Research Square

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

Isolation of the major wheat domestication gene Tenacious glume1 reveals its multiple phenotypic effects and sheds light on evolution of wheat

Jan Dvorak

jdvorak@ucdavis.edu

University of California, Davis https://orcid.org/0000-0002-0427-6528

Mengyuan Xiao

University of California, Davis

Le Wang

Jilin University

Juan Debernardi

UCDavis https://orcid.org/0000-0002-4591-7061

Karin Deal

Department of Plant Sciences, University of California, Davis CA 95616

Tingting Zhu

University of California, Davis

Ming-Cheng Luo

University of California Davis https://orcid.org/0000-0002-9744-5887

Patrick McGuire

University of California, Davis

Jacob Yousif

University of California, Davis https://orcid.org/0009-0009-5087-9395

Eduard Akhunov

Kansas State University https://orcid.org/0000-0002-0416-5211

Fei He

Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing

Yael Lev-Mirom

University of Haifa https://orcid.org/0009-0008-8139-3286

Tamar Eilam

Institute for Cereal Crops Improvement and School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv, 6997801 Israel

Assaf Distelfeld

University of Haifa

Yidong Zhou
University of California, Davis https://orcid.org/0000-0003-1423-1857
Hans-Georg Müller
University of California, Davis
Jorge Dubcovsky
University of California, Davis and Howard Hughes Medical Institute https://orcid.org/0000-0002-7571-4345

Article

Keywords:

Posted Date: April 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9315249/v1

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

Additional Declarations: There is **NO** Competing Interest.

Isolation of the major wheat domestication gene *Tenacious glume1* reveals its multiple phenotypic effects and sheds light on evolution of wheat

Mengyuan Xiao[1,†], Le Wang[1,3,†], Juan M. Debernardi[1,2], Karin R. Deal[1], Tingting Zhu[1], Ming-Cheng Luo[1], Patrick E. McGuire[1], Jacob N. Yousif[1], Fei He[4,5], Eduard Akhunov[4], Yael Lev-Mirom[6], Tamar Eilam[6], Assaf Distelfeld[6], Yidong Zhou[7], Hans-Georg Müller[7], Jorge Dubcovsky[1], and Jan Dvorak[1]

[1]Department of Plant Sciences, University of California, Davis, CA 95616, USA

[2]College of Agriculture and Environmental Science, University of California, Davis, CA 95616, USA

[3]College of Plant Science, Jilin University, Changchun 130062, China

[4]Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA

[5]Laboratory of Advanced Breeding Technologies, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, 100101, China

[6]Department of Evolutionary and Environmental Biology and the Institute of Evolution, University of Haifa, Haifa 3498838, Israel

[7]Department of Statistics, University of California, Davis

[†]These authors contributed equally to the execution of the project.

Corresponding author: jdvorak@ucdavis.edu

Running head: Wheat *Tg1* locus

Mutations at the *Tenacious glume1* (*Tg1*) loci on chromosome 2A, 2B, and 2D were pivotal for evolution of free-threshing wheat. Here, we identified and validated a *Tg1* candidate gene. The wheat 2A- and 2B-loci each contained a single copy of the gene but the 2D-locus was multigenic with at least one of the genes having a frameshift mutation. We found no fixed mutation in CDSs of the 2A- and 2B-genes and suggest that *cis*-regulatory mutations have led to evolution of free-threshing tetraploid wheat. *Tg1* is equivalent to barley *ELI-A* gene controlling the development of leaf ligule and to the bread wheat reduced-height locus *Rht8.* In addition, mutations of *Tg1* affected spike length, density, shape, and fertility, leaf rigidity, and lodging. Fluorescence *in situ* hybridization revealed association of *Tg1* expression with vascular bundles and tissue boundaries. Durum wheat near isogenic lines harboring non-free-threshing and free-threshing 2B-alleles differed in the abundance of glume sclerenchyma. Our search for differentially expressed genes in the comparison of CRISPR/Cas9 free-threshing mutants vs the non-free-threshing parental line yielded genes for biogenesis of cell wall and cell wall functions. Past selection was centered on the 2A- and 2B-loci in durum wheat, but selection was primarily centered on the 2D-locus in bread wheat. This is consistent with the tetraploid parent of

hexaploid wheat being free-threshing. Our haplotype analyses pinpointed the origin of hexaploid wheat in Iran, showed that European and Asian spelt originated separately and confirmed that European spelt and most Asian spelt were derived from free-threshing hexaploid wheat.

Domestication of wild einkorn (*Triticum monococcum* ssp. *aegilopoides*), wild emmer (*T. turgidum* ssp. *dicoccoides*), and wild barley (*Hordeum vulgare* ssp. *spontaneum*) in the Fertile Crescent about 10,000 years ago marked the beginnings of agriculture in western Asia[1]. Domesticated emmer (*T. turgidum* ssp. *dicoccon*) was the dominant food cereal until it was replaced by durum wheat (*T. turgidum* ssp. *durum*) during the Greco-Roman period[2].

The primary difference between wild and domesticated emmer is that spike rachis is brittle in the mature spikes of wild emmer but solid in the spikes of domesticated emmer. In both taxa, however, the hulls (glumes, lemmas, and paleas) tenaciously encapsulate seeds and interfere with threshing. Evolution of free-threshing wheat, such as durum wheat and bread wheat (*T. aestivum* ssp. *aestivum*), with soft hulls was critical for wheat becoming the pillar of global food supply as it is today.

Emmer and durum wheat are tetraploid (subgenomes AABB) and bread wheat is hexaploid (subgenomes AABBDD). Hexaploid wheat evolved by hybridization of tetraploid wheat with diploid goatgrass *Aegilops tauschii* (genomes DD)[3,4]. The exact nature of the tetraploid parent of hexaploid wheat is uncertain. It was earlier thought to be domesticated emmer[4] but new evidence suggests that it may have been a free-threshing tetraploid wheat[5]. The diploid parent of bread wheat consists of two main phylogenetic lineages, L1 and L2, which loosely approximate *Ae. tauschii* ssp. *tauschii* and *Ae. tauschii* ssp. *strangulata*¸ respectively[6,7] and a minor phylogenetic lineage L3[8]. The bulk of the wheat D subgenome was contributed by lineage L2 in the southwestern Caspian[7] with a small fraction introgressed from lineage L3 in what is today Georgia[8].

The rachis disarticulates in mature spikes of wild emmer and *Ae. tauschii* at an abscission zone in the rachis axis that is either above (wild emmer) or below (*Ae. tauschii*) the glume attachment to the rachis[9] (**Fig. 1a**). Glumes remain permanently attached to the disarticulated rachis internodes in mature spikelets and tenaciously hold the spikelet together (**Figs. 1b, c**). The spikelet is the propagule in both species. In free-threshing durum and bread wheat, the rachis does not disarticulate but the attachments of glume bases to the rachis node are brittle in maturity (**Figs. 1d, e**). Glumes fall off and the spikelet disintegrates during threshing. Seeds are the propagules in durum and bread wheat.

The spikes of hexaploid amphiploids produced by hybridization of free-threshing tetraploid wheat with *Ae. tauschii* (synthetic wheat) are non-free-threshing due to expression of the *Tenacious glume1* (*Tg1*) gene located on the short arm of *Ae. tauschii* chromosome 2D[10] (hereinafter *Tg-D1*). The same is true for bread wheat disomic substitution (DS) lines in which *Ae. tauschii* chromosome 2D is substituted for bread wheat cv Chinese Spring (CS) chromosome 2D[5].

Because synthetic wheat is non-free-threshing[10], nascent hexaploid wheat must also have been non-free-threshing. Spelt (*T. aestivum* ssp. *spelta*), a non-free-threshing hexaploid wheat cultivated in the past in mountainous regions of Europe, with elongated and tapered (speltoid) spikes resembling those of synthetic wheat, was suggested to be the progenitor of free-threshing hexaploid wheat[4]. Since hexaploid wheat evolved in Asia, it is difficult to envision European spelt to be the progenitor of free-threshing hexaploid wheat. The discovery of spelt in Iran[11] and a few other locations in Asia seemed to resolve this dilemma.

The non-free-threshing phenotype in wild and domesticated emmer is primarily determined by the recessive allele *q* on the long arm of chromosome 5A[12] and the dominant alleles at the *Tg-A1* and *Tg-B1* loci on the short arms of wheat chromosomes 2A and 2B, respectively[13-15]. The *q* gene is a homologue of the *Arabidopsis thaliana APETALA2* gene family[16]. Homozygosity for *q* produces speltoid spikes with tenacious glumes, whereas the presence of *Q* (*AP2L5*) reduces plant height and spike length, and generates square-head spikes with softer glumes (due to glumes taking on the developmental identity of lemmas)[17,18]. The domesticated *Q* allele originated by a mutation in the microRNA172 (miR172) binding site located in the last exon[18]. The mutation elevates *AP2L5* transcript levels in immature spikes[18].

The *Tg-A1, Tg-B1,* and *Tg-D1* loci were mapped in populations derived from non-free-threshing × free-threshing wheat crosses [13,15,19,20], but the effort did not result in the isolation of the gene. We followed an alternative approach. We first developed *Ae. tauschii* 2D DS lines in the genetic background of bread wheat cv Chinese Spring (CS)[5] from synthetic wheat produced from four different *Ae. tauschii* accessions[10]. Since the genetic background of the DS lines was nearly isogenic with that of CS, only genes and markers located on chromosome 2D were expected to segregate in the cross of a DS line with CS. We used DS line DSAt5403(CS2D)[5] (abbreviated as $DS2D_3$) in low-resolution mapping of *Tg-D1*[5].

Here, we deployed DSAt5402(CS2D)[5] (abbreviated as $DS2D_2$) to generate a high-density map of the *Tg-D1* region. We identified a candidate gene and validated it by (1) overexpressing it in bread wheat, (2) characterizing nonsense ethyl methanesulfonate (EMS) mutations of the 2A and 2B *Tg1* homoeoalleles in durum wheat, and (3) knocking out the candidate gene in DSAt5405(CS2D)[5] (henceforth $DS2D_5$), using the CRISPR/Cas9 editing technology. We profiled RNA-Seq gene expression in developing spikes in durum wheat and bread wheat and analyzed single-cell expression of the candidate gene with single-molecule fluoprescence *in situ* hybridization (smFISH). We analyzed evidence of past selection in the vicinity of the 2A-, 2B-, and 2D-loci, proposed the potential mechanism of evolution of free-threshing tetraploid wheat, and revisited the evolution of hexaploid wheat.

Genes orthologous to our candidate gene were independently identified as the liguleless *ELIGULUM-A* (*ELI-A*) in barley[21] and reduced height *Rht8* in bread wheat[22,23]. We provided evidence that free-threshing, liguleless, reduced height, and other traits are pleiotropic effects of mutations at same or orthologous loci.

Results

Identification of the *Tg1* candidate gene. We developed an F_2 map of the *Tg-D1* region utilizing 2,849 F_2 plants from the cross CS × $DS2D_2$ (**Supplementary Fig. 1**). We selected 95 F_2 plants (**Supplementary Fig. 1**) for high-resolution mapping and placed *Tg-D1* within a region flanked by gene *AET2Gv20108000* and DNA marker *Aet_25293813*. *Tg-D1* co-segregated with a marker derived from gene *AET2Gv20109500* in all 95 families (**Supplementary Data 1**). We therefore considered the *AET2Gv20109500* gene as a candidate for *Tg-D1*.

We used glume tenacity scores in the population of 95 $F_{2:3}$ families (**Supplementary Data 1**) to determine the dominance interaction between the CS *tg-D1* and *Ae. tauschii Tg-D1* alleles by testing the H_0 hypothesis: The glume tenacity score of the heterozygote=glume tenacity score of the mid-parent of CS and $DS2D_2$. As p=0.10 (quadratic contrast, N=95) (**Supplementary Data 2**), which was above the significance threshold α=0.025, we accepted H_0

and concluded that the *Ae. tauschii* and wheat 2D-alleles interacted largely in an additive manner.

***AET2Gv20109500* and its homologues in CS**. *AET2Gv20109500* was 2,871 bp long and had four alternative transcripts in Aet v4.0[24]. Transcript *AET2Gv20109500.1* was the longest, 2,788 bp. No gene homologous to *AET2Gv20109500* was on pseudomolecule 2D in the CS genome sequence assemblies IWGSC RefSeq v1.0[25] and v2.1[26] but a 99.7% identical gene was on an unassigned scaffold. As other assemblies of the CS genome sequence varied in the number of genes homologous to *AET2Gv20109500* on pseudomolecule 2D, we isolated bacterial artificial chromosome (BAC) clone 83J08 from a CS 2DS BAC library[27] with *AET2Gv20109500* primers and sequenced it (deposited at https://doi.org/10.25338/B8R658). The BAC clone contained four 14-kb sequences in a tandem (**Supplementary Fig. 1f)**, each sequence containing a single gene >99% identical to *AET2Gv20109500*. The same four-gene tandem was present in CS genome assembly CS-CAU[28]. We therefore assumed that the CS 2D-locus consists of four genes.

We found an open reading frame (ORF) of 2,427 bp (**Supplementary Fig. 2**) in *AET2Gv20109500*. As the gene had no intron the ORF was equivalent to the coding sequence (CDS) of the gene. A CDS of this length was annotated in CS IWGSC RefSeq v2.1 and CS-CAU. We used the first nucleotide of this CDS as the +1 position in a description of its sequence.

In BAC clone 83J08, the CDS of each of the four genes differed from the CDS of *AET2Gv20109500* by the same six synonymous mutations. In addition, the three proximal copies differed from *AET2Gv20109500* in positions 1,649 and 1,650 bp (**Supplementary Fig. 2)** in which sequence CG was mutated into T followed by a single base deletion (henceforth T-) causing a frameshift mutation. The genes with the frameshift mutation had a premature termination codon (PTC) at 1,662 bp.

We found a single homologue of *AET2Gv20109500* on pseudomolecules 2A and 2B in the CS IWGSC RefSeq v1.0, v2.1, and CS-CAU genome assemblies in regions overlapping with the genetically mapped *Tg-A1* and *Tg-B1* loci[5,14,15]. The CDSs of both genes were 2,415 bp long and were 96% and 97% identical to the *AET2Gv20109500* CDS, respectively.

