nature genetics

Article

https://doi.org/10.1038/s41588-025-02294-z

Super-pangenome analyses across 35 accessions of 23 *Avena* species highlight their complex evolutionary history and extensive genomic diversity

Received: 31 May 2024

Accepted: 14 July 2025

Published online: 20 August 2025

Check for updates

Hongyu Zhang[1,4], Ningkun Liu[1,2,4], Yaru Wang[1,4], Xinyuan Zheng[1,4], Wei Li[1,2,4], Ze Liu[1], Jianan Liu[1], Yu Wang[1], Longsheng Xing[1], Tao Li[1], Yange Yun[1], Qinghong Zhou[1], Meijia Wang[1], Yujie Qin[1], Jinjiang Yan[1], Zhizhong Gong[1,3] ✉, Qiang He[1,2] ✉ & Huilong Du[1,2] ✉

Common oat, belonging to the genus *Avena* with 30 recognized species, is a nutritionally important cereal crop and high-quality forage worldwide. Here, we construct a genus-level super-pangenome of *Avena* comprising 35 high-quality genomes from 14 cultivated oat accessions and 21 wild species. The fully resolved phylogenomic analysis unveils the origin and evolutionary scenario of *Avena* species, and the super-pangenome analysis identifies 26.62% and 59.93% specific genes and haplotypes in wild species. We delineate the landscape of structural variations (SVs) and the transcriptome profile based 1,401 RNA-sequencing (RNA-seq) samples from diverse abiotic stress treatments in oat. We highlight the crucial role of SVs in modulating gene expression and shaping adaptation to diverse stresses. Further combining SV-based genome-wide association studies (GWASs), we characterize 13 candidate genes associated with drought resistance such as *AsARF7*, validated by transgenic oat lines. Our study provides unprecedented genomic resources to facilitate genomic, evolution and molecular breeding research in oat.

Common oat (*Avena sativa* L., 2n = 6x = 42, AACCDD), the production of which ranks seventh among cereals (FAOSTAT, accessed May 2021), is an economically important food and feed crop worldwide and is highly adaptable to various climatic conditions, especially drought[1]. Currently, advances in sequencing technology have accelerated genomic studies for agriculturally important crops such as rice, soybean and maize, but similar efforts in oat were more challenging[2]. This is largely owing to its considerable genome size (~11 Gb), highly repetitive sequences (~85%) and high ploidy[3,4]. Recently, three chromosome-level assemblies of cultivated oats were reported, which greatly promote the genomic study of oats[3,4]. However, the minor portion of diversity captured by these genomes is far from sufficient to represent genetic diversity in cultivated oats.

Common oat belongs to *Avena*, which contains 30 recognized species including diploids (AA or CC), tetraploids (CCDD or AABB) and hexaploids (AACCDD)[5,6], exhibiting phenotypic, ecogeographical and genetic diversity[7]. Previous studies based on chloroplast and mitochondrial genomes have revealed the major evolutionary relationship of

[1]College of Life Sciences, Institute of Life Science and Green Development, Hebei University, Baoding, China. [2]Hebei Basic Science Center for Biotic Interaction, Hebei University, Baoding, China. [3]State Key Laboratory of Plant Environmental Resilience, College of Biological Sciences, China Agricultural University, Beijing, China. [4]These authors contributed equally: Hongyu Zhang, Ningkun Liu, Yaru Wang, Xinyuan Zheng, Wei Li. ✉e-mail: gongzz@cau.edu.cn; heqiang_cqnu@163.com; huilongdu@hbu.edu.cn

Avena species and determined possible ancestral donors of polyploid oats[8–11]. However, for polyploids, phylogenies only based on chloroplast and mitochondrial genomes blocked exploration of the evolutionary position for the paternal subgenome. Furthermore, frequent incongruence of phylogenetic relationships between species trees and maternally inherited trees has been documented in many genera such as *Malus*[12], *Sphagnum*[13] and *Quercus*[14]. The shortage of genomes in *Avena* greatly limited the comprehensive study of evolutionary history, especially for the B lineage.

Pangenomes, which represent all DNA sequence diversity in a species[2,15], are becoming the new reference for genomic research in many crops, such as soybean, rice, wheat and tomato[2]. However, the domestication bottleneck and artificial selection have caused the loss of genetic diversity in many cultivated crops, which has hindered crop improvement and breeding[16–18]. Effective utilization of wild relatives is becoming key to provide various desirable traits[19]. Therefore, the application of pangenomes has recently moved to a higher taxonomic unit, including diverse wild relatives within a genus for super-pangenome construction[19–21].

Here, we constructed a genus-level super-pangenome of *Avena* comprising 35 high-quality genomes from 23 species. Our phylogenomic analysis elucidates the origin and evolutionary trajectory of *Avena*. The super-pangenome analysis uncovers abundant specific genetic resources in wild species and delineates the landscape of SVs related to abiotic stress resistance. Collectively, our findings offer valuable genomic resources, providing a robust foundation for oat functional genomics and molecular breeding.

Results

The collection of 14 representative cultivated oats and 21 wild oat species

To select the representative accessions of cultivated oats, we constructed a phylogenetic tree based on SNPs from 1,079 globally distributed oat accessions (Fig. 1a). Based on the phylogenetic tree, population structure, principal-component analysis (PCA) and geographical distributions, these oat accessions were classified into six major groups (Fig. 1a,b and Extended Data Fig. 1a). Further combining morphological diversity and breeding research utility, we selected 14 representative cultivated oat accessions, including *A. sativa* and *A. sativa* ssp. *nuda* (Fig. 1a–c,e, Extended Data Fig. 1b and Supplementary Table 1). These accessions, originating from nine countries, encompassed 83.67% of the SNP variation observed across the 1,079 oat accessions. To obtain comprehensive genetic information of *Avena*, we also collected 21 wild species, including 11 diploids (AA and CC genomes), seven tetraploids (AABB and CCDD genomes) and three hexaploids (AACCDD genomes), which covers all different ploidy levels and lineages (Table 1). High phenotypic diversity was observed, such as grain size, tiller number and plant height as well as the morphology of plant and spike (Fig. 1c,e and Extended Data Fig. 1c–e). Finally, a total of 35 oat accessions from 23 species were collected, representing 77% of the species in *Avena* and spanning cultivated oats and their wild relatives.

Genome assembly and annotation of 28 newly sequenced oat accessions

Among the 35 oat accessions, 28 were de novo assembled in this study (Table 1). For the 27 newly sequenced accessions, chromosome-level assemblies were achieved by integrating PacBio HiFi long-read sequences, high-throughput chromosome conformation capture (Hi-C) data and Illumina paired-end sequences (Extended Data Fig. 2a and Supplementary Table 2). The assembly sizes of the 27 diploid, tetraploid and hexaploid genomes ranged from 3,405 Mb to 4,190 Mb, from 6,843 Mb to 7,829 Mb and from 10,702 Mb to 11,013 Mb with average contig N_{50} values of 92.77 Mb, 79.13 Mb and 49.43 Mb, respectively (Fig. 1d, Table 1 and Supplementary Tables 2–4). To generate a high-quality reference genome of the cultivar C0191, we integrated an additional 263.5 Gb of Oxford Nanopore Technologies ultralong reads, obtaining the final assembly with a genome size of 10.94 Gb and a contig N_{50} of 429.52 Mb, leaving only 19 gaps (Fig. 1d and Table 1). Quality assessment revealed an average 99.89% mapping rate of Illumina reads and 99.1% benchmarking universal single-copy ortholog (BUSCO) completeness, indicating high accuracy, completeness and continuity of our assemblies (Extended Data Fig. 2b and Supplementary Tables 3 and 5).

To minimize the deviation that could arise from different prediction approaches, the repeat sequences and gene models of these 35 genomes were annotated using the same pipeline. Transposable elements (TEs) constituted 82.73–89.32% of each genome, averaging 87.36%, with long terminal repeat (LTR) being the most abundant, accounting for 46.34% to 66.02% of the genomes (Table 1, Supplementary Fig. 3a and Supplementary Table 6). Finally, we obtained 40,834–59,955, 91,635–96,971 and 131,823–134,713 protein-coding genes in diploid, tetraploid and hexaploid genomes (Table 1 and Supplementary Tables 7 and 8).

Fully resolved subgenome origin, evolution and polyploidization history of *Avena*

The subgenomes of polyploid oats were identified by integrating evidence from phylogenetic analysis, specific *k*-mer patterns and sequence similarity (Supplementary Figs. 4–6). This evidence enables the classification of the 40 subgenomes, sourced from 21 wild species and two cultivated oat species (C0191 and C1071), into four lineages: A, B, C and D (Fig. 2a,b). Taking *Triticum aestivum* and *Lolium perenne* as the outgroup, we constructed a deep phylogenetic tree for the 40 subgenomes from 23 species by using 2,456 single-copy genes (Fig. 2a). The C lineage was the first to diverge from the common ancestor of *Avena* approximately 10.87 (~9.87–11.81) million years ago (Mya). The other lineages, comprising different A subtypes (Ac, Ad, Al and As)[22] as well as the B and D subgenomes, formed a single large clade. Furthermore, the B lineage diverged from the common ancestor of the A, B and D lineages approximately 3.50 (~3.22–3.85) Mya[23], which was supported by the contrasting gene tree topologies and synonymous substitution (K_s) analysis (Extended Data Fig. 3a and Supplementary Fig. 7). Subsequently, the Ac/Ad subclade diverged with the subclades of As/Al and D around 3.11 (~3.00–3.39) Mya, followed by divergence between the D subclade and the As/Al subclade occurring 2.75 (~2.62–2.99) Mya. Notably, phylogenetic trees based on both nuclear and mitochondrial genomes showed that *Avena hirtula*, previously classified into As subtypes[23,24], was more closely related to the Ad/Ac species, which was further supported by coalescent-based topologies and fluorescence in situ hybridization (FISH) analyses (Extended Data Fig. 3b,c and Supplementary Fig. 8). Our densely sampled phylogeny of *Avena* clarified the evolutionary relationships of all subgenomes and lineages, especially for the B lineage.

To further assess genome evolution in *Avena*, we reconstructed the ancestral karyotype of the four lineages (Fig. 2b and Extended Data Fig. 3d). Extensive large-scale rearrangements were observed within and among the A, B, C and D lineages, especially between species with different ploidy levels, which suggested that polyploidization could greatly reshape genome architecture (Fig. 2b)[25,26].

Notably, a major conflict was observed for the classification of *Avena agadiriana*, which was previously designated as an AABB species[8,27,28]. Phylogenomic analyses revealed that the subgenomes of AABB species including *Avena barbata*, *Avena abyssinica* and *Avena vaviloviana* were separately grouped together, while two subgenomes of *A. agadiriana* exhibited closer relationships with *Avena canariensis* (Ac) and the As subtype species, respectively (Fig. 2a). Specific *k*-mer distribution, genomic synteny and sequence similarity analyses all demonstrated obviously diverse patterns between *A. agadiriana* and the other three AABB species (Extended Data Fig. 4a–d and Supplementary Table 9). Our coalescent-based phylogeny analysis and FISH analysis also confirmed that *A. agadiriana* is not an AABB species (Fig. 2c),

https://doi.org/10.1038/s41588-025-02294-z

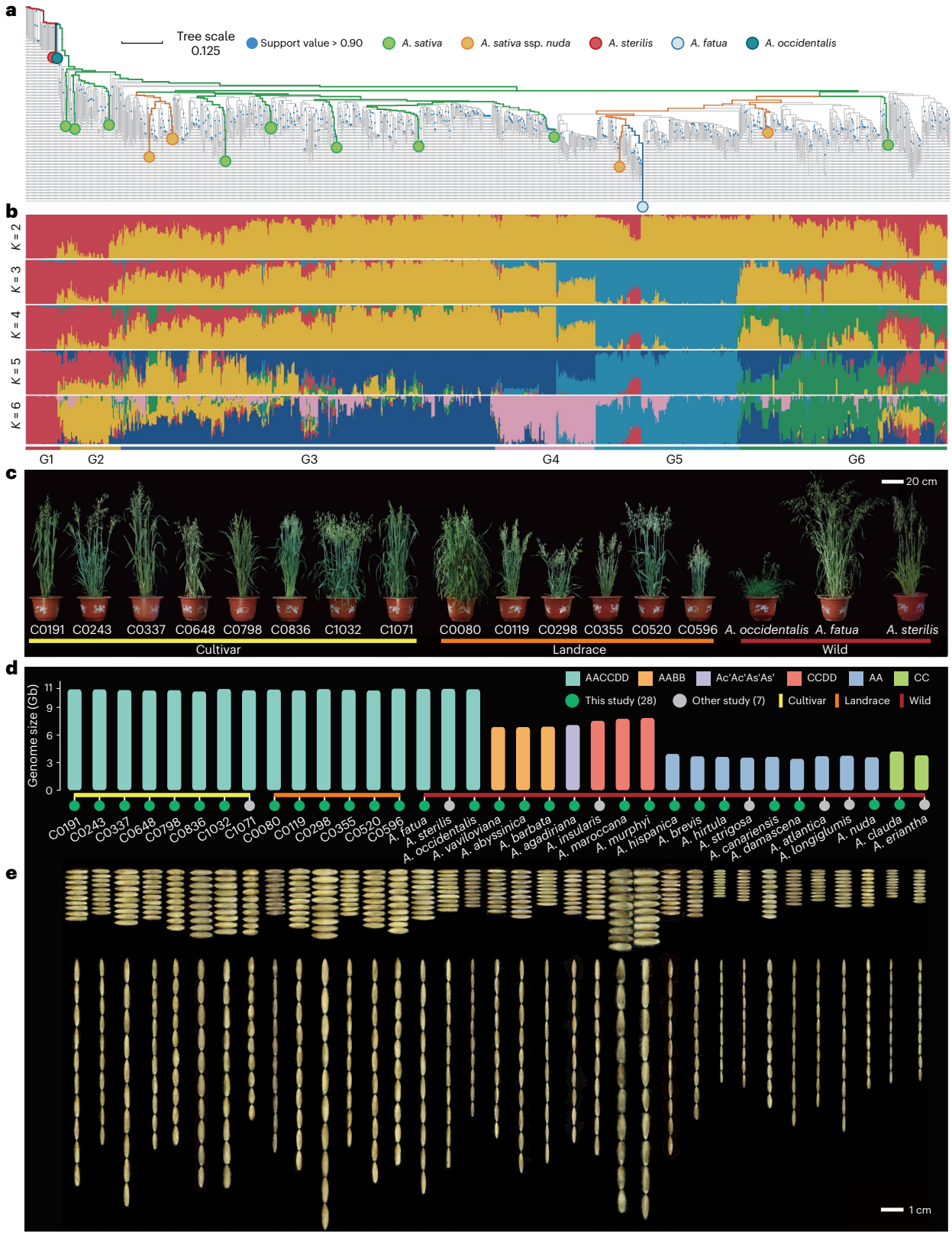


Fig. 1 | Phylogeny, population structure and diverse phenotypes of cultivated oats and wild species in *Avena*. **a**, Phylogenetic tree of the 1,079 hexaploid oat accessions with C0191 as the reference genome. Lines with different colors represent the selected 16 hexaploid oats as follows: wild species including *A. occidentalis* (cyan), *A. sterilis* (red), *A. fatua* (blue); cultivated oats (*A. sativa* (green), *A. sativa* ssp. *nuda* (orange)). Bootstrap support values are indicated with blue dots. **b**, Model-based clustering analysis with different numbers of ancestry kinship ($K = 2–6$). **c**, Highly diverse plant architectures of the 17 wild, cultivar and landrace varieties of hexaploid oats. **d**, Genome size, genomic compositions and domestication status of the 35 oat accessions. Yellow, orange and red lines represent cultivated, landrace and wild oats, respectively. Green and gray circles represent genomes from this and other studies, respectively. Reseda, orange, lilac, pink, blue and prasinous columns represent genomes with different genomic compositions of AACCDD, AABB, Ac'Ac'As'As', CCDD, AA and CC, respectively. **e**, Diversity of grain width and length of the 35 oat accessions with order corresponding to that in **d**.

https://doi.org/10.1038/s41588-025-02294-z

Table 1 | Summary of the assembly and annotation of 35 oat genomes

Accession	Genome constitution	Assembly length (Gb)	Chromosome loading ratio (%)	Contig N_{50} (Mb)	Gene number	Repetitive content (%)	BUSCO (%)	Source
***A. sativa* (C0191)**	AACCDD	10.94	99.94	429.52	131,823	88.08	98.4	This study
***A. sativa* (C0243)**	AACCDD	10.93	99.03	21.98	133,978	88.39	98.4	This study
***A. sativa* (C0298)**	AACCDD	10.97	99.66	69.83	133,773	88.59	98.4	This study
***A. sativa* (C0337)**	AACCDD	10.85	99.73	63.32	133,855	88.08	98.6	This study
***A. sativa* (C0355)**	AACCDD	10.88	99.73	32.45	134,062	88.37	98.4	This study
***A. sativa* (C0520)**	AACCDD	10.81	99.85	61.59	133,586	88.07	98.5	This study
***A. sativa* (C0596)**	AACCDD	11.01	99.07	28.79	133,835	88.83	98.5	This study
***A. sativa* (C0648)**	AACCDD	10.81	99.81	68.21	133,991	88.21	98.6	This study
***A. sativa* (C0836)**	AACCDD	10.70	99.91	11.29	133,998	88.91	98.6	This study
***A. sativa* (C1032)**	AACCDD	10.97	98.94	46.54	133,918	88.11	98.5	This study
***A. sativa* ssp. *nuda* (C0080)**	AACCDD	10.91	99.73	42.89	133,883	88.26	98.5	This study
***A. sativa* ssp. *nuda* (C0119)**	AACCDD	10.82	99.73	67.29	133,981	88.11	98.6	This study
***A. sativa* ssp. *nuda* (C0798)**	AACCDD	10.84	99.76	6.51	133,977	88.37	98.5	This study
***A. sativa* ssp. *nuda* (C1071)**	AACCDD	10.83	99.84	104.56	134,047	89.08	99.5	Other study[48]
***Avena fatua* (W0020)**	AACCDD	10.98	99.40	159.31	134,713	88.45	98.4	This study
***Avena sterilis* (W0038)**	AACCDD	10.99	99.93	473.40	133,364	89.32	98.4	Other study[49]
***Avena occidentalis* (W0039)**	AACCDD	10.94	99.46	12.00	132,046	88.24	98.3	This study
Avena vaviloviana	AABB	6.85	99.58	41.22	91,635	87.34	98.1	This study
Avena abyssinica	AABB	6.84	99.66	42.06	92,661	87.32	98.1	This study
Avena barbata	AABB	6.89	99.94	53.74	93,821	87.8	97.9	This study
Avena agadiriana	Ac'Ac'As'As'	7.06	99.73	166.56	93,848	87.43	98.1	This study
Avena insularis	CCDD	7.52	95.12	5.64	92,973	86.43	98.5	Other study[4]
Avena maroccana	CCDD	7.74	99.92	113.31	96,131	88.3	98.4	This study
Avena murphyi	CCDD	7.83	99.78	131.39	96,971	88.33	98.4	This study
Avena hispanica	AsAs	3.94	92.8	108.30	59,955	82.73	98.6	This study
Avena brevis	AsAs	3.67	99.23	121.12	44,693	86.21	97.1	This study
Avena hirtula	AsAs	3.61	99.84	66.25	43,905	85.82	97.9	This study
Avena strigosa	AsAs	3.53	99.89	4.65	44,132	84.88	97.3	Other study[50]
Avena nuda	AsAs	3.57	99.92	201.42	47,342	86.33	97.2	This study
Avena canariensis	AcAc	3.61	99.81	330.51	45,916	85.97	97.3	This study
Avena atlantica	AsAs	3.69	96.38	4.31	46,943	86.14	97.2	Other study[51]
Avena damascena	AdAd	3.40	99.84	83.77	43,732	85.6	97.3	This study
Avena longiglumis	AlAl	3.74	99.26	7.30	46,240	87.05	97.1	Other study[4]
Avena clauda	CpCp	4.19	99.96	91.54	42,171	86.24	97.5	This study
Avena eriantha	CpCp	3.78	97.28	1.31	40,834	84.12	97.5	Other study[51]

supporting that *A. agadiriana* has undergone a separate and distinct evolutionary history. Finally, based on the phylogenetic relationship, sequence similarity and K_s analysis of the subgenomes (Fig. 2a and Extended Data Fig. 4d,e), we designated the divergent karyotype of *A. agadiriana* as Ac'Ac'As'As'.

