St haplomes share a homologous origin with a recent ancestor[11,12]. For the rest haplomes, R haplome emerged at 6.94 Mya and E haplome emerged at 6.11 Mya, and they successively sister to the cluster containing A, B, D and Jr. Within this group, Jr and D showed the closest affinity and diverged at 4.75 Mya. In addition, three ancestral lineages represented by H, St and Y haplomes diverged from 10.22 Mya to 8.02 Mya.

Reticulate allopolyploidization histories of the *Elymus* species involving St, Y and H haplomes

To improve our understanding of the polyploidization history of *Elymus*, we resequenced 28 species from the genera *Pseudoroegneria* (St),

https://doi.org/10.1038/s41467-025-58341-0

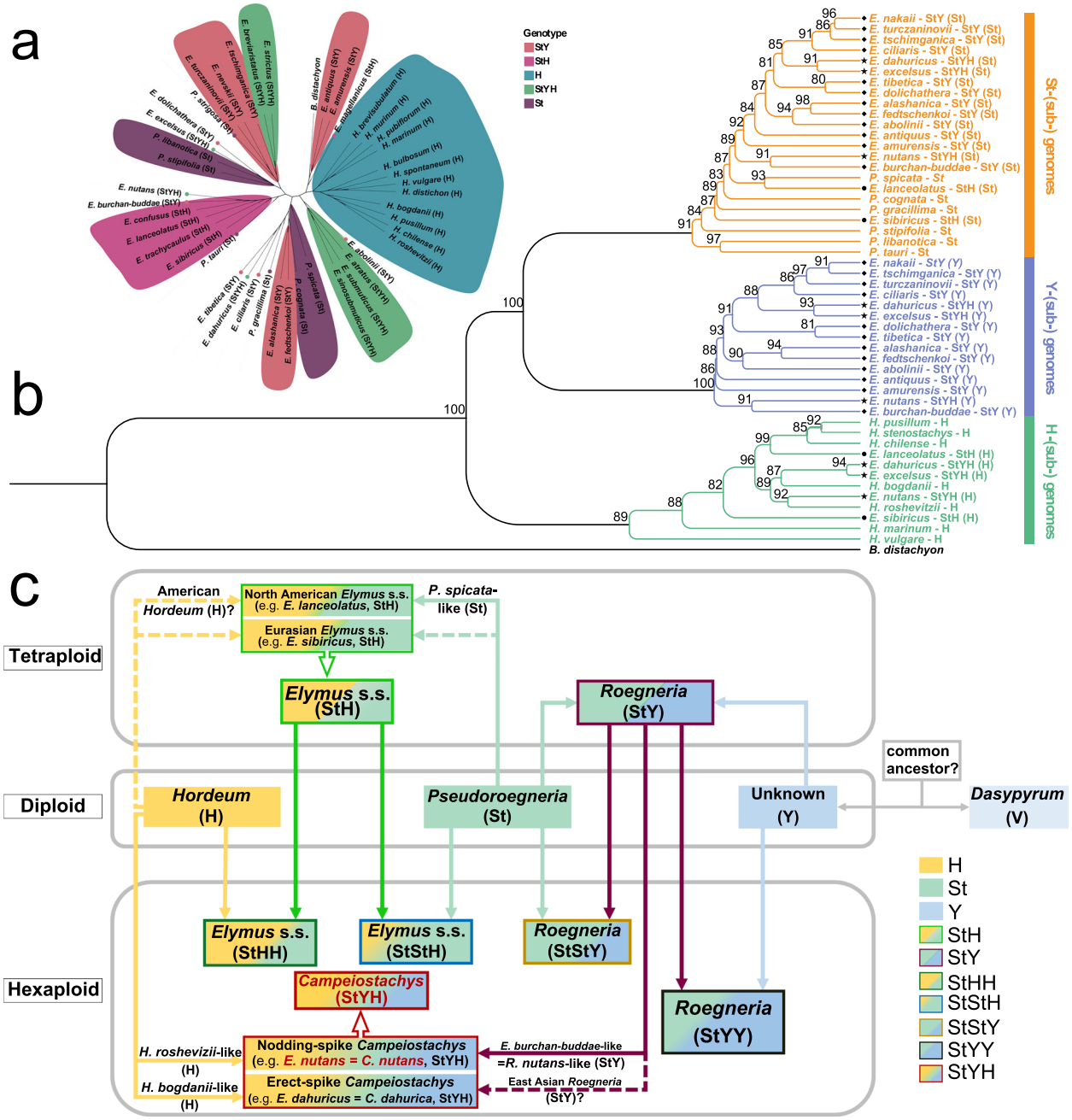


Fig. 3 | Phylogenetics and polyploidization of *Elymus* allopolyploid species involving the St, Y, and H subgenomes. a Phylogenetic tree constructed based on the common plastome genes of 44 relatives of *E. nutans*, with *B. distachyon* as an outgroup. **b** Phylogenetic tree based on 11,607 high quality single copy genes, with *B. distachyon* as an outgroup. Diamond, circle and star represent the genotypes StY, StH and StYH, respectively. **c** Diagram of allopolyploid reticulate evolution model in *Elymus*. Different proportions of background color in each rectangle represent lineages derived from *Hordeum* (H), *Pseudoroegneria* (St), or an unknown (Y) diploid donor species. The color of the rectangular border corresponds to the genome composition of the different polyploid *Elymus* species, including StH (*Elymus* s.s.), StY (*Roegneria*), StStH (*Elymus* s.s.), StHH (*Elymus* s.s.), StStY (*Roegneria*), StYY (*Roegneria*), and StYH (*Campeiostachys*). The arrow points from the donor species to the recipient species, with the dotted arrow representing the undetermined donor species. Source data are provided as a Source Data file.

Hordeum (H), *Roegneria* (StY) and *Elymus* (StH and StYH), with an average sequencing depth of 20×. These resequencing datasets were mapped onto the *E. nutans* genome, and the results confirmed the precise genomic composition of these species, which is consistent with previous reports[9] (Fig. 2b). In particular, we found that *E. alashanica* (=*Roegneria alashanica*) actually has a StY genome constitution, correcting the long doubted haplome composition of this species (St[?] or $StSt^{Y}$)[39]. Mapping statistics revealed that *H. roshevitzii* exhibits the highest homology with the H subgenome (74.74%) of *E. nutans*, while *E. burchan-buddae* (=*Roegneria nutans*, StY) demonstrates the highest homology with the St (90.01%) and Y (85.76%) subgenomes of *E. nutans* (Fig. 2b).

The single copy orthologous genes between St, Y and H were firstly identified in *E. nutans* genome, and then a total of 2046 homologous gene sequences between the different genomes/subgenomes/haplomes of these 28 species were extracted. The phylogenetic results clearly divided these genomes/subgenomes/haplomes into three groups corresponding to the St, Y and H haplomes (Fig. 3). Within each group, allohexaploids or allotetraploids clustered with the different diploids or allotetraploids, suggesting the presence of intricate and

diverse reticulate allopolyploidization histories among these species. We further confirmed that *H. roshevitzii* and the H subgenome of *E. nutans* form a distinct clade, while *E. burchan-buddae* and the St and Y subgenomes of *E. nutans* form another distinct clade, which suggest that the ancestors like them may serve as the potential donors of *E. nutans* (Fig. 3b; Supplementary Fig. 13c). The ancestor like *H. bogdanii* may serve as a donor of the H haplome for erect-spike *Campeiostachys* (such as *E. dahuricus* = *C. dahurica*, StYH), although its donor species of StY was not determined, which differs from the above-mentioned scenario where the ancestor like *H. roshevitzii* may act as a donor of the H haplome for nodding-spike *Campeiostachys* (such as *E. nutans* = *C. nutans*, StYH) (Fig. 3b). Additionally, our findings suggest that the St haplomes of the North American *Elymus* s.s. species, exemplified by *E. lanceolatus* (StH), is likely derived from the ancestor like *P. spicata*, and their H haplome possibly originated from the ancestor similar to American perennial *Hordeum* species, which is consistent with previous implications[40]. However, the ancestor-like donors of St and H haplomes of the *Elymus* s.s., typified by *E. sibiricus* (StH), still remain challenging to ascertain.

The chloroplast genome (plastome) is often used to determine the maternal donor of a plant species because of its highly conservative and maternal inheritance[41]. The newly assembled 29 plastomes and 16 previously reported plastomes[42,43] of *Elymus* s.l. were used to trace the maternal ancestral donors of these species (Supplementary Data 13). Phylogenetic analysis based on the common plastome genes revealed that species with identical genomic compositions exhibited variation in their ancestral donors (Fig. 3a; Supplementary Fig. 19), consistent with the occurrence of multiple origins and extensive reticulate evolution within the *Elymus* s.l.[6,10]. The most striking examples are *E. dahuricus* and *E. excelsus*, both members of the *E. dahuricus* complex[44], clustered together based on the nuclear single-copy gene tree, yet demonstrated different potential maternal donors in the plastome tree. Interestingly, all StYH allohexaploids have a closer relationship with the StY allotetraploids rather than diploid St/H species. These results are consistent with previous study, which suggested that the StY allotetraploids is the direct maternal donor of the StYH allohexaploids, with the diploid perennial *Hordeum* species being the likely paternal donor[6]. Notably, *E. burchan-buddae* is closest to the potential maternal ancestor of *E. nutans* according to the plastome tree (Fig. 3a). This inference is supported by the similar phenotypic morphology, such as nodding spikes, and the overlapping habitats, elevations, and distributions between *E. burchan-buddae* and *E. nutans*. Taken together, these results suggest that the distant ancestors like *E. burchan-buddae* and *H. roshevitzii* may serve as the potential maternal and paternal donors of *E. nutans* through distant hybridization, respectively.

Building on the results presented above and incorporating previous reports[6–9], we have developed a reticulate evolutionary model for allopolyploid species containing the St, Y, and H haplomes within the genus *Elymus* s.l. (Fig. 3c). Specifically, the tetraploid species *Elymus* s.s. (StH) and *Roegneria* (StY), both sharing a common diploid ancestor, namely the St haplome in *Pseudoroegneria*, originated through a process of hybridization and subsequent polyploidization with the H and Y haplomes, respectively. Subsequently, the hexaploid *Elymus* species, characterized by genomic constitutions such as StHH, StStH, StStY, StYY, and StYH, were formed through a similar allopolyploidization process, utilizing the aforementioned tetraploid *Elymus* species (StY and StH) as a foundation. It is important to note that this model calls for further refinement with more detailed information to be completely comprehensive.