The 5' portion of the *AET2Gv20109500* CDS was repetitive (**Supplementary Fig. 3),** had high GC content (**Supplementary Figs. 2, 4),** and was less conserved than the 3' portion when compared to the CDSs of the CS 2A and 2B genes (**Supplementary Fig. 4**). Two diverged copies missing 800 bp of the 5' CDS sequence were on pseudomolecules 1D and 7D in *Ae. tauschii* and CS (**Supplementary Fig. 4**).

Overexpression in bread wheat cv Fielder. To validate the *Tg1* candidate gene by overexpressing it in wheat, we incorporated a 2,427 bp-construct derived from the *Ae. tauschii AET2Gv20109500* gene sequence and driven by the maize ubiquitin promoter into bread wheat cv. Fielder via *Agrobacterium* transformation. Of ten independent T_0 transgenic plants, we confirmed the presence of a transgene in nine. Of them, eight harbored full-length (FL) CDS and one harbored a partial CDS missing 108 bp in the 5' region (PL). The CDS was absent from FL5, which we used as a control.

We grew 10 to 13 T_1 plants per family in the greenhouse. Five of the families segregated. In all five families, progeny plants with the transgene were less threshable than siblings devoid of a transgene (**Fig. 2a**), which is what we expected if the candidate gene was *Tg1*. We considered the co-segregation of the threshing phenotype with the presence of a transgene in

progeny plants as evidence of transgene expression but did not pursue molecular verification of transgene expression. Progeny plants harboring the transgene had more lax spikes than progeny without the transgene (**Fig. 2b, Supplementary Fig. 5**) and headed about three days earlier (**Supplementary Fig. 6**), but they did not differ in plant height and number of tillers with spikes (**Supplementary Figs. 7 and 8, Supplementary Data 2)** compared to siblings without a transgene and Fielder.

EMS nonsense mutations in durum wheat cv Kronos. To pursue further the validation process, we searched the Kronos EMS Targeting Induced Local Lesions In Genomes (TILLING) database[29] for mutations at the 2A- and 2B-homoeologues. We identified a nonsense mutation in the tilling line Kronos3628 at position 1562 for the former gene (henceforth line 1) and in tilling line Kronos2314 at position 1374 for the latter gene (henceforth line 2). We backcrossed the two lines twice to Kronos and crossed a BC_1F_1 plant of line 1 with a BC_2F_1 plant of line 2 to produce a double mutant (henceforth line 3).

In line 1, BC_2F_2 plants homozygous for the mutation were more threshable (see **Supplementary Data 2 for *p* values)** than plants homozygous for the WT allele or heterozygous (**Fig. 2c).** We found no significant difference among the three genotypes in line 2 (**Fig. 2c**). We could not quantify threshability of plants homozygous for both mutations (line 3) as they were nearly sterile. As plants homozygous for one mutation while disregarding the genotype of the other gene in line 3 were fertile, we used these genotypes to quantify threshability in line 3. For both genes, plants homozygous for a mutation were more threshable than their heterozygous and WT-homozygous siblings, although the difference was statistically significant only for the mutation in the 2B-gene (**Fig. 2d**). The fact that loss-of-function mutations in the 2A- and 2B-genes produced plants more threshable than the parental Kronos was consistent with the candidate gene being *Tg1*.

In addition, the mutants had pleiotropic effects on other traits, such as the development of leaf ligule, spike density, plant height, and plant support structures. The mutation in the 2A-gene made spikes more compact (lines 1 and 3) (**Supplementary Figs. 9, 10)** and mutation in the 2B-gene reduced plant height (lines 2 and 3) (**Supplementary Fig. 11, see Supplementary Data 2 for *P* values**). Neither mutation affected heading date or the number of spikes per plant (**Supplementary Figs. 12, 13**). Plants homozygous for both mutations (line 3) were liguleless (**Fig. 2e)**, short, lodging, and had droopy leaves (**Fig. 2f)**.

CRISPR/Cas9 gene editing in $DS2D_5$. The third validation approach we pursued was loss-of-function mutations generated with the CRISPR/Cas9 genome editing technology. We designed *Tg-D1*-specific gRNA for four sites within the 5' region of the *AET2Gv20109500* gene sequence (**Supplementary Fig. 2, Supplementary Data 4**), multiplexed them in pairs in the binary *Agrobacterium* transformation vector JD633[30], and transformed DS line $DS2D_5$.

We grew T_0 plants, harvested seeds from individual spikes, and grew and phenotyped 185 T_1 greenhouse-grown plants derived from 22 T_0 plants (**Supplementary Data 5**). All 22 plants were either homozygous or segregated for the construct in T_1 generation (**Supplementary Data 5**). Mean threshability of free-threshing CS was 0.99 with a 99% confidence interval (CI) of 0.98 to 1.0, whereas that of the non-free-threshing $DS2D_5$ was 0.37 with the 99% CI 0.25 to 0.50 (**Supplementary Data 5**). The density distribution of threshability was bimodal among the 185 T_1 plants (**Fig. 3a**). The threshability scores of 67 plants were within or below the $DS2D_5$ CI, and those plants were non-free-threshing. The scores of 52 plants were within the CS CI, and

those were free-threshing. The remaining T_1 plants were intermediate (**Fig. 3a).** Plant height in the T_1 population ranged from 34.3 to 117.0 cm (**Fig. 3b**) and was negatively correlated with threshability (r = -0.219, for P value see **Fig. 3d**). Spike length ranged from 4.0 to 10.3 cm (**Fig. 3c**) and was also negatively correlated with threshability (r= -0.691**).** The shape of free-threshing spikes was square-head, club, or compactoid whereas the non-free-threshing spikes were speltoid, like those of $DS2D_5$ (**Fig. 3c**). The average number of seeds per spikelet varied from 0.0 to 3.8 seeds per spikelet and was also negatively correlated with threshability (r= -0.238**)**. The number of seeds per spikelet also correlated with spike length. The number of days to flowering ranged from 71 to 104 days, but the trait did not correlate with threshability.

The 5' region of *AET2Gv20109500* where the gRNA sites were located had high GC content and contained many repeats (**Supplementary Figs. 2, 3**), which interfered with PCR sequencing of T_1 plants. We therefore employed RNA-Seq to sequence three free-threshing CRISPR/Cas9 mutants and parental $DS2D_5$. We grew T_2 progeny of free-threshing T_1 plants GH67316, GH67336, GH67379 (**Supplementary Data 5**) (henceforth families CRISPR-16, -17, and -6, respectively) and parental $DS2D_5$. For completeness, we also included plants of CS, $DS2D_2$, and $DS2D_3$. We isolated mRNAs from spikes 8-15 mm long, generated RNA-Seq read populations, and mapped the reads to the CS genome sequences CS-CAU and IWGSC RefSeq v2.1.

The *Tg-D1* reads in the three edited families shared a single base deletion located 3 bp upstream of the PAM site, the expected site of DNA cleavage by Cas9[31], in the gRNA #90 site. The reads of $DS2D_5$ had the WT sequence (**Supplementary Figs. 14)**. There was no difference between the mutants and the parental DS line at the remaining three gRNA sites. Threshability of the T_2 families was significantly higher and plants and spikes were significantly shorter than those of $DS2D_5$ (**Supplementary Table 1**). Thus, the changes in threshability, plant height, and spike length correlated with the frameshift mutation at gRNA site #90 in the three T_2 families.

Thus, we obtained three independent lines of evidence consistently showing that *AET2Gv20109500* was equivalent to *Tg1.* We therefore concluded that *AET2Gv20109500* was the *Ae. tauschii Tg1* gene and will refer to it as such.

Gene expression analysis in edited lines. To compare expression of the three CS *Tg1* homoeologous loci*,* we analyzed RNA-Seq reads (**see previous section**) mapped to the CS-CAU and independently to IWGSC RefSeq v2.1 genome sequences. In all seven lines and both mapping approaches, expression of the CS 2A-gene was about threefold higher than that of the CS 2B-gene (**Supplementary Fig. 15a, b**). Expression of the *Ae. tauschii Tg1* gene in $DS2D_5$ and expression of CS 2B-gene was similar (**Supplementary Fig. 15a, b**).

Next, we exploited the fact that the CS-CAU genome sequence contained all four 2D-genes and used reads mapped to CS-CAU to compare expression of the CS WT and mutant genes. There were approximately equal numbers of WT and mutant reads mapped to the 2D genes (p=0.72, paired t-test, N=5) (**Supplementary Fig. 15a)**, which indicated that both WT and mutant genes were expressed in CS.

As some of the phenotypic effects of CRISPR/Cas9 mutations in *Tg1* and mutations of the wheat *AP2L5* (*Q*) gene[17,18] were similar, we were compelled to address the possibility of interaction between *Tg1* and *AP2-like* (*AP2L*) genes. The number of reads mapped to the *Q* gene on chromosome 5A in mutants and $DS2D_5$ were similar (**Supplementary Table 2**). We found additional *AP2L* genes on chromosomes of homoeologous groups 1, 2, 5, and 7, with an orphan on chromosome 6B in the CS genome. The numbers of reads mapped to these genes in mutants

and $DS2D_5$ also were similar, except for the *AP2L-2B* locus, which in edited lines was slightly more expressed (P=0.04) than in $DS2D_5$. We therefore concluded that *Tg1* did not interact with *AP2L* genes.

Finally, we searched the RNA-Seq reads from 60,971 genes mapped to CS-CAU for differentially expressed genes (DEGs) in the CRISPR/Cas9 mutant families compared to the parental $DS2D_5$ plants. The three mutant families clustered together in a PCA plot (**Supplementary Fig. 16a**). We therefore combined them in our search. We found 145 DEGs in the CRISPR vs $DS2D_5$ comparison using the >log2-fold change statistic[32] (**Supplementary Fig. 16b)**. One gene (TraesCSC1B01G619600) among these 145 DEGs encoded a protein related to hormone biology, but 10 genes encoded proteins related to cell wall biology (highlighted in red in column A of **Supplementary Data 6**). We reasoned that DEGs related to threshability should also be differentially expressed in DS lines compared to CS. We found 898 DEGs in that comparison (**Supplementary Fig. 16c)**. We then searched for genes downregulated in edited lines compared to $DS2D_5$ and simultaneously upregulated in DS lines compared to CS or *vice versa,* upregulated in the former comparison and downregulated in the latter one. We found 15 such genes (**Supplementary Fig. 16d,** highlighted yellow in column B in **Supplementary Data 6**. Most of them were related to plant defense. Of the ten genes related to cell-wall biology, one gene, TraesCSC3D01G729600, annotated as cell wall-associated receptor kinase 5-like was among this subset. Another cell wall-related gene, TraesCSC2D01G140400, annotated as cell wall-associated receptor kinase galacturonan-binding-like, had the required expression pattern, but was not included among the 15 because its expression was zero in all 12 DS plants. Finding cell wall-related genes among DEGs accompanying *Tg-D1* CRISPR/Cas9 mutations is consistent with the importance of *Tg1* for sclerenchyma development and cell wall biogenesis described in the next section.

Developmental and spatial expression patterns. To preliminary characterize developmental expression of *Tg1*, we performed quantitative reverse transcription PCR (qRT-PCR) using *Tg-D1* specific primers and RNA from eight tissues in CS and $DS2D_2$. The highest signal was in the earliest development stages we analyzed (Zadoks[33] stage Z32), at which the immature spikes were about 5 to 8 mm long (**Supplementary Fig. 17**). Expression was generally higher (**see Supplementary Data 2 for *p* values)** in CS than in $DS2D_2$, possibly due to the greater number of 2D genes in CS than in $DS2D_2$.

To characterize developmental expression in immature spikes in a greater detail, we compared transcripts per million (TPM) of the 2A- and 2B-homoeologues in developing spikes of durum wheat cv Kronos at Waddington (W)[34] developmental stages W1.0 to W3.5. Expression peaked at W3.0 (**Supplementary Fig. 18a**), at which the immature spikes were about 2 mm long.

To visualize spatial distribution of transcription, we performed smFISH on sections of developping Kronos spikes at W1.5, W2.5, and W3.5 stages (**Fig. 4a, Supplementary Figs. 18b, c)** using probes derived from the Kronos *Tg1* homoeologues. We detected a strong signal at leaf primordia, the center of the spike, the cortex tissue below the developing spikelets, and nodes in developing stems below the spike (**Fig. 4a, Supplementary Figs. 18b, c**).

In the developing spike at W3.5, the *Tg1* hybridization signal was strong in the central part of the spike, where future vascular tissue will form at later stages, but was not detected in the early glume primordia (**Fig. 4a**). In contrast, the *Q* gene was strongly expressed in both the glume and lemma primordia (**Fig. 4b**), which is consistent with the lack of interaction between

the *Tg1* and *Q* (**Supplementary Table 2**). At a slightly more advanced stage, a low level of *Tg1* expression was detected in the glumes (**Fig. 4c**). We unfortunately do not have hybridization data for later stages of spikelet development. Since both glumes and lemmas are modified leaves, expression in leaf primordia provided indirect information about the potential spatial expression of *Tg1* in more mature glumes. In a transverse section of sheath and leaf primordia, the hybridization signal was strong in the mid rib, around the main vascular bundles and at the borders of the lamina tissue (**Fig. 4d**). Vascular bundles showed particularly intense expression in the mestome sheath surrounding the xylem and phloem and in the bundle sheath extension (BSE) cells, between the vascular bundle and the epidermis (**Fig. 4e**, **Supplementary Fig. 18d**). Most cell types where Kronos *Tg1* homoeologues were preferentially expressed belonged to sclerenchymatous tissues, in which cells formed strong secondary cell walls (sclerenchymatous tissues associated with support functions). These expression profiles were consistent with the drooping leaves observed in the Kronos double *tg-A1, tg-B1* mutant. We also detected *Tg1* expression in preligule cells (**Fig. 4f**), which was consistent with the absence of ligules in the Kronos double *tg-A1, tg-B1* mutant (both phenotypes are described in **Fig. 2**).

Expression of wild emmer *Tg-B1* in durum wheat cv Svevo. To explore regulation of threshability at the tetraploid level, we examined the effects of wild emmer *Tg-B1* on threshability and glume anatomy by introgressing a 21 Mb region including a *Tg-B1* gene from the wild emmer acc. Zavitan into the free-threshing durum wheat cv Svevo[35]. In the BC_4F_2 segregating population, threshability was significantly lower in plants homozygous for the wild emmer *Tg-B1* allele compared to sister lines homozygous for the Svevo allele (**Fig. 5a, b).** Reduced threshability was accompanied by greatly altered glume anatomy. NILs with Zavitan *Tg-B1* had more extensive sclerenchymatous tissue than NILs with the Svevo *tg-B1* (**Fig. 5c**).