Finally, combining the phylogenetic topologies using chloroplast and mitochondrial genomes, sequence similarity, divergence time estimation and K_s analysis, we proposed a comprehensively refined model for the origins, polyploidizations and evolutionary trajectories of *Avena* species (Fig. 2d, Extended Data Fig. 5a–f and Supplementary Fig. 11). Tetraploid AABB ancestors originated around ~0.02–1.24 Mya from the hybridization between a paternal BB diploid and a maternal AsAs diploid (Extended Data Fig. 5c). *A. agadiriana* originated from the hybridization between a paternal AcAc diploid and a maternal AsAs diploid around ~1.56–2.95 Mya (Extended Data Fig. 5d). The ancestor of tetraploid CCDD species originated around ~0.43–4.99 Mya from the hybridization between a paternal CC diploid and a maternal DD diploid and then diverged into *Avena murphyi*, *Avena maroccana* and *Avena insularis* successively at 0.8 Mya and 0.6 Mya, respectively (Extended Data Fig. 5b). The wild hexaploid oat originated from the hybridization between a paternal As/Al ancestor and a maternal CCDD tetraploid that is closely related to *A. insularis* around ~0.23–0.73 Mya (Extended Data

https://doi.org/10.1038/s41588-025-02294-z

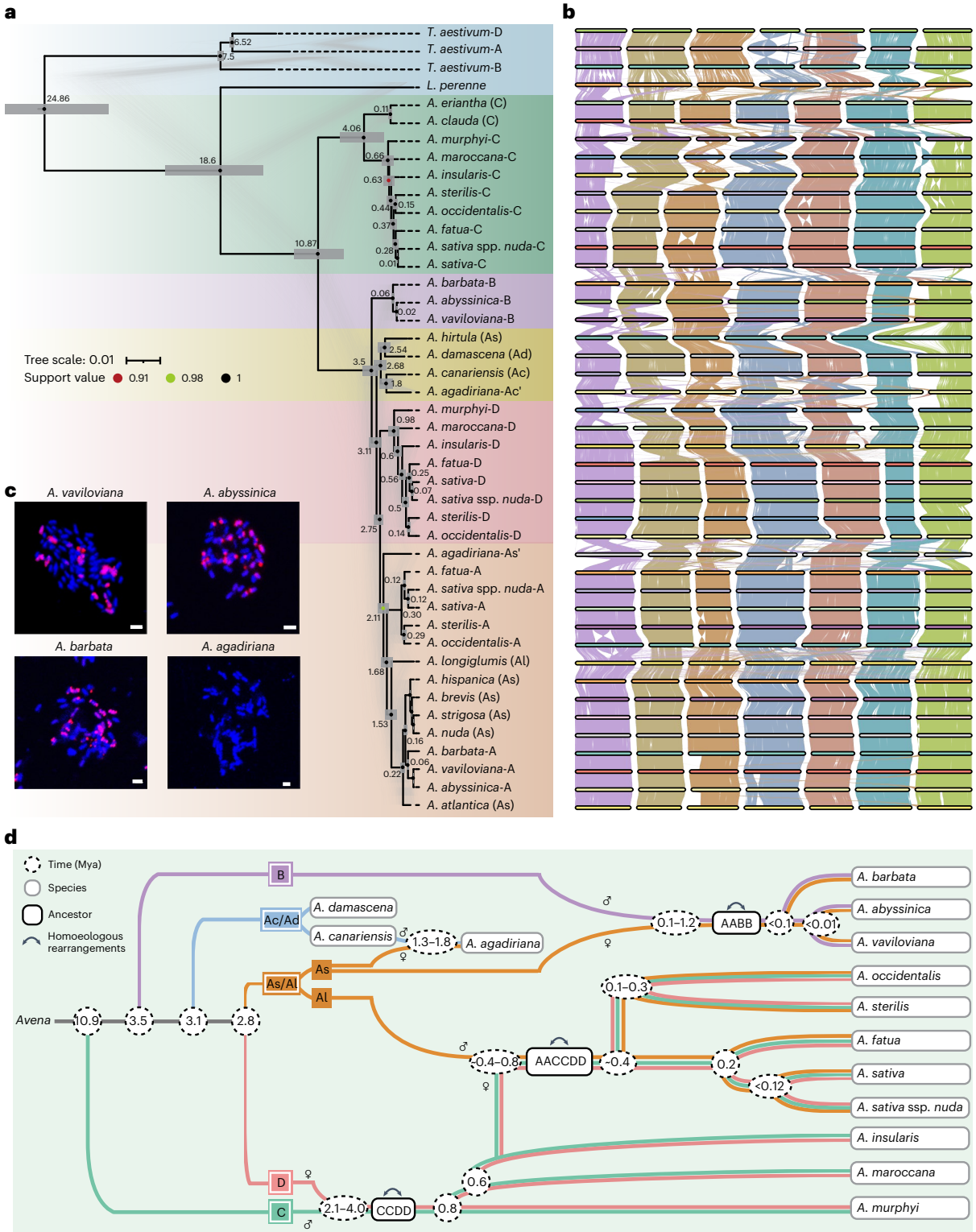


Fig. 2 | Inferred phylogeny, origin and evolution of *Avena* species. **a**, Phylogenetic relationships and timescale of 44 subgenomes in 25 species (40 subgenomes in 23 species are from *Avena*) inferred on the basis of 2,456 single-copy orthologous groups using RAxML taking *T. aestivum* and *L. perenne* as outgroup species. Orange, purple, green, red and blue backgrounds represent A, B, C and D lineages and outgroups, respectively. Bootstrap support values are indicated with black, red and green dots. The scale bar represents branch lengths, which correspond to the mean number of substitutions per site in the alignments. The gray bar indicates the confidence interval of differentiation time for each node. Trees for each single-copy gene are depicted by gray lines. **b**, Homoeologous synteny blocks among the 44 subgenomes correspond to **a**. Conserved syntenic blocks of different chromosomes (1–7) are presented with different colors. **c**, A specific FISH probe targeting the B subgenome can successfully label *A. barbata*, *A. abyssinica* and *A. vaviloviana* but not *A. agadiriana*, as observed in cells at the metaphase of mitosis. Scale bars, 2 µm. **d**, Model for the origins and evolutionary history of diploid oat ancestors and polyploidization events. Diploid progenitors (Ac/Ad, As/Al and B–D genomes) are indicated by different colors. Approximate dates of hybridization and differentiation events in Mya are given in the circles. Bidirectional arrows indicate that intersubgenomic exchange events occurred in allopolyploids. The tree also illustrates the maternal and paternal donors of the major clades.

Fig. 5e), consistent with a previous study[4]. *Avena sterilis* and *Avena occidentalis* were placed as a sister group of *Avena fatua* and *A. sativa* with a divergence time of ~0.15–0.79 Mya (Extended Data Fig. 5f). Notably, *A. sterilis* was known as the likely ancestral wild species of cultivated oat, while our results indicated that *A. fatua* was phylogenetically closest to cultivated oats (Extended Data Fig. 3e,f)[29–31]. These results demonstrate the necessity of the nuclear genome in exploring the subgenomic origin, polyploidization and evolution for polyploid species and provide detailed clues to the complicated evolution history of *Avena* species.

Super-pangenome construction for the genus *Avena*

To obtain a comprehensive gene repertoire for the *Avena* genus, we constructed a super-pangenome based on 76 (sub)genomes from 35 accessions and obtained 75,830 gene families (Fig. 3a). The super-pangenome contained 5,796 (8.03%), 12,428 (16.04%), 46,210 (60.92%) and 11,396 (15.10%) core, softcore, dispensable and specific gene families, respectively (Fig. 3b and Supplementary Table 10). Additionally, we also constructed pangenomes based on the subgenomes from cultivated and wild oats, identifying 55,633 and 70,158 gene families in cultivated and wild pangenomes, respectively (Extended Data Fig. 6a). Notably, wild oats (92.52%, 70,158 of 75,830) contributed 20,197 new gene families that were absent in cultivated oats (73.37%, 55,633 of 75,830) (Fig. 3a and Extended Data Fig. 6a). Distinct characterizations between the wild pangenome and the cultivated pangenome were observed in terms of coding sequence length, annotated Pfam domain rate, K_a/K_s ratio and gene expression level (Extended Data Fig. 6b–d). These results reveal higher genetic diversity in wild species and provide a comprehensive genus-wide gene repertoire for potential usage in oat biology and breeding.

To explore subgenome divergence in oat, we compared the features of genomic sequences and gene families among the A, C and D subgenomes of hexaploid oats and diploid–tetraploid oats (Fig. 3c). The results showed that the A, C and D subgenomes of hexaploid oats contained 3,194, 5,172 and 2,729 unique gene families, respectively. The A, C and D subgenomes of diploid–tetraploid oats contained 5,919, 2,752 and 1,033 unique gene families, respectively. Furthermore, differences were also observed in enriched gene ontology (GO) terms across the A, C and D subgenomes in both hexaploid and diploid–tetraploid oats. Unique gene families in the A subgenome of hexaploid oats were primarily associated with disease resistance and adaptability, whereas those in the D subgenome were mainly involved in organ development of the plant (Fig. 3d and Supplementary Table 11). For example, the number of genes encoding cellulose synthase, seed storage protein (SSP) and nucleotide-binding site (NBS) in the A and D subgenomes were much more than those in the C subgenomes in hexaploid oats (Fig. 3e, Extended Data Fig. 6h and Supplementary Table 12). These findings reveal a degree of functional differentiation among the A, C and D subgenomes in hexaploid oats and within the *Avena* genus, supported by differences in TE coverage, conservative gene loss ratio and K_a/K_s analysis (Extended Data Fig. 6e–g and Supplementary Fig. 16).

To further capture the genetic diversity in wild oats, we performed a comparative analysis on haplotypes and gene families between wild and cultivated oats. Wild oats harbored much more haplotypes and protein sequence variations than cultivated oats (Fig. 3f,g). Specifically, wild oats increased by 26.63% for new gene families and 59.93% for new haplotypes (Fig. 3g and Extended Data Fig. 6i). For example, the gene families related to NBS encoding, SSP and cellulose synthase in wild oats increased 10.60%, 27.42% and 21.86% of new genes and 59.56%, 59.09% and 58.83% of new genotypes, respectively (Fig. 3i and Supplementary Fig. 17). Notably, *A. fatua* contributed the most (11.2%) to the added genes and *A. vaviloviana* contributed the most (7.13%) to the added genotypes (Fig. 3h and Supplementary Table 13). These findings highlight the rich genetic resources in wild species and provide valuable insights for improving cultivated oats.

Previous studies have reported the use of wild oats to improve protein content, yield and resistance to powdery mildew and stem rust in cultivated oats[32–35]. We detected several large-scale introgression regions from wild oats into cultivated oats (Fig. 3j and Extended Data Fig. 6j), indicating that hybridization with wild oats has occurred in cultivated oats to improve their quality and adaptability. For instance, a region on chromosome 2A in C1032 and C0648 showed the same haplotype as wild species (Fig. 3j), and multiple genes related to environmental adaptability and stress response were identified in this region (Supplementary Table 14). In *AsPik2*, homologous to the rice blast resistance gene *Pik2* (National Center for Biotechnology Information (NCBI) 123147882)[36], four wild-derived nonsynonymous mutations were identified located in the exon regions of C1032 and C0648 (Fig. 3k). These results provide additional genetic evidence for using wild oats to improve cultivated oats, highlighting the importance of the rich genetic resources in our super-pangenome.

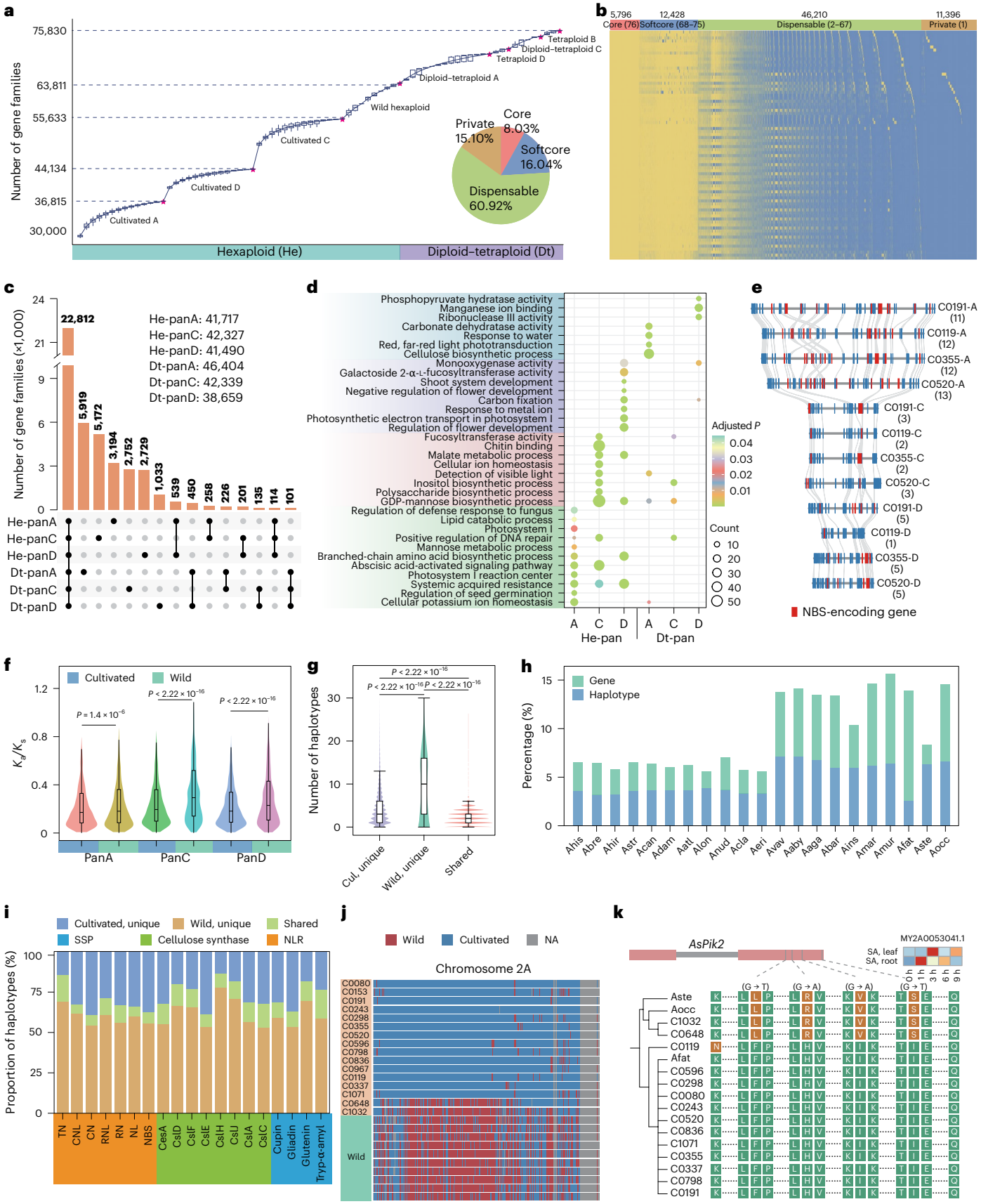
Landscape of SVs among 17 hexaploid oats

Taking the C0191 genome as the reference, we identified 1,377,775 SVs across 16 hexaploid genomes, and the number of SVs ranged from 58,817 to 127,679 per accession (Fig. 4b and Supplementary Table 15). These SVs were further merged into 555,764 nonredundant SVs, including 280,480 insertions, 267,883 deletions, 2,656 inversions and 4,745 translocations (Fig. 4a,b, Extended Data Fig. 7e and Supplementary Table 16). Fifty SVs were randomly selected for validation by read mapping, Hi-C heatmap and PCR. Almost all these SVs (45 of 50) proved to be correct (Extended Data Fig. 7a–d and Supplementary Table 17). SV-based phylogenetic

Fig. 3 | Super-pangenome of *Avena* and rich specific genetic resources in wild species. a, Modeling of pangenome sizes when incorporating additional subgenomes from all species in *Avena*. The pie chart shows the proportion of different groups of gene families. The red pentagrams represent the number of nonredundant genes in the pangenome constructed in a fixed increasing order. **b**, Presence and absence information of pan gene families in 76 subgenomes. Core, softcore, dispensable and private gene families are represented, which were present in 76 subgenomes, 68–75 subgenomes, 2–67 subgenomes and one subgenome, respectively. **c**, Upset plot of shared and unique gene families in panA, panC and panD of hexaploid (He) oats as well as tetraploid and diploid (Dt) oats. Numbers in the top right represent the number of nonredundant gene families for each pangenome. **d**, Functional analysis of unique genes in panA, panC and panD of 17 hexaploid oats as well as tetraploid and diploid oats, respectively. **e**, Synteny plots of a region showing the obvious expansion of resistance genes in A subgenomes for the 12 representative subgenomes in the A, C and D lineages. Red boxes indicate NBS-encoding genes. Blue boxes denote other genes. Gray lines identify syntenic gene pairs. Numbers in parentheses indicate NBS-encoding gene numbers in these regions. **f**, Comparison of K_a/K_s values of genes in panA, panC and panD in cultivated and wild oats. $n = 10{,}000$. **g**, Comparison of the number of haplotypes in each homologous gene group in cultivated (cul) and wild oats. $n = 13{,}1823$. In **f** and **g**, box plots show the 25th (lower edges) to 75th (upper edges) percentiles with median lines, and whiskers extending to 1.5× the interquartile range. **h**, Percentage of unique genes and genotypes in each wild species. The four letters on the *x* axis represent the abbreviation of the Latin name of these species. **i**, Comparison of the proportion of haplotypes for cellulose synthase, NBS-encoding and SSP gene families in cultivated and wild oats. NBS, nucleotide-binding site; CC, coiled-coil; LRR, leucine-rich repeat; TIR, toll-interleukin-1 receptor; RPW8, resistance to powdery mildew locus 8; CN, CC-NBS; CNL, CC-NBS-LRR; NL, NBS-LRR; TN, TIR-NBS; RNL, RPW8-NBS-LRR; RN, RPW8-NBS **j**, The genotype of chromosome 2A of hexaploid oats. NA represents the absence of SNP. **k**, A detailed alignment of the sequence of *AsPik2*, homologous to the rice blast resistance gene, in cultivated and wild oats. The heatmap in the top right corner shows transcriptome expression of the homologous gene of *AsPik2* at 0, 1, 3, 6 and 9 h after salicylic acid (SA) treatment of the root and leaf in C0836.

https://doi.org/10.1038/s41588-025-02294-z

analysis revealed the great impacts of SVs on genome divergence during the process of oat domestication and improvement (Fig. 4b). Curiously, C0298, one of the cultivars, was clustered into the wild oat clade, and haplotype analysis showed that it had multiple large-scale introgression regions from wild accessions, indicating its hybridization events with wild oats (Fig. 4b and Extended Data Fig. 7f). Additionally, we found that wild oats harbored significantly larger amounts of SVs than cultivated oats, while no significant difference was observed between hulled and



https://doi.org/10.1038/s41588-025-02294-z

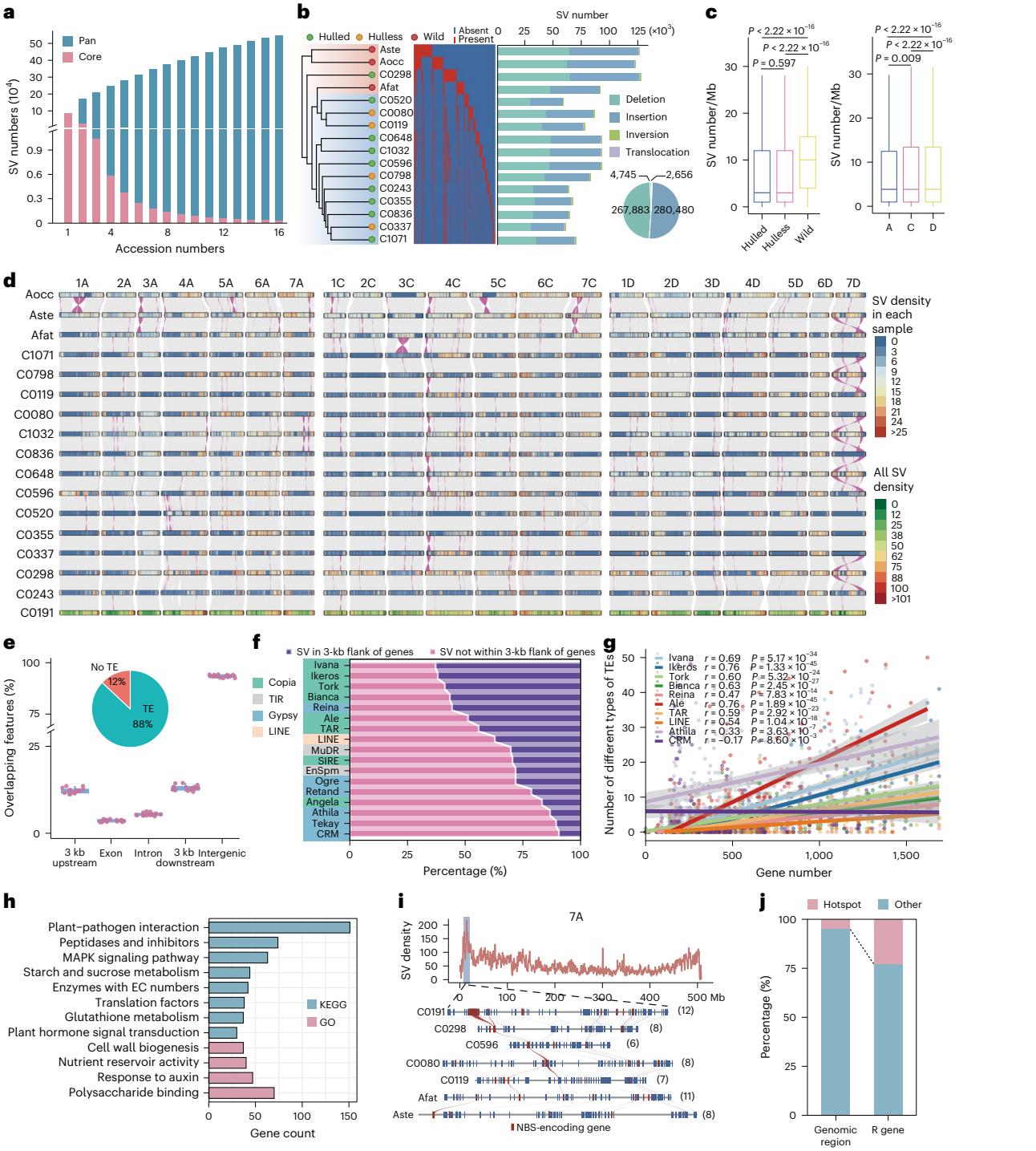


Fig. 4 | Characteristics of SVs among the 17 hexaploid oats. a, Accumulated number of pan and core SVs from each sample. These SVs were merged into 555,391 nonredundant SVs. **b**, Phylogenetic tree and distribution of species-specific and shared SVs across 16 oat accessions. The histogram represents the number of different types of SVs. The pie chart indicates the proportion of different types of nonredundant SVs. **c**, The number of SVs per 1-Mb window in wild (n = 32,817), hulled (n = 98,451) and hulless (n = 43,756) oats and the number of SVs per 1 Mb in the A (n = 54,912), C (n = 67,952) and D (n = 52,160) subgenomes. Box plots show the 25th (lower edges) to 75th (upper edges) percentiles with median lines, and whiskers extending to 1.5× the interquartile range. **d**, Chromosomal synteny and landscape of SV density for 16 oat accessions. Purple blocks indicate inversion. Heatmap tracks for each chromosome (16 accessions) represent SV density per 1 Mb in each accession. Heatmap tracks for each chromosome of the reference genome (C0191) represent the density of merged SVs per 1 Mb. **e**, Percentage of SVs in the various genomic features, including gene bodies and flanking regions. The number of samples in each feature is 16. Each dot represents one accession. Proportion of SVs in repeat and nonrepeat regions is shown in the pie chart. **f**, Presence preference of different TE families in SVs within genes and 3-kb flanking regulatory regions. The proportion of each type of LTR located in the nearby gene and its 3-kb flanking regions to the respective total LTRs within SV regions. **g**, Correlation analysis of gene density and the distribution of different TE families per 50 Mb. The gray region in the background represents the 95% confidence interval. **h**, Kyoto Encyclopedia of Genes and Genomes (KEGG) and GO enrichment analysis of genes in SV hotspot regions. EC, Enzyme Commission. **i**, A representative SV hotspot region on chromosome 7A. Seven representative accessions were used here for displaying the SVs and NBS-encoding genes in this hotspot region. Numbers in parentheses indicate the number of NBS-encoding genes in this region. **j**, Comparison of the percentage of the number of R genes between SV hotspot regions and whole-genomic regions.

hulless oats, which was consistent with the mosaic genetic composition due to sufficient hybridization (Fig. 4c).