Population structure and demographic history

We collected 197 *E. nutans* wild accessions from the QTP and Central Asia across its natural distribution (Fig. 4; Supplementary Data 14). Resequencing these accessions generated totally 21.87 Tb of clean data with an effective depth of 12× (Supplementary Data 14). We identified a set of 43,637,347 high quality SNPs evenly distributed across the genome (Supplementary Fig. 20). Most SNPs were "missense variant" (Supplementary Fig. 21; Supplementary Table 6) and the short-range LD (Supplementary Fig. 22) is less extensive than that of other Poaceae species such as oat[41] (2.29 Mb), wheat[45] (4.2 Mb), and barley[46] (2.97 Mb). NJ tree, PCA and pairwise F_{st} analyses identified three clusters with the low- (Cluster I), medium- (Cluster III), and high- altitude (Cluster II) regions (Fig. 4a–c). We also detected a significant pattern of isolation by distance (IBD) and isolation by environment (IBE) for all Clusters (Supplementary Fig. 23).

We examined demographic divergences of three clusters in *E. nutans* and found the medium-altitude group located in the Hengduan Mountains (Cluster III: SC, GS, and XZ_EAST) firstly diverged from the common ancestor of Cluster I (QH_EAST) and Cluster II (QH_WEST, XZ_WEST, and CA) (Supplementary Figs. 24 and 25). Among 16 possible demographic models, the scenario 16 had the highest importance and post-possibility (Supplementary Fig. 26). The result similarly showed that the medium-altitude Cluster III diverged firstly and the other two high/low-altitude clusters derived later (Fig. 4e; Supplementary Fig. 27). We also found strong gene flow between three clusters (Fig. 4d–f; Supplementary Fig. 27). Thus, we deduced that the medium-altitude Cluster III from Hengduan Mountains region with the highest species diversity[47], might act as the primary origin center of *E. nutans*. This is also supported by its highest nucleotide diversity (π) (Fig. 4d). In addition, the Hengduan Mountains of the QTP, have served as refugia for many plants during glacial ages[47] and they are likely one of the glacial refugia for *E. nutans*, as the cluster there shows an increase in Ne size during both glacial and interglacial periods (Supplementary Fig. 24).

Genetic footprints of environmental adaptation in *E. nutans*

The high-quality reference genome for *E. nutans*, combined with the deep resequencing data generated in this study, enabled the accurate identification of high-confidence genomic SNPs for adaptation locus mining. Using GEA approaches, we identified complex signatures indicative of local adaptation in *E. nutans* based on nine key environmental factors (Supplementary Data 15; Supplementary Fig. 28). A total of 2950 SNPs, involving 340 genes, were found to be significantly associated with one or more environmental factors, widely and evenly distributed across the genome (Supplementary Fig. 29). Of the SNPs significantly associated with environmental factors within the LFMM, 541 (involving 1281 genes) showed a pronounced load along one or more RDA axes (Supplementary Data 16). These SNPs were designated as "key adaptive SNPs" in *E. nutans*. The variants of these genes related to the key environmental factors exhibited varied geographic distributions in allele frequency. To exemplify their geographic distribution patterns, we selected three representative environmental variables to employ the GEA analysis (Fig. 5). For instance, a SNP located within the promoter region of a *NAC* gene (*Nascent polypeptide-associated complex*, *Enu.ptg000038l.G2830*) exhibited a significant association with bio1 (annual mean temperature; Fig. 5a). The low level of LD (Supplementary Fig. 30) in this region suggests that this locus may represent the causal variant. Subsequent analysis revealed that this locus harbors two alleles: the A allele was predominantly observed in the QH_WEST region, characterized by a relatively lower annual mean temperature, whereas the G allele was nearly fixed in regions with a higher annual mean temperature (Fig. 5b). This gene has been reported to preserve the stability of CBF (C-repeat binding factor) protein, thereby enhancing the cold stress resistance of plants[48]. The qRT-PCR result indicated that both alleles responded to cold stress, with allele A demonstrating significantly elevated expression levels at every time point (Fig. 5d), suggesting that individuals carrying allele A in the gene's promoter region may possess stronger tolerance to cold stress. The significant selective signal

https://doi.org/10.1038/s41467-025-58341-0

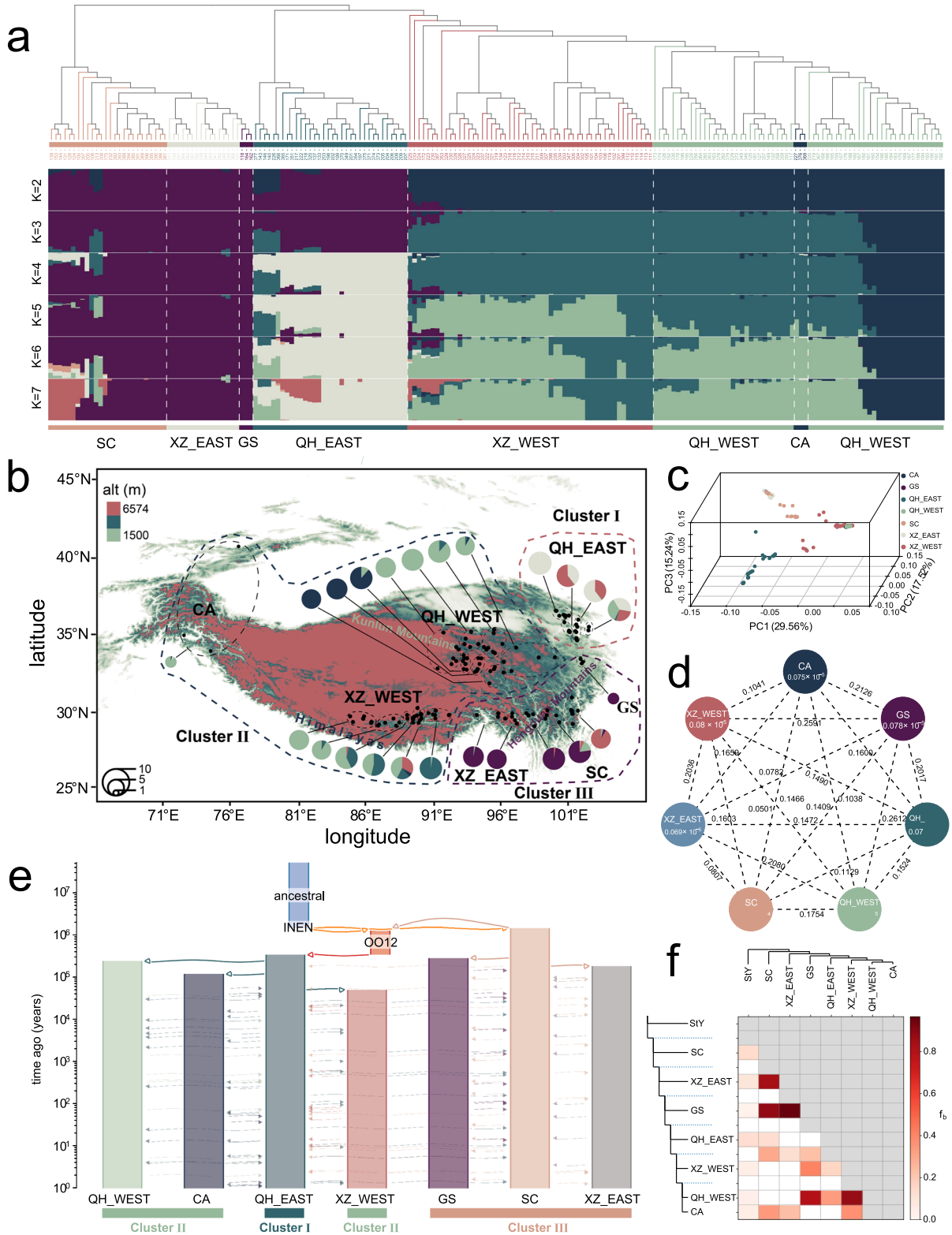
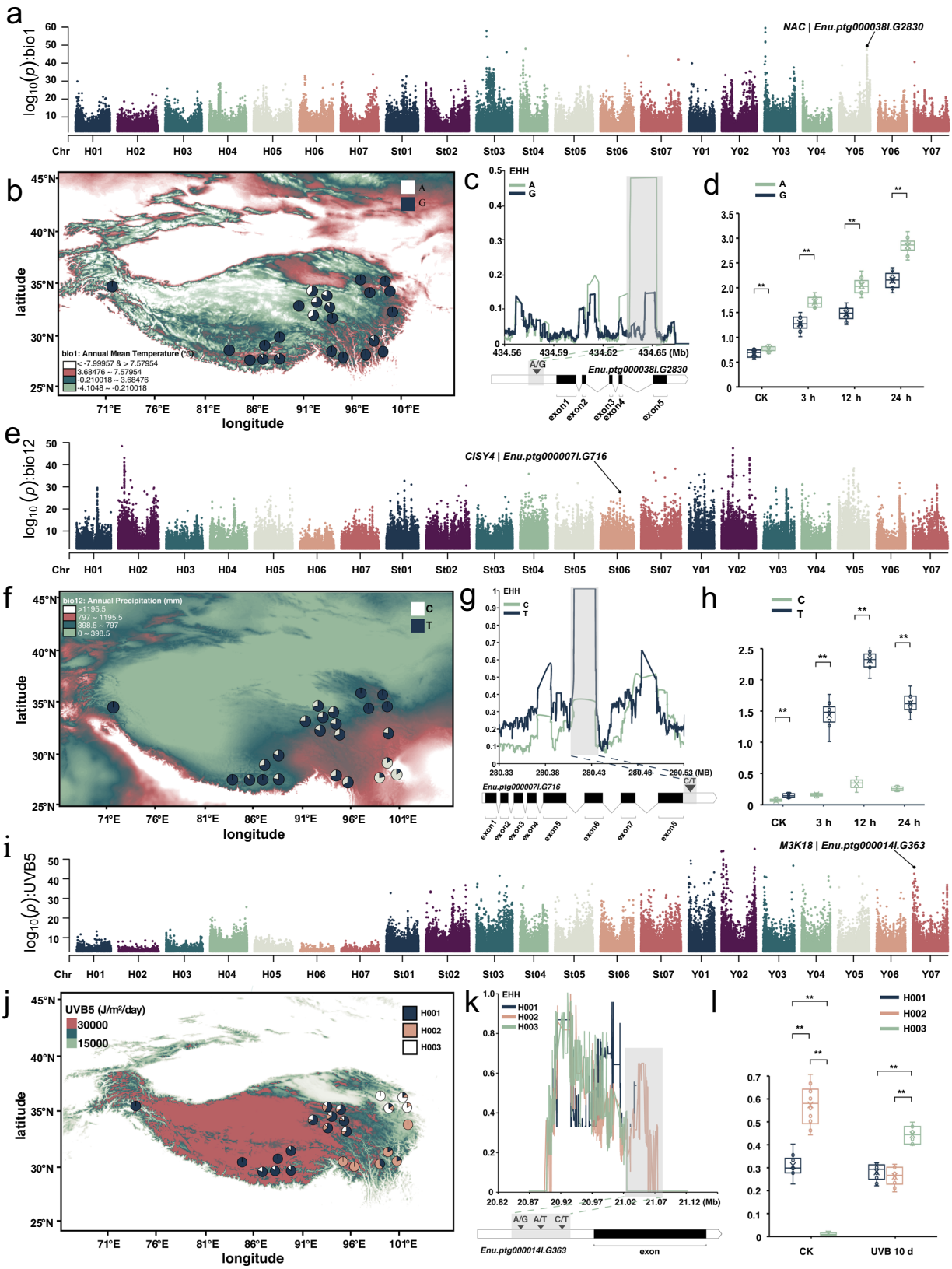


Fig. 4 | Population genomic analyses and demographic history simulation of *E. nutans*. **a** Population relationships illustrated using a phylogenetic tree and population genetic structure with K ranged from 2 to 7; **b** Population ancestral components inferred by ADMIXTURE ($K = 7$). The dark points show the natural distribution of 197 *E. nutans* accessions, and these accessions were divided into 24 clades according to their location. The elevation range of QTP is represented by color gradient; **c** PCA analysis based on genome wide SNPs; **d** Summary of nucleotide diversity (π) and population divergence (F_{st}) among geographic groups; **e** Population historical demography inferred based on the most possible scenario. INEN: initial population of *E. nutans*, OO12: bottleneck out of Cluster I and Cluster II. The arrows between the geographic groups represent the gene flow; **f** Results of Fbranch with the reference of StY species. The values in the matrix refer to excess allele sharing between the branch on the *y* axis (relative to its sister branch) and the P3 value identified on the *x*-axis. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-025-58341-0



detected between the two haplotypes (Fig. 5c) underscored the adaptive significance of temperature-related traits.