To find if the differences in threshability between the two sets of NILs were accompanied by differences in expression of the 2B-gene, we quantified RNA-Seq reads in NIL-65, one of the *Tg-B1* NILs above, and Svevo at three spike developmental stages. Based on the >log2-fold change statistic followed by *t*-test, expression of the 2B-gene was significantly lower in Svevo than in NIL-65 in all three developmental stages (**Supplementary Fig. 19**). Expression of the 2A-gene did not differ, which was expected as the two lines were isogenic at the 2A-gene. Thus, low threshability, extensive glume sclerenchyma, and high levels of expression of the wild emmer 2B-gene were associated.

Haplotype variation. We analyzed *Tg1* CDS haplotypes in the genome sequences of 130 accessions of *Ae. tauschii* from lineage L2[8] and 959 accessions of free-threshing hexaploid wheat (**Fig. 6, Supplementary Data 7**). All *Ae. tauschii* accessions had the sequence CG at the CDS 1,649 and 1,650 bp site. We detected no other fixed polymorphism between *Ae. tauschii* and wheat. Twelve of the 130 *Ae. tauschii* accessions had haplotype 26 (**Fig. 6a**), which was identical to haplotype 27 present in CS except for the 1,649 and 1,650 bp site (**Fig. 6a**). Haplotype 26 is therefore an *Ae. tauschii* precursor of wheat haplotype 27. These 12 *Ae. tauschii* accessions were collected in Azerbaijan and Caspian Iran (**Fig. 6b**). We expanded the length of the compared sequence to 2,000 bp upstream and 2,000 bp downstream of *Tg1* for these 12 accessions to render the comparisons more discerning. *Ae. tauschii* haplotypes in accessions TA2470, TA2471, TA2474, and TA1581 were most closely related to the expanded CS haplotype (red stars in **Fig. 6b**). Accessions TA2470 and TA2471 were collected in Iranian province Mazandaran and accessions TA2474 and TA1581 were collected in Iranian province

Gilan (Guilan in **Fig. 6b**). We then asked if the *Tg-D1* locus was multigenic in these four *Ae. tauschii* accessions, using the depths of sequencing reads[8]. First, we estimated sequencing depth for single-copy genes. Using these as a reference, *Tg1* read depths were 9X, 18X, 21X, and 22X in TA2470, TA2471, TA2474, and TA1581, respectively. The former read depth was within a 99% CI for a single-copy gene, but the latter three were outside of it. We therefore concluded that *Tg1* was a single-copy locus in TA2470 but was most likely comprised of two gene copies in TA2471, TA2474, and TA1581.

Among the 959 hexaploid wheat accessions, the 1,649 and 1,650 bp sequence was available for 156 bread wheat accessions (**Fig. 6a**). Most wheat accessions had genes with both the CG and T- sequences (CG, T-) on chromosome 2D, which was apparent as a double C/T peak in their Sanger electropherograms (**Supplementary Fig. 20**). Two accessions had only gene(s) with the T- sequence (haplotype 31) and one (PI 628252) had only gene(s) with the CG sequence (haplotype 28) (**Supplementary Fig. 21**). PI 628252 was non-free-threshing (**Supplementary Fig. 21**). Observations summarized in **Supplementary Fig. 21** suggest that the PI 628252 haplotype was a reversion to the WT caused by CNV within the multigene locus.

We also analyzed *Tg-D1* haplotypes in spelt, Georgian brittle-rachis macha wheat (*T. aestivum* ssp. *macha*), and *T. aestivum* ssp. *vavilovii*, which is a non-free-threshing spelt-like wheat with elongated rachillas (**Supplementary Data 8**). All accessions of European spelt, two accessions of macha, and one accession of Vavilov wheat we analyzed had only the T- genes, indicating that they were derived from free-threshing hexaploid wheat.

Asian spelt accessions were polymorphic. Five accessions had the (CG, T-) sequence, like bread wheat. They included spelt from Azerbaijan, Tadjikistan, Afghanistan, and Iranian spelt accessions 417a and 405a (**Supplementary Data 8**). Three accessions, PI 367202 from Afghanistan and acc. 407a and 77d from the Isfahan province in Iran, had only genes with the CG sequence (**Supplementary Data 8**). As no Asian spelt analyzed in this study had only the T- sequence, European and Asian spelt likely originated separately.

We also performed haplotype analyses for the 2A-gene in diploid, tetraploid, and hexaploid wheat (**Supplementary Data 9**) and 2B-gene in tetraploid and hexaploid wheat (**Supplementary Data 10**). No mutation appeared to be fixed among the CDS haplotypes.

Past selection at *Tg-A1*, *Tg-B1*, and *Tg-D1*. We computed nucleotide diversity π[36] within 0 to 50 Mb of *Tg1* homoeologues on pseudomolecules 2A and 2D and 10 to 60 Mb on pseudomolecule 2B in wild emmer, domesticated emmer, durum wheat, bread wheat, and *Ae. tauschii* using data from Zhou et al.[37]. We also calculated a cross-population composite likelihood ratio (XP-CLR)[38] and fixation index F_{ST}[39] in pairs of taxa.

For chromosome 2A, π was low in the vicinity of the 2A-locus compared to other regions in free-threshing durum wheat but not in wild and domesticated emmer and bread wheat (**Supplementary Fig. 22**). XP-CLR peaks surrounded the 2A-locus in the domesticated emmer-durum wheat and in domesticated emmer-bread wheat comparisons but not in wild emmer-domesticated emmer and durum-bread wheat comparisons (**Supplementary Fig. 23**). These peaks indicated selection in durum wheat and week selection in bread wheat but not in domesticated emmer. F_{ST} values were consistent with this conclusion (**Supplementary Fig. 24**).

For chromosome 2B, π dropped close to zero for 10 Mb on each side of the 2B-locus in durum wheat. This drop of π was absent from wild and domesticated emmer and bread wheat (**Supplementary Fig. 25**). A similar pattern was observed in XP-CLR values, which were high on each side of *Tg1* only in the domesticated emmer-durum wheat comparison (**Supplementary**

Fig. 26). F_{ST} values around *Tg1* were higher in domesticated emmer-durum wheat and domesticated emmer-bread wheat pairs than in wild emmer-domesticated emmer and durum wheat-bread wheat (**Supplementary Fig. 27**).

For chromosome 2D, π in bread wheat, but not in *Ae. tauschii,* dropped to low values around the *Tg1* locus (**Supplementary Fig. 28**). Both XP-CLR and F_{ST} scans in the *Ae. tauschii*-bread wheat comparison were consistent with selection in the vicinity of *Tg1* (**Supplementary Figs. 29 and 30**).

Thus, we detected past selection at the 2A- and 2B-loci in free-threshing durum wheat but not in the non-free-threshing domesticated emmer. We detected strong selection at the 2D-locus but weak selection at the 2A-locus and none at the 2B-locus in free-threshing bread wheat.

Discussion

Structure of the *Tg1* region. Employing a map-based approach, we identified gene *AET2Gv20109500* on *Ae. tauschii* pseudomolecule 2D as a candidate for *Tg1,* the principal genetic determinant of the free-threshing trait in wheat. We validated it by overexpressing it in bread wheat and knocking it out in EMS and CRISPR/Cas9 mutants in durum and bread wheat, respectively.

The *Tg1* locus is near the site of a nested chromosome insertion that produced Triticeae chromosome 2[24,40]. Blocks of telomeric and centromeric satellite DNA are still present in that region[24] causing sequencing difficulties. We encountered a gap in the CS 2D pseudomolecule where *Tg-D1* should have been in the CS IWGSC RefSeq v1.0[25] and v2.1[26] genome assemblies. *Tg-D1* was present on the pseudomolecule in other CS genome assemblies but they differed from each other in the number of genes[28,41,42]. We found a four-gene tandem on a BAC clone. The same four-gene tandem was on pseudomolecule 2D in the CS genome sequence CS-CAU[28]. We therefore concluded that the CS locus contained four genes. Another discrepancy we encountered in the *Tg1* region was scaffold orientation, which was inverted in our *Ae. tauschii* genome assemblies[24,43,44] compared to that suggested by the genetic map. Caution should therefore be exercised in interpreting sequence variation in the *Tg1* region.

***Tg1* has extensive pleiotropic effects**. A search for the genetic causes of a liguleless mutation on the short arm of barley chromosome 2H identified *ELI-A* as the candidate gene[21]. *ELI-A* and *Tg1* are in the same regions of the 2H and 2D pseudomolecules and have similar sequences. We confirmed functional equivalence of the genes by showing that a double nonsense *Tg-A1* and *Tg-B1* mutant of durum wheat was liguleless, like the barley mutant. Expression of both *ELI-A* and *Tg1* effects sclerenchymatous tissues. We also detected *Tg1* transcripts in a ligule primordium. Thus, free-threshing and liguleless are pleiotropic effects of mutations of orthologous genes. Another example of pleiotropy associated with genetic variation at *Tg1* is reduced plant height. Two independent attempts were reported to isolate the reduced height gene *Rht8* located on wheat chromosome 2D[22,23]. The sequence, location, and structure of the *Rht8* candidate locus indicated that *Rht8* is a variant of the wheat free-threshing *tg-D1* locus. CRISPR/Cas9 nonsense mutations of the *Ae. tauschii Tg1* gene produced plants with reduced height, in addition to other pleotropic effects. Likewise durum wheat Kronos nonsense mutants had reduced height. In total, we described pleiotropic effects of *Tg1* mutations on plant height, spike length, spike density, spike shape, seed fertility, lodging, leaf and glume rigidity, sclerenchyma development, and ligule development.

Molecular and biological functions of *Tg1*. A protein encoded by *Tg1* contains an RNase H-like domain in its second half, based on predictions we made using InterProScan 5[45,46]. Similar observations were reported in the previous studies[21-23]. Genes homologous to *Tg1* are widespread among vascular plants and were annotated as coding for a protein of unknown function, RNase H, or as transposase-like. The product of the gene was localized in the nucleus[23], but it is uncertain if that is due to its RNase H-like function or if the gene has a gene regulatory function.

Our smFISH localization of *Tg1* transcripts revealed that the gene is primarily expressed in vascular tissues in developing spikes, stem nodes, and leaves, and is preferentially expressed in the mestome sheath and the sclerenchyma flanking the main vascular bundles. Anatomical studies of the barley *ELI-A* mutant showed among other effects reduced thickness of cell walls in the mutant[21]. Our anatomical comparison of glume cross-sections in durum wheat isogenic lines, one with wild emmer non-free-threshing *Tg-B1* allele and the other with durum wheat free-threshing *tg-B1* allele, revealed a dramatic difference in the amount of schlerenchymatous cells in the middle of the glume. Wheat *Tg1*, like the barley orthologue, is expressed in the leaf mid rib and margins. The barley study proposed that the gene defined tissue boundaries[21].

Analyses of gene expression in *Rht8* mutants led to the suggestion that the locus regulates GA biosynthesis[23], specifically that it represses *GA3ox2* transcription and promotes expression of *AP2/ERF*[22]. We failed to find DEGs related to GA biosynthesis or GA responsive DEGs among 145 DEGs detected in immature spikes of CRISPR/Cas9 mutants compared to the parental $DS2D_5$. We detected one ABA biosynthesis gene but ten genes for cell wall biogenesis or cell wall-related processes among the 145 DEGs, which is consistent with the importance of *Tg1* for cell-wall related processes. We cannot exclude the possibility that the lack of agreement between our results and those of the *Rht8* studies is due to DEG searches in different organs: immature spikes here but stem elongation zones in the *Rht8* studies.

As we pointed out in Introduction, non-free-threshing wild emmer and *Ae. tauschii* differ from free-threshing wheat in the location of the abscission zone in mature spikes. In wild emmer and *Ae. tauschii*, the abscission zone develops respectively in rachis either above or below the spikelet attachment[9], but no abscission zone develops between mature glumes and rachis; mature glumes remain permanently attached to the rachis and hold the spikelet together during spikelet dispersal. In free-threshing wheat and in intermediate non-free-threshing phenotypes such as those imparted by *Tg1* on synthetic wheat and DS lines, the glume contains less sclerenchyma and its attachment to the rachis is fragile, leading to the disintegration of the spikelet during threshing.

***Tg1* and evolution of free-threshing wheat**. The free-threshing trait must have evolved twice during wheat evolution. The first time, it evolved at the tetraploid level by mutations at the *q*, *Tg-A1,* and *Tg-B1* loci. As synthetic wheat produced from free-threshing tetraploid wheat × *Ae. tauschii* hybrids is non-free-threshing[10], nascent hexaploid wheat must have been non-free-threshing. Thus, the free-threshing trait must have evolved the second time at the hexaploid level. We found that the free-threshing *tg-D1* locus harbored at least one gene with a frameshift mutation in all bread wheat cultivars we examined. The sole exception was Iranian acc. PI 628252, which had only the WT gene(s), and we found it to be non-free-threshing. Thus, the free-threshing phenotype in hexaploid wheat is always accompanied by the presence of one or more mutated genes. It needs to be determined if the expression of mutant protein is functionally significant for the free-threshing phenotype.

The DS lines we used here were isogenic with CS and each possessed the free-threshing *Q* allele on chromosome 5A. Yet, all DS lines were non-free-threshing and all had speltoid spikes. This fact makes it very unlikely that the mutation of *q* to *Q* was sufficient to produce free-threshing, square-head bread wheat. That almost certainly required mutations at both the *q* and *Tg-D1* loci. CRISPR/Cas9 frameshift mutations of the *Ae. tauschii Tg1* locus converted spikes of $DS2D_5$ from non-free-threshing into free-threshing and from speltoid into square-head or compactoid, thus confirming this hypothesis.

We detected selection centered on both 2A- and 2B-loci in free-threshing durum wheat but not in non-free-threshing domesticated emmer. In free-threshing bread wheat, we detected weak selection centered on the 2A-locus, none on the 2B-locus, and strong on the 2D-locus. This selection pattern is inconsistent with domesticated emmer being the tetraploid parent of bread wheat[4]. If that were the case, we should have detected selection footprints in bread wheat at both 2A and 2B-loci, like in durum wheat. The selection footprints in bread wheat are consistent with the tetraploid parent of bread wheat already being free-threshing.