Most SVs were located in intergenic and gene-flanking regions, with density increasing with distance in gene-flanking regions (Fig. 4e and Extended Data Fig. 7g). Furthermore, we found that 88% of SVs overlapped with TEs, with Gypsy being the most abundant in oat (Fig. 4e and Supplementary Fig. 18). Many TE families, such as Ivana (62.84%), Ikeros (61.88%), Tork (58.81%) and Bianca (56.63%), tended to be located in SVs near genes compared to other TE families such as Angela (16.72%), Athila (13.22%), Tekay (10.59%) and CRM (9.39%) (Fig. 4f). Further combing the correlation analysis, we found that gene density was significantly related to the distribution of these TE families, which showed strong enrichment nearby and within genes (Fig. 4g, Supplementary Fig. 19 and Supplementary Table 18). These results indicate that diverse TE families with presence preference may have played different roles in driving genetic variation formation.

To further investigate the highly diversified genomic regions, 427 SV hotspot regions were identified (Fig. 4d and Supplementary Table 19), and genes in these regions were mainly related to environmental adaptation such as 'plant–pathogen interaction', 'MAPK signaling pathway' and 'nutrient reservoir activity' (Fig. 4h). Notably, one SV hotspot region containing 545 independent SVs and harboring 17 NBS-encoding genes was identified, among which four were reported to be related to resistance to powdery mildew (Fig. 4i). Aside from this, we found that 5% of genomic hotspot regions harbored 22.86% of NBS-encoding genes (Fig. 4j), which further suggested that SVs in oat accessions may play important roles in adaptation to different environments. The landscape of SVs from these 17 wild and cultivated oats provides a valuable resource for exploring the effects of SVs on genomic evolution, genetic diversity and phenotypic variation in oat.

The impact of SVs on modulating gene expression under stress treatment in oat

A total of 1,401 RNA-seq samples covering diverse tissues (root, stem, leaf, flower and seed) and stress treatments (drought, heat, salt and alkali) from 17 hexaploid oat accessions were collected to explore the impact of SVs on gene expression and environmental adaptation (Fig. 5a, Extended Data Fig. 8a,b and Supplementary Table 7). According to previously reported approaches in wheat[37], oat[3,4] and bamboo[38], we evaluated the subgenome dominance of our accessions. First, we defined 13,374–14,386 homoeologous gene triads with a 1:1:1 correspondence across three subgenomes in 17 hexaploid oats and 16,151–15,533 homoeologous gene pairs in tetraploid oats (Supplementary Table 20). Our results indicated that, compared to A and D subgenomes, the expression of C subgenomes was suppressed in almost all accessions including hexaploid oats and tetraploid oats, showing a subtly yet significantly lower relative abundance (Fig. 5b, Extended Data Fig. 8c and Supplementary Table 20). Additionally, dynamic expression dominance between A and D subgenomes was observed, with the dominant subgenome shifting across the 17 hexaploid oats (dominant in A, 12; dominant in D, three) (Fig. 5b). Given that A subgenomes were newly introduced into hexaploid oats and were closely related to D subgenomes, dynamic subgenome dominance potentially represents the first step toward functional innovation through neofunctionalization or subfunctionalization.

To investigate the impact of SVs on gene expression and phenotypic variation, a total of 70,132 SV genes were determined. Gene expression analysis showed that SV genes had significantly lower expression levels than no-SV genes (Fig. 5c and Extended Data Fig. 8d). We found that SVs located in coding regions of genes were more likely to affect gene expression than those located in introns and upstream or downstream of genes, and SV genes had more silenced genes than no-SV genes (Fig. 5d and Extended Data Fig. 8d). Given that SVs can either promote or suppress gene expression, we refined the effects of SVs on gene expression patterns. The 17 hexaploid oat accessions were categorized into two groups (presence and absence) based on the SV genotype for each SV gene. Our results revealed that the number of SV genes with suppressed expression (38,302 SV genes) were much more than those with promoted expression (22,210 SV genes) (Fig. 5e and Extended Data Fig. 8e). Furthermore, a greater proportion of SV genes exhibited more than twofold expression changes following stress treatments compared to no-SV genes (Fig. 5f,g), indicating that SV genes were more sensitive to environmental stresses[39,40].

A total of 14,348, 10,140, 16,629 and 11,939 differentially expressed genes were detected after drought, heat, salt and alkaline stresses, respectively (Fig. 5h and Extended Data Fig. 9a). Functional analyses indicated that these genes were involved in defense response, abscisic acid binding, response to heat and heat shock protein binding (Extended Data Fig. 9b and Supplementary Table 21). We further identified SVs significantly associated with gene expression levels (SV-eGenes, Wilcoxon signed rank test) in hexaploid oats under drought (494), heat (309), salt (460) and alkaline (337) treatments (Fig. 5h). Many of these SV-eGenes have been reported to be involved in the stress response (Fig. 5i and Supplementary Table 22), such as *DRO1* in drought stress[41], *OsGA2ox6* in heat stress[42] and *OsPRR73* and *OsVP1* in salt and alkaline stress[43,44]. Quantitative PCR with reverse transcription (RT–qPCR) results further proved that these genes were upregulated or downregulated under stress treatment and also showed significant expression differences among accessions with and without SVs (Extended Data Fig. 9c–f). Overall, our study highlights the crucial role of SVs in modulating gene expression under stress treatment and presents a substantial set of candidate genes associated with the response to diverse stresses in oat.

Contributions of SVs to drought tolerance in oat

Oat is well-known for its strong drought tolerance compared to major crops such as rice, wheat, corn and barley[45]. To characterize the SVs underlying drought tolerance in oat, a GWAS based on both SVs and SNPs was conducted with 186 oat accessions (Fig. 6a and Supplementary Table 23). We obtained 2,223 candidate SVs with $P < 1.0 \times 10^{-5}$ and 3,154 related genes (SVs within gene bodies or 1-Mb flanking regions). To further identify drought resistance-related genes with expression affected by SVs, we combined the results of SV-based GWAS, differential expression and SV-eGene analyses (Extended Data Fig. 10a), identifying many previously reported drought tolerance-related genes (Supplementary Table 24). Finally, we obtained 13 candidate genes with SVs significantly associated with drought tolerance and gene expression under drought treatment (Fig. 6b).

Notably, the gene *A105907C064235.1* (*AsARF7*), orthologous to *AtARF7* (NCBI 100191131), encoding a putative auxin response factor (ARF), was identified in the significant peak on chromosome 7C. The leading SV, located 456 bp upstream of this gene's translation start site (Fig. 6c), included a WRKY transcription factor binding site and was significantly correlated with drought tolerance ($P = 7.76 \times 10^{-23}$). This SV was present in 80.00% of high-drought-resistant accessions and only in 10.89% of low-drought-resistant accessions (Fig. 6d). Phenotypic observation further confirmed that oat accessions with this SV (haplotype 2) showed significantly higher drought tolerance than those without this SV (haplotype 1) (Fig. 6e). Collinearity analysis with wild species indicated that this SV originated from ancestral diploid species (*Avena eriantha*) (Fig. 6f). Transcriptome and RT–qPCR analyses revealed significantly higher expression of *AsARF7* in oat accessions with the haplotype 1 genotype than in those with the haplotype 2 genotype (Fig. 6g,h and Supplementary Tables 7 and 26). Additionally, in *Nicotiana benthamiana* plants, the expression of *AsARF7* with haplotype 1 was also significantly higher than that with haplotype 2, and both genotypes showed downregulated expression of *AsARF7* under drought treatment (Fig. 6i). These results indicate that the SV is significantly associated with drought tolerance in oat and affects the expression of *AsARF7* under drought treatment.

https://doi.org/10.1038/s41588-025-02294-z

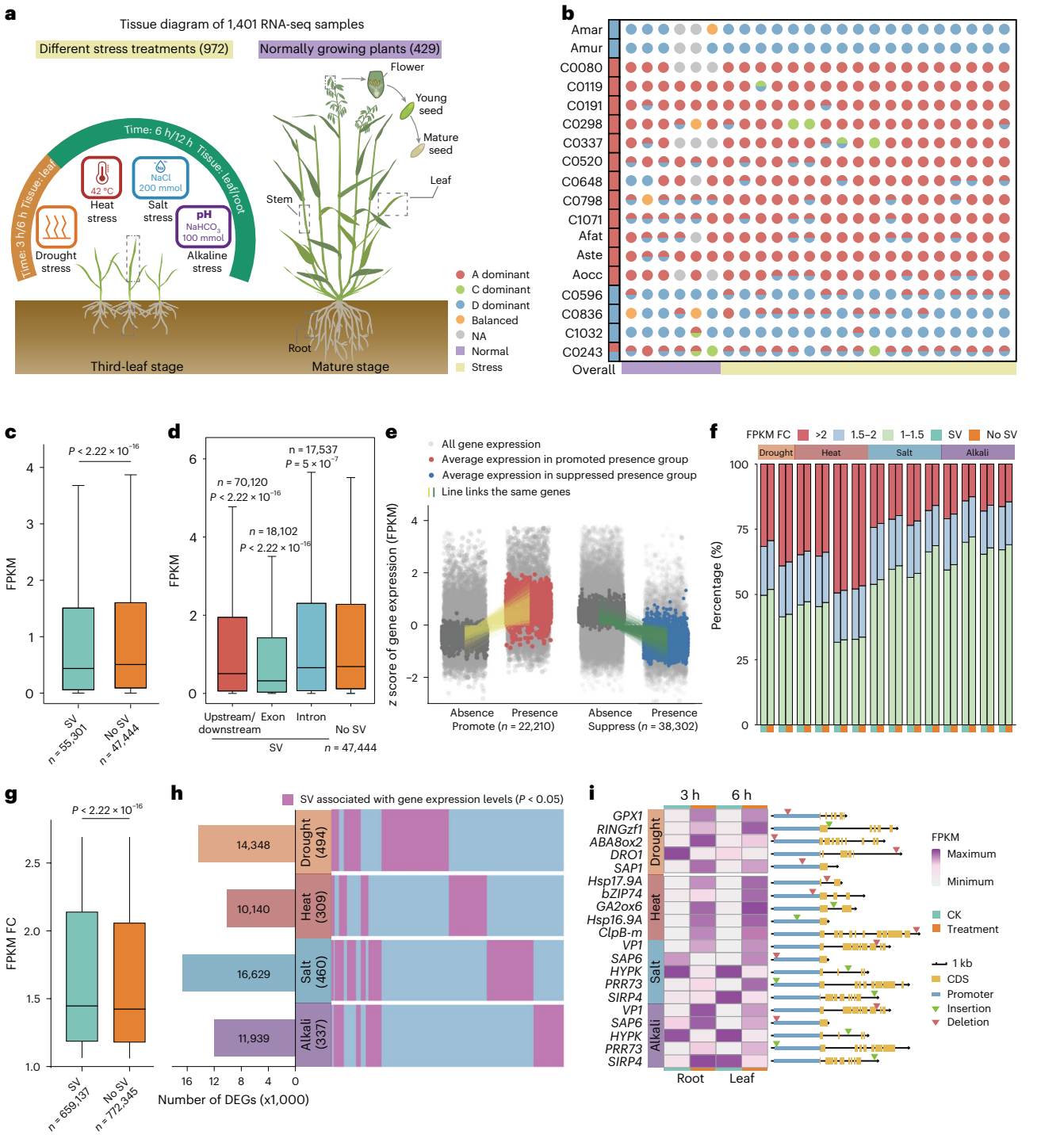


Fig. 5 | The impact of SVs on gene expression alteration. a, Schematic overview of RNA-seq samples across multiple tissues with different stress treatments (left) and normally growing plants (right). **b**, The dominant subgenome of the 18 oat accessions (16 hexaploid oats and two tetraploid oats) using RNA-seq samples shown in **a**. The four letters on the *y* axis represent the abbreviation of the Latin name of these species. **c**, Comparison of expression levels of SV genes and no-SV genes. FPKM, fragments per kilobase of transcript per million fragments mapped. **d**, Comparison of expression levels of SV genes with SVs located in different genic regions (including exon, intron and flanking regions) and no-SV genes. **e**, Expression of SV genes with separated expression values for the absence and presence genotype groups. The *x* axis showed two groups, of which 38,302 and 22,210 genes are associated with suppression and promotion affected by SVs, respectively. The *y* axis showed normalized (*z* score) expression values. Green and yellow lines link the average expression values from each gene for their presence and absence of genotype groups. **f**, Proportion comparisons between SV genes and no-SV genes with different expression fold change after diverse stress treatments. FC, fold change. **g**, Comparison of expression fold change of SV genes and no-SV genes. In **c**, **d** and **g**, box plots show the 25th (lower edges) to 75th (upper edges) percentiles with median lines, and whiskers extending to 1.5× the interquartile range. **h**, Overall performance of SV-eGenes with differential expression under the four stress treatments is shown. The number of differentially expressed genes (DEGs) across four stress treatments (left). SV-eGenes were defined as genes with expression levels significantly affected by the presence of SVs under four stress treatments (Wilcoxon signed rank test, $P < 0.05$) (right). Numbers in parentheses are the number of differentially expressed SV-eGenes under the four stress treatments. **i**, Expression and structure of several previously reported genes associated with the four stress treatments are shown. Triangles marked with different colors represent insertions and deletions. CDS, coding sequence. CK, control group (no stress treatment).

https://doi.org/10.1038/s41588-025-02294-z

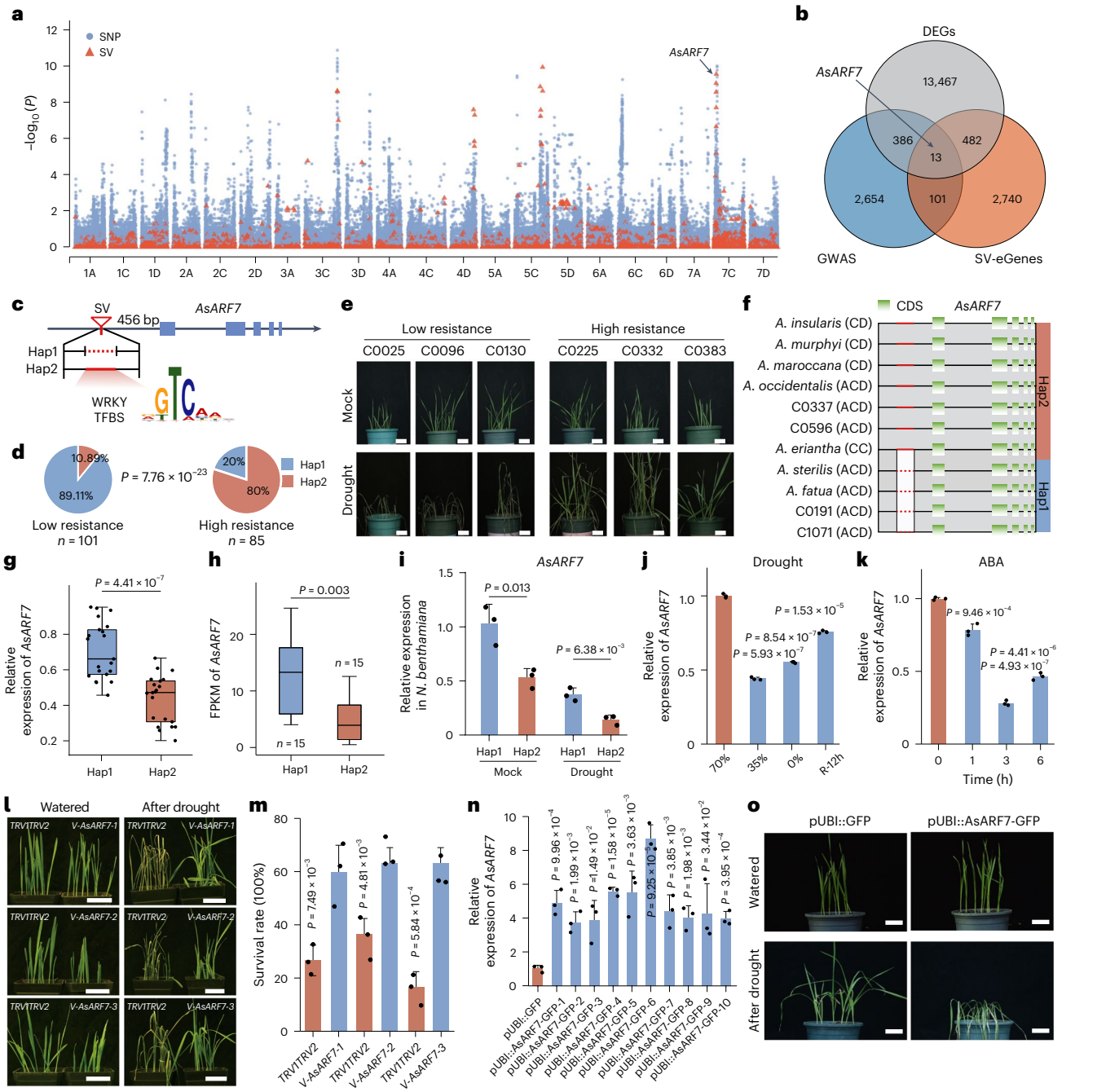


Fig. 6 | Identification and validation of key SVs and genes related to oat drought tolerance based on the SV-based GWAS. a, GWAS signals associated with drought tolerance including SNP (blue circles) and SV (red triangles) markers. Candidate genes are highlighted with black arrows. **b**, Venn diagram showing the numbers of common and unique genes identified by differential expression, SV-eGene analysis and SV-based GWAS. **c**, A significant peak associated with drought tolerance on chromosome 7C, the leading SV that harbored a WRKY binding site, was located 456 bp upstream of the translation start site of *AsARF7*. Hap, haplotype; TFBS, transcription factor binding site. **d**, Percent of accessions with present (haplotype 2) or absent (haplotype 1) SV genotype for high- and low-resistance accessions (two-tailed Fisher's exact test). **e**, Phenotype validation of oat accessions with two different haplotypes under normal and drought conditions. Scale bars, 5 cm. **f**, Schematic diagram of the candidate SV affecting the *AsARF7* gene in cultivated and wild oats. **g**,**h**, RT–qPCR and transcriptome analysis of *AsARF7* expression levels of the two haplotypes under drought treatment. RT–qPCR was performed in each of the seven accessions with haplotypes 1 and 2, respectively, and three biological replicates were made (**g**). Transcriptome analysis was performed in accessions with haplotypes 1 and 2 (**h**). In **g** and **h**, box plots show the 25th (lower edges) to 75th (upper edges) percentiles with median lines, and whiskers extending to 1.5× the interquartile range. **i**, RT–qPCR detection of *AsARF7* expression levels of the two haplotypes under drought treatment and mock treatment in transgenic tobacco plants. **j**, RT–qPCR detection of *AsARF7* expression levels before and after drought treatments. 70% represents 70% soil moisture content. 35% represents 35% soil moisture content. 0% represents 0% soil moisture content. R-12h, recovered for 12 h. **k**, RT–qPCR detection of AsARF7 expression levels before and after ABA treatments. The *x* axis represents the time course of ABA treatment. **l**, Drought tolerance assessment of *AsARF7*-knockdown lines. Photographs were taken under well-watered conditions and after a 2-d period of recovery with full irrigation after drought treatment. Whole-plant images of 5-week-old plants (**l**). Scale bars, 5 cm. **m**, Survival rate statistics of *AsARF7*-knockdown lines. **n**, RT–qPCR detection of *AsARF7* expression levels in pUBI::AsARF7-GFP transgenic plants. **o**, Overexpression of *AsARF7* decreased plant tolerance to drought. 'Watered' represents well-watered conditions. 'After drought' represents a 2-d period of recovery with full irrigation after drought treatment. Scale bars, 5 cm. In **i**–**k**, **m** and **n**, error bar represents the standard deviation of three biological replicates.