We also identified a set of genes associated with precipitation factors (Supplementary Data 16), encompassing well-documented genes such as *CYP450*, *CHS*, and *GH* (*Glycoside hydrolase*). A notable example is the *CISY4* (*citrate synthase*) gene, *Enu.ptgOOOOO7l.G716* (Fig. 5e), which participates in the tricarboxylic acid cycle and the plant's resilience mechanisms to drought stress[49], exhibiting allele frequency distribution patterns comparable to those of genes associated with temperature (Fig. 5f). We detected a C/T allele variation within the 3'-UTR sequences, with individuals possessing the C allele predominantly originating from regions characterized by higher

Fig. 5 | Genome scale identification of adaptive loci associated with key environmental factors. a–i Manhattan plot for variants associated with the annual mean temperature (bio1) (**a**), annual precipitation (bio12) (**e**), and sum of monthly mean UV-B during highest quarter (UVB5) (**i**) (two-sided *Z*-test, unadjusted for multiple tests); **b–j** the allele frequency structure of *Enu.ptg000038l.G2830* (**b**) and *Enu.ptg000007l.G716* (**f**), and haplotype frequency structure of *Enu.ptg000014l.G363* (**j**) associated with bio1, bio12, and UVB5 across the 24 clades, the temperature, precipitation and UVB range of QTP is represented by color gradient; **c–k** Decay of EHH for alternative alleles around *Enu.ptg000038l.G2830* (**c**), *Enu.ptg000007l.G716* (**g**), and *Enu.ptg000014l.G363* (**k**); **d–l** Time dynamic relative expression level of *Enu.ptg000038l.G2830* (**d**), *Enu.ptg000007l.G716* (**h**), and *Enu.ptg000014l.G363* (**l**) genes between different genotypes using qRT-PCR under cold, drought, and UV-B treatments, respectively. $n = 10$ biologically independent samples. In the box plots, the central box delineates the interquartile range, which is defined by the first quartile (25th percentile) and the third quartile (75th percentile). A horizontal line within the box marks the median (50th percentile). The whiskers represent the standard deviation of the data. Group comparisons were conducted using Welch's two-sided *t*-tests, with statistical significance indicated as follows: $^{*}P < 0.05$, $^{**}P < 0.01$. Source data are provided as a Source Data file.

annual precipitation (Fig. 5f). The qRT-PCR results confirmed that this gene indeed responds to drought stress, with the T allele exhibiting greater expression levels compared to the C allele (Fig. 5h), indicating a heightened response to drought stress. Likewise, robust selection pressure was discerned between the two alleles of this gene (Fig. 5g). In addition, we also found that several genes including *MYB*[50], *CHI*[51] (*chalcone isomerase*), *M3K*[52] (*MAPKKK*), *CHS*[53] (*Chalcone synthase*) and chloroplast transport proteins[54] (translocon of the inner membrane of chloroplasts, TIC), were associated with the UV-B adaptation of *E. nutans* (Supplementary Data 16). Here, a mitogen-activated protein kinase kinase kinase gene *MAPKKK18* (*Enu.ptg000014l.G363*) was found to be significantly associated with six UV-B related environmental factors and exhibited a robust selective signal (Fig. 5i; Supplementary Fig. 31). Subsequent analysis elucidated three SNPs within its promoter region, that could categorize all *E. nutans* accessions into three distinct haplotypes, characterized by significant discrepancies in geographic distribution (Fig. 5j; Supplementary Fig. 31). Additionally, a pronounced selective signal was observed between haplotypes H003 and H001/H002 in the vicinity of this locus (Fig. 5k). The qRT-PCR results showed that this gene responds to UV-B treatment, with haplotypes H001 and H002 exhibiting down-regulation, whereas haplotype H003, predominantly represented by the QH_EAST group with the lowest local UV-B pressure, showed a significant up-regulation (Fig. 5j, l).

GWAS analysis identified candidate genes related to UV-B resistance

To investigate the genomic loci associated with UV-B resistance in *E. nutans*, we conducted a common garden study and performed the GWAS analysis utilizing 43,637,347 SNPs from 197 accessions (Supplementary Data 17). High intensity UV-B radiation has the potential to induce DNA damage in plants, trigger the accumulation of reactive oxygen species, and impair the photosystem PSII, which is integral to the process of photosynthesis[55]. Plants have the capacity to mitigate UV-B stress through the accumulation of flavonoids (Flavo) and total anthocyanins (TA). Consequently, the levels of chlorophyll A (chlA), chlorophyll B (chlB), total chlorophyll (chlT), Flavo, malondialdehyde (MDA) and TA can serve as indicative metrics for evaluating the capacity of plants to resist UV-B resistance[55].

We identified a total of 633 unique loci encompassing 1606 genes, among which 465 genes exhibited responsiveness to UV-B stress as indicated by the transcriptome data ($|\log FC| \geq 1.0$, Supplementary Data 18), with 88 of these genes also identified in the GEA association results. We also found that the majority of these genes were located at different gene (syntenic, 1:1:1 or 1:1) regions (Supplementary Table 5). Numerous genes implicated in growth and development, sugar transport, energy metabolism, and antioxidant process were also identified. Notably, the association scans for all six traits revealed a pronounced peak on the anterior segment of chromosome *En*-Y07, which colocalized with the GEA locus discussed previously (*En*-Y07: 20900000 - 21000000) and once again highlighted the gene *MAPKKK18* (*Enu.ptg000014l.G363*) (Fig. 6a, b; Supplementary Fig. 31). MAPK signaling cascades can be selectively activated in response to UV-B-induced DNA damage, as well as in the context of DNA replication stress. *MKP1* and its substrates *MPK3* and *MPK6* operate in an antagonistic fashion to modulate UV-B stress tolerance, with *MPK3* and *MPK6* serving a negative regulatory role in this process[56]. Here we found that the first SNP in the promoter sequence of *MAPKKK18* showed the most substantial GWAS effect across all six UV-B related traits, with H003 possessing the A allele while H001/H002 possessing the G allele in this locus (Fig. 6c–e; Supplementary Fig. 32). The H003 of the *MAPKKK18* exhibited a significantly elevated content of negative traits (MDA) and a notably reduced content of other positive traits following UV-B treatment (Fig. 6f). Given that *E. nutans* accessions with haplotype H003 exhibit significantly heightened expression following UV-B treatment, as indicated in the GEA section (Fig. 5l), it is inferred that this gene may function as a negative regulator of UV-B resistance. The dual-luciferase reporter assay revealed that the *LUC* gene in leaves transiently expressing the promoters of haplotypes H001, H002 and H003 exhibited significant differences under both CK and UV-B stress conditions. Notably, the *LUC* expression levels in leaves expressing haplotype H003 were significantly greater than those in H001 and H002 under UV-B condition (Fig. 6g), providing additional evidence for this inference. Most importantly, upon heterologous expression of *MAPKKK18* in *Arabidopsis thaliana*, the three over-expression (OE) lines exhibited reduced resistance to UV-B stress relative to the wild type (WT), as reflected by its sensitive phenotype, lower chlA, chlB, Flavo and TA content and higher MDA content after UV-B treatment (Fig. 6h; Supplementary Fig. 33). In addition, the *Atm3k18* mutant lines showed significantly improved resistance to UV-B radiation compared to the WT. Notably, despite UV-B treatment, the *LUC* expression levels for each haplotype remained consistently low, in agreement with the transcriptome analysis results (Figs. 6g, 5l). Collectively, these data suggest that the variations within the promoter of the *MAPKKK18* gene play a role in the adaptation of *E. nutans* to UV-B stress by maintaining a low level of gene expression. Interestingly, our findings suggest that *MAPKKK18* may serve as a distinct, *UVR8*-independent key player[56] in the UV-B stress response and adaptation of high-altitude plants (Supplementary Fig. 34), while the detailed molecular underpinnings await future investigation.

