

Complete genome and indexed EMS mutant library of einkorn wheat to advance polyploid wheat functional genomics

Shisheng Chen

shisheng.chen@pku-iaas.edu.cn

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang
<https://orcid.org/0000-0002-8617-4356>

Hongna Li

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Shikai Lyu

Qinghai University <https://orcid.org/0000-0001-8019-8704>

Wanyi Hu

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Pengliang An

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Tao Shen

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Lei Hua

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Mengkai Li

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Shams Rehman

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Hensen Li

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Shengliang Cao

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Lintong Song

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Kai Wang

Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Zhongxu Chen

Tcuni Bioscience Inc.

Bosheng Li

Peking University Institute of Advanced Agricultural Sciences <https://orcid.org/0000-0002-1816-7007>

Li Guo

Peking University <https://orcid.org/0000-0001-6100-3481>

Hao Li

Henan University <https://orcid.org/0000-0001-8271-9163>

Xingwang Deng

Peking University <https://orcid.org/0000-0001-8709-1467>

Yongming Chen

Peking University Institute of Advanced Agricultural Sciences <https://orcid.org/0000-0002-2143-3134>

Article

Keywords: einkorn, T2T genome, EMS mutant library, functional genomics, gene redundancy, exome capture

Posted Date: August 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10177678/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.

Abstract

Polyploidy complicates functional genomics in common wheat (*Triticum aestivum* L.) due to homoeologous gene redundancy. Here, we present a telomere-to-telomere (T2T) genome assembly of domesticated einkorn wheat (*Triticum monococcum*) accession PI 306540 and a corresponding indexed ethyl methanesulfonate (EMS) mutant library. The 5.11 Gb assembly fills all gaps, representing the most contiguous A-genome reference to date. We develop a targeted exome capture array for *T. monococcum* and generate an EMS-mutagenized population exhibiting diverse phenotypic variation. Exome sequencing of 3,777 independent mutant lines identified 1,338,432 mutations across 36,787 high-confidence genes, averaging 36.38 mutations per gene. This diploid platform bypasses polyploid redundancy, enabling rapid gene-to-trait associations. We demonstrate its utility by characterizing mutations in four genes (*SR60*, *MLO*, *TmPDS*, and *TmGBSSI*) and identifying causal genes for brown-midrib and yellow-green leaf mutants. Together, the genome and mutant library overcome polyploid redundancy to unlock rapid functional genetics in wheat.

INTRODUCTION

Common wheat originated ~10,000 years ago through natural hybridization and polyploidization of three diploid ancestral species, resulting in a complex genome^{1,2}. As a staple crop providing ~20% of global calories and protein, continuous enhancement of wheat productivity and nutritional quality is essential to meet the demands of a growing human population^{3,4}. However, progress in wheat functional genomics remains constrained relative to other cereals⁵, largely due to limited mutational resources and genomic tools.

Crop improvement often exploits natural or induced genetic variation. Early mutagenesis efforts in wheat began in the 1920s using ionizing radiation⁶, followed in the 1960s by chemical mutagens such as EMS. Unlike radiation, which frequently causes large deletions and chromosomal rearrangements, EMS predominantly induces G/C-to-A/T point mutations that are generally less deleterious and more tractable for functional studies⁷. Over the past several decades, induced mutagenesis has played an essential role in gene discovery, functional analysis, and crop breeding.

Numerous EMS-mutagenized populations have been developed in polyploid wheat. These populations were initially screened using pooled DNA and the endonuclease *CelI* in a method known as Targeting Induced Local Lesions in Genomes (TILLING)⁸. *CelI*-based assays were later replaced by exome capture sequencing⁷, a robust, cost-effective, and high-throughput approach⁹. This strategy identified over 10 million mutations in protein-coding regions across 2,735 mutant lines of tetraploid and hexaploid wheat⁷. Similar resources were subsequently developed in the Chinese hexaploid wheat cultivars Jing411 and KN9204^{10,11}.