Despite evidence for selection centered on the 2A- and 2B-loci in free-threshing durum wheat we failed to find loss-of-function mutations in the CDSs of either gene at both tetraploid and hexaploid levels. The detrimental pleiotropic effects of the nonsense mutations of the 2A- and 2B-genes in durum wheat cv Kronos make it unlikely that free-threshing tetraploid wheat could have evolved by loss-of-function mutations. We suggest that the free-threshing trait evolved by *cis*-regulation mutations reducing transcript levels. This hypothesis is consistent with the following observations. CDS of *AET2Gv20109500* translates into the same amino acid sequence as the WT gene at the *tg-D1* multigene locus. Overexpressing *AET2Gv20109500* is therefore the same as overexpressing the wheat WT gene. The overexpression of the *AET2Gv20109500* in transgenic plants converted free-threshing cv Fielder into non-free-threshing. Equally illuminating were comparisons of NILs harboring a wild emmer *Tg-B1* allele with sibling NILs harboring the durum wheat allele. The presence of a wild emmer allele in isogenic durum wheat genetic background increased the level of transcription and converted the free-threshing durum phenotype into a non-free-threshing phenotype.

Spelt and its role in wheat evolution. European spelt was once considered to be the non-free-threshing progenitor of bread wheat[4]. MacKey[47] pointed out that European spelt resembles plants that appeared in the progeny of domesticated emmer × bread wheat hybrids and hypothesized that spelt was derived from bread wheat, rather than being its ancestor. Experimental evidence obtained more recently is fully consistent with his hypothesis[48,5,37]. Here, we sequenced the 2D-locus in 10 accessions of European spelt. All had the same T- mutation we found in free-threshing hexaploid wheat, providing further evidence in favor of his hypothesis. Moreover, we showed that other forms of hexaploid wheat with primitive traits, such as *T. aestivum* ssp. *macha* and ssp. *vavilovii,* also contained the T- sequence and also were likely derived from free-threshing wheat.

We previously placed the origin of hexaploid wheat into southwestern Caspian Iran[5]. Our search in *Ae. tauschii* for a precursor for the wheat *tg-D1* haplotype present in bread wheat confirmed that and further narrowed the putative place of hexaploid wheat origin into eastern Gilan and western Mazandaran. To assess the positions of spelt in hexaploid wheat evolution in Iran, we analyzed the *Tg-D1* haplotypes in four accessions of Iranian spelt in addition to four accessions of spelt collected in Azerbaijan, Tadjikistan, and Afghanistan. Five of these accessions, including Iranian spelt acc. 417a and 405a, had the free-threshing haplotype (CG, T-)

but three, Afghan acc PI 367202 and Iranian acc. 407a and 77d contained a haplotype with only the WT gene(s). The possession of the WT haplotype at the *Tg-D1* locus is a necessary but not sufficient condition for an ancestral position. Accessions 407a and 77d were respectively collected in the Bakhtiari province and neighboring Isfahan province[49], which are geographically removed from the putative area of hexaploid wheat origin in Casian Iran. Moreover, we cannot exclude that the WT haplotype encountered in acc. 407a and 77d is not a reversion generated by CNV, like that encountered in non-free-threshing Iranian bread wheat acc. PI 628252. This possibility is real, as spelt 417a and 405a harboring the (CG, T-) haplotype also were collected in Bakhtiari province[49]. More work is clearly needed before the ancestral position of Iranian spelt in bread wheat evolution is accepted.

Methods

Genetic mapping. We used markers *XBF201235*, *AET2Gv20102000*, *XCA658378,* and *XBE488719* to genotype 2,849 F_2 plants from the cross CS × $DS2D_2$. They were genotyped with the KASP assay[50] by LGC Biosearch Technologies on contractual basis. Based on the map, we selected 95 F_2 plants from the CS × $DS2D_2$ F_2 population, grew $F_{2:3}$ families comprised of 16 to 34 F_3 plants per family in the greenhouse, and assessed their glume tenacity by manually evaluating resistance of glumes to deformation on a 1 to 5 scale, with 1 being most resistant and 5 the least (to parallel the threshability quantifications). The measurements were conducted using a randomized, double-blind design. Families with low mean and low standard deviation in the F_3 generation were genotyped as *Tg-D1* homozygous in F_2, those with high mean and low standard deviation in the F_3 generation were genotyped as *tg-D1* homozygous in F_2, and those with medium mean and high standard deviation (due to segregation) in the F_3 generation were genotyped as heterozygous in F_2**.** To confirm the phenotyping accuracy, we grew 91 of the 95 $F_{2:3}$ families in the field and determined glume tenacity in 19 to 44 mature F_3 plants per family. We used CS and $DS2D_2$ spikes as controls in each phenotyping session.

Threshability. We quantified threshability by rolling a 2.4-kg muller over a single mature spike in a threshing box. The shape of spikes in CRISPR/Cas9 edited plants varied from speltoid to compactoid. To minimize the confounding effects of this variation on the assay we removed the bottom and top of the spike and used only the middle section comprised of 10 to14 spikelets for the assay. The surface of the muller and threshing box were covered with corrugated rubber. The muller was pushed horizontally without application of vertical force, liberating seeds from the spike during rolling. Threshability was expressed as the number of liberated seeds/total number of seeds in the spike or its section. The length of the path that the muller traveled over a spike determined the threshing intensity. To standardize threshing intensity across different batches, we adjusted the length of the path that the muller traveled over a spike for each phenotyping session. Importantly, the measurement was relative and adjusted with a standard relevant to the analyzed population only. For example, threshability of the transgenic plants was relative only to that of bread wheat cv Fielder, and threshability of Kronos mutant plants was relative only to durum wheat cv Kronos. Consequently, threshability cannot be compared between populations from different experiments. The measurements were conducted using a randomized, double-blind design.

Marker development and genotyping for mapping. We developed and validated 19 chromosome 2D-specific molecular markers for mapping. We designed primers with Primer3Plus[51,52] at http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi. The primers are listed in **Supplementary Data 4**.

Wmc112 was an SSR marker from the Wheat Microsatellite Consortium [53] (https://graingenes.org/cgi-bin/GG3/report.cgi?class=reference;id=14340). Marker AET2Gv20109100 was an SSR marker developed by us from the AET2Gv20109100 gene. We genotyped the SSR markers by PCR followed by product sizing by gel electrophoresis or on an ABI 3730XL Capillary Electrophoresis Genetic Analyzer (Applied Biosystems). The remaining 17 markers were SNP or indel markers.

BAC clone 83J08. The French Plant Genomic Resource Centre (INRAE-CNRGV, http://cnrgv.toulouse.inrae.fr/) isolated the CS BAC clone 83J08 from a 2DS chromosome arm library[27] and sequenced it on the Pacific BioSciences (PacBio) High-Fidelity (HiFi) sequencing platform on a contractual basis. The mean size of reads was 15.72 kb, the read accuracy was >99%, and coverage was 500X. We assembled the BAC sequence from the HiFi reads using Hifiasm[54] with the default parameters. We manually ordered contigs based on a CS BioNano optical map[26].

Transgenic plants. Synbio Technologies, Inc synthesized DNA molecules of *AET2Gv20109500* CDS on a contractual basis. We used Gateway cloning[55] to transfer the vectors into *Agrobacterium* for plant transformation. We used pLC41 (Japan Tobacco, Inc.) as the destination vector and maize ubiquitin gene promoter[56,57] upstream of the target gene on pLC41.

The UC Davis Plant Transformation Facility (https://ptf.ucdavis.edu) inoculated free-threshing hexaploid wheat cv Fielder with *Agrobacterium* by using a Japan Tobacco, Inc. proprietary technology licensed to UC Davis and selected transgenic plants on a hygromycin-containing tissue-culture medium. We obtained ten transgenic lines and determined the presence of the transgene by PCR in the T_0 and T_1 generations, with one primer located within the gene sequence and the paired primer located on the plasmid downstream of the gene. The primers can be found in **Supplementary Data 4**. Since we obtained phenotypic data consistent with transgenes being expressed, we did not pursue molecular proof for transgene expression.

Kronos mutants. We searched the TILLING population[29] database of durum wheat cv Kronos EMS-treated M_4 mutant lines for mutations in *TrturKRN2A01G009670* and *TrturKRN2B01G014080.* To screen progeny of mutants we selected for mutations, we PCR-amplified the target genomic DNA and sequenced the amplicons on an ABI 3730XL analyzer (Applied Biosystems).

CRISPR mutants. We designed four gRNAs (#52, #58, #90, #94) within the 5' end of the *AET2Gv20109500* gene using a CRISPR gRNA design portal[58]. We list the sequences of the CRISPR gRNAs, the adjacent protospacer adjacent motifs (PAM), and their locations in the gene in **Supplementary Data 4 and Supplementary Fig. 2**. For wheat transformation, we used the binary vector JD633 (Addgene Plasmid #160393). This vector contains a hygromycin cassette to select transformed plants, a GRF-GIF cassette to enhance regeneration, a wheat codon optimized Cas9 cassette and two AarI restriction sites that allow cloning one or multiple gRNAs under a single TaU6 promoter. The four CRISPR gRNAs were cloned in two pairs (Pair 1: gRNAs #52

and #94; Pair 2: gRNAs #58 and #90) using the tRNA-based multiplexing approach {https://doi.org/10.1073/pnas.1420294112}. We performed cloning of the pairs of gRNAs with the Golden Gate method following a protocol developed by the Daniel Voytas group {https://doi.org/10.1105/tpc.16.00922}. We validated the vector sequences by Sanger sequencing. The two binary vectors containing the pairs of gRNAs were incorporated by electroporation into an agrobacterium strain AGL1 carrying a ternary vector for enhanced virulence {https://doi.org/10.1104/pp.19.00767}. Agrobacterium cultures were grown separately and then combined in an equal proportion to transform two-week old immature embryos of $DS2D_5$ plants following a previously published protocol (https://doi.org/10.1007/978-1-4939-1695-5_15). The CS DS line harbored the *Ae. tauschii* ssp. *strangulata* 2D disome from synthetic wheat RL5405[10]. We backcrossed RL5405 six times to the CS genetic background[5] and expect the line to be nearly isogenic with CS. We grew 33 independent transgenic plants in the greenhouse, harvested three mature spikes from each T_0 plant, planted ten T_1 seeds per spike in the greenhouse, and collected data on days to anthesis, plant height, spike length, the number of seeds per spikelet, and threshability, from 182 plants derived from 22 T_0 plants and 28 individual spikes. We simultaneously grew and collected data from CS and $DS2D_5$.

Quantitative RT-PCR. We collected eight different tissues (**Supplementary Data 3**) from CS and $DS2D_2$ for qRT-PCR. We froze tissues in liquid nitrogen immediately after harvest, stored them at -80°C, and isolated RNA using the Applied Biosystems MagMAX Plant RNA Isolation Kit. We determined RNA concentration and quality with the Implen NanoPhotometer P 360. In addition to the DNase treatment included in the RNA isolation kit, we included another round of DNase treatment with Promega RQ1 RNase-Free DNase to maximally remove residual DNA. We then performed single-stranded complementary DNA (cDNA) synthesis using Applied Biosystems High-Capacity cDNA Reverse Transcription Kit with 1 µg of total RNA.

We performed qRT-PCR reactions on an Applied Biosystems StepOnePlus Real-Time PCR System with the QuantiNova SYBR Green PCR Kit, using about 100 ng of cDNA per reaction. We used the Comparative Cycle Threshold (C_T) method[59] with ACTIN transcript level as the reference. We obtained ACTIN cDNA primers from Debernardi et al.[18], while we designed the *AET2Gv20109500* chromosome 2D-specific primers (**Supplementary Data 4**). We estimated PCR efficiencies using calibration curves with five concentrations that were fourfold apart (1:1, 1:4, 1:16, 1:64, and 1:256). The PCR efficiencies of *ACTIN* cDNA and *AET2Gv20109500* primers were 87. 5% and 91.8%, respectively.

We used two to three biological replicates for each of the eight tissues, with four technical replicates per biological replicate, and averaged the technical replicates if variance was <1. For each biological replicate, we used an RNA sample without reverse transcription as a negative control in qPCR to ensure that samples were free of DNA contamination. We performed melting-curve analysis after each reaction to ensure that only the target sequences were amplified. We defined samples with C_T values >35 and variances >1 among technical replicates as too low to be detected.

Dominance. We investigated the interaction (dominance) between the $DS2D_2$ *Tg-D1* and CS *tg*-D1 alleles in the population of 95 $F_{2:3}$ families from the CS × $DS2D_2$ cross**.** We inferred plant genotypes from segregation at the *AET2Gv20109500* locus and used glume tenacity scores (**Supplementary Data 1)** as phenotypes. We analyzed the data with analysis of variance (ANOVA), using the 95 glume tenacity scores as variables.

Haplotype analyses. *Tg1* haplotype analyses were based on DNA and cDNA sequences. We either generated the sequences or obtained the sequences from the following sources: (1) We obtained about 10X genome coverage of *Ae. tauschii* whole genome shotgun (WGS) Illumina reads of 130 *Ae. tauschii* accessions from the *Ae. tauschii* pangenome project[8]. We mapped reads to the *Ae. tauschii* reference genome Aet v4.0[24] using BWA-MEM[60] and detected SNPs and indels with SAMtools and BCFtools[61]. (2) We sequenced a 512 bp-region (CDS 1,378 bp to 1,889 bp) surrounding the *AET2Gv20109500* CDS 1,649-1,650 bp mutation site in wheat 2D genes for 102 hexaploid wheat accessions with Sanger sequencing technology. We amplified the region by PCR and sequenced the amplicons with the ABI 3730XL analyzer (Applied Biosystems) using the manufacturer's protocol. (3) We downloaded bread wheat RNA-Seq data produced from spike RNAs of 90 bread wheat accessions[62] from the National Center for Biotechnology Information (NCBI) short-read archive (accession no. SRP091625). We aligned FASTQ files to CS IWGSC RefSeq v1.0[25] using HISAT2[63]. We generated BAM files and used them for SNP calling with the HISAT2 and GATK pipeline (https://gatk.broadinstitute.org/hc/en-us). (4) He et al. [50] generated exome sequences of 811 hexaploid and tetraploid wheat accessions. We obtained the SNP data from the CS IWGSC RefSeq v1.0 browser[25]. (5) We generated the *AET2Gv20109500* CS 2D sequence from BAC clone 83J08 and obtained the *TraesCS2A02G059900* and *TraesCS2B02G073600* sequences from CS IWGSC RefSeq v1.0[25]. (6) Zhou et al.[37] generated Illumina whole-genome-shotgun (WGS) sequences with 6X genome coverage of 384 accessions from various *Triticum* taxa. We downloaded the genotype file in VCF format from https://bigd.big.ac.cn/gvm with ID GVM000082.