Discussion

In this study, we present a chromosome-scale, annotation-based reference genome for *E. nutans*, encompassing a genome size of 10.47 Gb that has been successfully anchored onto 21 pseudo-chromosomes corresponding to three subgenomes or haplomes (St, Y, and H), achieving an anchoring rate of 95.05% (Supplementary Fig. 3a; Supplementary Table 2). This represents a significant advancement in hexaploid species characterization, being only the second such comprehensive analysis after bread wheat and notably incorporating the Y haplome within the Triticeae tribe. Despite the typical possession of large, complex, and highly repetitive genomes of the Triticeae, several genomes have been sequenced at the chromosome-level, with N50 values ranging from 29 kb–55.01 Mb[32]. In the present study, our assembly yielded a contig N50 of 76.25 Mb, a BUSCO completeness score of 99.50% and LAI scores of 13.97, respectively, which significantly improved the quality of the chromosome assembly and ensured a high level of genome contiguity

https://doi.org/10.1038/s41467-025-58341-0

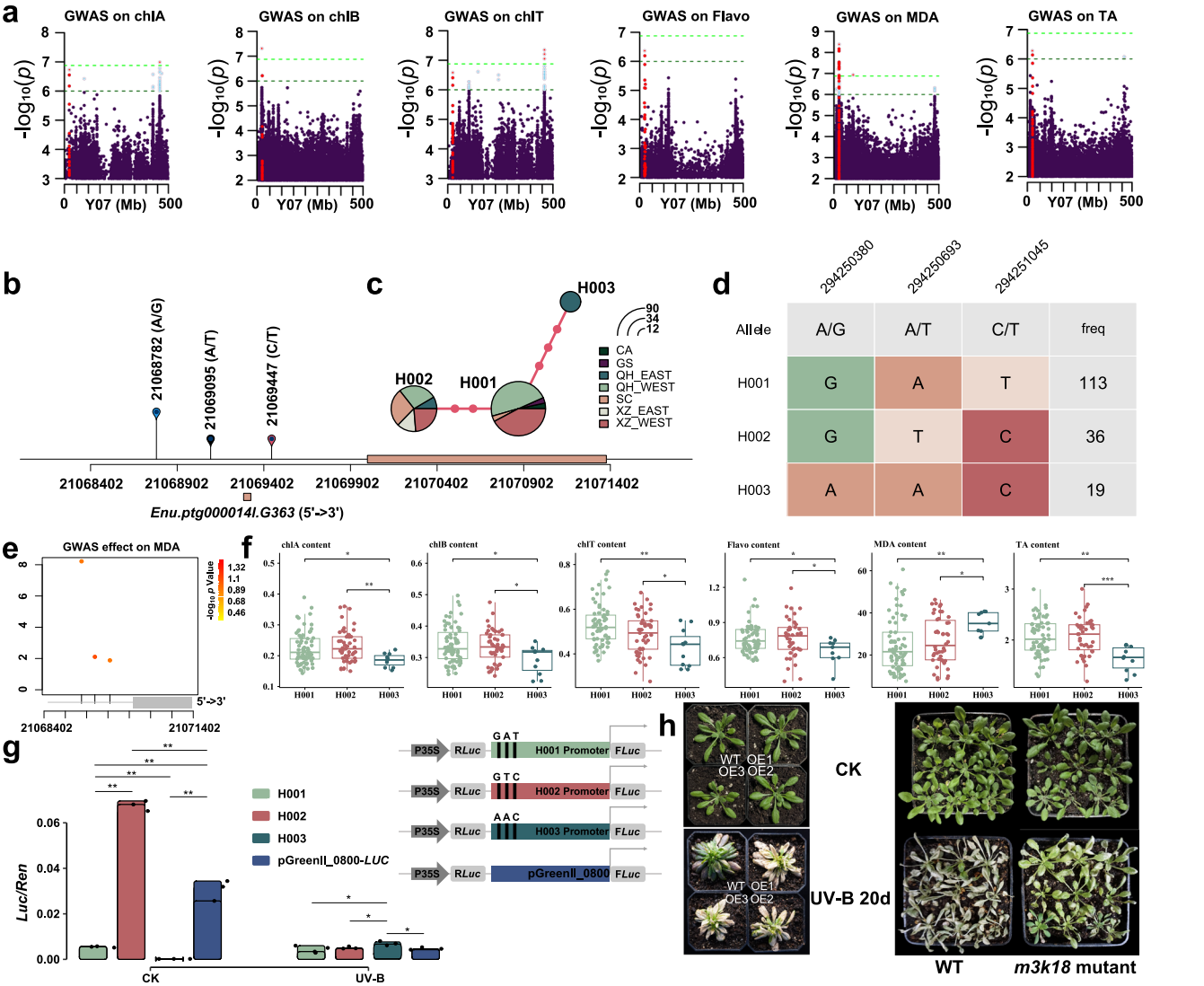


Fig. 6 | Genome-wide scan for the UV-B resistance traits and candidate gene function validation. a GWAS analysis on UV-B resistance traits (chlA, chlB, chlT, MDA, TA, and Flavo) co-localized the gene *Enu.ptg000014l.G363* (two-sided *Z*-test, unadjusted for multiple tests); **b** Three SNPs detected within the promoter and gene regions of *Enu.ptg000014l.G363*; **c** Haplotype network constructed based on the three SNPs; **d** Three SNPs in the *Enu.ptg000014l.G363* promoter divided all accessions into three haplotypes; **e** GWAS effect estimation of MDA; **f** UV-B resistance traits content (μg·L^{-1}) comparison between three haplotypes. The central box denotes the interquartile range (25th to 75th percentiles), with the median marked by an internal line. Whiskers indicate the standard deviation. Group differences were assessed using Welch two-sided *t*-tests, with significance levels: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Sample sizes: H001 ($n = 76$), H002 ($n = 52$), H003 ($n = 12$); **g** Promoter activity of three promoter haplotypes measured based on Dual-Luciferase reporter assay at CK and UV-B treatment with the mean and standard deviation of three biological replicates ($n = 3$), on the right is a schematic diagram of the vector construction of different haplotypes and control. Differences between groups are compared using two-sided Welch *t*-tests. Significant differences between the lines are indicated as follows: $^{*}P < 0.05$, $^{**}P < 0.01$; **h** Phenotypic differences between wild type and over-expression *Arabidopsis thaliana* lines (left), and between wild type and mutant lines (right) before and after UV-B treatment. Source data are provided as a Source Data file.

(Supplementary Table 2). This genome contributes to the enrichment of the genomic repository for the wheat tribe and serves as a pivotal reference for the molecular breeding of wheatgrass species, as well as for the genetic improvement of forage and cereal crops such as wheat and barley. Our results will enable the identification of genes associated with crucial agronomic traits and stress resilience, thereby facilitating their utilization in breeding forage and cereal programs.

Through the execution of deep resequencing on 18 Triticeae species representing the potential ancestral lineages (comprising 13 StStYY and 5 HH species), the chronological allohexaploid origin of *E. nutans* and its putative ancestor-like species, *H. roshevitzii* and *E. burchan-buddae*, were successfully elucidated, providing a foundational framework for revealing speciation processes of Triticeae allopolyploids (Fig. 3). The formation of *E. burchan-buddae*-like ancestor (StStYY) may arise via an allotetraploidization event from diploid StSt and YY ancestors. Ultimately, the allohexaploidization event of *E. nutans* from ancestors like *H. roshevitzii* and *E. burchan-buddae* took place around 1.5 Mya (Supplementary Fig. 13c), predating the polyploidy events observed in crops such as peanut[57], oat[41], and wheat[45]. The StSt diploid species are now found for the genus *Pseudoroegneria* while the diploid YY species have not been reported up to now or they have been extinguished. However, our phylogenomic analyses suggested that Y haplome within *E. nutans* exhibits the closest similarity to the V and Jv haplomes from genera *Dasypyrum* and *Thinopyum* within Triticeae (Fig. 2), challenging the previous reports that it is closely related to the St haplome[7,11,12]. The diploid YY lineage may have diverged from the common ancestor of these two genera around 4 Mya. Our studies also suggest that the diploid ancestral HH, StSt and

YY lineages successively diverged from 12 to 8 Mya (Fig. 2). The short divergences may give rise to all likely allopolyploid combinations from three lineages as found in the diverse *Elymus* s.l. species (Fig. 3). Such discoveries enhance our understanding of the phylogenetic relationships among St, H and Y genomes/haplomes as well as reticulate evolutions through repeated allopolyploidization events between them within Triticae.

Reticulate polyploidization is typically accompanied by the pseudogenization, sub-functionalization, and neo-functionalization of orthologous genes. These processes confer greater genetic diversity and plasticity to polyploids and co-evolve with subgenomic dynamics, collectively shaping the adaptability of polyploid species[58]. In this study, our extensive analyses based on homologous gene triads did not suggested a clear global subgenome dominance pattern in aspect of gene expression, ontology enrichment, and co-expression regulation under the cold, drought and UV-B stress conditions, which similar to that of *Avena sativa*[36], *Brassica napus*[59], and *Triticum aestivum*[37,38], although some non-global subgenome bias were found in those species. However, the response of star gene triads showed differential expression in coping with these stress conditions and also different expression patterns to the same stress. In addition, the "homologous gene function change" on particular gene triads such as the *RUG3-UVR8-HERC2* (St-H-Y) shift, the *BHLH56-PIF3-BHLH119* (St-H-Y) shift, and the *FACR5-FAR-FACR5* (St-H-Y) shift were also found in response to UV-B, cold, and drought stress, respectively (Supplementary Data 12). These results suggest that the sub-functionalization and neo-functionalization of orthologues, which found in other polyploid species that have undergone intense domestication such as *Gossypium hirsutum*[60] and *Triticum aestivum*[37], also appears to be occurring among the three *E. nutans* subgenomes. However, genomic regional and developmental stage-specific gene expression, often influenced by transposon insertion, DNA methylation, homoeologous exchange, cell type, etc., are also critical factors contributing to the establishment of subgenome dominance, but remain challenging with our current methodologies and warrant further investigation[56,61].

Utilizing whole-genome population genomic data, we identified the key genes and allelic variations underlying high-altitude adaptation of *E. nutans* to the QTP (Figs. 5, 6). Especially, we found that allelic mutations in the promoter region of the gene *MAPKKK18* have altered its expression that enhances the adaptation of this species to UV-B radiation in the high-altitude populations. The high-quality genome represented here as well as the developed comprehensive database (https://elymus.genome.org.cn/) will act as a repository for the continuous dissemination and analysis of the fundamental dissections of reticulate evolutions within the genus *Elymus* and the tribe Triticeae. These findings will help researchers to breed forage and cereal crops through distant hybridization between genera of the Triticeae based on their solid evolutionary relationships.

Methods

Plant materials and DNA sequencing

This study utilized 198 wild accessions of *E. nutans* (Supplementary Data 14), among which accession FS20130 was selected for genome sequencing and assembly due to its unique characteristic of being collected from the highest recorded altitude, while the remaining 197 accessions were designated for subsequent genome resequencing analysis. Each accession was germinated from seed, and then transplanted to a common garden located in both field and greenhouse conditions. Furthermore, 28 related species, including 5 *Hrodeum* (HH), 6 *Pseudoroegneria* (StSt), 2 *Elymus* s.s. (StStHH), 13 *Roegneria* (StStYY), and 2 *Campeiostachys* (StStYYHH) species were also included in the genome resequencing analysis, generating approximately 22.78 Tb of raw sequencing data (Supplementary Data 13).

Young leaves from selected accessions were collected for DNA extraction using a modified CTAB method, with quantification by 1× dsDNA HS kit. For PacBio sequencing, SMRTbell libraries were prepared following standard protocols, excluding fragments <5 kb, and sequenced on the Revio platform. Additionally, 350 bp insert-size Illumina paired-end libraries were constructed using the Proximo Hi-C plant protocol (Phase Genomics) and sequenced on the NovaSeq 6000 system. Hi-C raw reads were filtered using the following criteria: (1) Q30 quality threshold; (2) exclusion of >10 bp adapter sequences (≤2 mismatches allowed) and duplicates; (3) removal of sequences with >5% ambiguous bases. For genome resequencing, paired-end libraries (2 × 150 bp) were prepared using the Truseq DNA PCR-free kit and sequenced on an Illumina HiSeq X Ten platform (Wuhan, China). Raw data were processed using fastp[62] with a 5 bp sliding window (1 bp step size), discarding reads with Phred scores <20 and adapter contamination.