Owing to recent polyploidization, most genes in tetraploid and hexaploid wheat are retained as multiple homoeologous copies. Induced mutations in polyploid wheat therefore often fail to manifest phenotypes because of functional redundancy¹². Consequently, achieving loss-of-function phenotypes requires stacking mutations across all homoeologs through extensive crossing, as demonstrated for *SBEII*, *BGI1*, *VRT2*, and *SVP1*¹³⁻¹⁵. These difficulties underscore the value of diploid wheat mutant libraries, which bypass redundancy and enable direct genotype-to-phenotype associations.

T. monococcum ($A^m A^m$) is one of the oldest domesticated cereals and has played a significant role in the emergence of agriculture. As a close relative of *T. urartu* ($A^u A^u$), it can serve as a model for studying traits, genes, and alleles relevant to bread wheat¹⁶. Today, *T. monococcum* is regarded as an important genomic resource for modern wheat improvement, particularly as a source of disease resistance genes, including *Sr21*, *Sr22a/Sr22b*, *Sr35*, *Sr60*, *SrTm4*, *Yr34/Yr48*, *LrTM16*, and *Lr63*¹⁷⁻¹⁹. Three draft genome assemblies of *T. monococcum* have been released, but their contiguity remains limited (contig N50 = 55.4-131.20 Mb), with 211-1,881 gaps^{20,21}. A more contiguous assembly would facilitate gene identification and studies of A-genome evolution.

In this study, we present a T2T genome assembly (5.11 Gb) of cultivated einkorn accession PI 306540, which carries multiple rust resistance genes^{17,22}. We developed an exome capture array for *T. monococcum*, generated an EMS-mutagenized population, and performed whole-exome capture sequencing (WES) on 3,777 mutant lines. The resulting 1,338,432 mutations are publicly accessible (<http://223.85.181.157:8182/>). To demonstrate the utility of this resource, we validated mutant phenotypes for four key genes (*SR60*, *MLO*, *TmPDS*, and *TmGBSSI*) and characterized the causal genes underlying brown-midrib and yellow-green leaf mutants. This work provides a complete genome assembly and a gene-indexed diploid mutant library to accelerate functional genomics and modern wheat improvement.

RESULTS

T2T assembly and annotation of the *T. monococcum* genome

The *T. monococcum* accession PI 306540 was self-pollinated for ten generations to ensure genetic homogeneity (Fig. 1a). To improve the previous v1.0 assembly²⁰, we generated 1.45 Tb of sequencing data (incorporating prior data), comprising 539.01 Gb (103.86×) PacBio HiFi reads, 200.53 Gb (38.64×) ONT ultra-long reads, and 705.53 Gb (135.94×) Hi-C reads (Supplementary Table 1). Using a hybrid assembly strategy², we generated a complete T2T genome assembly of PI 306540 (Fig. 1b, Supplementary Fig. 1, and Supplementary Table 2). The final assembly spans 5.11 Gb with a contig N50 of 729.67 Mb, representing a substantial improvement over the earlier version (contig N50 = 131.20 Mb)²⁰. This assembly closed all 211 gaps present in v1.0 (Fig. 1c) and surpassed all published genomes of *T. urartu*, *T. monococcum*, and other diploid wheat relatives in both contiguity and completeness (Fig. 1d and Supplementary Table 3). The near-perfect mapping rates of the raw sequencing data (99.99% for HiFi, ONT, and Hi-C reads; Supplementary Fig. 1b) supported genome completeness. Additional validation metrics, including a Benchmarking Universal Single-Copy Orthologs (BUSCO) score of 97.30%, a consensus quality value (QV) of 74.40, and a Long Terminal Repeat Assembly Index (LAI) of 21.50 (Supplementary Table 2), further confirmed the high accuracy and completeness of the assembly.

We annotated 39,423 high-confidence genes using transcriptomic data from five tissues across six developmental stages and homologous protein evidence, achieving a BUSCO completeness score of 99.04% (Supplementary Table 2). Repetitive sequence analysis revealed that 4.56 Gb (89.25%) of the genome comprises interspersed repeats, predominantly Gypsy (48.64%) and Copia (22.62%) long terminal repeat (LTR) retrotransposons. Using CENH3 ChIP-seq, we identified all seven centromeres in PI 306540 (Fig. 1e, Supplementary Fig. 2, and Supplementary Table 4), with an average length of 8.91 Mb (Supplementary Table 5). By contrast, the centromeres of two previously sequenced *T. monococcum* accessions, TA299 (wild) and TA10622 (domesticated), are shorter, averaging 5.46 Mb and 5.24 Mb, respectively.