Furthermore, we conducted functional validation of *AsARF7* in drought resistance. RT–qPCR results showed that the expression level of *AsARF7* was significantly downregulated after drought and abscisic acid (ABA) treatments (Fig. 6j,k and Extended Data Fig. 10b,c). Using the tobacco rattle virus (TRV)-mediated gene-silencing system, we successfully generated oat lines with over 50% reduced *AsARF7* expression (Extended Data Fig. 10d). Knockdown oat lines showed stronger drought tolerance, a higher survival rate, a slower water loss rate and higher chlorophyll content under drought stress (Fig. 6l,m and Extended Data Fig. 10e–g). Additionally, we constructed pUBI::AsARF7-GFP overexpression transgenic oat plants (Extended Data Fig. 10h), with pUBI::GFP as a negative control (Fig. 6n). After drought treatment, pUBI::AsARF7-GFP plants exhibited weaker drought resistance than controls (Fig. 6o). Altogether, these results indicate that the SV near *AsARF7* contributes to oat drought tolerance by affecting its expression, which further emphasizes the important effects of SVs on gene expression and the necessity of SV-based GWAS for identifying trait-associated genetic variants in oat.

Discussion

Common oat is a globally important food and feed crop[4]. However, the complex and large genome size makes its genomic research and breeding lag far behind other economically important crops[3,4]. Aside from this, the serious shortage of genomes in *Avena* greatly limited the comprehensive studies of genetic diversity and evolutionary history and obstructed the utilization of excellent alleles in wild species. While pangenomes for numerous crops have been reported, most were focused on species with simple genomes or involving only one or a few species[2,15]. In this study, we constructed a genus-level super-pangenome including 35 oat accessions from 23 species of *Avena*. Compared to previous studies, our pangenome incorporated 21 wild oat species with different ploidy levels, rendering it the most representative germplasm resource for a pangenome to date and capturing nearly complete genetic resources of *Avena*.

Super-pangenome construction enables in-depth phylogenomic analysis, thereby clarifying the evolutionary position of all subgenomes and lineages and revealing reticulate evolutionary trajectories of *Avena* species. For instance, the evolutionary status of the B subgenome has been a subject of ongoing controversy in previous studies[10,23,29,46]. Our results showed that the B lineage diverged earliest from the common ancestor of the A, B and D lineages, resolving previous debates regarding evolutionary trajectories. Moreover, the fully resolved evolutionary position of *Avena* contributed to correcting misclassifications of some species. For example, our results demonstrated the special karyotype and evolutionary history of *A. agadiriana* (Ac'Ac'As'As') and supported that *A. agadiriana* was not an AABB species[8]. In addition, our findings revealed that wild oats had 26.63% new gene families and 59.93% new gene haplotypes compared to cultivated oats, further corroborating the importance of the rich genetic resources in wild species and facilitating their application in oat breeding.

SVs frequently introduce genetic diversity, which can profoundly affect gene expression and function, leading to phenotypic changes[2,47]. These high-quality genomes offered an opportunity to construct a comprehensive SV catalog for understanding genomic and phenotypic changes in oat. We highlighted the crucial role of SVs in modulating gene expression and shaping adaptation to diverse stresses. The integrated analyses of SV-eGene identification, differential expression analysis and SV-based GWAS enabled us to mine key SVs and candidate genes associated with stress resistance. These discoveries provided insights into the important and complex regulatory roles of SVs on the variations of gene expression and extensive adaptability to different environmental conditions, thereby establishing a robust and reliable oat genotyping platform for facilitating fine mapping of key genes and genomics-assisted breeding.

Collectively, our super-pangenome generated a comprehensive dataset to explore and use the entire spectrum of genetic diversity present in *Avena*, which will greatly expedite genomics, evolution and molecular breeding research in oat.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-025-02294-z.

References

1. Rasane, P., Jha, A., Sabikhi, L., Kumar, A. & Unnikrishnan, V. S. Nutritional advantages of oats and opportunities for its processing as value added foods — a review. *J. Food Sci. Technol.* **52**, 662–675 (2015).
2. Shi, J., Tian, Z., Lai, J. & Huang, X. Plant pan-genomics and its applications. *Mol. Plant* **16**, 168–186 (2023).
3. Kamal, N. et al. The mosaic oat genome gives insights into a uniquely healthy cereal crop. *Nature* **606**, 113–119 (2022).
4. Peng, Y. et al. Reference genome assemblies reveal the origin and evolution of allohexaploid oat. *Nat. Genet.* **54**, 1248–1258 (2022).
5. Liu, Q., Lin, L., Zhou, X., Peterson, P. M. & Wen, J. Unraveling the evolutionary dynamics of ancient and recent polyploidization events in *Avena* (Poaceae). *Sci. Rep.* **7**, 41944 (2017).
6. Yan, H. et al. High-density marker profiling confirms ancestral genomes of *Avena* species and identifies D-genome chromosomes of hexaploid oat. *Theor. Appl. Genet.* **129**, 2133–2149 (2016).
7. Loskutov, I. G., Gnutikov, A. A., Blinova, E. V. & Rodionov, A. V. The application of Vavilov's approaches to the phylogeny and evolution of cultivated species of the genus *Avena* L. *Vavilovskii Zhurnal Genet. Selektsii* **27**, 921–932 (2023).
8. Fu, Y. B. Oat evolution revealed in the maternal lineages of 25 *Avena* species. *Sci. Rep.* **8**, 4252 (2018).
9. Liu, Q. et al. Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. *BMC Plant Biol.* **20**, 406 (2020).
10. Morikawa, T. & Nishihara, M. Genomic and polyploid evolution in genus *Avena* as revealed by RFLPs of repeated DNA sequences. *Genes Genet. Syst.* **84**, 199–208 (2009).
11. Yan, H. H. et al. Phylogenetic analysis of the genus *Avena* based on chloroplast intergenic spacer *psb*A-*trn*H and single-copy nuclear gene *Acc1*. *Genome* **57**, 267–277 (2014).
12. Liu, B. B. et al. Phylogenomic conflict analyses in the apple genus *Malus* s.l. reveal widespread hybridization and allopolyploidy driving diversification, with insights into the complex biogeographic history in the Northern Hemisphere. *J. Integr. Plant Biol.* **64**, 1020–1043 (2022).
13. Meleshko, O. et al. Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus. *Mol. Biol. Evol.* **38**, 2750–2766 (2021).
14. Liu, S. Y. et al. An integrative framework reveals widespread gene flow during the early radiation of oaks and relatives in Quercoideae (Fagaceae). *J. Integr. Plant Biol.* **67**, 1119–1141 (2025).
15. Li, W. et al. Plant pan-genomics: recent advances, new challenges, and roads ahead. *J. Genet. Genomics* **49**, 833–846 (2022).
16. Tieman, D. et al. A chemical genetic roadmap to improved tomato flavor. *Science* **355**, 391–394 (2017).
17. Zhuang, Y. et al. Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nat. Plants* **8**, 233–244 (2022).

18. Yu, H. et al. A route to de novo domestication of wild allotetraploid rice. *Cell* **184**, 1156–1170 (2021).
19. Li, N. et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet.* **55**, 852–860 (2023).
20. Tang, D. et al. Genome evolution and diversity of wild and cultivated potatoes. *Nature* **606**, 535–541 (2022).
21. Shi, T. et al. The super-pangenome of *Populus* unveils genomic facets for its adaptation and diversification in widespread forest trees. *Mol. Plant* **17**, 725–746 (2024).
22. Fominaya, A., Vega, C. & Ferrer, E. Giemsa C-banded karyotypes of *Avena* species. *Genome* **30**, 627–632 (1988).
23. Latta, R. G., Bekele, W. A., Wight, C. P. & Tinker, N. A. Comparative linkage mapping of diploid, tetraploid, and hexaploid *Avena* species suggests extensive chromosome rearrangement in ancestral diploids. *Sci. Rep.* **9**, 12298 (2019).
24. Yan, H. et al. Genome size variation in the genus *Avena*. *Genome* **59**, 209–220 (2016).
25. Xiong, Z., Gaeta, R. T. & Pires, J. C. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl Acad. Sci. USA* **108**, 7908–7913 (2011).
26. Chester, M. et al. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl Acad. Sci. USA* **109**, 1176–1181 (2012).
27. Peng, Y. et al. Phylogenetic relationships in the genus *Avena* based on the nuclear *Pgk1* gene. *PLoS ONE* **13**, e0200047 (2018).
28. Morikawa, T., Gushiken, Y. & Tsurukawa, N. Chromosomal diversity and morphological dimorphism in Moroccan wild oat, *Avena agadiriana*. *Plant Syst. Evol.* **281**, 107–113 (2009).
29. Paczos-Grzęda, E. & Bednarek, P. Comparative analysis of hexaploid *Avena* species using REMAP and ISSR methods. *Turk. J. Bot.* **38**, 1103–1111 (2014).
30. Loskutov, I. G. On evolutionary pathways of *Avena* species. *Genet. Resour. Crop Evol.* **55**, 211–220 (2008).
31. Badaeva, E. D., Shelukhina, O. Y., Dedkova, O. S., Loskutov, I. G. & Pukhalskyi, V. A. Comparative cytogenetic analysis of hexaploid *Avena* L. species. *Russ. J. Genet.* **47**, 691–702 (2011).
32. Ociepa, T. The oat gene pools—review about the use of wild species in improving cultivated oat/Pule genowe owsa—przeglad informacji na temat wykorzystania dzikich gatunkow w ulepszaniu owsa uprawnego. *J. Cent. Eur. Agric.* **20**, 251–261 (2019).
33. Ladizinsky, G. & Fainstein, R. Domestication of the protein-rich tetraploid wild oats *Avena magna* and *A. murphyi*. *Euphytica* **26**, 221–223 (1977).
34. Ladizinsky, G. Domestication via hybridization of the wild tetraploid oats *Avena magna* and *A. murphyi*. *Theor. Appl. Genet.* **91**, 639–646 (1995).
35. Okoń, S., Ociepa, T., Nucia, A., Cieplak, M. & Kowalczyk, K. Is every wild species a rich source of disease resistance? *Avena fatua* L.—potential donor of resistance to powdery mildew. *Plants* **10**, 560 (2021).
36. Zhai, C. et al. The isolation and characterization of *Pik*, a rice blast resistance gene which emerged after rice domestication. *New Phytol.* **189**, 321–334 (2011).
37. Ramírez-González, R. H. et al. The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018).
38. Ma, P.-F. et al. Genome assemblies of 11 bamboo species highlight diversification induced by dynamic subgenome dominance. *Nat. Genet.* **56**, 710–720 (2024).
39. Qin, P. et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**, 3542–3558 (2021).
40. Yan, H. et al. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nat. Genet.* **55**, 507–518 (2023).
41. Uga, Y. et al. Control of root system architecture by *DEEPER ROOTING 1* increases rice yield under drought conditions. *Nat. Genet.* **45**, 1097–1102 (2013).
42. Lo, S. F. et al. Ectopic expression of specific GA2 oxidase mutants promotes yield and stress tolerance in rice. *Plant Biotechnol. J.* **15**, 850–864 (2017).
43. Wei, H., Wang, X., He, Y., Xu, H. & Wang, L. Clock component OsPRR73 positively regulates rice salt tolerance by modulating *OsHKT2;1*-mediated sodium homeostasis. *EMBO J.* **40**, e105086 (2021).
44. Kim, J.-J. et al. Overexpression of a proton pumping gene *OVP1* enhances salt stress tolerance, root growth and biomass yield by regulating ion balance in rice (*Oryza sativa* L.). *Environ. Exp. Bot.* **175**, 104033 (2020).
45. An, X., Zhang, L., Luo, J., Zhao, S. & Jiao, T. Effects of oat hay content in diets on nutrient metabolism and the rumen microflora in sheep. *Animals* **10**, 2341 (2020).
46. Katsiotis, A., Hagidimitriou, M. & Heslop-Harrison, J. S. The close relationship between the A and B genomes in *Avena* L. (Poaceae) determined by molecular cytogenetic analysis of total genomic, tandemly and dispersed repetitive DNA sequences. *Ann. Bot.* **79**, 103–109 (1997).
47. Zhang, C. et al. High-quality genome of a modern soybean cultivar and resequencing of 547 accessions provide insights into the role of structural variation. *Nat. Genet.* **56**, 2247–2258 (2024).
48. Li, W. et al. Genomic insights into the divergence between hulled and hulless oats. *Cell Rep.* **44**, 116055 (2025).
49. He, Q. et al. The near-complete genome assembly of hexaploid wild oat reveals its genome evolution and divergence with cultivated oats. *Nat. Plants* **10**, 2062–2078 (2024).
50. Li, Y. et al. Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat. Commun.* **12**, 2563 (2021).
51. Maughan, P. J. et al. Genomic insights from the first chromosome-scale assemblies of oat (*Avena* spp.) diploid species. *BMC Biol.* **17**, 92 (2019).

Methods

SNP calling and population analysis

Whole-genome resequencing data of 1,078 hexaploid oat accessions obtained from the OatOmics database (http://www.oatomics.com) and whole-genome resequencing data of *A. sativa* cv. Sang obtained from a previous study[3] were aligned to the reference genome using BWA (version 0.7.17)[52] with the MEM algorithm. The alignment results were filtered and sorted using SAMtools (version 1.7)[53]. BCFtools (version 1.10.2) was used to perform genomic variant detection. The SNPs were further filtered with the following parameters: 'QD < 2.0||FS > 60.0||MQ < 40.0||SOR > 3.0||MQRankSum < -12.5||ReadPosRankSum < -8.0'. We generated a distance matrix of nonsynonymous SNPs using VCF2Dis (version 1.46) and constructed an NJ phylogenetic tree based on the FastME server (http://www.atgc-montpellier.fr/fastme). Multiple iterations were performed using a replacement sampling method through VCF2Dis. A subset of SNPs was randomly selected, and a new NJ tree was constructed as described above, repeating this process 1,000 times with the parameters -Rand 0.25. Finally, 1,000 bootstrap trees were integrated into the above NJ tree using nw_support in the Newick Utilities Tutorial (version 1.6.0) (https://www.ezlab.org/). The results were visualized with the iTOL (https://itol.embl.de).

PCA was performed with these SNPs using PLINK (version 1.90b6.26)[54]. ADMIXTURE (version 1.3.0)[55] was used to conduct a population structure analysis for K ranging from 2 to 9. Here, $K = 6$ was subsequently chosen for its smaller cross-validation error, combining with the results of the phylogenetic tree, PCA and geographical distributions. The world map was generated using WorldClim (https://worldclim.org/) where data are freely available for academic use (https://worldclim.org/about.html).

Plant materials and growth conditions

Seventeen hexaploid oat accessions were selected for de novo assembly, consisting of two cultivated species (*A. sativa*, *A. sativa* ssp. *nuda*) and three wild species (AACCDD: *A. fatua*, *A. occidentalis* and *A. sterilis*). Furthermore, 18 wild species including 11 diploids (AA: *A. hispanica*, *A. brevis*, *A. hirtula*, *A. strigosa*, *A. canariensis*, *Avena damascene*, *A. atlantica*, *A. longiglumis*, *A. nuda*; CC: *A. clauda*, *A. eriantha*) and seven tetraploids (AABB: *A. vaviloviana*, *A. abyssinica*, *A. agadiriana*, *A. barbata*; CCDD: *A. insularis*, *A. maroccana*, *A. murphyi*) were collected from the National Crop Genebank of China. These oat accessions were grown in the greenhouse at 25 °C/18 °C under a 12-h light/12-h dark photoperiod. *N. benthamiana* was grown in at 22 °C under a 12-h light–12-h dark photoperiod.

Genome sequencing and assembly

Among the 35 oat accessions, 28 accessions were sequenced and de novo assembled in this study. The other seven genomes were reported in other studies[4,49–51], including C1071 (ref. 48). Fresh young leaves from the tillering stage were collected and sequenced on the PacBio Revio platform and the DNBSEQ-T7 platform (MGI Tech). For Hi-C sequencing, fresh leaves were fixed in formaldehyde, and DNA was digested with the restriction enzyme MboI. These digested DNA fragments were sequenced on the Illumina NovaSeq 6000 platform at Annoroad Gene Technology.

For the 27 newly sequenced oat accessions, PacBio HiFi long reads were used for assembly using hifiasm (version 0.19.7-r598)[56] with parameters --hg-size 11g. Subsequently, the Hi-C data were aligned to the contigs by Juicer (version 1.5)[57] and then clustered into chromosomes by 3D-DNA. Juicebox (version 1.11.08)[58] was used for manual inspection and adjustment. The reference genome C0191 was assembled using the above-mentioned strategy integrating ~263.5 Gb of ONT ultralong reads. PacBio HiFi long reads, Hi-C and Illumina data of C0191 were obtained from another study[49]. The completeness of genome assembly was evaluated using the embryophyta_odb10 database of BUSCO (version 5.2.2)[59], and the LAI value was calculated with LTR_retriever (version 1.9)[60].

Genome annotation

RepeatModeler (version 1.0.11)[61], LTR_FINDER (version 1.05)[62], LTRharvest (version 1.5.11)[63] and LTR_retriever (version 2.9.0)[60] were used to construct de novo repeat libraries. RepeatMasker (version 4.1.1)[64] was used to identify genome-wide repeat sequences and repeat elements with the library combined by the de novo library and a known repeat library (Repbase 15.02). TEsorter (version 1.4.6)[65] was used to classify full-length LTRs, including class I and class II elements covered by the REXdb database.

A combination of ab initio prediction, homology search and transcriptional evidence was used to predict coding gene models. Using transcripts assembled by Trinity (version 2.12)[66], gene structure was predicted with PASA (version 2.3)[67], and then the gene model was trained in AUGUSTUS (version 3.2.3)[68]. GenomeThreader (version 1.7.3)[69] was used to search against protein sequences of closely related species. RNA-seq reads were mapped to the assembled genome using HISAT2 (version 2.2.1)[70] with default parameters, followed by genome-guided transcript assembly with StringTie (version 1.13)[71]. TransDecoder (version 5.1.0) (https://github.com/TransDecoder/TransDecoder) was used to predict open reading frames using the assembled transcripts. Finally, EVidenceModeler[72] was used to integrate all predicted gene structures into a nonredundant gene set. The completeness of the final gene set was evaluated using BUSCO with the embryophyta_odb10 dataset. InterProScan (version 5.48-83.0) was used to functionally annotate the gene models of the resulting gene set. KofamScan (version 1.3.0)[73] with default parameters was used to assign each gene KEGG orthology terms.

Subgenome identification of polyploids

To identify the subgenomes of polyploids, subgenome-specific *k*-mers were identified using SubPhaser (version 1.2.6)[74]. Next, we split the genomic sequence of the A, C and D subgenomes from all polyploids into 200-kb non-overlapping fragments and aligned them to three subgenomes of C0191. Only the reads with the best match were used to calculate genomic identity and coverage. The syntenic blocks between published genomes and our polyploids were generated by jcvi (version 1.2.1).

Phylogenetic tree construction and divergence time estimation

A total of 2,456 single-copy genes in 44 subgenomes from 25 species (*T. aestivum* and *L. perenne* as the outgroup with data downloaded from the NCBI) were identified using OrthoFinder (version 2.3.8)[75] with default parameters. These genes were passed to MUSCLE (version 3.8.1551)[76] to create the concatenated amino acid sequence alignment. Maximum likelihood inference of phylogenetic relationships was performed using RAxML (version 8.2.12)[77] with the PROTGAMMAIJTTF model. One hundred single-copy genes were randomly selected to align and then used to build a phylogenetic tree (repeated 1,000 times) to generate the consensus tree. The distribution of gene trees was visualized with DensiTree (version 3.0.2) (https://www.cs.auckland.ac.nz/~remco/DensiTree/).