Genome assembly and subgenome identification

Pacbio HiFi reads were used for de novo assembly using the Hifiasm v.0.19.4[28], following the default parameter settings. The Hi-C reads were aligned to the primary contig assembly using BWA v.0.7.17[63], with selection criteria focusing on uniquely aligned paired-end reads for subsequent analyses. The contigs were subjected to clustering, ordering, and orientation procedures using Juice v.1.6 (https://github.com/Automattic/juice) and 3D-DNA[64], after which they were integrated into pseudochromosomes. The Juicebox v2.20 software[65] visualizes the chromatin interaction matrix derived from the Hi-C data in the form of a heat map, with the 'mismatch assembly' manually adjusted according to the observed adjacent interactions.

A reference-guided approach using sequence similarity was performed to distinguish the three subgenomes within *E. nutans*. Whole-Genome Duplication Integrated analysis[66] (WGDI v.0.6.5) was initially used to assess genomic synteny between the chromosomes of *E. sibiricus* (which harbors the St and H subgenomes) and *E. nutans* (which comprises the St, H, and Y subgenomes). In addition, SubPhaser v1.2.6[33] was employed to facilitate accurate subgenome identification and ensure accurate differentiation of each subgenome. This software uses subgenome-specific k-mers ($K = 15$) to cluster chromosomes, effectively separating them into three distinct groups. Moreover, using the unique patterns of differential k-mers and chromosome homology, all chromosomes were accurately assigned to their respective subgenomes using unsupervised hierarchical clustering and PCA analysis. Ultimately, the chromosomes were renamed to reflect their alignment to barley (H), designated as *En*-St01 to *En*-St07 for the St subgenome, *En*-Y01 to *En*-Y07 for the Y subgenome, and *En*-H01 to *En*-H07 for the H subgenome.

Genome quality assessment

The quality of the *E. nutans* genome was assessed in three dimensions: continuity, accuracy, and completeness. First, the contig N50 length was calculated to assess the continuity of the assembly. Subsequently, the resequenced short reads from each accession were aligned to the assembly using BWA v.0.7.17[63] with standard default parameters. Reads mapping rate, genomic coverage, and effective mean depth were used to assess the accuracy of the assembly. Simultaneously, the HaplotypeCaller module within GATK v4.1.9.0[67] was used to analyze SNPs to estimate base level accuracy. Ultimately, the BUSCO v5.2.2[68] pipeline was applied to assess gene completeness, while the LAI index was used to evaluate the completeness within the more highly repetitive genomic regions.

Repeat and gene annotation

Repetitive elements within the genome were annotated using a combination of evidence-based and de novo strategies. Initially, RepeatMasker v.4.1.0[69] was used to identify homologous transposable elements (TEs) in the genome, using the Repbase database (http://www.girinst.org/repbase) as a reference. Subsequently, a non-

redundant database of long-end-repeating retrotransposons (LTR-RTs) specific to *E. nutans* was created using LTRharvest v.2.9.8[70] and LTR_Finder 1.0.7[71], and was further consolidated using LTR_retriever v.2.8.5[72], a tool capable of effectively excluding false-positive elements. Finally, the data from these two approaches were integrated to generate the final predictions of repetitive elements within the *E. nutans* genome.

Predictions of protein-coding genes was achieved by a combination of transcriptome data, homologous gene annotation and ab initio gene prediction methods. Prior to gene annotation, the TE sequences were masked using RepeatMasker v4.1.0[69] with the TE library above mentioned. To enhance the accuracy of the *E. nutans* gene annotation, RNA sequencing was performed on three tissues - roots, stems, and leaves - to provide comprehensive transcriptome data. Total RNA was extracted from each of the tissue using the TRIzol reagent. A total of 26.49 Gb of raw sequencing reads were generated by PE150 sequencing on the MGI platform. Low quality reads were filtered out using Trimmamatic v.0.38[73], and the subsequent transcripts were assembled using Trinity v.2.6.5[74]. Following transcript assembly, transcripts were mapped to the assembled genome using the Program to Assemble Spliced Alignment PASA v.2.5.3[75], and transcript alignments were then clustered based on their mapping positions. The gene structure identified in the assembly was then interactively refined using Augustus v.3.5.0[76]. Annotations from three related species of *E. nutans*, namely *E. sibiricus*, *H. vulgare*, and *T. aestivum*, were integrated to facilitate the homologous gene annotation of *E. nutans* using GeMoMa v.1.7.1[77]. The masked genome was then subjected to ab initio gene prediction using Augustus and Helixer v.0.3.2[78]. Finally, all gene predictions were combined using EvidenceModeler v.1.1.1[79], thereby generating a consensus gene set.

Moreover, to annotate non-coding RNA, including tRNA, rRNA, small nuclear RNA (snRNA), and microRNA (miRNA), we employed a hybrid search strategy integrating de novo prediction and homology-based methods. Predictions of tRNA and rRNA genes were carried out using tRNAscan-SE v.2.0.12[80] and the barrnap program v.0.8[81], respectively. Prediction of snRNA and miRNA genes was performed using a combination of Infernal software v.1.1.5[82] and the Rfam database (ftp://ftp.ebi.ac.uk/pub/databases/Rfam/) for comprehensive annotation.

Gene function annotation

To determine the functional role of the protein-coding genes, comprehensive searches were performed in multiple databases, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Cluster of Orthologous Groups of proteins (KOG/COG/eggNOG) based on eggnog-mapper v2.1.6[83]; SwissProt, TrEMBL and Non-Redundant Proteins (NR) by using DIAMOND v2.0.14[84]; and Pfam by applying InterProScan v5.54-87.0[85]. Furthermore, the DeepGO tool[86]was used to predict protein functions based on both sequence and protein-protein interaction data, using a deep ontology-aware classifier.

WGD, Ka/Ks, and collinearity analysis

Collinearity between the three subgenomes of *E. nutans* was assessed using the JCVI program[87]. The WGDI pipeline[66] was utilized to calculate the non-synonymous (Ka) and synonymous (Ks) substitution rates between the rice and *E. nutans* subgenomes, in addition to characterizing the whole-genome duplication (WGD) events and estimating the divergence time of the *E. nutans* subgenomes.

Identification of pseudogenes

The pseudogenization of genes in *E. nutans* was identified utilizing two complementary strategies: (1) The WGDI pipeline[66] was employed to determine collinear genes among three subgenomes, specifically identifying instances where two subgenomes harbored a homologous gene while the third subgenome lacked this gene; (2) The pseudogenization of all extant genes was ascertained using the PseudogenePipeline v1.0.0[88].

Evolution and allopolyploidization history of *E. nutans*

To investigate the phylogenetic relationships and divergence time of *E. nutans* within the Triticae family, the OrthoFinder software v.2.5.5[89] was used to analyze the 20 subgenomes of 12 species with available genome sequences and annotations, yielding a dataset of 305 single-copy genes. The coding sequences (CDS) of the identified single-copy orthologs were then extracted, and the aligned protein and CDS alignments were concatenated to form a supermatrix. Transversion rates at fourfold degenerate sites (4DTv) were then extracted from the constructed supermatrix. Finally, an evolutionary tree was constructed using iqtree2[90] in combination with the astral.jar script (https://github.com/smirarab/ASTRAL). Divergence time was calibrated using data from the TimeTree online website[91] and then estimated using the MCMCTree method within the PAML v4.7 package[92].

To elucidate the evolutionary position of *E. nutans* within the genus *Elymus*, determine its exact divergence time, and identify its ancestral species, we performed an analysis of the 50 subgenomes of 28 representative species of *Elymus* and its related species using the -at parameter of the WGDI software[66]. A collection of 11,607 high quality collinear genes was compiled, with the criterion that at least 80% of the species were present. Subsequent alignment of these genes was performed using Muscle v.5.1.0[93], and the resulting phylogenetic tree was constructed using iqtree2[90]. Integration of multiple phylogenetic trees was achieved through the application of the astral.jar script (https://github.com/smirarab/ASTRAL). This method was similarly employed for the estimation of divergence time. Furthermore, the plastomes of 45 species within *Elymus* s.l., *Hordeum*, and *Pseudoroegneria* were assembled and annotated to construct a common gene phylogenetic tree, with *Brachypodium distachyon* as the outgroup. Of these, the plastomes for 28 species were assembled using GetOrganelle v17.7.1[94] from the raw resequencing reads. The remaining species were retrieved from the NCBI database (Supplementary Data 13).

Collinearity and FISH analysis

To identify syntenic blocks across the three subgenomes of *E. nutans*, we performed alignments between subgenomes St and Y and H using minimap2[95], followed by sorting of the resulting alignments using samtools v1.8. Subsequent analysis of the identified syntenic blocks and structural rearrangements within the three subgenomes was conducted with SyRi[96], with results visualized using plotsr[97]. Furthermore, an all-against-all reciprocal sequence similarity search was performed across all annotated proteins to delineate gene density landscape similarities, using DIAMOND v2.0.14[84] for the search. Pairwise synteny block detection was then performed using MCScanX[98] with the top homologous pairs of 5, the minimally matched syntenic anchors of 5, and the maximum number of gene gaps of 25.

A significant inter-chromosomal exchange event between subgenomes Y03 and H04 of *E. nutans* was verified using non-denaturing FISH (ND-FISH) and Oligo-FISH painting techniques[99]. Metaphase chromosomes from root tips of germinated seeds were prepared following the protocol described by Han et al.[100] with modification. Specifically, root tips were fixed in pre-chilled fixative solution (methanol/acetic acid = 3:1) at 4 °C for 18–24 h, then transferred to 70% ethanol for storage at 4 °C. The fixed root tips were washed three times with deionized water and subsequently incubated in enzyme solution (2% cellulase + 1% pectinase) at 37 °C for 35–45 min. The enzyme solution was gently removed to avoid tissue damage. Following enzyme treatment, root tips were incubated in pre-warmed 0.075 M KCl solution at 37 °C for 25–30 min. The processed root tips were then placed on clean microscope slides, treated with 45% acetic acid, and covered with

coverslips. Gentle pressure was applied to disperse the cells evenly. The quality of chromosome spreads was examined under a microscope to ensure proper cell dispersion and intact chromosome morphology before proceeding to FISH hybridization. Synthetic oligonucleotides Oligo-pSc119.2, Oligo-pTa535, and Oligo-HvCSR were utilized for ND-FISH analysis, and their details are presented in Supplementary Table 7. Each oligonucleotide probe was labeled at the 5′ end with either 6-carboxyfluorescein (6-FAM) for generating green fluorescence or 6-carboxytetramethylrhodamine (Tamra) for producing red fluorescence. The FISH results were captured photographically under an Olympus BX-53 microscope using a DP-70 CCD camera.