Sequence coverage depth of *Ae. tauschii* accessions TA1581, TA2470, TA2471, and TA2474. We mapped clean whole-genome-shotgun Illumina reads of *Ae. tauschii* acc. TA1581, TA2470, TA2471, and TA2474[8] to the *Ae. tauschii* genome sequence Aet v5.0[43] using BWA-MEM with default parameters[64], filtered the aligned reads using SAMtools[65] using mapping quality of 50, and removed duplicated sequences with SAMtools. We then computed the read coverage depth of single-copy genes *AET2Gv20028600*, *AET2Gv20079900*, *AET2Gv20132500*, and *AET2Gv20120100* with BEDTools[66] in the sequences of the CS A, B, and D subgenomes[26] and Aet v4.0 and v5.0 genome sequences[24,43]. For each accession, we estimated the mean and 99% CI of read depth for the four genes and computed mean read depth from the estimates.

RNA-Seq in CRISPR/Cas9 mutans, CS, and *Ae. tauschii* DS lines. We collected immature spikes about 0.8 to 1.5 cm long from CS and $DS2D_2$, $DS2D_3$, and $DS2D_5$[5]. We also collected immature spikes from T_2 families derived from free-threshing CRISPR/Cas9 T_1 plants GH67316, GH67336, and GH67379 (**Supplementary Data 5**)**.** We removed immature spikes from two or three tillers per plant and combined them. We sampled three plants per line. The tissues were flash frozen in liquid nitrogen and stored at -80C until RNA isolation.

We isolated total RNA using the RNeasy Plant mini kit (Part #74904) with on-column DNase treatment according to manufacturer's instructions (Qiagen, Redwood City, CA). We confirmed RNA integrity and quality with the Plant RNA 6000 Nano Kit (Part# 5067-1511) for the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA).

Sequencing libraries were constructed and sequenced by Novogene Inc (Sacramento, CA, USA) on a contractual basis. They purified messenger RNA from total RNA using oligo-T attached to magnetic beads. They synthesized the first strand of cDNA using random hexamer primers and the second strand of cDNA using dTTP and dUTP as primers. They then repaired ends, ligated adapters, and size selected and purified the libraries. They sized the libraries with Qubit and real-time PCR and pooled and sequenced the quantified libraries on the Illumina platform. They converted images into reads using CASAVA (Illumina Inc.) base-recognition software and stored raw reads in FASTQ format. They filtered the raw data by removing reads containing adaptors, >10% Ns, and low quality bases (base quality less than 5 in >50% bases of the read)[67]. The number of clean bases per library ranged from 8.25 to 6.05 Gb, with an error rate of 0.01 and Q20 scores ranging from 98.75 to 97.8.

Novogene Inc. aligned reads to the CS IWGSC RefSeq v2.1 [26] genome sequence using HISAT2[68]. They quantified expression of each gene in a sample as the fragments per kilobase of transcript sequence per million bp (FPKM)[69]. This measure took into consideration sequencing depth and gene length[68].

We independently aligned clean reads to the CS genome sequence CS-CAU[28] using HISAT2 [68]. Gene-level read counts for each sample were obtained with featureCounts using the parameters "-C -B"[70]. Gene expression levels were subsequently normalized and quantified (TPM). Read alignments on a gene across all samples were visualized using the Integrative Genomics Viewer (IGV)[71].

A total of 1,031,887,526 RNA-Seq fragments were mapped to CS IWGSC RefSeq v2.1[26]. We computed Spearman correlations between the samples to assess their homogeneity and removed one CRISPR/Cas9 sample from the dataset. The number of bp per sample ranged among the remaining 23 samples from 47,889,158 in CRISPR_17_2 to 34,615,869 in CRISPR_6_1. The total average number of bp uniquely mapped per CRISPR sample was similar (P=0.93, two tailed t-test) to the average number of reads uniquely mapped per $DS2D_5$ (38,697,818 bp). We used a total of 106,914 wheat genes in mRNA fragment mapping. After filtering genes that had fragments mapped in only two samples, we were left with 60,969 genes.

To identify DEGs, we mapped reads onto the HC gene models in CS-CAU genome sequence. We filtered out genes with extremely low expression (non-zero counts, in ≤2 samples across all libraries). We then searched for DEGs focusing on the comparison of (1) mutants vs parental $DS2D_5$ and (2) all DS lines vs CS. For each comparison, we used a simple two-group design, disregarding family structure. We assessed differential expression using a **negative binomial generalized linear model (GLM)** implemented in the **DESeq2** package in R[72]**.** The model appropriately accounts for the discrete nature of sequencing data and the mean–variance relationship characteristic of RNA-seq experiments. For each gene g and sample i, the observed read count Y_{gi} was assumed to follow a negative binomial distribution with mean μ_{gi} and gene-specific dispersion parameter α_g. Differences in sequencing depth across samples were normalized using DESeq2 **median-of-ratios size factor estimation.** We modeled the expected mean count as:

$$\log(\mu_{gi}) = \log(s_i) + \beta_{g,0} + \beta_{g,1} \cdot I(\text{group}_i),$$

where s_i denotes the size factor for sample i, $\beta_{g,0}$ is the baseline expression level, and $\beta_{g,1}$ represents the log2 fold change between two compared groups.

For both comparisons, we estimated gene-specific dispersion parameters using a DESeq2 empirical Bayes approach, which borrows information across genes to stabilize variance

estimates in the presence of limited replication. For each gene, Wald statistics were used to assess evidence against the null hypothesis of no differential expression between groups.

We performed principal component analysis (PCA) on variance-stabilized transformed counts to assess sample relationships and overall expression patterns. Volcano plots were generated to visualize the relationship between effect size (log2FC) and statistical significance.

RNA-Seq of durum wheat Svevo and wild emmer acc. Zavitan transcripts. We grew and isolated RNA from durum wheat cv. Svevo and BC_4F_2 near-isogenic line NIL-65 in the Svevo background. NIL-65 harbored a small chromosome segment including the *Tg-B1* locus derived from wild emmer acc. Zavitan.

We grew Svevo and NIL-65 plants in the greenhouse with a 16h light/8h dark cycle at 23°C and collected developing spikes at three stages: (1) booting stage (Early), (2) when the top awn appeared (Mid), and (3) heading stage (Late). We collected four biological replicates at each developmental stage and genotype. Tissues were immediately frozen in liquid nitrogen and stored at -80°. We isolated RNA using a modified phenol-chloroform protocol based on [73] and assessed RNA integrity with the Agilent 2200 TapeStation system, prepared sequencing libraries with Illumina TruSeq Stranded mRNA, following the manufacturer's instructions, and sequenced them on an Illumina NovaSeq platform to generate paired-end 150-bp reads. We checked raw sequencing reads for quality and mapped them to the wild emmer (WEW_v2.1, Zavitan) reference genome sequence and quantified transcripts with Salmon[74].

We conducted differential expression analysis with the DESeq2 R package[72]. We imported raw count data from Salmon into DESeq2 to identify DEGs between Svevo and NIL-65 across the three developmental stages. We considered genes with an adjusted *p*-value (padj) < 0.05 and |log2FoldChange| > 1 to be differentially expressed.

Other data analysis tools. We searched for homologs of *AET2Gv20109500* in grass family (taxid:4479) through the NCBI Web BLAST service (BLAST+ 2.10.1)[75] at https://blast.ncbi.nlm.nih.gov/Blast.cgi. We searched for gene annotation of hits in Phytozome v12.1[76] (https://phytozome.jgi.doe.gov/pz/portal.html),. We used ORFfinder[77] at https://www.ncbi.nlm.nih.gov/orffinder/ and Augustus[78] to identify ORFs. We generated the dot plot in **Supplementary Fig. 3** by YASS web server[79] at https://bioinfo.lifl.fr/yass/index.php. We generated the VISTA plot (**Supplementary Fig. 4**) with mVISTA[80] using the LAGAN alignment program[81] at http://genome.lbl.gov/vista/mvista/submit.shtml. We performed searches for wheat homologs of *AET2Gv20109500* with the IWGSC online BLAST tool at https://urgi.versailles.inra.fr/blast_iwgsc/blast.php and the Triticeae Multi-omics Center online BLAST tool at http://202.194.139.32/blast/viroblast.php. We performed significance analyses with R and displayed the results in **Supplementary Data 2.** We constructed the dendogram in **Supplementary Fig. 22c** with NGPhylogeny.fr[82], using default settings.

Legends to Figures

Fig. 1. Spikelet and glume morphology. **a** A mature spikelet of non-free-threshing wild emmer (bottom) and mature spikelets of *Ae. tauschii* (top). In wild emmer, spikes disarticulate at abscission zones located in the rachis distal to the attachment of the spikelet to the rachis. The proximal rachis internode is permanently attached to the spikelet and forms a wedge at the spikelet base facilitating spikelet dispersal. In *Ae. tauschii,* spikes disarticulate at abscission

zones located in the rachis proximal to the attachment of the spikelet to the rachis. The disarticulated spikelets have a barrel shape. **b** Glumes are permanently attached to the top of the proximal rachis internode in the mature spikelet of wild emmer. **c** Glumes are permanently attached to the base of the distal rachis internode in the mature spikelet of *Ae. tauschii.* **d-e** Fragile zones develop at glume bases in free-threshing durum wheat cv Kronos (**d**) and bread wheat cv Chinese Spring (**e**), and glumes readily fall off and spikelets disintegrate during threshing. The white horizontal bar in each image is 5 mm.

Fig. 2. *AET2Gv20109500* overexpression in bread wheat cv Fielder and EMS nonsense mutations in durum wheat cv Kronos. a Threshability of $T_{0:1}$ families transformed with the full length CDS (FL) and partial length CDS missing 108 bp at the 5' end (PL) of *AET2Gv20109500*. In each family, open rectangles show mean threshability of progeny devoid of a transgene and solid rectangles show mean threshability of progeny with a transgene. Only families that segregated and had numbers of progeny sufficient for statistical analysis are shown. Threshability of all families including cv Fielder are at the extreme right. * indicates a significant difference ($\alpha = 0.05$, *t*-test) within a family. The error bars represent the standard errors calculated from least-square means. **b** Spike morphology of Fielder and transgenic progeny with and without a transgene. Scale bars: 1 cm. Palea, lemma, and glume are indicated. **c** Mean threshability (a white number within a column) of Kronos and BC_2F_2 progeny of Kronos line 1 segregating for the 2A-gene nonsense mutation (left panel) and BC_1F_2 progeny of Kronos line 2 segregating for the 2B-gene nonsense mutation (right panel). Mean threshability is given for homozygous wild type (WT-homo), heterozygous (Hetero), and homozygous mutant (Mut-homo) genotypes. Scores with a common letter (above columns) are not significantly different from each other (ANOVA and Tukey's test, α=0.05). Threshability of Kronos is shown as a reference. **d** Mean threshability in the F_2 generation of line 3. Shown is mean threshability of homozygous WT plants (WT-homo), heterozygous (hetero), and homozygous plants for the mutant 2A-gene (Mut-homo) while disregarding the status of the 2B-gene (left panel), and homozygous WT plants (WT-homo), heterozygous (Hetero), and homozygous plants for mutant 2B-gene (Mut-homo), disregarding the status of the 2A-gene (right panel). **e** The ligule and (**f)** plant morphology of Kronos, homozygotes for the 2A- or 2B-gene mutations, and double mutant homozygous for both 2A- and 2B-gene mutations. Scale bars: 1 cm for (**e**) and 10 cm for (**f**).

Fig. 3. CRISPR/Cas9 mutants. a Numbers plants in indicated ranges of threshability scores in the population of 185 T_1 transgenic CRISPR/Cas9 plants. **b** Variation in plant height among T_1 plants. **c.** Variation in spike length and shape among T_1 plants. From the left: CS, $DS2D_5$, and T_1 plants GH67220, GH67221, GH67183, GH67207, GH67298, GH67295, and GH67187 (**Supplementary Data 5**). T1 plants GH67220 and GH67221 have speltoid spike shape, GH67183, GH67207, and GH67298 have a square-head spike shape, and GH67295 and GH67187 have a club and compactoid spike shape, respectively. **d** Correlations between plant height, spike length, the number of seeds per spikelet, days to anthesis, and spike threshability. Correlation coefficients are above the diagonal and their *p*-values are below the diagonal. Significant coefficients (α=0.05) are in yellow.

Fig. 4. **Single-molecule fluorescence in situ hybridization (smFISH) of probes derived from the durum wheat cv Kronos *Tg1* homoeologues to Kronos primordia. a** smFISH to the Kronos developing spike at W3.5. **b** Relative expression of *Q* (blue) and *Tg1* (red) in early glume

and lemma primordia. **c** Expression of *Tg1* in a slightly later stage of spikelet development showing incipient expression in glume primordia. **d** A transverse section at W1.5 showing sheath and two leaf primordia LP1 and LP2 (center is tip of the inflorescence meristem). The squares mark the lamina borders and the oval marks the mid rib of LP2. **e** A transverse LP2 section showing *Tg1* expression in the mestome sheath and the bundle sheath extension (BSE) cells. Manual color coding is based on the known anatomical distribution of these tissues. **f** Expression of *Tg1* in the preligule cells in a longitudinal section of two leaf primordia. Cells are stained with calcofluor white. Data were obtained from[83]. SM: spikelet meristem, FM= floral meristem. Cell walls in **a, b, c,** and **f** are stained with calcofluor. Images for **a, b**, and **f** were obtained using Molecular Cartographer[83] and for **c**, **d**, and **e** from a separate 10X experiment visualized with Xenium Explorer v4.0 (cell segmentation was provided by 10X).

Fig. 5. Effect of introgression of the wild emmer *Tg-B1* allele on glumes in NIL-65 and durum wheat cv Svevo. a Threshed spikes of NIL-65 (left) and Svevo (right). The NIL-65 spikes exhibit lower threshability than Svevo. **b** Mean percentages of threshed seeds per spike in NILs (*n*=11) harboring the wild emmer *Tg-B1* allele compared to sister lines (*n*=15) harboring the Svevo *tg-B1* allele. **c** Transverse sections of glumes (750 µm from the rachis base) from a *Tg-B1* NIL and a *tg-B1* sister line. The *Tg-B1* NIL exhibits more extensive sclerenchyma tissue, contributing to the tough-glume phenotype. For more details visit[35].