Divergence times were estimated using MCMCtree (version 4.9) in the PAML[78] package with parameters 'model = 4, burnin = 2,000, sampfreq = 10, nsample = 20,000'. Polyploidization and divergence time was also estimated with the formula $T = K_s/2\mu$ ($\mu = 6.5 \times 10^{-9}$)[79], in which T is the divergence time. Finally, CAFE (version 4.2.1)[80] was used to identify expansion and contraction gene families. Karyotypes were constructed through the –km and –k subroutines of WGDI (version 0.6.3)[81] based on protein sequences of rice. Rice data (IRGSP-1.0) were downloaded from Ensembl Plants. According to previous studies[38,82,83], the timescales of polyploidization were bracketed from the species

divergence as the lower limit to the divergence time of parental lineages as the upper limit.

Interchromosomal exchanges in polyploid oat were identified with SubPhaser.

Chloroplast genome assembly and phylogenetic analysis

Based on the resequencing data of 22 oat accessions in this study, along with four other sets of published resequencing data (*Avena wiestii*, *Avena lusitanica*, *T. aestivum*, *L. perenne*), the 26 chloroplast genomes were assembled using the GetOrganelle (version 1.7.7.0)[84] toolkit with the parameters '-R 21 -k 21,45,65,85,105,117 -F embplant_pt -t 5 -R 25 -w 68'. Multiple-sequence alignments were performed using MUSCLE, and ML analysis was performed using RAxML with 100 bootstrap replicates under the GTRGAMMAX evolutionary model. Illumina data for *T. aestivum* and *L. perenne* were downloaded from the NCBI.

Mitochondrial phylogenetic analysis

First, resequencing data of 26 species were aligned to the mitochondrial genome of *A. longiglumis* downloaded from the NCBI using BWA and then filtered with SAMtools[85] (view -h -b -q30). SNP calling was carried out using BCFtools (version 1.12)[86] -mpileup with default parameters. Next, VCF2Dis (version 1.47) (https://github.com/BGI-shenzhen/VCF2Dis) was used to generate the distance matrix and passed to mat-2nwk (http://www.atgc-montpellier.fr/fastme/).

Karyotypes of *A. agadiriana*

Chromosome phylogenetic trees were constructed based on single-copy genes for two subgenomes of *A. agadiriana* and ten other subgenomes in A and B lineages. We split the sequences of the A and B subgenomes from all diploids and tetraploids (except *A. wiestii*) into 200-bp non-overlapping fragments and aligned them to both subgenomes of *A. vaviloviana*, *A. abyssinica* and *A. agadiriana*. Illumina reads for *A. wiestii* were downloaded from the NCBI. Only the reads with the best match were used to calculate genomic identity and coverage. Next, we calculated K_s using KaKs_Calculator (version 2.0)[87] between two subgenomes.

Super-pangenome construction and analysis

Orthologous gene families in *Avena* were clustered using OrthoFinder with default parameters, referring to recently published genus-level pangenome studies[21,88,89]. Functional enrichment analysis of GO terms and KEGG categories was performed using the 'clusterProfiler' R (version 4.4.2) package with $P \leq 0.05$. K_a/K_s values were calculated using KaKs_Calculator based on multiple-sequence alignments performed with ParaAT (version 2.0)[90] with syntenic gene pairs identified using McscanX (https://github.com/wyp1125/MCScanX).

Identification of SSP, NBS-encoding and cellulose synthase gene families

HMMER (version 4.0.0)[91] with *an E* value of 1×10^{-5} was used to identify candidate cellulose synthase genes with the conserved CesA (PF03552.9) and Csl (PF13641.1) protein domains. The homology sequences of rice and *Arabidopsis* were downloaded from the NCBI, and these candidate genes in oats were aligned with MAFFT to assign them into the seven subfamilies CslA, CslC, CslD, CslE, CslF, CslH and CslJ. The best model of the phylogenetic relationships was estimated by IQ-TREE. For SSP, the conserved protein domains of HMW glutenin (PF03157), gliadin (PF13016), Tryp-α-amyl domains (PF00234) and the cupin_1 domain (PF00190) were used to identify candidate genes in oats with HMMER using *an E* value of 1×10^{-5}. Lastly, candidate proteins of NBS-encoding families were retrieved using RGAugury[92] with default parameters.

Statistics of genotypes

All-to-all alignments were performed with diamond (version 0.9.14) between the proteins of C0191 and proteins from other species. The best matches in other species using each of the proteins from C0191 as a reference were screened to form homozygous gene groups, and we counted the number of specific and shared haplotypes within these homozygous groups in both wild and cultivated oats.

Transcriptome sequencing and analyses

Among the 1,401 transcriptome samples, 1,365 transcriptome samples including three biologically independent experiments were collected and sequenced in this study. The other 36 transcriptome samples were obtained from other studies[48,49].

After total RNA extraction, the RNA libraries were constructed according to the manufacturer's instructions, and samples from these accessions were sequenced using the Illumina platform. fastp (version 0.20.1)[93] with '-e 20 -w 10' was used to trimmed raw RNA-seq reads. For each tissue, the clean reads were aligned to the reference genome using HISAT2 with default settings. Gene expression levels were calculated using StringTie, measured as FPKM values. The DESeq2 package was used for identifying differentially expressed genes (P value < 0.05 and $|\log_2(\text{fold change})| > 1$).

Principal-component analysis of the 1,401 RNA-seq samples

For PCA, FPKM values of expressed genes were transformed using $\log_2(\text{FPKM} + 1)$. A distance matrix was then constructed with the dist function, and, based on the matrix, the neighbor-joining tree was clustered for different oat tissues and treatments using the hclust function. PCA analysis was performed with the prcomp function in R (version 4.1.1) (https://www.r-project.org/).

Expression divergence among subgenomes

Methods for subgenome expression divergence in this study refer to previously published studies in wheats[37], oats[3,4] and bamboos[38]. First, OrthoFinder was employed to identify 1:1:1 homoeologous genes across the three subgenomes from 17 hexaploid oat accessions. Additionally, we defined a triad as expressed when the sum of the A, C and D subgenome homoeologs had FPKM > 0.5 and standardized the relative expression of each homoeolog across the triad. The final dominant subgenome was determined based on the standardized relative expression of the A, C and D subgenomes using a Wilcoxon rank-sum test across all samples in each accession. A total of 1,143 samples were used to evaluate subgenome dominance in hexaploid oats. We also evaluated the subgenome dominance status in tetraploid CCDD species using the same approach based on 132 transcriptome samples.

Structural variation detection

We first aligned query genomes to the C0191 reference genome using MUMmer[94]. Raw alignment results were further picked using delta-filter with the parameter '-m -i 90 -l 100'. SyRI (version 1.6.3)[95] with default parameters was used to identify SVs. In this study, we focus on SVs ranging from 50 bp to 1 Mb in size. Two SVs with overlapping genomic coordinates exceeding 90% were merged.

To obtain presence–absence variations (PAVs) in oat populations for GWAS, the coverage of nonredundant PAVs in the genome was used to identify PAVs of each accession. PAVs with coverage less than 70% were marked as 'missing'.

Identification of structural variation hotspot regions

The number of SVs within 1-Mb windows (step size of 500 kb) for each chromosome was calculated. Subsequently, based on the count of SVs within these windows, the top 5% of windows with the highest SV frequency were defined as SV hotspots. Following this, contiguous windows identified as hotspots were merged to form 'hotspot regions'.

SV-eGene identification

Based on the recent study of *Brassica oleracea*[96], we identified 'SV genes' as genes containing SVs within gene bodies or the regions 3 kb

upstream or downstream. The effects of SVs on gene expression were classified into two types: promoting and suppressing gene expression. We defined genotypes with SVs as 'presence' and genotypes sharing the same sequence with the reference genome as 'absence'. Among these, SV genes that had different SV genotypes in at least three genomes were selected for gene expression analysis. Promoting SVs were defined if the mean FPKM of a gene with SV presence was at least 1.5-fold higher than that with SV absence.

Genome-wide association study analysis

Based on the drought tolerance phenotypes of 186 oat accessions, SNP-based and SV-based GWASs were performed using SNPs or SVs with minor allele frequency < 0.05 and missing call rate < 0.1 with GEMMA (version 0.98.5)[97] software. Genome-wide significance thresholds (1.15×10^{-6}) were determined using a uniform threshold of $1/N$ (N refers to the effective number of SNPs and SVs calculated by the Genetic Type I error calculator). To examine whether presence of the SV was correlated with drought tolerance, we compared the differences in the proportions of the two haplotypes in high (survival rate > 50%) and low (survival rate ≤ 50%) drought-resistant accessions.

Plasmids and cloning procedures

For the VIGS assay, different-length open reading frame fragment regions of *AsARF7* were amplified from C0836 cDNA. These fragments were fused into the pTRV2 virus vector with the Nimble Cloning kit (NC Biotech, NC001) to generate pTRV2::AsARF7 constructs. A 3,000-bp genomic DNA fragment containing the *AsARF7* promoter was amplified from C0836 genomic DNA and cloned into pCambia 1305.2, generating the construct pHAP1::AsARF7. We conducted base insertion using $pAsARF7_{SV1}$::AsARF7 (pHAP1::AsARF7) as a template to generate $pAsARF7_{SV2}$::AsARF7 (pHAP2::AsARF7) via PCR. To generate the pUBI::AsARF7-GFP construct, a 1,982-bp genomic DNA fragment containing the *AsARF7* gene was amplified from C0836 genomic DNA and cloned into the pCambia 3300 vector by using the ClonExpress II One Step Cloning kit (Vazyme, C112-02).

Virus-induced gene silencing mediated by tobacco rattle virus

TRV infection experiments were performed as previously described[98,99]. To initiate the experiment, the fragments of *AsARF7* used for VIGS were assembled into the pTRV2 virus vector. *Agrobacterium tumefaciens* strain GV3101 with a TRV infectious clone was cultured at 28 °C. After centrifugation, pTRV1 and pTRV2 were supplemented with ACT buffer (acetosyringone (Coolaber, SL95131, 19.62 mg l^{-1}), cysteine (Amethyst, 52-90-4, 400 mg l^{-1}) and Tween-20 (Sigma, P2287, 5 ml l^{-1})) and mixed at a 1:1 ratio.

Plant material preparation for PCR validation

Drought, simulated drought with polyethylene glycol 6000 (PEG 6000; Sangon Biotech, A630434) and ABA hormone (Harvey, PGS1349) treatment were integrated to verify whether the candidate genes respond to drought stress at the expression level. Sterilized seeds were carefully placed on 1/2 MS medium supplemented with 1% (wt/vol) agar and adjusted to pH 5.8. Once germinated, the seeds were transferred to pots filled with a 1/2 Hoagland solution. When the seedlings reached the trifoliate stage, they were subjected to different treatments: drought, ABA (spraying and soaking) and 20% PEG 6000 solution. Plant aerial parts were collected at specific time intervals (70% soil moisture content, 35% soil moisture content, 0% soil moisture content and recovered for 12 h) after the initiation of drought treatment. Treatments were performed using solutions containing 20% PEG 6000 for drought simulation. Aerial plant parts and roots were collected at 12 h after the initiation of PEG 6000 treatment. Aerial plant parts and roots were collected at specific time intervals (0, 1, 3, 6 and 9 h) after the initiation of treatment with ABA at 250 µmol l^{-1} (spraying) or 125 µmol l^{-1} (soaking).

Reverse transcription PCR analysis

Total RNA samples from oat organs were prepared using TRIzol Reagent (Thermo Fisher Scientific, 15596018). To synthesize first-strand cDNA, reverse transcription PCR was performed on 1 µg of total RNA using the PrimeScript RT Reagent Kit with gDNA Eraser for qPCR (Takara, RR047A). RT–qPCR assays were performed using the SYBR Green Premix Pro Taq HS qPCR Kit (Takara, RR820A). *AsACTIN2* was used as an internal control. All primers used in the qPCR assays were synthesized by Sangon Biotech and are listed in Supplementary Table 25.

Water loss rate assay

The leaves were taken from six plants in the same state, which were cut off and weighed immediately. The leaves were placed on filter paper at room temperature for 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 h and were weighed at each time point. Three biological replicates were performed for each line, and the water loss rate was calculated.

Chlorophyll measurement

The chlorophyll measurement assay was performed as previously described[100]. The leaves were incubated in 95% (vol/vol) ethanol for 5 d in the dark. The absorbance was measured at 665 and 649 nm. Chlorophyll contents were calculated according to the following ratio: $(6.63A_{665} + 18.08A_{649})$/g fresh weight.

Nuclear run-on assay

The nuclear run-on assay was performed as previously described[101–103]. Three grams of one-tip three-leaf oat leaves were collected to determine transcription levels after drought treatment. The run-on reaction was performed at 37 °C for 40 min. Anti-BrdU beads (80 µl, Santa Cruz) were used to purify RNA at 4 °C for 2 h. The newly transcribed RNA was extracted with TRIzol Reagent and used for RT–qPCR analysis.

Transient expression in *N. benthamiana*

To verify the effect of SVs on the expression level of *AsARF7*, the transient expression assay was performed in *N. benthamiana*[104]. *A. tumefaciens* GV3101 containing pHAP1::AsARF7-GFP and pHAP2::AsARF7-GFP vectors was cultivated overnight at 28 °C and then suspended in MMG buffer (10 mM $MgCl_2$ (Sigma, M9272), 10 mM 2-(*N*-morpholino)ethanesulfonic acid (MES; Sigma, M8250, pH 5.7) and 150 µM acetosyringone) at an optical density at 600 nm of 1.2. Four-week-old *N. benthamiana* was injected, and further studies were performed at 3 dpi. The leaves were placed on filter paper at room temperature for 3 h and then were collected as materials after drought treatment.

Genetic transformation in *A. sativa*

A. sativa transformation was performed following previously published protocols[105–108]. Mature embryos from healthy plants of 'cv. Bayou No. 18' were collected and cultured on L3-M medium (4.6 g l^{-1} L3 Base Salts with vitamins (Coolaber, PM1621), 30 g l^{-1} maltose (Coolaber, CM7181), 4 g l^{-1} Phytagel (Coolaber, CP8581Z), 2 mg l^{-1} 2,4-D (Coolaber, PH105), 1 mg l^{-1} Dicamba (Coolaber, PH113)) until an embryonic callus was produced. *A. tumefaciens* strain GV3101 containing the pUBI::AsARF7-GFP construct was cultured overnight at 28 °C in YEP medium. After centrifugation at 25 °C and 3,438*g* for 10 min, the precipitate was resuspended in WLS solution (4.30209 g Linsmaier & Skoog Base Salts (Coolaber, PM1470), 100 µl 1,000× MS vitamins (Coolaber, PML1780), 10 g glucose (Sangon Biotech, A501991-0500), 0.5 g MES (Sigma, M8250), add water to 1 l, pH 5.8) to an optical density of 0.5. Equal proportions of *A. tumefaciens* strain GV3101 containing the constructs pUBI::AsARF7-GFP and pUBI::TaWOX5 were mixed. The selected embryonic callus was immersed in a mixed solution of *A. tumefaciens* strains for 30 min. Next, the embryogenic callus was removed and cultured on filter paper containing 75 µmol acetosyringone in the dark for 3 d. Subsequently, embryogenic calli were cultured on WLS-RES medium (4.6 g l^{-1} L3 Base Salts with vitamins (Coolaber, PM1621), 0.1 g l^{-1} picloram (Coolaber,

CP8601), 0.5 g l^{-1} L-glutamic acid (Coolaber, CG5771), 0.1 g l^{-1} casein acid hydrolysate (Coolaber, CC3191), 0.75 g l^{-1} $MgCl_2 \cdot 6H_2O$ (Sangon Biotech, A610328-0500), 40 g l^{-1} maltose (Coolaber, CM7181), 1.95 g l^{-1} MES (Sigma, M8250), 4 g l^{-1} Phytagel (Coolaber, CP8581Z), 25 µmol $AgNO_3$, 5 mg l^{-1} vitamin C, 0.2 mg ml^{-1} Timentin) for a duration of 5 d. WLS-P5 medium (WLS-RES medium containing 5 mg l^{-1} Basta (Sangon Biotech, A356357)) was used to screen the successfully surviving callus, which was then transferred to a regeneration medium (4.6 g L3 Base Salts with vitamins, 5 mg zeatin (Coolaber, PH110), 20 g sucrose (Diamond, A100335-0250), 0.5 g MES, 200 µl of 12.5 g l^{-1} $CuSO_4 \cdot 5H_2O$ (Sangon Biotech, A600063), 4 g Phytagel, add water to 1 l, pH 5.8) to promote growth until it reached a size of 3–5 cm. Finally, the roots were cultivated in rooting medium (4.6 g L3 Base Salts with vitamins, 0.2 mg ml^{-1} IBA (Coolaber, PH1031), 15 g sucrose, 0.5 g MES, 4 g Phytagel, add water to 1 l, pH 5.8). *AsARF7* expression was detected by RT–qPCR.

Fluorescent in situ hybridization

Primers were designed on the basis of the consensus sequences of the B subgenome and the As genome.

In situ hybridization assays were performed following the previous protocol[49,109,110]. Oat seeds were germinated and subjected to a 24-h treatment at 4 °C. The seeds were incubated for 12 h at 24 °C, followed by a high-temperature stimulation at 30 °C for 4 h. Root tips were excised and soaked in distilled water, followed by a 24-h pretreatment at 4 °C. Subsequently, the root tips were transferred to a solution of 0.075 mol l^{-1} KCl and incubated at 4 °C for 45 min. Next, the root tips were fixed in Carnoy's fixative for 24 h. After fixation, the root tips were subjected to hydrolysis using a mixed enzyme solution (1.5% Cellulase R-10, 0.4% Macerozyme R-10, 0.4 M mannitol, 20 mM KCl, 20 mM MES, pH 5.7, 10 mM $CaCl_2$, 0.1% BSA and 5 mM β-mercaptoethanol). Hydrolyzed samples were centrifuged and precipitated with 45% acetic acid. The resulting suspension droplets were heated on glass slides for preparation. Probes were labeled with biotin-dUTP; primers are listed in Supplementary Table 25. Hybridization was performed overnight at a temperature of 37 °C. Alexa Fluor 594-conjugated streptavidin (Invitrogen) was used for immunofluorescence. FISH experiments were repeated independently three times with similar results.

Statistics and reproducibility

To assess the significance of the correlations between (sub)genome size (Mb) and TE content (Mb), we calculated Pearson's correlation coefficient (r) with Student's t-test (two tailed). This was also used to evaluate the correlations of the distribution patterns between genes and different types of TEs. To test for differences in Figs. 3f,g, 4c and 5c,d,g, Extended Data Figs. 6e–h and 9c–f, Supplementary Fig. 16e–g, Extended Data Fig. 6b and Supplementary Fig. 2, we used a two-sided Wilcoxon rank-sum test (the wilcox.test function in R with continuity correction). To test for differences in Fig. 6g–n and Extended Data Figs. 9c–f and 10d–g, we used Student's t-test (two tailed). The significance in Extended Data Fig. 10c and Supplementary Fig. 3b was tested using one-way ANOVA with Turkey's multiple-comparison test ($\alpha = 0.05$).

Each experiment involved at least three biological replicates.

Reporting summary

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

Data availability

The raw data (CRA015654 and CRA015559) and genome assemblies (GWHEROY00000000-GWHGDJH00000000.1) in this study have been deposited at the Chinese National Genomics Data Center (https://bigd.big.ac.cn/) under the BioProject accession number PRJCA024628. Raw sequencing data (SRR31832837–SRR31832882) and genome assemblies (JBKZJG000000000–JBKZJZ000000000 and JBKZKA000000000–JBKZKH000000000) have been deposited at the NCBI under the BioProject PRJNA1200805. Transcriptome sequencing data under normal growth (SRR31798633–SRR31798686 and SRR31814753–SRR31815091) have been deposited at the NCBI under the BioProject PRJNA1201252. Transcriptome sequencing data under different stresses (SRR31832975–SRR31833946) have been deposited at the NCBI under the BioProject PRJNA1203203. Whole-genome resequencing data of 1,078 hexaploid oat accessions were obtained from the OatOmics database (http://www.oatomics.com). Source data are provided with this paper.

Code availability

The scripts used in this study are available via GitHub (https://github.com/HongyuZhang-HBU/oat-pangenome) and Zenodo[111] (https://doi.org/10.5281/zenodo.15761906).

References

52. Li, H. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
54. 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).
55. Alexander, D. H., Novembre, J. & Lange, K. L. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
56. 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**, 170–175 (2021).
57. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
58. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
59. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
60. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
61. Robert Hubley, A. S. *RepeatModeler Open-1.0* www.repeatmasker.org/RepeatModeler/ (2010).
62. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
63. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
64. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **Chapter 4**, 4.10.1–4.10.14 (2009).
65. Zhang, R.-G. et al. TEsorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. *Hortic. Res.* **9**, uhac017 (2022).
66. 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).
67. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).

68. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–ii225 (2003).
69. Gremme, G., Brendel, V., Sparks, M. E. & Kurtz, S. Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* **47**, 965–978 (2005).
70. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
71. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
72. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
73. Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 7 (2020).
74. Jia, K. H. et al. SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific *k*-mers. *New Phytol.* **235**, 801–809 (2022).
75. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
76. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
77. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
78. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
79. He, Q. et al. High-quality genome of allotetraploid *Avena barbata* provides insights into the origin and evolution of B subgenome in *Avena*. *J. Integr. Plant Biol.* **67**, 1515–1532 (2025).
80. Han, M. V., Thomas, G. W., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
81. Sun, P. et al. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* **15**, 1841–1851 (2022).
82. Zhang, R.-G. et al. Subgenome-aware analyses suggest a reticulate allopolyploidization origin in three *Papaver* genomes. *Nat. Commun.* **14**, 2204 (2023).
83. Doyle, J. & Egan, A. Dating the origins of polyploidy events. *New Phytol.* **186**, 73–85 (2009).
84. Jin, J. J. et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).
85. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
86. Danecek, P. & McCarthy, S. BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* **33**, 2037–2039 (2017).
87. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: a toolkit incorporating γ-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**, 77–80 (2010).
88. Zhang, Y. et al. Telomere-to-telomere *Citrullus* super-pangenome provides direction for watermelon breeding. *Nat. Genet.* **56**, 1750–1761 (2024).
89. Huang, Y. et al. Pangenome analysis provides insight into the evolution of the orange subfamily and a key gene for citric acid accumulation in citrus fruits. *Nat. Genet.* **55**, 1964–1975 (2023).
90. Zhang, Z. et al. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **419**, 779–781 (2012).
91. Potter, S. C. et al. HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
92. Li, P. et al. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* **17**, 852 (2016).
93. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
94. Marçais, G. et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
95. Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).
96. Li, X. et al. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in *Brassica oleracea*. *Nat. Genet.* **56**, 517–529 (2024).
97. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
98. Zhang, J. et al. Vacuum and co-cultivation agroinfiltration of (germinated) seeds results in tobacco rattle virus (TRV) mediated whole-plant virus-induced gene silencing (VIGS) in wheat and maize. *Front. Plant Sci.* **8**, 393 (2017).
99. Liu, N. et al. Comprehensive co-expression network reveals the fine-tuning of *AsHSFA2c* in balancing drought tolerance and growth in oat. *Commun. Biol.* **8**, 393 (2025).
100. Tian, T. et al. *Arabidopsis* FAR-RED ELONGATED HYPOCOTYL3 integrates age and light signals to negatively regulate leaf senescence. *Plant Cell* **32**, 1574–1588 (2020).
101. Liu, N. et al. A lncRNA fine-tunes salicylic acid biosynthesis to balance plant immunity and growth. *Cell Host Microbe* **30**, 1124–1138 (2022).
102. Zhao, X. et al. Global identification of *Arabidopsis* lncRNAs reveals the regulation of *MAF4* by a natural antisense RNA. *Nat. Commun.* **9**, 5056 (2018).
103. Zhou, J. et al. Intronic heterochromatin prevents cryptic transcription initiation in *Arabidopsis*. *Plant J.* **101**, 1185–1197 (2020).
104. Li, Q. et al. DEAD-box helicases modulate dicing body formation in *Arabidopsis*. *Sci. Adv.* **7**, eabc6266 (2021).
105. Wang, K. et al. Author Correction: The gene *TaWOX5* overcomes genotype dependency in wheat genetic transformation. *Nat. Plants* **8**, 717–720 (2022).
106. Liu, X. et al. Uncovering the transcriptional regulatory network involved in boosting wheat regeneration and transformation. *Nat. Plants* **9**, 908–925 (2023).
107. Yu, Y. et al. Enhancing wheat regeneration and genetic transformation through overexpression of *TaLAX1*. *Plant Commun.* **5**, 100738 (2024).
108. Ishida, Y., Tsunashima, M., Hiei, Y. & Komari, T. Wheat (*Triticum aestivum* L.) transformation using immature embryos. *Methods Mol. Biol.* **1223**, 189–198 (2015).
109. Said, M. et al. The *Agropyron cristatum* karyotype, chromosome structure and cross-genome homoeology as revealed by fluorescence in situ hybridization with tandem repeats and wheat single-gene probes. *Theor. Appl. Genet.* **131**, 2213–2227 (2018).
110. Xi, W. et al. New ND-FISH-positive oligo probes for identifying *Thinopyrum* chromosomes in wheat backgrounds. *Int. J. Mol. Sci.* **20**, 2031 (2019).
111. Zhang, H. et al. Super-pangenome analyses across 35 accessions of 23 *Avena* species highlight their complex evolutionary history and extensive genomic diversity. *Zenodo* https://doi.org/10.5281/zenodo.15761906 (2025).

Acknowledgements

This work was supported by the Natural Science Foundation of Hebei Province (grant nos. C2023201074 and C2021201048 to H.D.),

the Young Elite Scientists Sponsorship Program by CAST (grant YESS20210080 to H.D.), the National Natural Science Foundation of China (32100500 to H.D.), open funds of the State Key Laboratory of Plant Environmental Resilience (grant no. SKLPERKF2406 to Z.G.), the Interdisciplinary Research Program of Natural Science of Hebei University (grant no. 513201422004 to H.D.) and Post-graduate's Innovation Fund Project of Hebei University (grant no. HBU2025BS011 to H.Z.). We sincerely thank Y. Fan at Chengdu University, Y. Liu at the Dingxi Academy of Agricultural Sciences and B. Wu at the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences for providing the oat accession resources. We also thank K. Wang at the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences for providing the TaWOX5 plasmid.

Author contributions

H.D. conceived and supervised the project; H.D., Q.H. and Z.G. designed the study. Yaru Wang, X.Z., Q.H., Yu Wang and H.Z. sequenced and processed the raw data. Q.H., H.Z., Yaru Wang, M.W. and X.Z. assembled and annotated the genome. H.Z., X.Z., Z.L., Q.Z. and L.X. performed phylogenetic and genome evolution analyses; Yaru Wang, Q.H. and H.Z. conducted the transcriptome analysis; H.Z., Yaru Wang, Q.H., X.Z. and J.L. conducted pangenome analyses. Q.H., Yaru Wang, T.L., W.L. and N.L. performed TE and SV analyses. W.L., J.L. and Q.H. performed the population genetics analysis. N.L., Q.H., Y.Y., T.L., Y.Q. and J.Y. conceived of and designed the experiments. H.D. composed the outline of the paper. Q.H., W.L. and H.Z. wrote the paper. H.D. revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

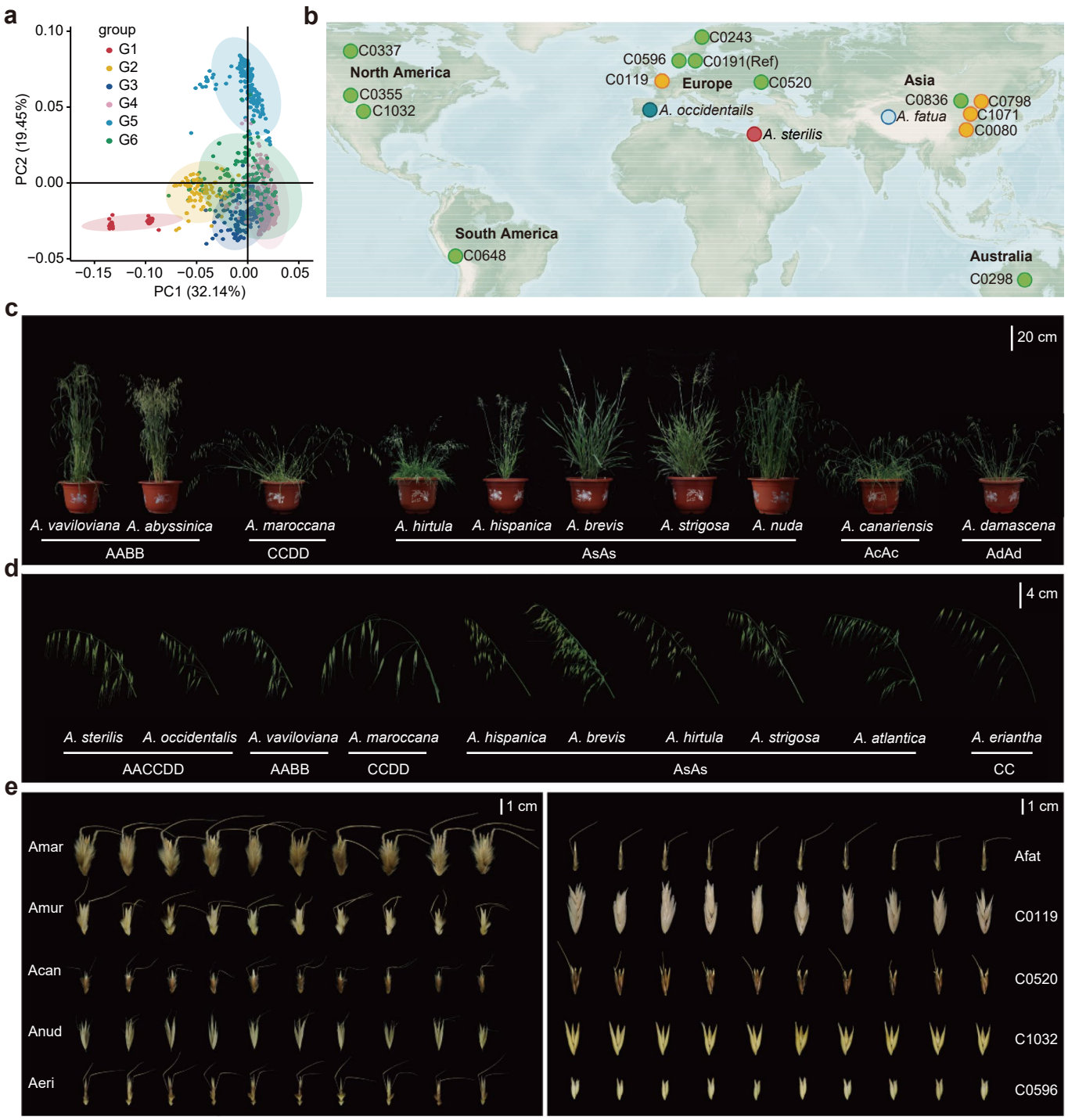
Extended data is available for this paper at https://doi.org/10.1038/s41588-025-02294-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02294-z.

Correspondence and requests for materials should be addressed to Zhizhong Gong, Qiang He or Huilong Du.

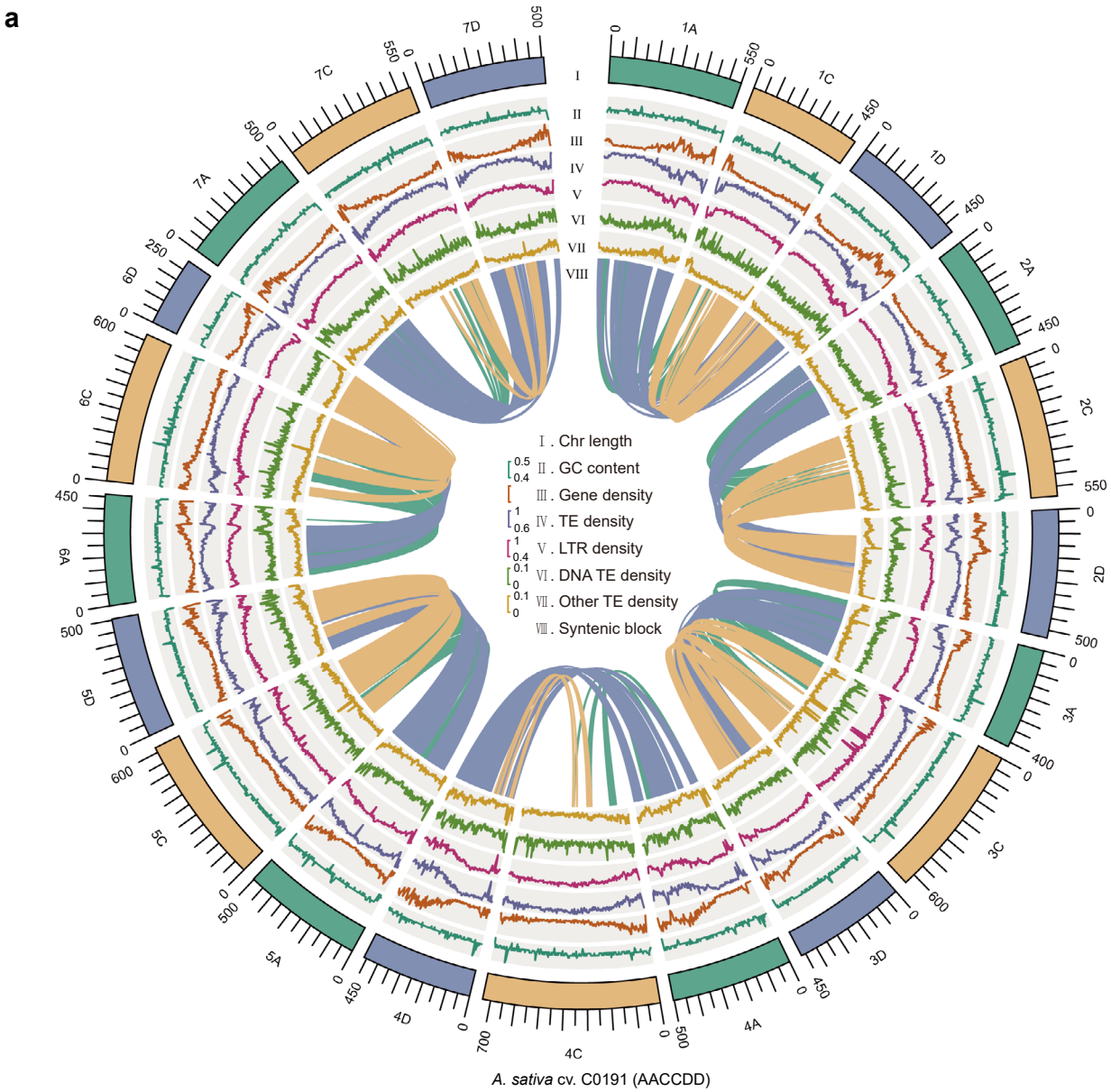
Peer review information *Nature Genetics* thanks Mark Chapman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

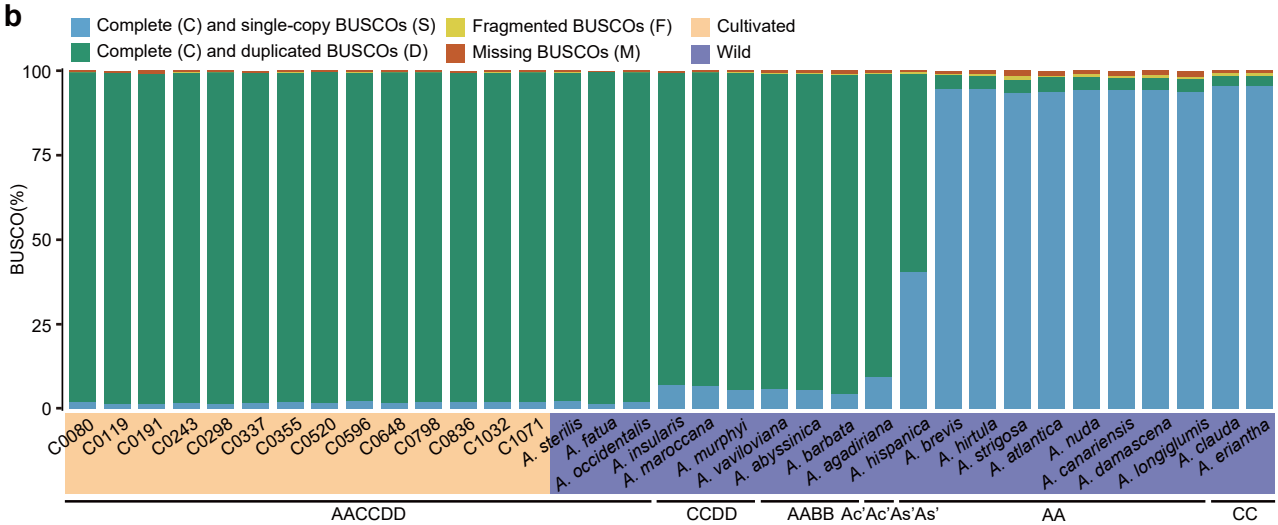
https://doi.org/10.1038/s41588-025-02294-z



Extended Data Fig. 1 | Geographical distribution and phenotype diversity of cultivated and wild oats accessions. a, Principal component analysis (PCA) of 1,079 globally collected hexaploid oat accessions. PC1 (32.14%) clearly separates wild oat accessions (G1) and cultivated oat accessions (G2-G6); PC2 (19.45%) clearly separates G5 group of hulless oats from other groups; **b**, Geographic distribution of the 17 diverse representative accessions among 1,079 hexaploid oat accessions. The world map was generated using the WorldClim (https://worldclim.org/) whose data are freely available for academic use (https://worldclim.org/about.html). The color of points corresponds to Fig. 1a. **c-e**, The highly diverse agronomic phenotypes of representative wild oat accessions, including plant architecture (**c**), spike architecture (**d**) and spikelet (**e**).

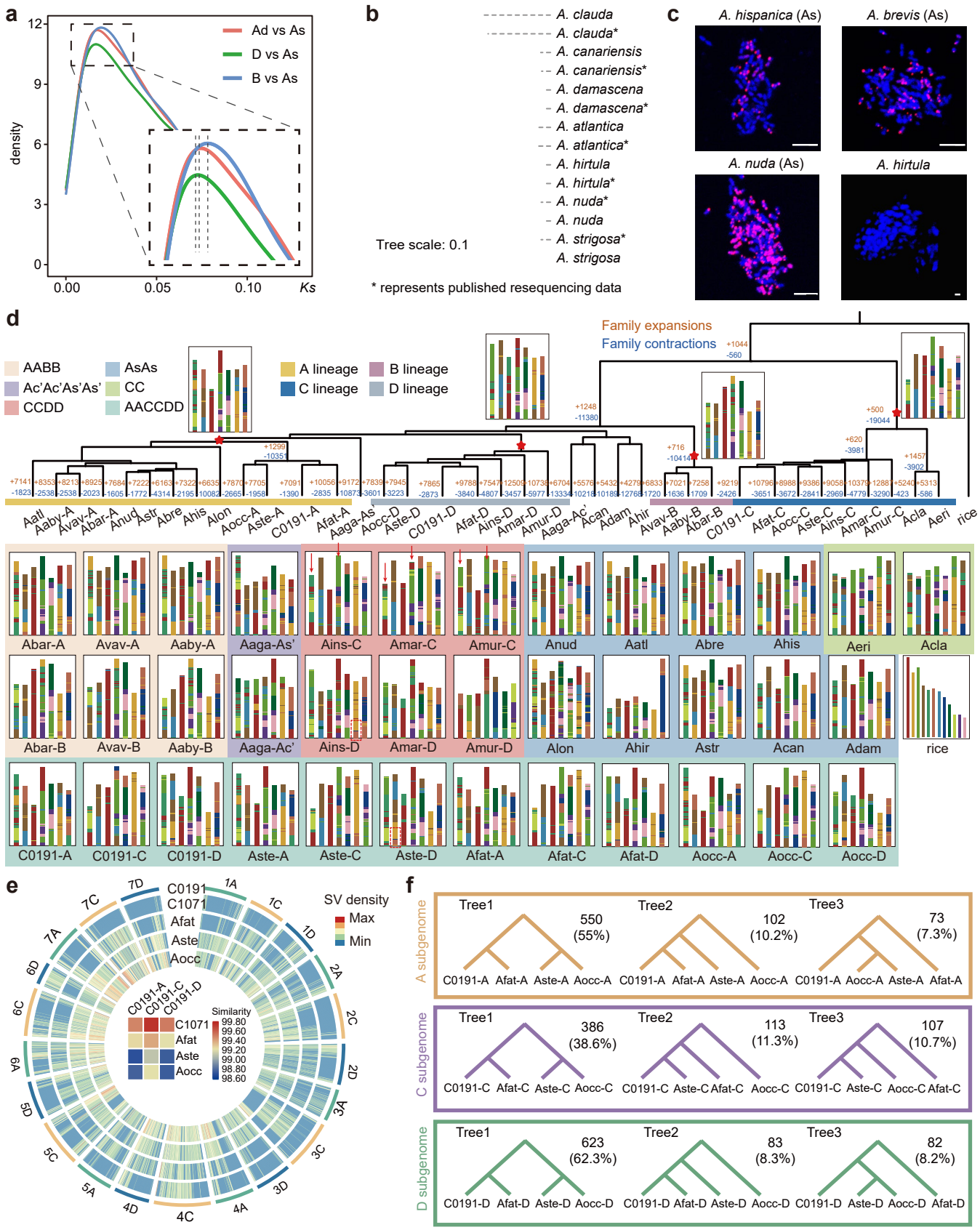
https://doi.org/10.1038/s41588-025-02294-z





Extended Data Fig. 2 | Circos plot of the reference genome C0191 and assembly assessment of genomes in our study. a, Circos plot showing chromosome-level features of C0191 reference genome. Tracks represent (I) chromosome length, (II) GC content, (III) gene density, (IV)TE density, (V) LTR density, (VI) DNA TE density, (VII) other TE density and (VIII) syntenic blocks. **b**, BUSCO completeness assessment for genome annotation of 35 oat accessions.