Whole-genome sequence alignment revealed that collinearity is largely conserved among PI 306540, TA10622, and TA299 (Supplementary Fig. 3). Nonetheless, we detected 986 and 9,696 structural variants (SVs) in comparisons of PI 306540 with TA10622 and TA299, respectively. These variants include 19 and 115 large inversions and translocations (>100 kb) in the respective comparisons (Fig. 1f and Supplementary Table 6). To elucidate the evolutionary relationships between *T. monococcum* and other *Triticum* species, we constructed a maximum-likelihood phylogenetic tree based on 818 single-copy genes, using *Oryza sativa* and *Brachypodium distachyon* as outgroups²³. Divergence time estimation indicated that wild and domesticated einkorn separated ~1.72 million years ago (Mya), while *T. monococcum* and *T. urartu* diverged around 5.51 Mya (Fig. 1g).

EMS-induced diploid wheat mutant population reveals diverse phenotypes

To establish a platform for functional genomics and trait discovery in diploid wheat, we generated an EMS-induced mutant population in PI 306540. Approximately 15,000 seeds were treated with 0.45% EMS, yielding 4,835 germinated M₁ plants, of which 3,777 were fertile. A single M₂ progeny from each fertile M₁ plant was advanced via single-seed descent, resulting in a population of 3,777 lines (Supplementary Figs. 4 and 5a-5d). Phenotypic evaluation was conducted during the 2023-2025 growing seasons (Supplementary Figs. 5c and 5d). Systematic field-based assessment throughout the entire developmental cycle revealed extensive phenotypic variation in both vegetative and reproductive traits (Fig. 2 and Supplementary Figs. 5e-5g).

At the seedling stage, mutant lines exhibited diverse alterations in leaf pigmentation, including chlorosis, albinism, longitudinally striped albino-variegated patterns, and anthocyanin overaccumulation (Fig. 2a and Supplementary Fig. 5e). Additional phenotypes included lesion-mimic and necrotic lesions, and leaf morphological changes such as altered width, modified lamina shape, and pronounced rolling (Figs. 2b-2d).

At the reproductive stage, we observed pronounced variation in plant height, from severe dwarfism to extreme tallness (Fig. 2e). We also found substantial variation in spike architecture, including elongated, compact, branched, shortened, and other structurally modified spikes (Figs. 2f and 2g). Whole-plant phenotypes further included altered stem pigmentation, heading date variation, and differences in grain morphology (Figs. 2h-2j and Supplementary Fig. 5g). Collectively, these observations demonstrate that the PI 306540 EMS mutant library encompasses a rich repertoire of phenotypic alterations.

Development of a targeted exome capture array for *T. monococcum*

To address the lack of an exome capture platform for *T. monococcum*, we designed a targeted exome capture array based on the PI 306540 v1.0 annotation (available before our complete T2T assembly). This array comprises 394,086 biotinylated 120-bp oligonucleotide probes distributed across all seven chromosomes, targeting 47.29 Mb of exonic sequences (including untranslated regions (UTRs); Supplementary Fig. 6 and Supplementary Data 1). These probes cover 36,996 protein-coding genes, accounting for 99.16% of the 37,310 high-confidence genes annotated in the PI 306540 v1.0 genome (Supplementary Table 7).

Identification and characterization of induced mutations

Using the custom-designed capture array, we performed whole-exome sequencing on the PI 306540 mutant library comprising 3,777 lines (one plant per line: 1,381 at M₂ and 2,396 at M₃; Supplementary Fig. 5). Sequencing yielded 36.38 Tb of raw data (range: 4.63-34.33 Gb per sample; Supplementary Fig. 7a and Supplementary Table 8), corresponding to an average sequencing depth of ~203.6× over the targeted regions.