Fig. 6. *Ae. tauschii* and wheat haplotypes. a The positions of polymorphic sites and mutation types (S, M, and F for synonymous, missense, or frameshift, respectively) in the 2,427 bp CDS of the *AET2Gv20109500* are indicated at the top. Bases identical to those in the CS reference sequence are green and variant bases are in other colors. Letters specify bases and numbers specify the number of simple sequence repeats (SSR). A slash indicates polymorphism. The symbol (CG, T-), as in CS (haplotype 27) indicates the simultaneous presence of a WT gene and the gene with the T- frameshift. The following code is used for data sources: 1. Illumina reads[8]; 2. Sanger sequence (our data); 3. RNA-Seq[62]; 4. Exome sequence [50]; 5. BAC clone 83J08 sequence (our data). **b** Geographic locations of *Ae. tauschii* accessions with haplotype 26. After expanding the length of the analyzed sequence by a total of 4 kb, haplotypes of four accessions (red stars) appeared most closely related to the wheat haplotype.

Data availability

All data used for the analyses that support the findings of this study are available within the article and its Supplementary files. The BAC 83J08 sequence is available at https://doi.org/10.25338/B8R658 (It was deposited to the datadryad repository and will be publicly available upon the publication of this manuscript. For now, the editor and reviewers can access the data at this temporary link: https://datadryad.org/stash/share/_-39GK2rPTrRf-qZD3onKhOAOjoVudlD7slmrSrUgpA).

Acknowledgements

This work was supported by the National Science Foundation of United States grants IOS-1238231, IOS-1929053, and IOS-2102953 (JDv and M-CL); grant from Agriculture and Food

Research Initiative Competitive Grant 2022-68013-36439 (WheatCAP) from the USDA National Institute of Food and Agriculture, and by the Howard Hughes Medical Institute (JDu). We also acknowledge support from the Israel Science Foundation grants 1137/17, 1134/22, and 2223/22 (AD) and the Henry A. Jastro Graduate Research Award (MX). We thank Dr. Brande Wulf for supplying us with the *Ae. tauschii* Illumina reads. We thank Jennifer Braggs, Matthew J. Milner, and Yong Q. Gu, Crop Improvement & Genetics Research, USDA-ARS, Albany CA USA, who produced RNA samples for RNA-Seq analyses of CRISPR/Cas9 mutants and DS lines.

Author contributions

MX, KRD, M-CL, and JDv conceived the project. MX mapped the gene, developed markers, generated and analyzed the transgenic plants and Kronos mutants, conducted the qRT-PCR tests, and, with LW, analyzed haplotype variation. KRD developed markers, genotyped the mapping population, and constructed maps. M-CL and LW assembled and analyzed the BAC clone sequence. LW analyzed the *Ae. tauschii* Illumina reads and read depths, analyzed wheat genomic sequences, constructed dot-plots, and performed diversity analyses. TZ performed optical map alignments. FH and EA performed past-selection analyses. JDv, KRD, and MX analyzed PI 628252. AD, YLM, and TE generated the RNA-Seq data for Svevo and Zavitan made anatomical comparisons of glumes. LW mapped RNA-Seq reads to CS-CAU and analyzed and quantified reads. JDv analyzed reads mapped to CS IWGSC v2.1. JNY and JDv designed CRISPR genome-specific gRNAs, JMD produced CRISPR/Cas9 transgenic plants, JDv and PEM grew and phenotyped them, and KRD determined the presence of the transgene by PCR. JB, MJM, and YQG isolated RNAs, characterized them, and submitted them to Novogene Inc. YZ and HGM conducted RNA-Seq statistical analyses. JDu and JMD assisted with the development and characterization of transgenics for overexpression. JDu made available the Kronos EMS mutants. He also performed smFISH and its analyses. JDv wrote the first draft of the paper. All authors edited and approved the final draft.

Competing interests

The authors declare no competing interests.

Supplementary Information Supplementary Figures 1 to 30. (Word files)

Description of Additional Supplementary Files

Supplementary Data 1. Graphical genotypes of 95 $F_{2:3}$ families. (Excel file)

Supplementary Data 2. Summary of statistical analyses. (Excel file)

Supplementary Data 3. Stages and tissues used for qRT-PCR. (Excel file)

Supplementary Data 4. Polymerase chain reaction (PCR) primers and gRNA sequences. (Excel file)

Supplementary Data 5. Spike length and shape, spikelet fertility, and spike threshability in 185 T_1 plants. (Excel file)

Supplementary Data 6. FPKM of differentially expressed genes (DEGs) in three families of free-threshing CRISPR/Cas9 mutants, CS, DS2D2, DS2D3, and DS2D5. (Excel file)

Supplementary Data 7. *Tg-D1* haplotypes detected in *Ae. tauschii* and *T. aestivum* accessions. (Excel file)

Supplementary Data 8. *Tg-D1* haplotypes detected in accessions of spelt, macha and Vavilov wheat. (Excel file)

Supplementary Data 9. *Tg-A1* haplotypes in diploid, tetraploid and hexaploid wheat. (Excel file)

Supplementary Data 10. *Tg-B1* haplotypes in accessions of tetraploid and hexaploid wheat. (Excel file)

1 Harlan, J. R. Agricultural origins: centers and noncenters. *Science* **174**, 468-474 (1971).

2 Nesbitt, M. & Samuel, D. in *Hulled Wheats. Promoting the conservation and use of underutilized and neglected crops. 4. Proc. 1st Internatl. Workshop on Hulled wheats.* (eds S. Padulosi, K. Hammer, & J. Heller) 41-100 (International Plant Genetic Resources Institute, Rome, Italy).

3 Kihara, H. Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare* (Japanese). *Agric. and Hort. (Tokyo)* **19**, 13-14 (1944).

4 McFadden, E. S. & Sears, E. R. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* **37**, 81-89, 107-116 (1946).

5 Dvorak, J. *et al.* The origin of spelt and free-threshing hexaploid wheat. *J. Hered.* **103**, 426-441, doi:doi:10.1093/jhered/esr152 (2012).

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

7 Wang, J. R. *et al. Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol.* **198**, 925-937, doi:10.1111/nph.12164 (2013).

8 Gaurav, K. *et al.* Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. *Nat. Biotechnol.* **40**, 422-431, doi:10.1038/s41587-021-01058-4 (2021).

9 Sakuma, S., Salomon, B. & Komatsuda, T. The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. *Plant Cell Physiol.* **52**, 738-749, doi:10.1093/pcp/pcr025 (2011)

10 Kerber, E. R. & Rowland, G. G. Origin of free threshing character in hexaploid wheat. *Canad. J. Genet. Cytol.* **16**, 145-154 (1974).

11 Kuckuck, H. On the findings of *Triticum Spelta* L. in Iran and on the arising of *Trititicum aestivum*-types through crossings of different *Spelta*-types. *Wheat Inf. Serv.* **9-10**, 1-2 (1959).

12 Sears, E. R. Cytogenetic studies with polyploid species of wheat II. *Genetics* **29**, 232-246 (1944).

13 Jantasuriyarat, C., Vales, M. I., Watson, C. J. W. & Riera-Lizarazu, O. Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (*Triticum aestivum* L.). *Theo. Appl. Genet.* **108**, 261-273 (2004).

14 Faris, J. D., Zhang, Z. C. & Chao, S. M. Map-based analysis of the tenacious glume gene *Tg-B1* of wild emmer and its role in wheat domestication. *Gene* **542**, 198-208, doi:10.1016/j.gene.2014.03.034 (2014).

15 Sharma, J. S. *et al.* Genetic analysis of threshability and other spike traits in the evolution of cultivated emmer to fully domesticated durum wheat. *Mol. Genet. Genomics* **294**, 757-771, doi:10.1007/s00438-019-01544-0 (2019).

16 Simons, K. J. *et al.* Molecular characterization of the major wheat domestication gene *Q*. *Genetics* **172**, 547-555 (2006).

17 Debernardi, J. M., Greenwood, J. R., Finnegan, E. J., Jernstedt, J. & Dubcovsky, J. APETALA 2-like genes AP2L2 and Q specify lemma identity and axillary floral meristem development in wheat. *Plant J.* **101**, 171-187, doi:10.1111/tpj.14528 (2020).

18 Debernardi, J. M., Lin, H. Q., Chuck, G., Faris, J. D. & Dubcovsky, J. microRNA172 plays a crucial role in wheat spike morphogenesis and grain threshability. *Development* **144**, 1966-1975, doi:10.1242/dev.146399 (2017).

19 Villareal, R. L., Mujeeb-Kazi, A. & Rajaram, S. Inheritance of threshability in synthetic hexaploid (Triticum turgidum X T. tauschii) by T. aestivum crosses. *Plant Breed.* **115**, 407-409 (1996).

20 Katkout, M. *et al.* QTL analysis of genetic loci affecting domestication-related spike characters in common wheat. *Genes Genet. Syst.* **89**, 121-131, doi:DOI 10.1266/ggs.89.121 (2014).

21 Okagaki, R. J. *et al.* ELIGULUM-A Regulates Lateral Branch and Leaf Development in Barley. *Plant Physiol.* **176**, 2750-2760, doi:10.1104/pp.17.01459 (2018).

22 Chai, L. *et al.* A natural variation in Ribonuclease H-like gene underlies *Rht8* to confer "Green Revolution" trait in wheat. *Mol. Plant.* **15**, 377-380, doi:10.1016/j.molp.2022.01.013 (2022).

23 Xiong, H. *et al.* Cloning and functional characterization of *Rht8*, a "Green Revolution" replacement gene in wheat. *Mol. Plant* **15**, 373-376, doi:10.1016/j.molp.2022.01.014 (2022).

24 Luo, M. C. *et al.* Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* **551**, 498-502, doi:10.1038/nature24486 (2017).

25 IWGSC. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191, doi:ARTN eaar7191 10.1126/science.aar7191 (2018).

26 Zhu, T. *et al.* Optical maps refine the bread wheat *Triticum aestivum* cv Chinese Spring genome assembly. *Plant J.* doi/10.1111/tpj.15289 (2021).

27 Safar, J. *et al.* Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet. Genome Res.* **129**, 211-223, doi:10.1159/000313072 (2010).

28 Wang, Z. *et al.* Near-complete assembly and comprehensive annotation of the wheat Chinese Spring genome. *Mol. Plant* **18**, 892-907, doi:10.1016/j.molp.2025.02.002 (2025).

29 Krasileva, K. V. *et al.* Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. USA.* **114**, E913-E921, doi:10.1073/pnas.1619268114 (2017).

30 Debernardi, J. M. *et al.* A GRF-GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nat. Biotechnol.* **38**, 1274-1279, doi:10.1038/s41587-020-0703-0 (2020).
31 Ceasar, S. A., Rajan, V., Prykhozhij, S. V., Berman, J. N. & Ignacimuthu, S. Insert, remove or replace: A highly advanced genome editing system using CRISPR/Cas9. *Bioch. Biophys. Acta. Mol. Cell. Res.* **1863**, 2333-2344, doi:10.1016/j.bbamcr.2016.06.009 (2016).
32 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
33 Zadoks, J. C., Chang, T. T. & Konzak, C. F. Decimal Code for Growth Stages of Cereals. *Weed Res.* **14**, 415-421, doi:DOI 10.1111/j.1365-3180.1974.tb01084.x (1974).
34 Waddington, S. R., Cartwright, P. M. & Wall, P. C. A Quantitative Scale of Spike Initial and Pistil Development in Barley and Wheat. *Ann Bot-London* **51**, 119-130, doi:DOI 10.1093/oxfordjournals.aob.a086434 (1983).
35 Lev-Mirom, Y. *et al.* Characterization of a major threshability locus in tetraploid wheat. . *bioRxiv* https://doi.org/10.64898/2026.03.30.715257 (2026).
36 Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press, 1987).
37 Zhou, Y. *et al.* Triticum population sequencing provides insights into wheat adaptation. *Nat. Genet.* **52**, 1412-1422, doi:10.1038/s41588-020-00722-w (2020).
38 Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Research* **20**, 393-402, doi:10.1101/gr.100545.109 (2010).
39 Wright, S. Genetical structure of populations. *Nature* **166**, 247-249 (1950).
40 Luo, M. C. *et al.* Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *P Natl Acad Sci USA* **106**, 15780–15785 (2009).
41 Alonge, M., Shumate, A., Puiu, D., Zimin, A. V. & Salzberg, S. L. Chromosome-scale assembly of the bread wheat genome reveals thousands of additional gene copies. *Genetics* **216**, 599-608, doi:10.1534/genetics.120.303501 (2020).
42 Liu, S. *et al.* A telomere-to-telomere genome assembly coupled with multi-omic data provides insights into the evolution of hexaploid bread wheat. *Nat Genet* **57**, 1008-1020, doi:10.1038/s41588-025-02137-x (2025).
43 Wang, L. *et al.* Aegilops tauschii genome assembly Aet v5.0 features greater sequence contiguity and improved annotation. *G3 (Bethesda)* **11**, doi:https://doi.org/10.1093/g3journal/jkab325 (2021).
44 Maratikyathanahalli Srikanta, R. *et al. Aegilops tauschii* genome assembly v6.0 with improved sequence contiguity differentiates assembly errors from genuine differences with the D subgenome of Chinese Spring wheat assembly IWGSC RefSeq v2.1. *G3 (Bethesda)* **15**, doi:10.1093/g3journal/jkaf042 (2025).
45 Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res* **33**, W116-120, doi:10.1093/nar/gki442 (2005).
46 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240, doi:10.1093/bioinformatics/btu031 (2014).
47 Mac Key, J. in *2nd International Wheat Genetics Symposium.* (ed J. M. Mac Key) 237-275 (Hereditas).

48 An, X. L. *et al.* Genetic diversity of European spelt wheat (Triticum aestivum ssp spelta L. em. Thell.) revealed by glutenin subunit variations at the Glu-1 and Glu-3 loci. *Euphytica* **146**, 193-201, doi:10.1007/s10681-005-9002-6 (2005).

49 Kuckuck, H. Experimentelle Untersuchungen zur Entstehung der Kulturweizen. I. Die Variation des iranischen Speltzweizen und seine genetischen Bezeihungen zu *Triticum aestivum* ssp. *vulgare* (Vill., Host) Mac Key, ssp. *spelta* (L.) Thell. und ssp. *macha* (Dek. et Men.) Mac Key mit einem Beitrag zur Genetik des *Spelta*-Komplexes. *Zeit. Pflanzenzuchtg.* **51**, 97-140 (1964).

50 He, F. *et al.* Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat Genet* **51**, 896-+, doi:10.1038/s41588-019-0382-2 (2019).