SNP calling and population structure analysis

The genomes of 197 *E. nutans* accessions and 28 accessions of related species were resequenced using the Illumina HiSeq X Ten sequencing platform. Raw sequencing data were processed with FQC software[101] to remove adapter sequences and bases with a quality score <20. Paired-end reads from each accession were aligned to the de novo assembled *E. nutans* reference genome using BWA v.0.7.17[63] with default parameters. Only uniquely mapped and correctly paired reads were retained for subsequent analyses. Following sorting of the aligned reads with samtools[102], PCR products were identified and marked. SNP calling was then performed using the GATK v4.1.9.0[67] to generate a merged VCF file. The SNPs were subjected to filtering in accordance with the following criteria: QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum <−12.5, ReadPosRankSum <−8.0 with the clusterWindowSize of 5 and clusterSize of 2[101]. The linkage-disequilibrium (LD), as inferred from squared allele-frequency correlations (r^2), was analyzed at both the species level and the level of geographic groups using PopLDdecay[103,104] with a window size of 100 kb and a step size of 10 kb. Utilizing LD-pruned SNP variation data, the genetic structure of the 197 *E. nutans* accessions was investigated via ADMIXTURE v.1.23[105], with K values (indicative of genetic memberships) evaluated across the range of 1–10. Principal component analysis (PCA) was conducted using plink2[106]. To elucidate the phylogenetic relationships between the 197 *E. nutans* accessions and their relative species, an individual-based NJ tree was constructed with fasttree v.2.1[107]. The snpEff[108] was employed to annotate the identified SNPs. Nucleotide diversity (π) was computed using VCFtools v.0.1.16[109], with the resulting π values for each geographic group determined as averages across the genome. Moreover, to expedite genomic research of *Elymus*, we proposed the *E. nutans* genome database (https://elymus.genome.org.cn/). This comprehensive resource incorporated the genome reference sequence, annotations (Supplementary Data 19) and all associated dataset and findings from this study. Additionally, the database offers Primer3 and KASP primer design tools, enabling researchers to easily develop SSR and KASP primers for genome-wide analysis, thereby streamlining germplasm screening and genetic mapping investigations not only for *E. nutans* but also for a broad range of other *Elymus* species.

IBD and IBE analysis

To investigate the impact of geographic and environmental factors on spatial genetic variation, total genetic variants and "key adaptive variants" were used to perform IBD and IBE analyses, respectively. Pairwise differentiation coefficient (F_{st}) values were computed using VCFtools v.0.1.16[109], with a window size of 100 kb and a step size of 10 kb. The Mantel test was employed to assess the association between the matrix of $F_{st}/(1\text{-}F_{st})$ and the matrices of geographic distance (km) and environmental distance, utilizing the *vegan* v.2.5 R package[110] with 9999 permutations. Environmental distance was calculated through the Euclidean distance, which integrated 19 bioclimatic environmental factors, altitude, and six factors related to UV-B radiation.

Demographic history and gene flow

Two methods were employed to estimate the demographic history of *E. nutans*. Initially, the PSMC[111] was utilized to explore historical fluctuations in the effective population size (Ne) of *E. nutans*, employing default parameters with resequencing data. Representative individuals were randomly selected from each geographical group to conduct PSMC analyses, with 100 bootstrap estimates executed for each individual to assess the robustness of the results. The mutation rate was calibrated at 3.75×10^{-8} per site per year, with a generation time of 1 year assumed for the analysis. Subsequently, the genetic algorithm for inferring demographic history (GADMA) tool[112] was employed, utilizing the Diffusion Approximation for Demographic Inference (dadi) engine[113], to automatically infer the joint demographic history of three Clusters from SNP data. Finally, based on the fundamental insights derived from the PSMC and GADMA analyses, a simulation of 16 potential demographic history models was conducted, followed by the application of the approximate Random Forest algorithm within DIYABC-RF v1.0 software[114] to identify the most plausible model. The inferred population origin and dispersal patterns were then visualized utilizing the DemesDraw python pipeline (https://github.com/grahamgower/demesdraw). To assess potential gene flow across seven distinct geographical groups, the ABBA-BABA methodology-based software Dsuite[115] was used to calculate Patterson's D-statistic. Based on the phylogenetic relationships at geographic groups level, D and f statistics for branches were used to reflect the degree of gene flow, including an outgroup of relative species with StStYY. The migrate-n software[116] was also utilized to estimate the historic gene flow dynamics between the seven geographic groups.

Transcripts assembly, gene expression quantification and co-expression analysis

To investigate the key genes of *E. nutans* that mediate responses to UV-B radiation, cold, and drought, RNA sequencing was conducted under different stress conditions. Samples exposed to UV-B radiation were collected at the control (CK) condition and 20 days post-exposure to UV-B radiation, using a wavelength of 311 nm and an intensity of 1200·mw·m^{-2}. RNA sequencing data for drought and cold stress conditions were retrieved from the NCBI with project IDs PRJNA596107 and PRJNA320799, respectively. For each stress condition, as well as the CK, three biological replicates were included. RNA sequencing was carried out on an Illumina Nova6000 platform. The raw RNA sequencing reads were initially processed through fastp software[62] to filter and obtain high-quality clean reads. Subsequent mapping of the clean reads to the *E. nutans* reference genome was achieved using Bowtie2[117], with transcripts quantification conducted via the RSEM software[118]. Differentially expressed genes (DEGs) were identified using DESeq2[119] with a criterion of $|\log_2(\text{D10/CK or D20/CK})| \geq 1$ and $P_{adj} < 0.05$. Transcript sequences were subjected to alignment against the Nr, SwissProt, KEGG, and COG/KOG databases using blastx[120], with a threshold of 1e-5 to ensure robust function annotation.

The gene co-expression network under three stress conditions was constructed using the multiscale embedded gene co-expression network analysis (MEGENA) method[121]. This method first identifies gene pairs that have a significant correlation (FDR < 0.05). These significant gene pairs are then sorted according to their absolute Pearson correlation coefficients and sequentially embedded in a three-dimensional topological sphere. The aim of this process is to determine whether each gene pair can be plotted on the sphere without any edges crossing each other. The specific implementation steps were modified according to Song et al.[121].

Subgenome expression bias

The WGDI v0.6.5[66] was used to identify gene triads exhibiting a 1:1:1 homologous relationship and positioned within a collinearity block across the Y, H, and St subgenomes of *E. nutans*. The total expression level of a gene triad was determined by summing the transcripts per million (tpm) values of its three homoeologs as follows:

$$\text{Total(tpm)} = \text{St(tpm)} + \text{H(tpm)} + \text{Y(tpm)} \quad (1)$$

To mitigate the influence of genes with negligible expression levels, potentially due to erroneous alignments, a gene triad was considered significant if its Total(tpm) >0.5. To normalize the expression levels of all gene triads within a consistent framework, the relative expression of each gene within a triad was computed using the following formulas:

$$\text{St(tpm)} = \text{St(tpm)}/\text{Total(tpm)} \quad (2)$$

$$\text{Y(tpm)} = \text{Y(tpm)}/\text{Total(tpm)} \quad (3)$$

$$\text{H(tpm)} = \text{H(tpm)}/\text{Total(tpm)} \quad (4)$$

This normalization was performed for gene expression data under each stress (UV-B, cold and drought) and their respective CKs. Subsequently, gene triads were categorized based on their expression patterns. Briefly, Euclidean distances were calculated between the normalized expression values of each gene triad and seven predefined ideal expression categories (St-dominant, St-suppressed, Y-dominant, Y-suppressed, H-dominant, H-suppressed and balanced) using the rdist function in R 4.3.3. The relative expression of each subgenome to the gene triads was visualized using the ggtern R package[122]. Each gene triad was assigned to an expression bias category based on the shortest Euclidean distance to these ideal categories.

GEA analysis identifying local adaptive SNPs

Two approaches – the univariate latent factor linear mixed model (LFMM) and redundancy analysis (RDA) - were used to identify genetic variants associated with environmental factors. As briefly, the LEA v.3.3.2 R package[123], which implements a least squares estimation – based method, was utilized to identify associations between SNPs and the 19 bioclimatic environmental variables (https://worldclim.org/), altitude, and six UV-B-related factors (https://www.ufz.de/gluv/index.php?en=32435) using the LFMM model. RDA analysis was performed by the *rda* function within the *vegan* package v.2.5[110]. We selected nine variables with pairwise correlation coefficients where $r < 0.80$ or >-0.80 (Supplementary Fig. 28) using the 'cor' function built-in in R. A standard deviation cut off of 3 was used to detect the significant environment-associated variants.

Common garden experiment, phenotyping and identification of UV-B resistance associated genes based on GWAS analysis

To investigate the genomic loci associated with UV-B resistance in *E. nutans*, we conducted a common garden study and performed the GWAS analysis. We grew 197 *E. nutans* accessions in the common garden and assessed their UV-B resistance phenotypes, which were then used for GWAS analyses. For each *E. nutans* accession, a total of 36 healthy, plump, tender and consistent tillers were evenly transplanted into six sand culture tubes, each with a diameter of 7 cm. These tubes were controlled using U-PVC material to regulate the cultivation conditions. Throughout the experimental period, the nutrient solution was administered regularly to ensure normal growth of the plants. The experimental design included two treatments: UV-B radiation stress and CK. Each treatment was replicated three times. An artificial irradiation system was established in the greenhouse using UV-B lamps (G8T5E 313 nm UV-B narrowband lamp, YIXIAN). The distances between the UV-B lamps to the tops of the plants was adjusted every 3 days to maintain a uniform UV-B intensity of 350 mw·m^{-2}. The actual UV-B intensity of each individual was measured repeatedly with a UV meter (ZJ-5800) equipped with an UV-B sensor. To filter UV-C radiation, the UV-B lamps were enveloped in a 0.125 mm thick cellulose diacetate film. To mitigate interference from ambient light, the surrounding area of the plants was covered with a black shading net perpendicular to the ground. UV-B irradiation was applied daily from 9:30 and 17:30. In the CK group, no UV-B radiation treatment was administered, allowing the plants to grow under natural conditions. After 20 days of irradiation, samples from both the experimental group (which was subjected to UV-B radiation) and the parallel CK group were collected between 10:00 to 12:00 a.m. These samples were immediately immersed in liquid nitrogen and stored at -80°C until they were ready for assay analysis. The concentrations of TA, Flavo, chlA, chlB, and MDA were quantified under both CK and stress conditions using spectrophotometric methods. The UV-B stress experiments were conducted three times between years 2021 and 2023. The best linear unbiased prediction (BLUP) values for each trait were then used for the subsequent association analysis. In the GWAS analysis, 43,637,347 high quality SNPs were used to predict genetic-trait associations. The mixed linear model (MLM) was utilized to conduct the trait-SNP association analysis, taking into account both population structure (Q) and individual genetic membership (K) to eliminate the influence of population genetic relationships on the association results. The Manhattan plot was plotted using the CMplot R package[124] with a cut-off *P*-value of 0.1/number of SNPs.