After quality filtering and sequence alignment to the PI 306540 reference genome, we identified 1,338,432 high-confidence mutations across the population, including 1,334,096 single nucleotide polymorphisms (SNPs) and 4,336 insertions/deletions (InDels) (Fig. 3a, Supplementary Fig. 7b, and Supplementary Table 9). The average mutation density was 28.30 per kilobase (kb) of targeted sequence, with mutations distributed across 36,787 high-confidence genes (99.44% of captured genes) and an average of 36.38 mutations per gene (Supplementary Table 7). The chromosomal distribution of mutations closely paralleled gene density (Fig. 3a). The average number of mutations per sample was 470.48 for M₂ individuals and 287.44 for M₃ individuals (Supplementary Table 10). Among EMS-induced SNPs, G/C-to-A/T transitions accounted for 97.92%, 97.38%, and 97.64% in the M₂, M₃, and combined populations, respectively (Fig. 3b; Supplementary Table 11). Progressive selfing fixation reduced heterozygosity from 63.64% in M₂ to 38.49% in M₃, while homozygosity increased correspondingly from 36.36% to 61.51% (Fig. 3c; Supplementary Table 10).

We next annotated all high-confidence SNPs relative to the PI 306540 gene annotations (v1.0 and T2T) to assess their genomic distribution and predicted functional impact. As expected for exome-capture sequencing, most variants resided within coding and flanking regulatory regions. In the M₂ population, 85.64% of SNPs mapped to genic regions (exons, introns, UTRs, and splice regions), 13.12% to upstream/downstream regulatory regions, and 1.25% to intergenic regions. The M₃ population exhibited similar proportions: 84.33% genic, 14.39% regulatory, and 1.28% intergenic (Fig. 3d; Supplementary Table 12).

Among SNPs located in genic regions, 64.17% were in exons, 21.30% in introns, 11.83% in UTRs, and 2.70% in splicing regions. Of the exonic SNPs, missense mutations constituted the dominant functional class, accounting for 61.46% of all exonic variants (Supplementary Table 13). High-impact variants, including stop-gained, start-lost, stop-lost, and mis-splicing alleles, collectively accounted for 1.79% (23,888 SNPs) of all variants, whereas synonymous substitutions comprised ~19.93% (Fig. 3e; Supplementary Table 13). The substitution frequency heatmap revealed threonine-to-alanine (Thr-to-Ala) and valine-to-alanine (Val-to-Ala) changes as the most prevalent non-synonymous amino acid changes (Fig. 3f).

Validation of mutations and establishment of a public database

To validate the identified mutations, we randomly selected 201 variants for Sanger sequencing, including 196 SNPs and five InDels located in 43 protein-coding genes across all seven chromosomes (Supplementary Table 14). For each mutant line, genomic DNA extracted from leaf tissue of 8-16 progeny was pooled in equimolar amounts and amplified using gene-specific primers flanking each mutation. Sanger sequencing confirmed the expected mutation at 194 of the 201 loci (96.52%) compared with the PI 306540 wild-type (WT) sequence (Fig. 3g; Supplementary Table 14).

To facilitate community access, we established a publicly accessible database containing the full set of 1,338,432 EMS-induced mutations, available at <http://223.85.181.157:8182/>. The dataset is accessible through a JBrowse interface based on the PI 306540 T2T genome. This platform enables users to retrieve information for mutations of interest, including predicted functional impacts, gene annotations, and nucleotide/amino acid changes.

Validation of the stem rust resistance (Sr) gene *SR60*

The resistance gene *SR60* (*WTK2*) in PI 306540 encodes a protein with two putative kinase domains and confers resistance against multiple *Puccinia graminis* f. sp. *tritici* (*Pgt*) races¹⁸. Although *SR60* was previously identified through map-based cloning and validated by transgenic complementation¹⁸, functional evidence from loss-of-function mutants remained lacking. To address this gap, we

screened the EMS mutant database of PI 306540 and identified 68 independent mutations in *SR60* (PI306540.5A02G00007400; Supplementary Table 15). Among these, a premature stop codon mutation (Q177*; line M1461) was selected for further study. Since PI 306540 carries other *Sr* genes (*Sr21*, *Sr22b*, and *SrTm4*)¹⁷, M1461 was crossed with *T. monococcum* accession PI 272557, which lacks known *Sr* genes²⁴. From the resulting F₂ population, we selected plants carrying either the WT *SR60* allele or the M1461 allele, while excluding other *Sr* genes. When challenged with *Pgt* race 34PKUSC, plants carrying the WT *SR60* allele exhibited intermediate resistance, whereas those homozygous for the M1461 allele were susceptible (Fig. 4a). These results demonstrate that PI306540.5A02G00007400 is required for *SR60*-mediated resistance to *Pgt*.