51 Rozen, S. & Skaletsky, H. J. in *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* (eds S. Krawet, S. Misener, & N.J. Totowa) 365-386 (Humana Press, 2000).

52 Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71-74, doi:10.1093/nar/gkm306 (2007).

53 Somers, D. J., Isaac, P. & Edwards, K. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **109**, 1105-1114 (2004).

54 Cheng, H. Y., Concepcion, G. T., Feng, X. W., Zhang, H. W. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170-175, doi:10.1038/s41592-020-01056-5 (2021).

55 Katzen, F. Gateway (R) recombinational cloning: a biological operating system. *Expert Opin. Drug Dis.* **2**, 571-589, doi:10.1517/17460441.2.4.571 (2007).

56 Christensen, A. H., Sharrock, R. A. & Quail, P. H. Maize Polyubiquitin Genes - Structure, Thermal Perturbation of Expression and Transcript Splicing, and Promoter Activity Following Transfer to Protoplasts by Electroporation. *Plant Mol. Biol.* **18**, 675-689, doi:Doi 10.1007/Bf00020010 (1992).

57 Weeks, J. T., Anderson, O. D. & Blechl, A. E. Rapid Production of Multiple Independent Lines of Fertile Transgenic Wheat (Triticum-Aestivum). *Plant Physiol.* **102**, 1077-1084, doi:DOI 10.1104/pp.102.4.1077 (1993).

58 Smedley, M. A., Hayta, S., Clarke, M. & Harwood, W. A. CRISPR-Cas9 based genome editing in wheat. Current Protocols, 1,. *Current Protocols* **1**, e65., doi:e65. doi: 10.1002/cpz1.65 (2021).

59 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).

60 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (arXiv Prepr arXi:13033997), (2013).

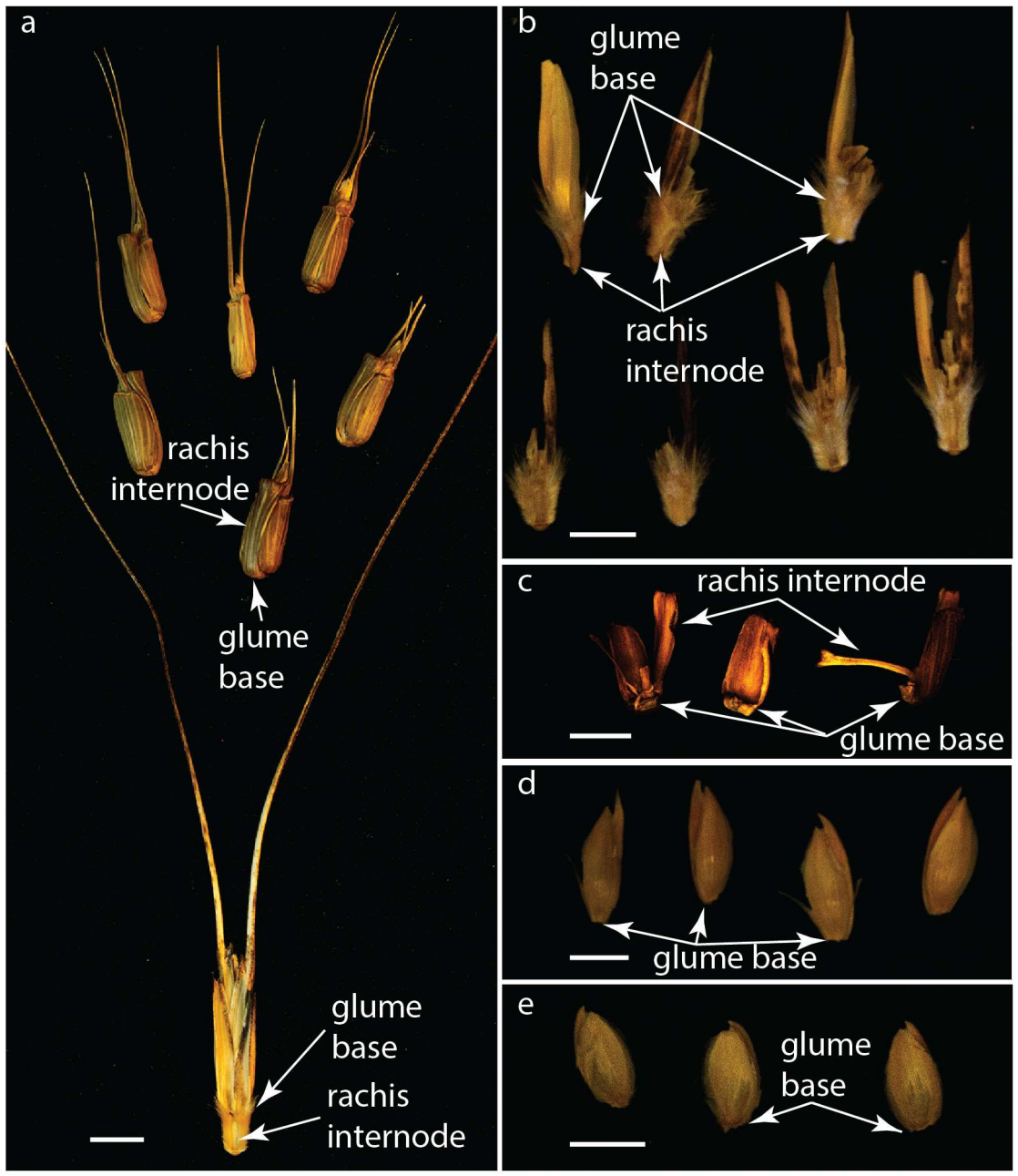
61 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinform.* **27**, 2987-2993, doi:10.1093/bioinformatics/btr509 (2011).

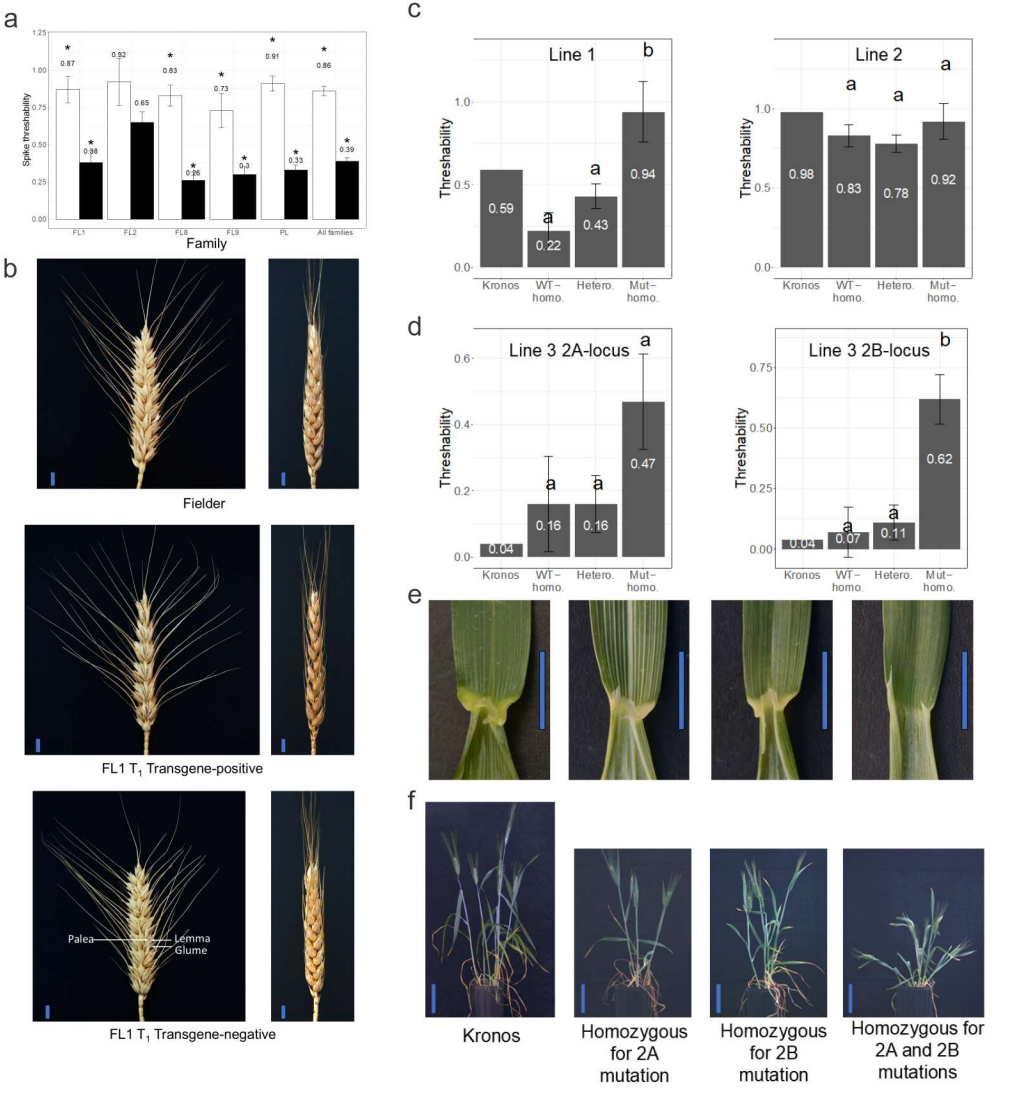
62 Wang, Y. *et al.* Transcriptome Association Identifies Regulators of Wheat Spike Architecture. *Plant Physiol.* **175**, 746-757, doi:10.1104/pp.17.00694 (2017).

63 Kim, D., Landmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357-U121, doi:10.1038/Nmeth.3317 (2015).

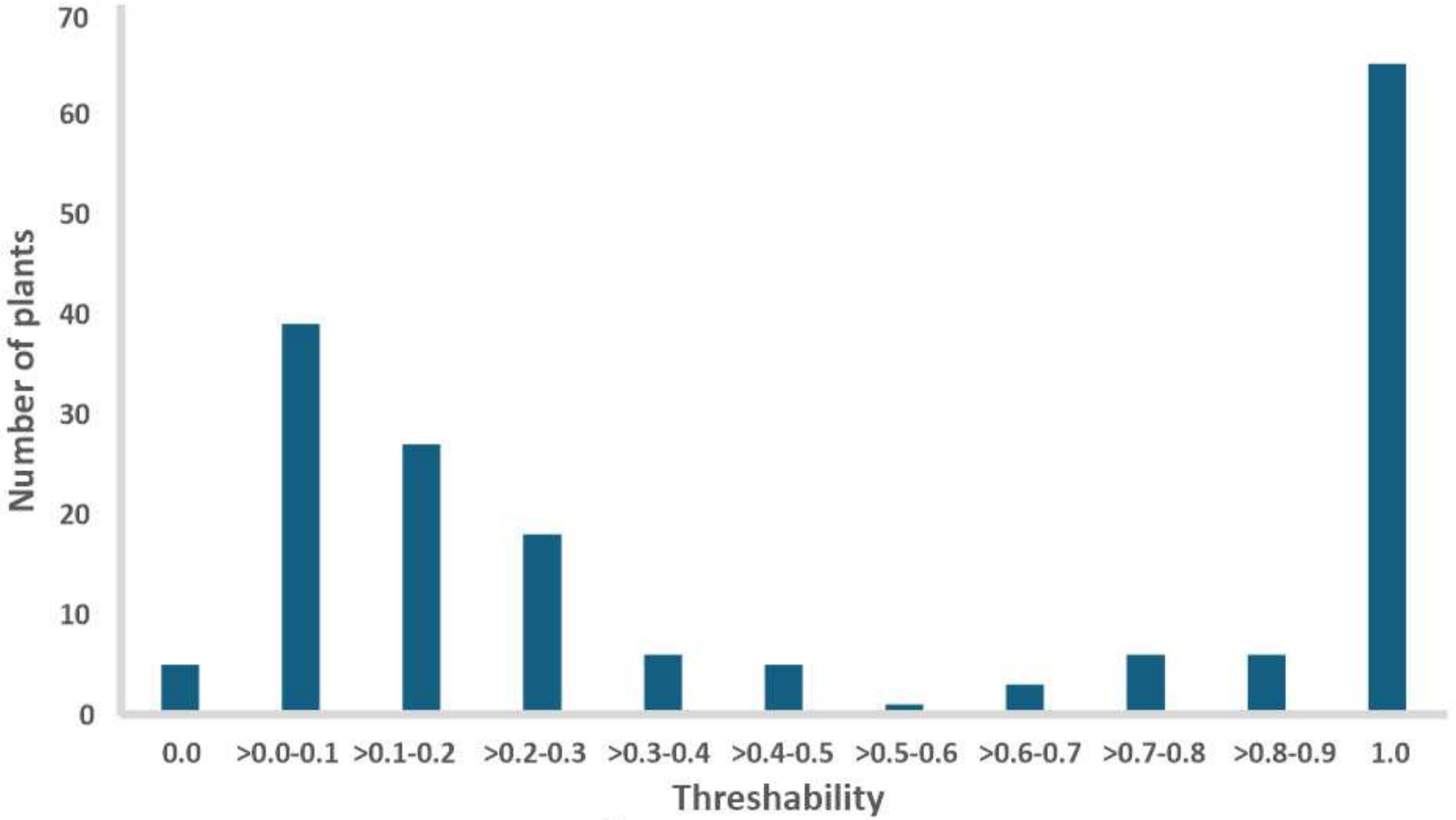
64 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinform.* **26**, 589-595, doi:btp698 [pii] 10.1093/bioinformatics/btp698 (2010).
65 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinform.* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
66 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinform.* **47**, 11 12 11-34, doi:10.1002/0471250953.bi1112s47 (2014).
67 Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131-1139, doi:10.1038/nsmb.2660 (2013).
68 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).
69 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotech.* **28**, 511-U174, doi:10.1038/nbt.1621 (2010).
70 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
71 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
72 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
73 Li, Z. W. & Trick, H. N. Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. *Biotechniques* **38**, 872-+, doi:Doi 10.2144/05386bm05 (2005).
74 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417-+, doi:10.1038/nmeth.4197 (2017).
75 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic logical aligment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
76 Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178-D1186, doi:10.1093/nar/gkr944 (2012).
77 Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* **31**, 28-33, doi:10.1093/nar/gkg033 (2003).
78 Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309-312, doi:10.1093/nar/gkh379 (2004).
79 Noe, L. & Kucherov, G. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res.* **33**, W540-W543, doi:10.1093/nar/gki478 (2005).
80 Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273-W279, doi:10.1093/nar/gkh458 (2004).