Dual-luciferase reporter assay

The promoter sequences of the *M3k18* gene (*Enu.ptg000014l.G363*) from representative haplotypes (Supplementary Data 20) H001, H002, and H003 accessions were cloned into the pGreenII 0800-*Luc* vector via the TA cloning method, which were subsequently introduced into the GV3101 (pSoup-P19) vector. Tobacco leaves, which had been injected with *Agrobacterium*, were exposed to UV-B irradiation at a dose of 0.3 w/m^2 daily between 9:00 - 15:00. It was imperative to monitor the condition of the tobacco leaves continuously, and protein extraction was to be performed promptly once the leaves commenced wilting. The fluorescence activity of the leaves, both with and without UV-B irradiation, was measured using a luciferase assay kit (cat. no. DL101, Vazyme Biotech).

Overexpression and mutant lines construction of *M3K18* in *Arabidopsis thaliana*

The open reading frame (ORF) of the *M3K18* gene (*Enu.ptg000014l.G363*), which was identified in both GEA and GWAS analyses, was amplified by PCR using specific forward and reverse primers (Supplementary Table 8). This amplification was performed on the accession 'FS 20130', which was utilized for genome sequencing and assembly. The amplified fragment was inserted into the SalI and KpnI restriction enzyme sites of the binary vector GV3101. The recombinant vector was then transformed into the *Arabidopsis thaliana* Col-0 ecotype through the floral dip method. Transgenic plants were selected on a 0.001% Basta (Phosphinothricin, PPT) solution. The integration of *Enu.ptg000014l.G363* into the transgenic lines was confirmed by genomic PCR (Supplementary Table 8). For the assessment of abiotic stress responses, T3 generation transgenic plants were subjected to UV-B radiation. The *Arabidopsis thaliana m3k18* mutant line was obtained from the *Arabidopsis* Biological Resource Center (Stock No: SAIL_154_H03). SALK T-DNA primers (LB, LP and RP) proposed by the Salk institute genomic analysis laboratory were used for the identification of homozygous mutant lines. Samples exposed to UV-B radiation were also collected at the control (CK) condition and 20 days post-exposure to UV-B radiation, using a wavelength of 311 nm and an intensity of 1200·mw·m^{-2}. After 20 days of exposure, MDA levels,

flavonoids and total anthocyanins, and chloroplast A/B content were assessed in both WT, mutant and transgenic plants with six biological replicates. RNAseq of WT and OE lines of *Arabidopsis thaliana* were performed with three biological replicates under 0 day (CK), 3 days, and 10 days UV-B treatment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The assembled genome and annotation file in this study have been deposited in the China National GeneBank DataBase (CNGB) under accession CNP0006257. The whole-genome resequencing data for *E. nutans* have been deposited in CNGB under accession CNP0004856. The whole-genome resequencing data for 28 related species of *E. nutans* have been deposited in CNGB under accession CNP0006294. All genome sequencing data, including the raw data for HiFi and HiC sequencing, and the genome assembly and annotation file have been deposited in CNGB under accession CNP0006708. The transcriptomes sequencing (RNAseq) data of *E. nutans* used for genome annotation have been deposited in CNGB under accession CNP0004714. RNAseq raw data of *Arabidopsis thaliana* under UV-B stress experiment have been deposited in CNGB under accession CNP0006839. All raw sequencing data files have also been deposited in the Genome Sequence Archive (GSA) database under the BioProject accession PRJCA036869. Source data are provided with this paper.

Code availability

All codes and scripts used in this study are available at Github [https://github.com/xiongyi95/Elymus_nutans-genome-origin-and-evolution.git]. Codes are also archived at Code Ocean [https://codeocean.com/capsule/4333238/tree].

References

1. Li, T. et al. Genome evolution and initial breeding of the Triticeae grass *Leymus chinensis* dominating the Eurasian Steppe. *Proc. Natl Acad. Sci. USA* **120**, e1985983176 (2023).
2. Dewey, D. R. The genomic system of classification as a guide to intergeneric hybridization with the perennial triticeae. In *Gene Manipulation in Plant Improvement* (eds. Gustafson, J. P.) 209–279 (Springer Nature Press, 1984).
3. Frawley, E. S., Ciotir, C., Micke, B., Rubin, M. J. & Miller, A. J. An ethnobotanical study of the genus *Elymus*. *Econ. Bot.* **74**, 159–177 (2020).
4. Kumar, A. et al. Morpho-physiological evaluation of *Elymus semicostatus* (Nees ex Steud.) Melderis as potential donor for drought tolerance in wheat (*Triticum aestivum* L.). *Genet. Resour. Crop Evol.* **69**, 411–430 (2022).
5. Cox, T. S. et al. Breeding perennial grain crops. *Crit. Rev. Plant Sci.* **21**, 59–91 (2002).
6. Fan, X. et al. Phylogenetic relationships and Y genome origin in *Elymus* L. sensu lato (Triticeae; Poaceae) based on single-copy nuclear *Acc1* and *Pgk1* gene sequences. *Mol. Phylogenet. Evol.* **69**, 919–928 (2013).
7. Liu, Q. L. et al. Endo-allopolyploidy of autopolyploids and recurrent hybridization—a possible mechanism to explain the unresolved Y-genome donor in polyploid *Elymus* species (Triticeae:Poaceae). *J. SYST Evol.* **60**, 344–360 (2022).
8. Petersen, G., Seberg, O. & Salomon, B. The origin of the H, St, W, and Y genomes in allotetraploid species of *Elymus* L. and *Stenostachys* Turcz. (Poaceae: Triticeae). *Plant Syst. Evol.* **291**, 197–210 (2011).
9. Yen, C. & Yang, J. L. Biosystematics of Triticeae: Volume V. Genera. In *Campeiostachys*, *Elymus*, *Pascopyrum*, *Lophopyrum*, *Trichopyrum*, *Hordelymus*, *Festucopsis*, *Peridictyon*, and *Psammopyrum* (eds. Yen, C. & Yang, J.) 712 (Springer Nature Press, 2022).
10. Leo, J., Bengtsson, T., Morales, A., Carlsson, A. S. & von Bothmer, R. Genetic structure analyses reveal multiple origins of *Elymus sensu* stricto (Poaceae). *Genet. Resour. Crop Evol.* **72**, 167–185 (2024).
11. Liu, Q. L. et al. Phylogenetic relationships in *Elymus* (Poaceae: Triticeae) based on the nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. *N. Phytol.* **170**, 411–420 (2006).
12. Chen, C. et al. Chromosome-specific painting reveals the Y genome origin and chromosome rearrangements of the St genome in Triticeae. *Plant Physiol.* **196**, 870–882 (2024).
13. Sun, G. L., Ni, Y. & Daley, T. Molecular phylogeny of *RPB2* gene reveals multiple origin, geographic differentiation of H genome, and the relationship of the Y genome to other genomes in *Elymus* species. *Mol. Phylogenet. Evol.* **46**, 897–907 (2008).
14. Wu, S. D., Wang, Y., Wang, Z. F., Shrestha, N. & Liu, J. Q. Species divergence with gene flow and hybrid speciation on the Qinghai–Tibet Plateau. *N. Phytol.* **234**, 392–404 (2022).
15. Liu, J. Q., Duan, Y. W., Hao, G., GE, X. J. & Sun, H. Evolutionary history and underlying adaptation of alpine plants on the Qinghai–Tibet Plateau. *J. SYST Evol.* **52**, 241–249 (2014).
16. Zhang, T. C. et al. Genome of *Crucihimalaya himalaica*, a close relative of *Arabidopsis*, shows ecological adaptation to high altitude. *Proc. Natl Acad. Sci. USA* **116**, 7137–7146 (2019).
17. An, Y. P. et al. Comparative physiological and transcriptomic analyses reveal genotype specific response to drought stress in Siberian wildrye (*Elymus sibiricus*). *Sci. Rep.* **14**, 21060 (2024).
18. Lu, B. R. Morphological identification of *Elymus sibiricus*, *E. nutans*, and *E. burchan-buddae*, and their genomic relationships. *J. SYST Evol.* **32**, 504–513 (1994).
19. Sun, G. L., Zhang, X. D. & Scoles, G. Origin of the H genome in StH-genomic *Elymus* species based on the single-copy nuclear gene *DMC1*. *Genome* **54**, 655–662 (2011).
20. Zhang, Z. Y. et al. Phenotype- and SSR-based estimates of genetic variation between and within two important *Elymus* species in western and northern China. *Genes* **9**, 147–165 (2018).
21. Dou, Q. W., Yu, F., Li, Y., Zhao, Y. Y. & Liu, R. J. High molecular karyotype variation revealed in indigenous *Elymus nutans* in the Qinghai Plateau. *Plant Divers.* **39**, 117–122 (2017).
22. Xiong, Y. L. et al. Discovery of the ethylene response factors in *Elymus sibiricus* on a transcriptome-wide scale and the beneficial function of *EsiERF285* in combating drought and heat stress. *Ind. Crops Prod.* **210**, 118170 (2024).
23. Cui, G. W. et al. Physiological adaptations of *Elymus dahuricus* to high altitude on the Qinghai–Tibetan Plateau. *Acta Physiol. Plant.* **41**, 115–124 (2019).
24. Lasky, J. R., Josephs, E. B. & Morris, G. P. Genotype–environment associations to reveal the molecular basis of environmental adaptation. *Plant Cell* **35**, 125–138 (2022).
25. Hou, Y. G. et al. Haplotype-based pangenomes reveal genetic variations and climate adaptations in moso bamboo populations. *Nat. Commun.* **15**, 8085 (2024).
26. Tan, X. Q., Huang, Y. W., Xiong, D. W., Lv, K. & Chen, F. Q. The effect of *Elymus nutans* sowing density on soil reinforcement and slope stabilization properties of vegetation-concrete structures. *Sci. Rep.* **10**, 20462 (2020).
27. Li, Y. Y., Dong, S. K., Wen, L., Wang, X. X. & Wu, Y. Soil seed banks in degraded and revegetated grasslands in the alpine region of the Qinghai-Tibetan Plateau. *Ecol. Eng.* **49**, 77–83 (2012).
28. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 1–6 (2021).