https://doi.org/10.1038/s41588-025-02294-z

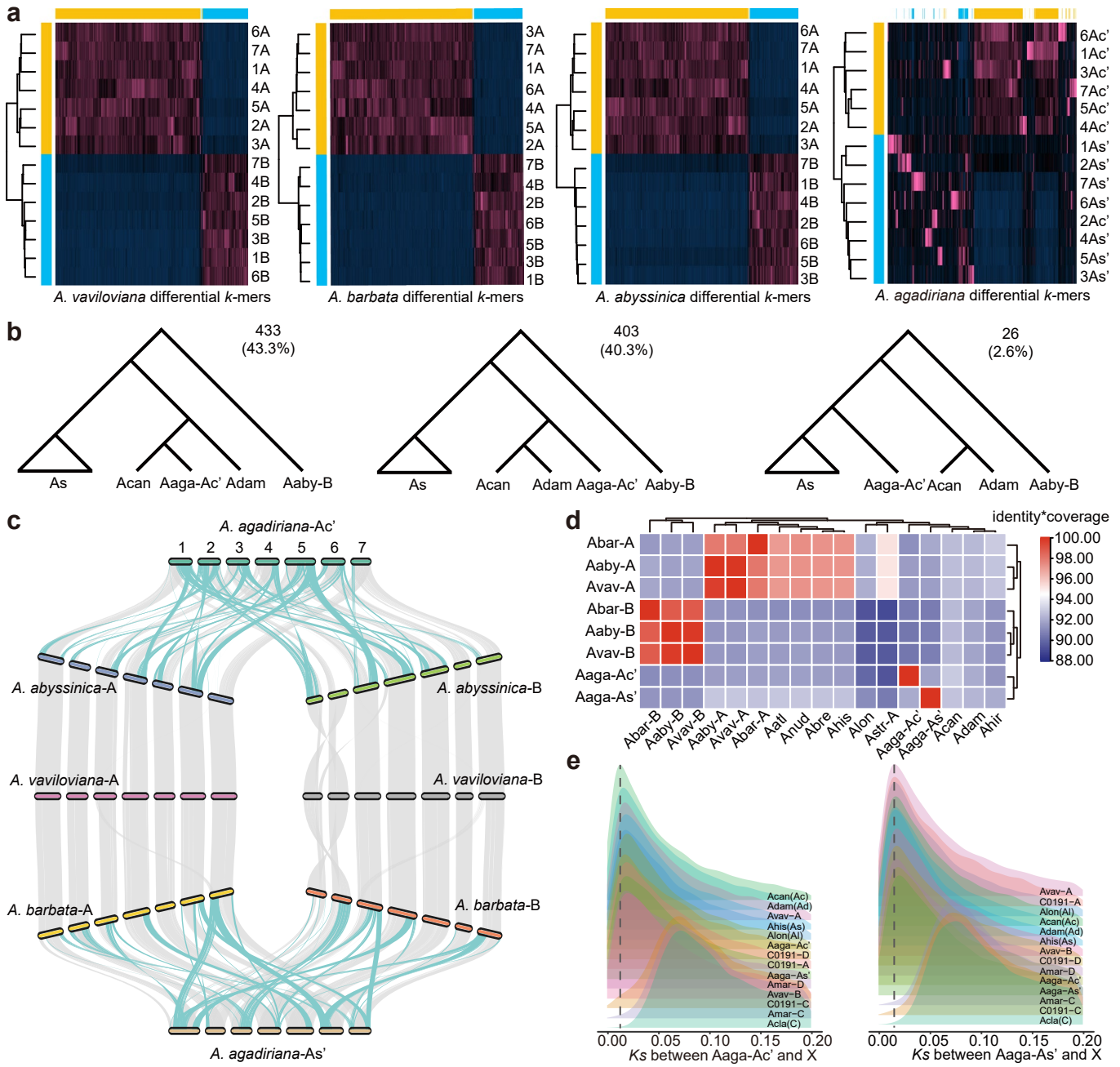


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41588-025-02294-z

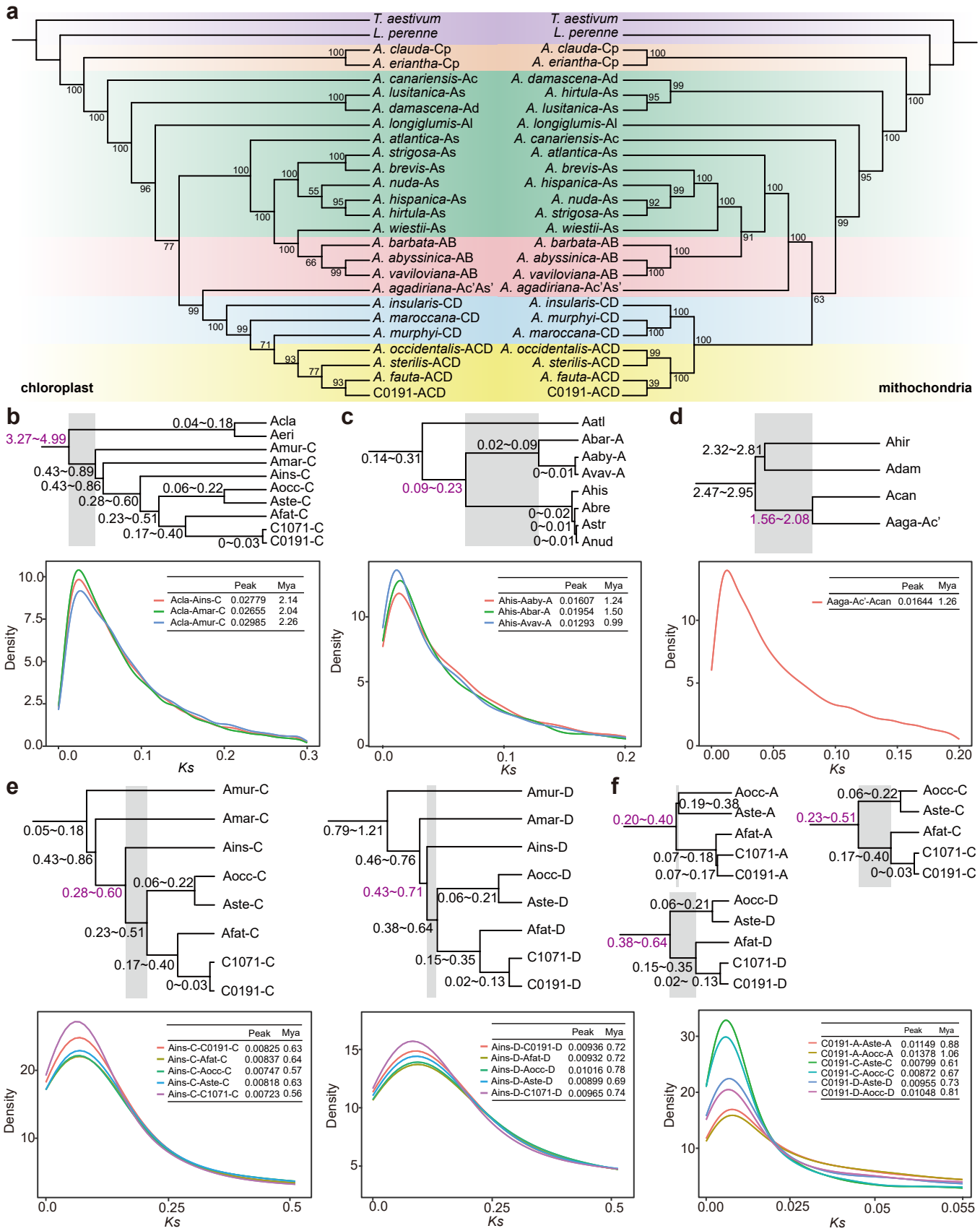
Extended Data Fig. 3 | Phylogenetic relationships and karyotypes evolution of *Avena* species. a, Distribution of *Ks* for orthologue duplicates identified between two (sub)genomes to estimate divergence time. Ad (*A. damascene*), As (*A. brevis*), D (*A. maroccana*-D), B (*A. vaviloviana*-B). **b**, SNP-based phylogenetic tree constructed with *A. longiglumis* as the reference genome based on our resequencing data and published resequencing data. **c**, A specific FISH probe targeting the As subtypes can label *A. hispanica*, *A. brevis* and *A. nuda*, but not *A. hirtula*, as observed in cells at the metaphase of mitosis. Scale bars, 5 μm. **d**, The reconstructed ancestral karyotype of the Avena genus using rice as the outgroup. Gene families with expansion and contraction in each species and node are highlighted in blue and orange, respectively. The red pentagrams represent the position of the inferred ancestor of A, B, C and D lineages. The arrows represented that the chromosome 1 and 4 in C subgenomes of tetraploids exhibited large-scale rearrangements. The red dashed boxes represent a large segment specific to the D subgenome of *A. insularis* but absent in that of *A. maroccana* and *A. murphyi*, then was translocated to the end of chromosome 2D during the process of polyploidization and retained in the hexaploid oats. **e**, The density of SVs in C1071, Afat, Aste, Aocc with the C0191 genome as reference. The average sequence similarity between subgenomes of C0191 and C1071, Afat, Aste, Aocc are shown in the heatmap. **f**, Proportions of contrasting gene tree topologies for 100 randomly selected single-copy genes (1,000 times) with regard to three major conflicting relationships. The values in the upper right corner of each tree show the percentage of windows that recovered the topology.

https://doi.org/10.1038/s41588-025-02294-z



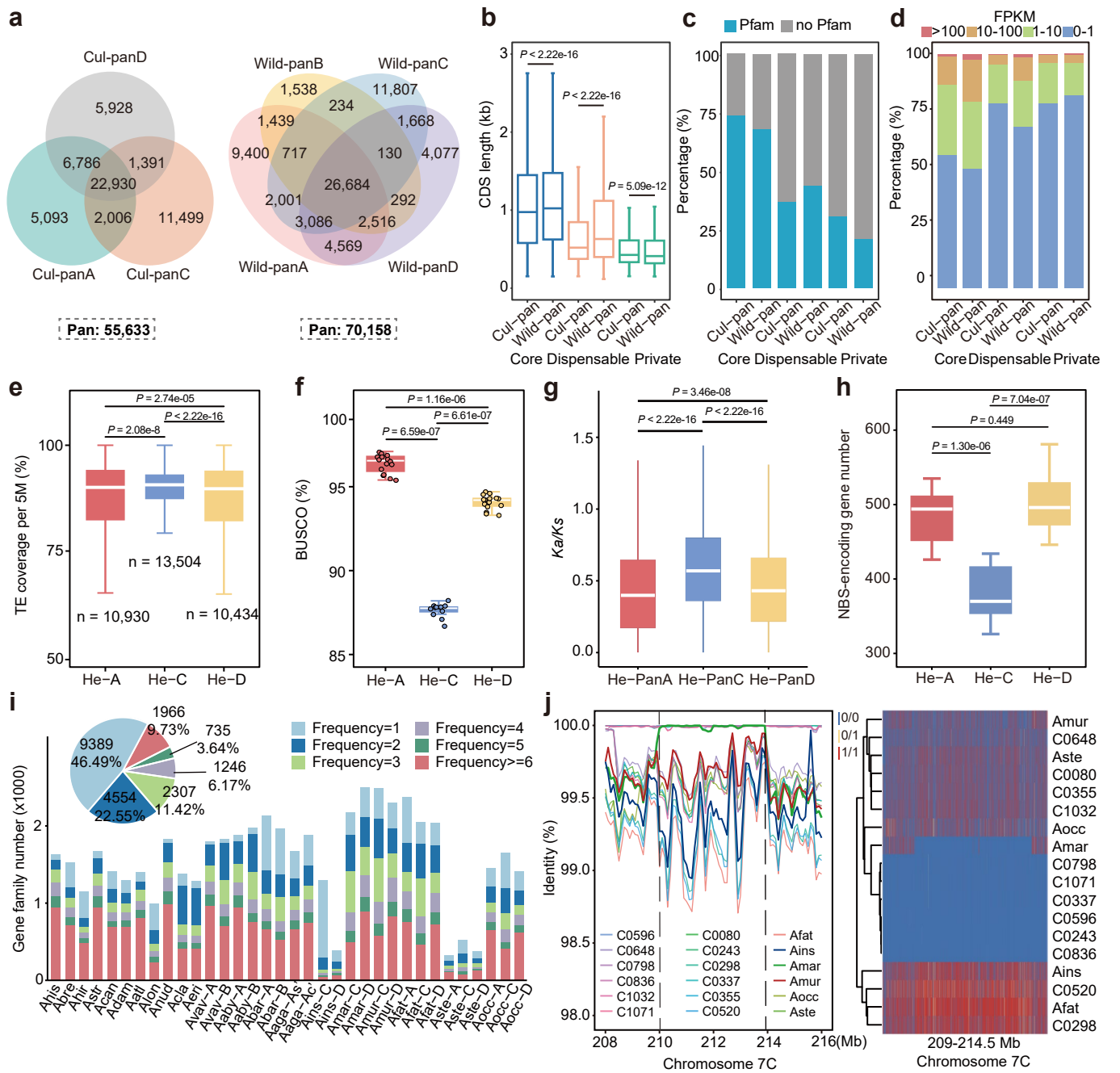
Extended Data Fig. 4 | The divergent genome composition between *A. agadiriana* and AABB tetraploid species. a, Unsupervised hierarchical clustering of differential 15-mers validates that *A. vaviloviana*, *A. barbata*, *A. abyssinica* are successfully phased into two subgenomes based on clearly distinct patterns, while the two subgenomes of *A. agadiriana* showed higher sequence identity, making the subgenomes phasing relatively challenging and easily confused. **b**, The coalescent-based tree reconstructed from 100 randomly selected single-copy genes (1,000 times). The values in the upper right corner of each tree show the percentage of windows that recovered the topology. As (*A. hispanica*, *A. brevis*, *A. atlantica*, *A. nuda*). **c**, Intergenomic synteny among *A. agadiriana*, *A. abyssinica*, *Avaviloviana* and *A. barbata*. Some large inversions and translocations between *A. agadiriana* and species with AABB genome constitution were shown in blue. **d**, The average sequence similarity between the subgenomes of *A. agadiriana*, and three AABB species, as well as other subgenomes. **e**, Distribution of synonymous divergence (*Ks*) for orthologue genes identified between each subgenome of *A. agadiriana* (Aaga-Ac' and Aaga-As') and other subgenomes.

https://doi.org/10.1038/s41588-025-02294-z



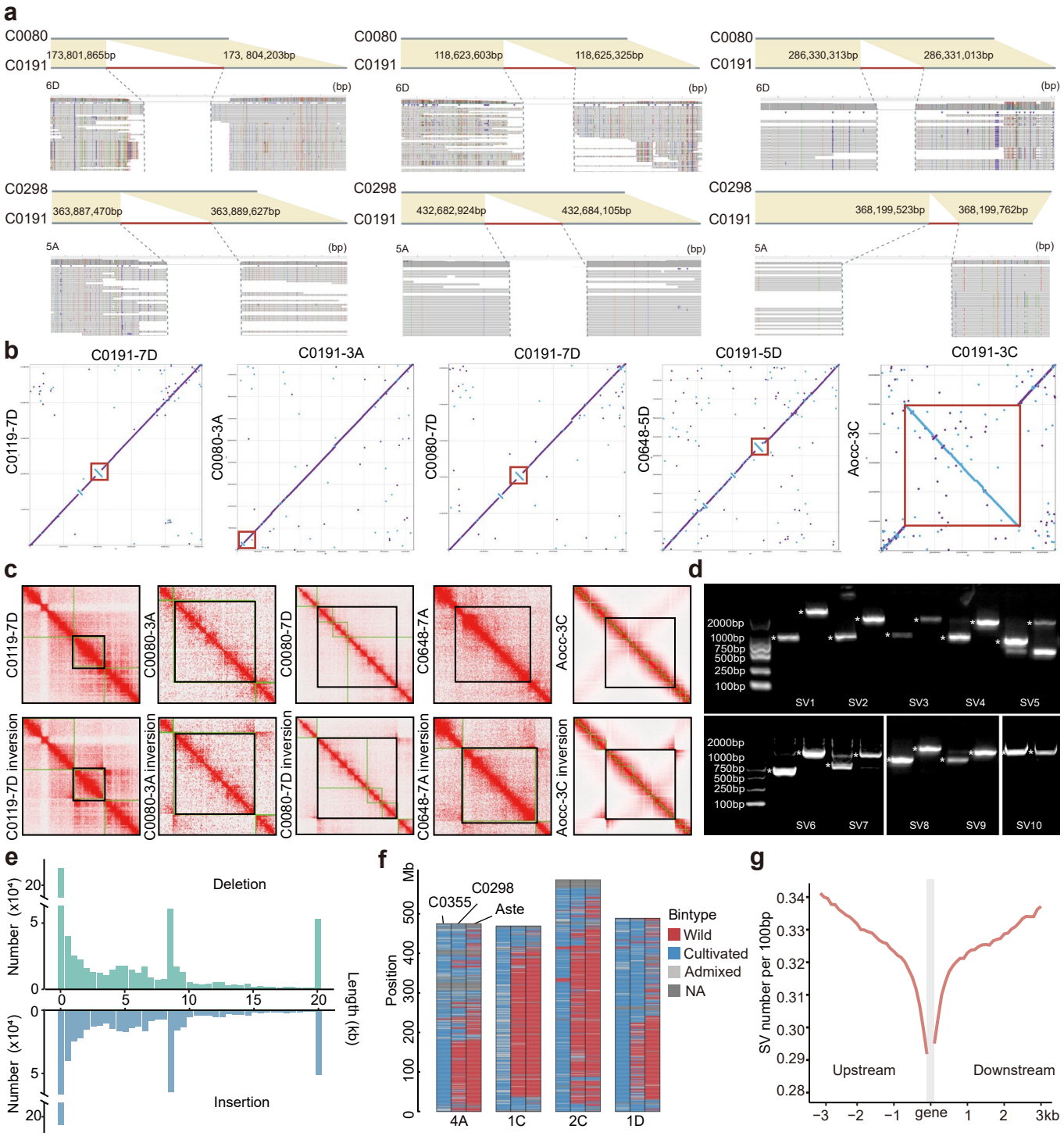
Extended Data Fig. 5 | Refined model for the origins, polyploidizations and evolutionary trajectories of *Avena* species. a, Phylogenetic trees of 24 oat species with *L. perenne* and *T. aestivum* as an outgroup based on 26 chloroplast (cp) genomes and 7,137 mitochondrial (mt) SNP data sets. **b-f**, Divergence and polyploidization time estimation based on the 2,456 single-copy orthologous groups using MCMCTree, and distribution of synonymous divergence (*Ks*) for orthologue duplicates identified between two subgenomes.

https://doi.org/10.1038/s41588-025-02294-z



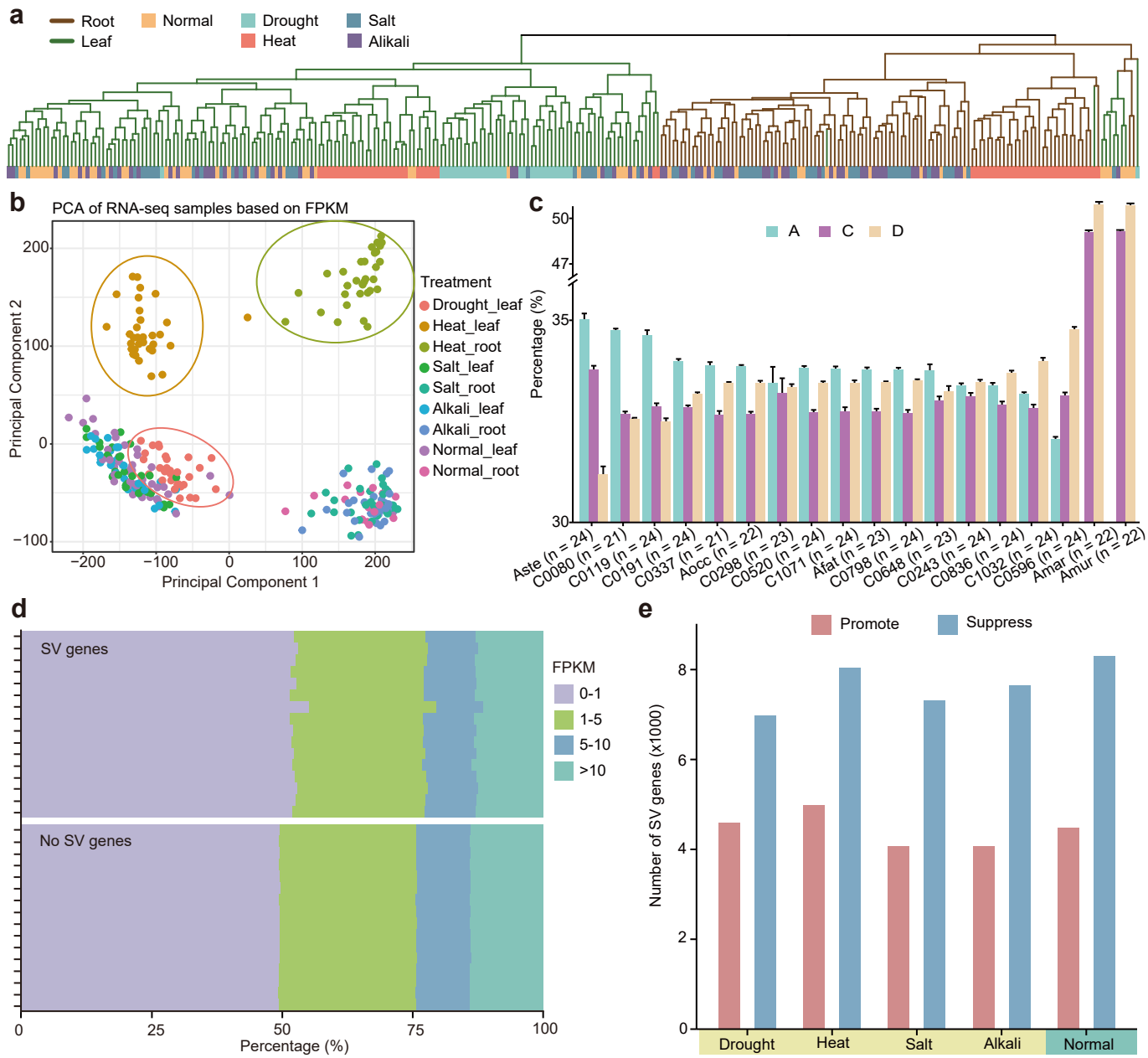
Extended Data Fig. 6 | Genetic diversity in cultivated and wild oats. a, Venn diagram showing the numbers of common and unique gene families identified in panA, panC and panD of cultivated oats and panA, panB, panC and panD of wild oats. **b**, The distribution of CDS length in core, dispensable and private genes in cultivated and wild pangenomes. Cul-core (n = 1,460,017), Cul-dispensable (n = 394,369), Cul-private (n = 8,665), Wild-core (n = 776,159), Wild-dispensable (n = 553,189), Wild-private (n = 35,786). **c**, Proportion of genes with Pfam domains in core, dispensable, and private genes in cultivated and wild oats. Blue histograms indicate the genes with Pfam domain annotation; gray histograms indicate the gene without Pfam domain annotation. **d**, Comparison of gene expression levels in core, dispensable, and private genes in cultivated and wild oats. **e-h**, Comparison of the TE coverage with a window of 5 Mb (**e**), complete BUSCOs (n = 17) (**f**), *Ka/Ks* (n = 10,000) (**g**), NBS-encoding gene number (n = 17) (**h**) among the three subgenomes of hexaploid oats. **i**, The number of specific gene families derived from each wild oats compared to cultivated oats. The pie chart represents the proportion of gene families unique to wild oats that are derived from different number of subgenomes. **j**, A candidate introgression region from *A. maroccana*. The sequence similarity and SNP dataset all supported that *A. maroccana* showed close relationship with cultivated oats in this region.