81 Brudno, M. *et al.* LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721-731, doi:10.1101/gr.926603 (2003).
82 Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465-469, doi:10.1093/nar/gkn180 (2008).
83 Xu, X. *et al.* Spatial and single-cell expression analyses reveal complex expression domains in early wheat spike development. *Genome Biol.* **26**, 352, doi:10.1186/s13059-025-03811-3 (2025).

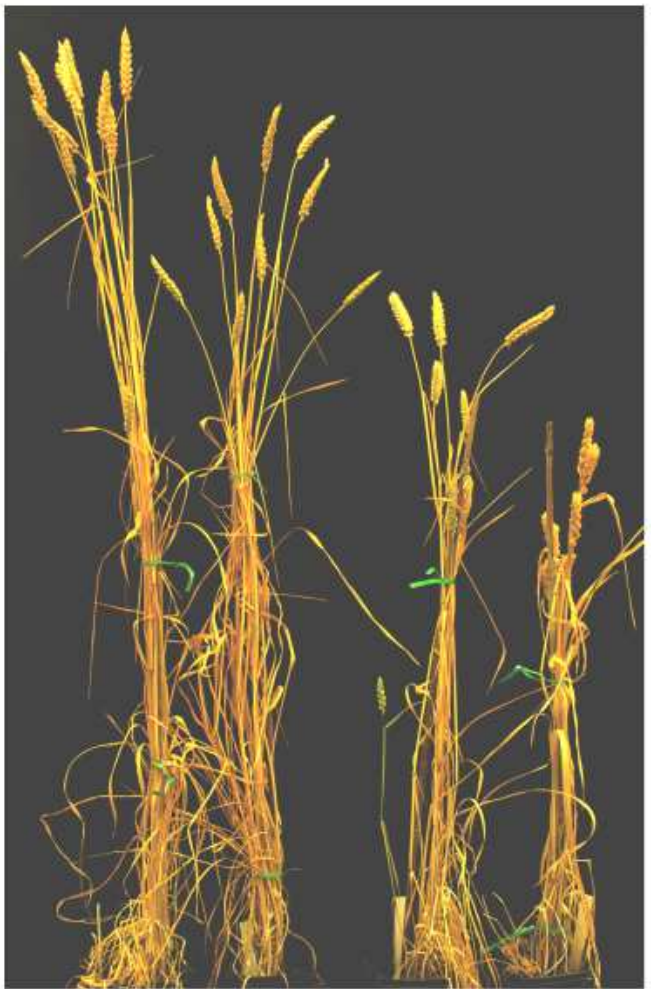

a
rachis
internode
glume
base
glume
base
rachis
internode
b
glume
base
rachis
internode
c
rachis internode
glume base
d
glume base
e
glume
base


a
Spike Threshability
1.25
1.00
0.75
0.50
0.25
0.00
0.87
0.38
0.92
0.65
0.83
0.26
0.73
0.3
0.91
0.33
0.86
0.39
FL1
FL2
FL8
FL9
PL
All families
Family
b
Fielder
FL1 T1 Transgene-positive
Palea
Lemma
Glume
FL1 T1 Transgene-negative
c
Line 1
Threshability
1.0
0.5
0.0
0.59
0.22
0.43
0.94
a
a
b
Kronos
WT−homo.
Hetero.
Mut−homo.
Line 2
0.98
0.83
0.78
0.92
a
a
a
d
Line 3 2A-locus
0.6
0.4
0.2
0.04
0.16
0.16
0.47
a
a
a
Line 3 2B-locus
0.75
0.50
0.25
0.00
0.04
0.07
0.11
0.62
a
a
a
b
e
f
Kronos
Homozygous for 2A mutation
Homozygous for 2B mutation
Homozygous for 2A and 2B mutations

a



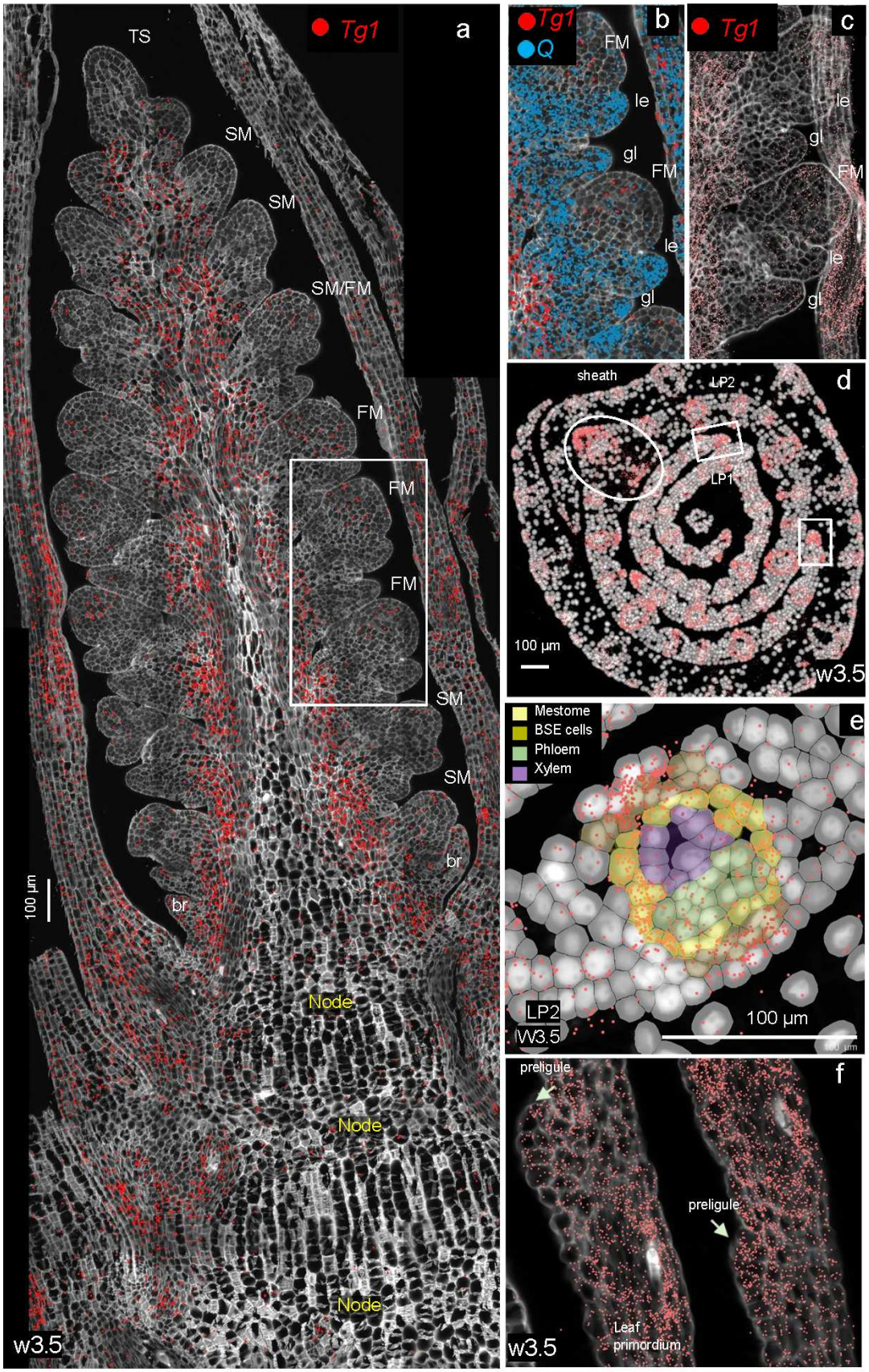
b



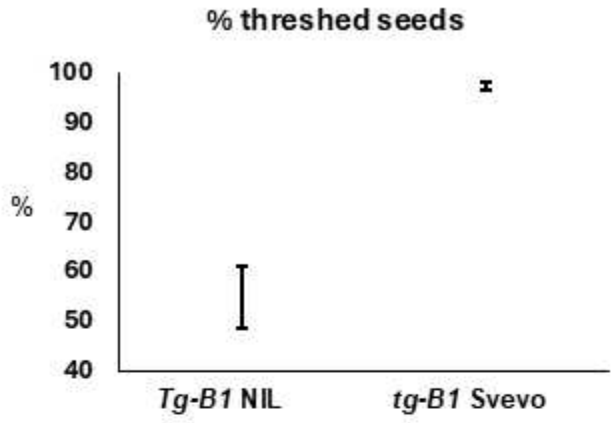
c

d

	Plant height	Spike length	Seeds per spikelet	Days to anthesis	Threshability
Plant height	X	0.367	0.533	0.183	-0.219
Spike length	0.0001	X	0.461	-0.157	-0.691
Seeds per spikelet	0.0001	0.0001	X	0.032	-0.238
Days to anthesis	0.0132	0.0338	0.6672	X	0.024
Threshability	0.0029	0.0001	0.0012	0.7471	X


a
Tg1
TS
SM
SM
SM/FM
FM
FM
FM
SM
SM
br
br
100 µm
Node
Node
Node
w3.5
b
Tg1
Q
FM
le
gl
FM
le
gl
c
Tg1
le
gl
FM
le
gl
d
sheath
LP2
LP1
100 µm
w3.5
e
Mestome
BSE cells
Phloem
Xylem
LP2
W3.5
100 µm
f
preligule
preligule
Leaf primordium
w3.5

Tg-B1 NIL *tg-B1* Svevo



Tg-B1 NIL

tg-B1 Svevo

a

Haplotype	Taxon	Source	Accessions (no.)	60-77 bp	125 bp	376-390 bp	402 bp	480 bp	532 bp	522-542 bp	846 bp	1086 bp	1152 bp	1247 bp	1320 bp	1397 bp	1449 bp	1467 bp	1491 bp	1570 bp	1602 bp	1617 bp	1649-1650 bp	2076 bp	2418 bp
				S	M	S	S	S	M	S	S	S	S	M	S	M	S	S	S	M	S	S	F	S	S
1	*Aegilops tauschii*	1	2	3,4	C	S	G	C	T	7	C	C			C	C	G				G			G	C
2		1	1	4	C	S	G	C	G	7	C	C			C	C	G				G			G	C
3		1	6	3,4	C	S	G	C	G	7	C	C			C	C	G				G			G	C
4		1	1	3	T	S	T	G	G	7,4	C	C		G	C	C	G				G			G	C
5		1	1	3	T	S	G/T	C/T	NA	4	C	C		G	C	C	G				G			G	C
6		1	2	3	T	S	T	C	NA	4	C	C		G	C	C	G				G			G	C
7		1	1	3	T	S	G/T	C	G	4	C	C		G	C	C	G				G			G	C
8		1	4	3	T	S	T	C	G	7,4	C	C		G	C	C	G				G			G	C
9		1	2	3	T	S	G/T	C	G	7,4	C	C		G	C	C	G				G			G	C
10		1	1	3	T	S	G	C	NA	4	C	C		G	C	C	G				G			T	C
11		1	3	3	T	S	G	C	G	7,4	C	C		G	C	C	G				G			T	C
12		1	1	3	T	S	G	C	NA	4	C	C		G	C	C	G				G			G	T
13		1	2	3	T	S	G	C	G	7,4	C	C		G	C	C	G				G			G	C
14		1	2	3	T	S	G	C	G	7,4	C	C			T	C	G				G			G	C
15		1	1	3	T	S	G	C	G	7,4	C	C		G	C/T	C	G				G			G	C
16		1	5	3	T	S	G	C	G	7,4	C	C		G	C	C	G				G			G	C
17		1	1	3	C/T	S	G	C	G	7,4	C	C		G	C	C	G				G			G	C
18		1	5	3	C	5,7	G	C	G	7	C	C		G	C	C	G				G			G	C
19		1	1	3	C	5,4	G	C	G	7	C	C		G	C	C	G				G			G	C
20		1	1	3,2	C	S	G	C	G	7	C	C		G	C	C	G				G			G	C
21		1	52	3	C	S	G	C	G	7	C	C		G	C	C	G				G			G	C
22		1	1	3	C	S	G	C	G	7,9	8	A		G	C	C	G	G	C	C	G	T	A	G	C
23		1	1	3	C	S	G	C	G	7	8	A		G	C	C	G	G	C	C	G	T	A	G	C
24		1	1	3	C	S	G	C	G	7	C	A		G	C	C	G	C/G	C	C	G	T	A	G	C
25		1	20	3	C	S	G	C	G	7	C	A		G	C	T	G	G	C	C	G	T	A	G	C
26		1	12	3	C	S	G	C	G	7	C	A		G	C	C	G	G	C	C	G	T	A	G	C
27	*Triticum aestivum* ssp. *aestivum* (CS)	5	1	3	C	S	G	C	G	7	C	A		G	C	C	G	G	C	C	G	T	A	(CG, T-)	G	C
28	*Triticum aestivum ssp. aestivum*	2	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	C	C	G	T	A	CG	NA	NA
29		2	147	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	C	C	G	T	A	(CG, T-)	NA	NA
30		3	5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	C	C	G/A	T	A	(CG, T-)	NA	NA
31		2	2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	C	C	G	T	A	T-	NA	NA
32		3	16	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	C	C	G	T	A	NA	NA	NA
33		4		NA	NA	NA	NA	NA	NA	NA	C	A	G	C	C	G	G	G	C	C	G	T	A	NA	G	C
34	*Triticum aestivum* ssp. *compactum*	4	784	NA	NA	NA	NA	NA	NA	NA	C	A	G	C	C	G	G	G	C	C	G	T	A	NA	G	C
35	*Triticum aestivum* ssp. *sphaerococcum*	4		NA	NA	NA	NA	NA	NA	NA	C	A	G	C	C	G	G	G	C	C	G	T	A	NA	G	C
36	*Triticum aestivum* ssp. *aestivum*	4	3	NA	NA	NA	NA	NA	NA	NA	C	A	G	C	C	G/A	G	G	C	C	G	T	A	NA	G	C

b

Supplementary Files

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

- SupplementaryData1.xlsx
- SupplementaryData2123022.xlsx
- SupplementaryData3032326.xlsx
- SupplementaryData4032426.xlsx
- Supplementarydata5032026.xlsx
- Supplementarydata6032126.xlsx
- SupplementaryData7033026taushwheat.xlsx
- SupplementaryData8033026nonfree.xlsx
- SupplementaryData9033026.xlsx
- SupplementaryData100330262B.xlsx
- SupplementaryInformationplaintext.docx