29. Shen, W. J. et al. Chromosome-scale assembly of the wild cereal relative *Elymus sibiricus*. *Sci. Data.* **11**, 1–8 (2024).
30. Anderson, S. N. et al. Transposable elements contribute to dynamic genome content in maize. *Plant J.* **100**, 1052–1065 (2019).
31. Zeng, X. et al. The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau. *Proc. Natl Acad. Sci. Usa.* **112**, 1095–1100 (2015).
32. Zhu, T. T. et al. Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant J.* **107**, 303–314 (2021).
33. Jia, K. H. et al. SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *N. Phytol.* **235**, 801–809 (2022).
34. Otto, S. P. The evolutionary consequences of polyploidy. *Cell* **131**, 452–462 (2007).
35. Grover, C. E. et al. Homoeolog expression bias and expression level dominance in allopolyploids. *N. Phytol.* **196**, 966–971 (2012).
36. Kamal, N. et al. The mosaic oat genome gives insights into a uniquely healthy cereal crop. *Nature* **606**, 113–119 (2022).
37. Pfeifer, M. et al. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091 (2014).
38. Harper, A. L. et al. Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat. *Plant Biotechnol. J.* **14**, 1207–1214 (2015).
39. Wang, R. R. C. & Jensen, K. B. *Roegneria alashanica* Keng: a species with the StStStYStY genome constitution. *Genome* **60**, 546–551 (2017).
40. Blattner, F. R. Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). *N. Phytol.* **169**, 603–614 (2006).
41. Peng, Y. Y. et al. Reference genome assemblies reveal the origin and evolution of allohexaploid oat. *Nat. Genet.* **54**, 1248–1258 (2022).
42. Yuan, S. et al. Complete chloroplast genomes of three wild perennial *Hordeum* species from central Asia: genome structure, mutation hotspot, phylogenetic relationships, and comparative analysis. *Front. Plant Sci.* **14**, 1170004 (2023).
43. Leo, J., Bengtsson, T., Morales, A., Carlsson, A. S. & von Bothmer, R. Genetic structure analyses reveal multiple origins of *Elymus* sensu stricto (Poaceae). *Genet Resour. Crop Evol.* **72**, 167–185 (2025).
44. Yang, C. R. et al. Genomic constitution and intergenomic translocations in *Elymus dahuricus* complex revealed by multicolor GISH. *Genome* **60**, 1–8 (2017).
45. Pang, Y. L. et al. High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. *Mol. Plant.* **13**, 1311–1327 (2020).
46. Kleinhofs, A. et al. A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. *Theor. Appl. Genet.* **86**, 705–712 (1993).
47. Stokstad, E. Mountains and monsoons created Tibetan biodiversity. *Science* **369**, 493–493 (2020).
48. Liu, Q. et al. Identification, characterization and functional differentiation of the *NAC* gene family and its roles in response to cold stress in ginseng, *Panax ginseng* C.A. Meyer. *PLoS ONE* **15**, e0234423 (2020).
49. Ma, X. C. et al. Seasonal drought promotes citrate accumulation in citrus fruit through the *CsABF3*-activated *CsAN1-CsPH8* pathway. *N. Phytol.* **242**, 1131–1145 (2024).
50. Hemm, M. R., Herrmann, K. M. & Chapple, C. *AtMYB4*: a transcription factor general in the battle against UV. *Trends Plant Sci.* **6**, 135–136 (2001).
51. Cheng, H. et al. Molecular cloning and function assay of a chalcone isomerase gene (*GbCHI*) from *Ginkgo biloba*. *Plant Cell Rep.* **30**, 49–62 (2011).
52. Zhang, S. Q. & Klessig, D. F. MAPK cascades in plant defense signaling. *Trends Plant Sci.* **6**, 520–527 (2001).
53. Wade, H. K., Bibikova, T. N., Valentine, W. J. & Jenkins, G. I. Interactions within a network of phytochrome, cryptochrome and UV-B phototransduction pathways regulate chalcone synthase gene expression in *Arabidopsis* leaf tissue. *Plant J.* **25**, 675–685 (2010).
54. Izumi, M., Ishida, H., Nakamura, S. & Hidema, J. Entire photodamaged chloroplasts are transported to the central vacuole by autophagy. *Plant Cell* **29**, 377–394 (2017).
55. Frohnmeyer, H. & Staiger, D. Ultraviolet-B radiation-mediated responses in plants. Balancing damage and protection. *Plant Physiol.* **133**, 1420–1428 (2003).
56. Chen, Z., Dong, Y. & Huang, X. Plant responses to UV-B radiation: signaling, acclimation and stress tolerance. *Stress Biol.* **2**, 51–61 (2022).
57. Zhuang, W. J. et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* **51**, 865–876 (2019).
58. Zhao, X. B., Fu, X. D., Yin, C. B. & Lu, F. Wheat speciation and adaptation: perspectives from reticulate evolution. *aBIOTECH* **2**, 386–402 (2021).
59. Chalhoub, B. et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953 (2014).
60. Huang, X. H. et al. Epigenomic and 3D genomic mapping reveals developmental dynamics and subgenomic asymmetry of transcriptional regulatory architecture in allotetraploid cotton. *Nat. Commun.* **15**, 10721 (2024).
61. González Besteiro, M. A. et al. *Arabidopsis* MAP kinase phosphatase 1 and its target MAP kinases 3 and 6 antagonistically determine UV-B stress tolerance, independent of the UVR8 photoreceptor pathway. *Plant J.* **68**, 727–737 (2011).
62. Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, 884–890 (2018).
63. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
64. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
65. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
66. Sun, P. C. et al. WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant.* **15**, 1841–1851 (2021).
67. Mckenna, A. et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
68. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
69. Graovac, M. T. & Chen, N. S. Using repeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* **25**, 4–10 (2009).
70. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinf.* **9**, 1–14 (2008).
71. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, 265–268 (2007).
72. Ou, S. J. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2017).

73. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
74. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
75. Engstrm, P. G. et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* **10**, 1185–1191 (2013).
76. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, 435–439 (2006).
77. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89 (2016).
78. Stiehler, F. et al. Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics* **36**, 5291–5298 (2020).
79. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, 1–11 (2008).
80. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
81. Bruno, F., Marinella, M. & Santamaria M. e-DNA meta-barcoding: from NGS raw data to taxonomic profiling. *Methods Mol Biol.* **69**, 257–278 (Humana Press, 2015).
82. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
83. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
84. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
85. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
86. Kulmanov, M., Khan, M. A. & Hoehndorf, R. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics* **34**, 660–668 (2017).
87. Tang, H. B. et al. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta* **3**, e211 (2024).
88. Karro, J. E. et al. Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res.* **35**, D55–D60 (2007).
89. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
90. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
91. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
92. Yang, Z. H. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
93. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
94. Jin, J. J. et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 1–31 (2020).
95. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
96. Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 1–13 (2019).
97. Goel, M. & Schneeberger, K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922–2926 (2022).
98. Wang, Y. P. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
99. Liu, B. et al. Single-gene FISH maps and major chromosomal rearrangements in *Elymus sibiricus* and *E. nutans*. *BMC Plant Biol.* **23**, 98–113 (2023).
100. Han, F. P., Lamb, J. C. & Birchler, J. A. High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc. Natl Acad. Sci. Usa.* **103**, 3238–3243 (2006).
101. Brown, J., Pirrung, M. & McCue, L. A. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **33**, 3137–3139 (2017).
102. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2013).
103. Zhang, K. et al. Resequencing of global Tartary buckwheat accessions reveals multiple domestication events and key loci associated with agronomic traits. *Genome Biol.* **22**, 1–17 (2021).
104. Zhang, C., Dong, S. S., Xu, J. Y., He, W. M. & Yang, T. L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).
105. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinf.* **12**, 1–6 (2011).
106. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
107. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *Plos One* **5**, e9490 (2010).
108. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
109. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
110. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
111. Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. PSMC-analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula flycatchers*. *Mol. Ecol.* **25**, 1058–1072 (2016).
112. Noskova, E., Ulyantsev, V., Koepfli, K. P., O'Brien, S. J. & Dobrynin, P. GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. *GigaScience* **9**, giaa005 (2020).
113. Gutenkunst, R., Hernandez, R., Williamson, S. & Bustamante, C. Diffusion approximations for demographic inference: DaDi. *Nat. Precedings.* **5**, 1 (2010).
114. Collin, F. D. et al. Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC random forest. *Mol. Ecol. Resour.* **21**, 2598–2613 (2021).
115. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
116. Beerli, P., Mashayekhi, S., Sadeghi, M., Khodaei, M. & Shaw, K. Population genetic inference with MIGRATE. *Curr. Protoc. Bioinf.* **68**, e87 (2019).
117. Langdon, B. W. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min.* **8**, 1–7 (2015).

118. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **12**, 323 (2011).
119. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
120. Ye, J., Mcginnis, S. & Madden, T. L. BLAST: improvements for better sequence analysis. *Nucleic Acids Res.* **34**, 6–9 (2006).
121. Song, W. M. & Zhang, B. Multiscale embedded gene co-expression network analysis. *Nucleic Acids Res.* **11**, e1004574 (2015).
122. Hamilton, N. E. & Ferry, M. ggtern: Ternary diagrams using ggplot2. *J. STAT Softw.* **87**, 1–17 (2018).
123. Frichot, E. & Francois, O. LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
124. Yin, L. L. et al. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteom. Bioinf.* **19**, 619–628 (2021).

Acknowledgements

This work was supported by funds from National Natural Science Foundation of China (32271753 & 32471764 to X.M.), the National Key R&D Program of China (2023YFF0805601 to Y.Y. & 2024YFD1301202 to X.M.), Sichuan Provincial Key Research and Development Project (2023YFSY0012 to S.J. & 2023YFN0087 to X.M.), the Key Science & Technology Project of Gansu Province (22ZD6NA007 to Y.Y.), the Fundamental Research Funds for the Central Universities (lzujbky-2022-ey07 to Y.Y.), the Gansu Province Postgraduate Innovation Star Program (No. 2025CXZX-099 to S.Y.), Sichuan Province's 14th Five-Year Plan Forage Breeding Breakthrough Project (2021YFYZ0013 to X.M.). We thank the High-performance Computing Platform of Sichuan Agricultural University for providing the analysis platform. All of the sequencing services were provided by the Haorui Genomics Technology Co., Ltd. (Xi' an, China).

Author contributions

X.M., J.L. and Y.Y. designed and coordinated the project. S.Y., J. Zhang and Yanli X. did the genome assembly. J.P. and S.Y. performed the genome annotation. L.L., Z.W. and X.P. did the transcriptome analysis. Y.X. and Yanli X. did the GEA and GWAS analysis and further gene function validation. Z.Z. and X.P. screened the homozygous overexpression and mutant lines of *Arabidopsis thaliana*. Q.Y., Y.X. and Yanli X. performed the UV-B radiation common garden experiment and traits measurement. C.J., G.L. and Z.Y. performed the FISH assay. Z.D. and Y.L. assembled the chloroplast genomes. D.L., X.F., L.S., X.L. and J. Zhao performed the phylogenetic and population genomic analysis. X.F. and L.S. contributed materials of *Elymus* and its related species. M.S., S.J. and S.B. performed DNA and RNA extraction. Y.X., Yanli X. and S.Y. wrote and edited the main manuscript. J.L., Y.Y. and X.M. revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58341-0.

Correspondence and requests for materials should be addressed to Shiqie Bai, Jianquan Liu, Yongzhi Yang or Xiao Ma.

Peer review information *Nature Communications* hanks Mario Caccamo, Ruohe Yin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.