https://doi.org/10.1038/s41588-025-02294-z



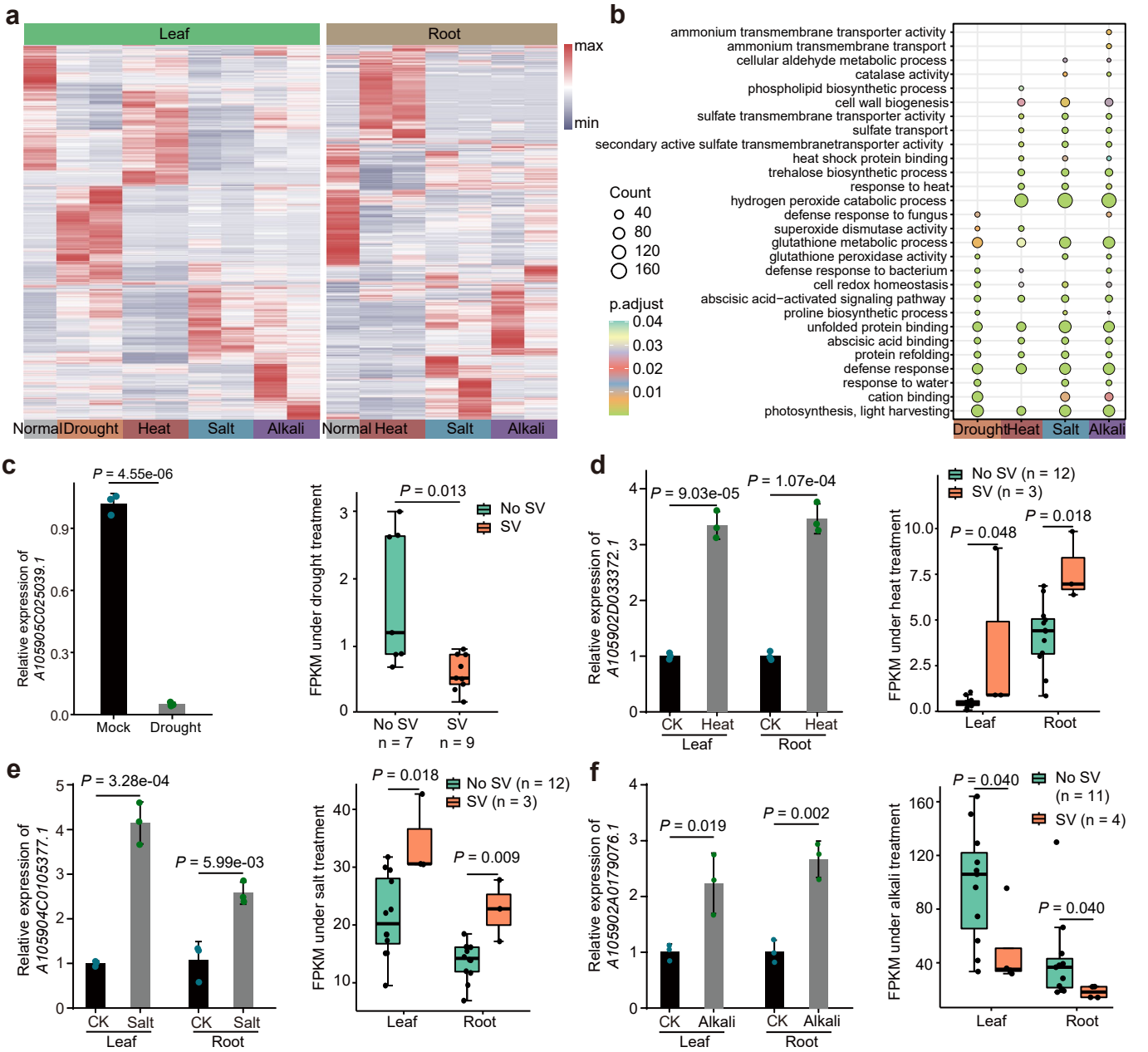
Extended Data Fig. 7 | Verification and analysis of structural variations. a, HiFi reads were used to validate the borders of 6 deletions randomly selected in C0080 and C0298. **b**, Schematic diagrams showed the distribution of inversion between two oat genomes (x-axis: C0191). Red box represented the validated inversion events relative to C0191. **c**, Illustration of inversion identified between two oat genomes by Hi-C contact map. Chromatin interaction heatmap revealed inversion signals appearing after manual flipping. These maps supported the inversions in chromosome 7D, 3 A, 5D and 3 C. **d**, Ten insertions were randomly selected for validation by PCR amplification. The detailed information of the validation of 10 SVs were presented in Supplementary Table 17. SV8, SV9 and SV10 came from other gels. Uncropped images are shown in Supplementary Figure 20. Different bands show the presence or absence of the SVs amplified in two oat genomes. The asterisk indicates the target band. The experiments were repeated independently three times, with similar results. **e**, Length distribution of deletions and insertions among 16 oat accessions. **f**, The haplotype of chromosome 4 A, 1 C, 2 C and 1D among C0355, C0298 and *A. sterilis* genomes indicated the similar genotype of C0298 with *A. sterilis*. **g**, Density of SV number per 100 bp in gene bodies and 3 kb flanking regions among 16 oat accessions.

https://doi.org/10.1038/s41588-025-02294-z



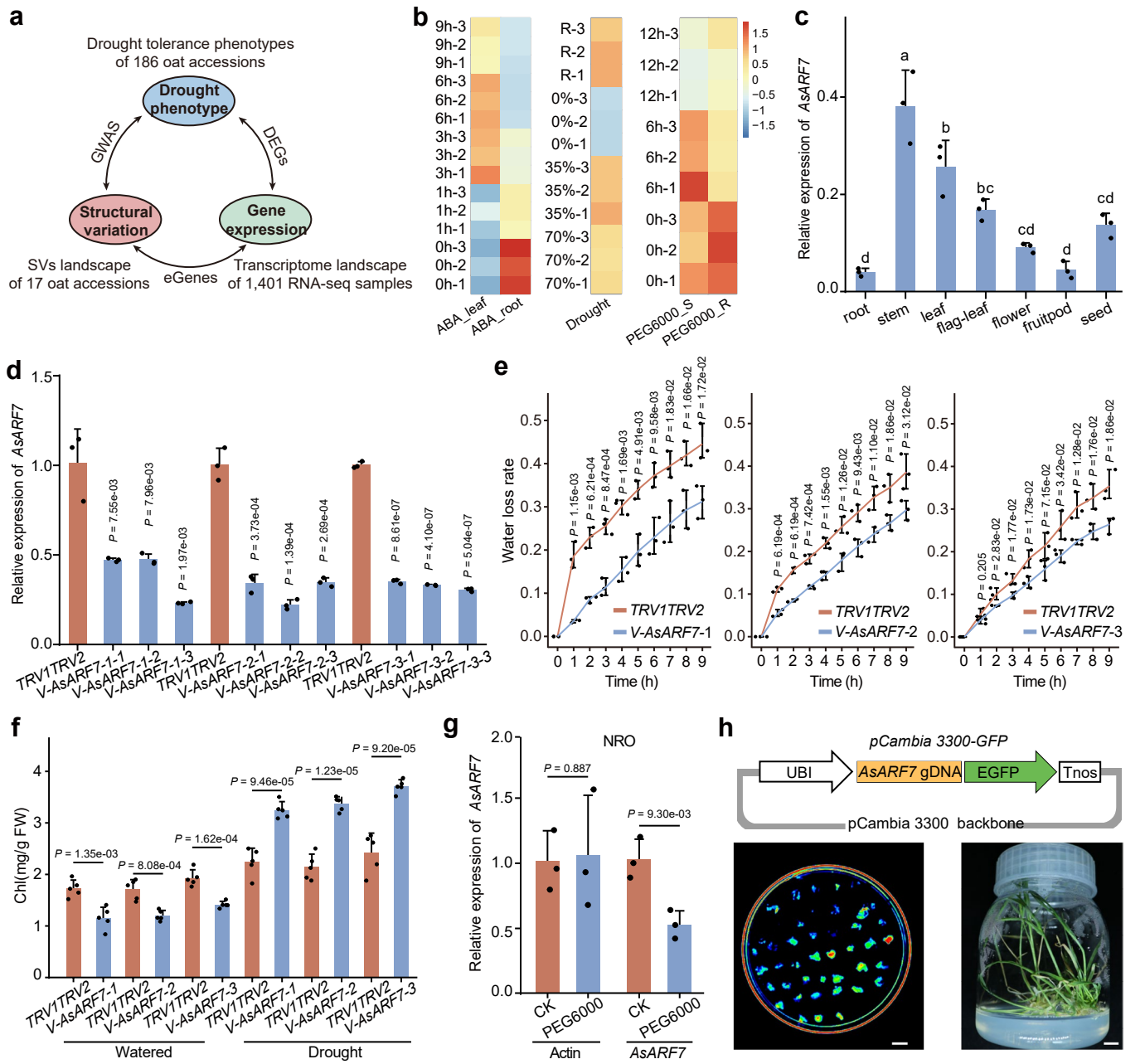
Extended Data Fig. 8 | The effects of SVs on the gene expression. **a**, Phylogenetic tree of RNA-seq samples. **b**, Principal-component analysis (PCA) for the RNA-seq samples based on the FPKM of each sample. **c**, The relative expression percentage of each subgenome based on the 1:1:1 (1:1) gene triads (pairs) in 18 oat accessions. The relative expression of different subgenomes of each accession summed to 100%. Error bar represents the SD of all RNA-seq samples of each accession. **d**, Comparison of average gene expressions between SV genes and no-SV genes among all RNA-seq samples in 16 oat accessions. **e**, The number of SV genes with suppression and promotion expression under normal growing and four different stress treatments.

https://doi.org/10.1038/s41588-025-02294-z



Extended Data Fig. 9 | Gene expression analysis under different stress treatments. a, Heatmap of the expression levels of the genes under normal and different stress treatments of both leaf and root. **b**, Functional analysis of differential expression genes from four different stress treatments. **c**, RT-qPCR detection of *A105905C025039.1* expression level in normal and drought conditions was shown in the left. Comparison of the expression level (FPKM) of *A105905C025039.1* between the two haplotypes was shown in the right. **d-f**, RT-qPCR detection of *A105902D033372.1*, *A105904C0105377.1* and *A105902A0179076.1* expression level in control and different stress treatments in two tissues were shown in the left (Two-tailed Student's *t* test). Error bar represents the SD of three biological replicates. Comparison of the expression level (FPKM) of *A105902D033372.1*, *A105904C0105377.1* and *A105902A0179076.1* between the two haplotypes in two tissues under corresponding stress treatment were shown in the right (two-sided Wilcoxon rank-sum test).

https://doi.org/10.1038/s41588-025-02294-z



Extended Data Fig. 10 | Functional validation of *AsARF7*. **a**, The relationship among drought phenotype, SVs and gene expression. **b**, The expression heatmap of *AsARF7* under ABA, drought and PEG6000 treatment. PEG600_S represents the aerial part after PEG600 treatment. PEG600_R represents the root part after PEG600 treatment. **c**, The expression of *AsARF7* in different tissues including root, stem, leaf, flag-leaf, flower, fruit pod and seed. **d**, Detection of *AsARF7* gene expression level in knockdown mutant strains. **e**, Water loss rate statistics of *AsARF7* knockdown strains with three biological replicates. **f**, Chlorophyll content detection statistics of *AsARF7* knockdown strains under drought treatment. **g**, Nuclear run-on assay showed that the transcription of *AsARF7* was decreased in PEG6000 treatment plants. **h**, Genetic transformation in *A. sativa*. The up part: schematic representation of the *pUBI::AsARF7-GFP* vector with the UBI promoter and NOS terminator; The down part: detection of GFP fluorescence signal in callus induced from mature embryos infected with the *pUBI::AsARF7-GFP* vector (leaf). Regeneration phenotypes of mature embryos infected with the *pUBI::AsARF7-GFP* vector. Scale bar, 1 cm (right). Error bar represents the SD of three (**c**,**d**,**g**) of five (**f**) biological replicates.

nature portfolio

Corresponding author(s): Huilong Du, Qiang He, Zhizhong Gong

Last updated by author(s): Jul 10, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Analysis was performed using custom R/ Perl and Python code. Additional open source libraries were used BWA (v0.7.17, https://sourceforge.net/projects/bio-bwa/), SAMtools (v1.7, https://sourceforge.net/projects/samtools/), bcftools (v1.10.2 & v1.12, https://github.com/samtools/bcftools), VCF2DIS (v1.46 & v1.47, https://github.com/BGI-shenzhen/VCF2Dis), Newick Utilities Tutorial (V1.6.0, https://www.ezlab.org/), PLINK (v1.90b6.26, https://www.cog-genomics.org/plink2/), ADMIXTURE (v1.3.0, http://dalexander.github.io/admixture/), Hifiasm (v0.19.7-r598, https://github.com/chhylp123/hifiasm), Juicer (v1.5, https://github.com/aidenlab/juicer), 3D-DNA, Juicebox (v1.11.08, https://github.com/aidenlab/Juicebox), BUSCO (v5.2.2, https://busco.ezlab.org/), LTR_retriever (v1.9, https://github.com/oushujun/LTR_retriever), RepeatModeler (v1.0.11, https://github.com/Dfam-consortium/RepeatModeler), LTR_FINDER (v1.05, https://github.com/oushujun/LTR_FINDER_parallel), LTRharvest (v1.5.11, hamburg.de/en/forschung/gi/software/ltrharvest.html), RepeatMasker (v4.1.1, http://repeatmasker.org/), TEsorter (v1.4.6, https://github.com/zhangrengang/Tesorter), Trinity (v2.12, https://github.com/trinityrnaseq/trinityrnaseq/wiki), PASA (v2.3, https://github.com/PASApipeline/PASApipeline), AUGUSTUS (v3.2.3, https://sourceforge.net/projects/augustus/), GenomeThreader (v1.7.3, https://genomethreader.org/), HISAT2 (v2.2.1, https://github.com/DaehwanKimLab/hisat2), StringTie (v1.13, https://ccb.jhu.edu/software/stringtie/), TransDecoder (v5.1.0, https://github.com/TransDecoder/TransDecoder), InterProScan (v5.48-83.0, https://www.msi.umn.edu/sw/interproscan), KofamScan (v1.3.0, https://github.com/takaram/kofam_scan), SubPhaser (v1.2.6, https://github.com/zhangrengang/SubPhaser), jcvi (v1.2.1, https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)), OrthoFinder (v2.3.8, https://github.com/davidemms/OrthoFinder), MUSCLE (v3.8.1551, http://drive5.com/muscle/), RAxML (v8.2.12, https://github.com/amkozlov/raxml-ng), DensiTree (v3.0.2, https://www.cs.auckland.ac.nz/~remco/DensiTree/), MCMCtree (v4.9, http://abacus.gene.ucl.ac.uk/software/paml.html), CAFÉ (v4.2.1, https://hahnlab.github.io/CAFE/), MCScanX (https://github.com/wyp1125/MCScanX), WGDI (v0.6.3, https://wgdi.readthedocs.io/en/latest/index.html), GetOrganelle (v1.7.7.0, https://github.com/Kinggerm/GetOrganelle), KaKs_Calculator (v2.0, https://github.com/kullrich/kakscalculator2), ParaAT (v2.0, https://github.com/wonaya/ParaAT), HMMER (v4.0.0, http://hmmer.org/),

diamond (v0.9.14, https://github.com/bbuchfink/diamond), Fastp (v0.20.1, https://github.com/OpenGene/fastp), SyRI (v1.6.3, https://schneebergerlab.github.io/syri/fileformat.html), GEMMA (v0.98.5, https://github.com/genetics statistics/GEMMA), PhyloNet (v3.8.2, https://github.com/NakhlehLab/PhyloNet), QuIBL (https://github.com/miriammiyagi/QuIBL), WorldClim (https://worldclim.org/). The custom scripts used in this study are available via GitHub(https://github.com/HongyuZhang-HBU/oat-pangenome) and Zenodo(https://doi.org/10.5281/zenodo.15761906).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw data and genome assemblies in this study have been deposited in the Chinese National Genomics Data Center (https://bigd.big.ac.cn/) under the BioProject accession number PRJCA024628. The raw sequencing data (SRR31832837-SRR31832882) and genome assemblies (JBKZJG000000000-JBKZJZ000000000 and JBKZKA000000000-JBKZKH000000000) have been deposited in the NCBI under the Bioproject (PRJNA1200805). The transcriptome sequencing data under normal growth (SRR31798633-SRR31798686 and SRR31814753-SRR31815091) have been deposited in the NCBI under the Bioproject (PRJNA1201252). The transcriptome sequencing data under different stresses (SRR31832975-SRR31833946) have been deposited in the NCBI under the Bioproject (PRJNA1203203). The whole genome resequencing data of 1,078 hexaploid oat accessions obtained from OatOmics database (http://www.oatomics.com). Source data are provided in this article and the associated supplementary materials.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This is not applicable.
Reporting on race, ethnicity, or other socially relevant groupings	This is not applicable.
Population characteristics	This is not applicable.
Recruitment	This is not applicable.
Ethics oversight	This is not applicable.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were required to establish sample size in this study. We performed RNA sequencing for 1,365 oat samples from 21 wild oat species and 14 cultivated oat accessions. We selected 28 oat accessions to de novo assemble the genome representing 77% of the species in Avena and spanning cultivated oats and their wild relatives, based on the phylogenetic tree, population structure, principal component analysis (PCA) , geographical distributions, and genetic background. The selection for the present work aimed at covering all genome types of the genus Avena.
Data exclusions	No data exclusions. Details of data collection and processing are provided in the Methods.
Replication	Three biological replicates are performed in RNA-seq experiments. Each experiment involves at least three biological replicates. (Replications for each experiment were clearly stated in figure legends or Methods section.)
Randomization	Roots, stems, Leaves, fowers and seeds were randomly collected from oat individuals with same growth stages. The sampling process for genome DNA/RNA sequencing was randomly conducted.

Blinding | Group allocation was not relevant to this study and therefore randomization was not necessary.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☒	☐ Demonstrate how to render a vaccine ineffective
☒	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐ Increase transmissibility of a pathogen
☒	☐ Alter the host range of a pathogen
☒	☐ Enable evasion of diagnostic/detection modalities
☒	☐ Enable the weaponization of a biological agent or toxin
☒	☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	Cultivated and wild oat accessions were provided by the National Crop Genebank of China, Dingxi Academy of Agricultural Sciences and School of Food and Biological Engineering, Chengdu University.
Novel plant genotypes	Virus-induced gene silencing mediated by TRV was used to generate the knock-down strains.
Authentication	Detection of TRV viral content in gene knockdown strains to confirm the success of TRV virus infection and the qRT-PCR assay was used to detecte the target gene expression level in knockdown mutant and overexpression lines.