PI 306540 mutant library enables rapid validation of multicopy genes in wheat

Unlike polyploid wheat, where inactivating all homoeologs requires extensive crossing, the diploid wheat mutant library allows direct functional validation. In hexaploid wheat, simultaneous inactivation of all three *TaGBSSI* homoeologs via EMS mutagenesis nearly eliminates amylose, but necessitates prolonged crossing and selfing to combine mutations across the three subgenomes²⁵. By contrast, PI 306540 carries a single copy (*TmGBSSI*; PI306540.7A02G00072400), which shares 96.20% protein identity with *TaGBSSI*-A1 (Chinese Spring CS-IAAS genome²). Screening the PI 306540 mutation database identified 71 independent mutations in *TmGBSSI*, including three premature stop codons (Supplementary Table 16). Mutant line M3261 (W90*) was selected for further analysis. Endosperms homozygous for this mutant allele stained reddish-brown with iodine, whereas WT endosperms exhibited dark blue (Fig. 4b).

Similarly, achieving robust and broad-spectrum powdery mildew resistance in hexaploid wheat requires simultaneous knockout of all three *MLO* homoeologs²⁶. The PI 306540 ortholog, *TmMLO* (PI306540.5A02G00545600), shares 99.63% protein identity with *TaMLO*-A1. From the PI 306540 mutant library, we found 43 independent mutations in *TmMLO*, including two premature stop codons and one splice donor mutation (Supplementary Table 17). Mutant line M1416 (W424*) was selected for further characterization. Screening against 25 isolates of *Blumeria graminis* f. sp. *tritici* (*Bgt*) identified several isolates that were virulent on PI 306540 (Supplementary Table 18). When challenged with a mixture of virulent isolates, plants homozygous for the mutant allele showed strong resistance, whereas WT plants were susceptible (Fig. 4c).

Finally, the wheat *phytoene desaturase* (*TaPDS*) gene²⁷ was used to further validate the library. The diploid ortholog *TmPDS* (PI306540.4A02G00330000) shares 99.83% protein identity with *TaPDS*-A1. Mutant line M2351 carries a G-to-A substitution at the splice acceptor site of the ninth intron in *TmPDS*, resulting in premature translation termination (Fig. 4d). Plants homozygous for this frameshift mutation displayed a fully albino phenotype, whereas segregating non-mutant plants developed normal green seedlings (Fig. 4d). Together, these examples demonstrate that the diploid wheat mutant library bypasses the genetic redundancy of polyploid wheat, thereby enabling efficient functional validation through a direct genotype-to-phenotype connection.

Identification of the causal gene underlying a brown-midrib mutant

From the PI306540 mutant library, we identified an M₂ mutant line, M1820, which exhibited reddish-brown pigmentation in multiple tissues at the flowering stage (Fig. 5a). When M1820 was crossed with WT PI 306540, all F₁ plants resembled the WT phenotype. Segregation in the F₂ generation (162 WT-like : 50 mutant-like plants) fit a 3:1 ratio ($\chi^2 = 0.23$, $P = 0.63$), indicating that this pigmentation trait is controlled by a single recessive gene, which we designated *bm1*.

To map the causal gene, we performed bulked segregant RNA sequencing (BSR-Seq) on the M1820 × PI306540 population. SNP index analysis identified 12 associated SNPs within a 217.51-Mb interval (39.60-257.11 Mb) on chromosome 6A (PI306540 T2T; Supplementary Table 19). Genotyping 212 F₂ plants with six PCR-based sequencing markers (Supplementary Table 20) delimited the gene to a 0.71-cM interval flanked by markers *pkw173* and *pkw257* (Fig. 5b). This interval corresponds to an 83.54-Mb region in the PI 306540 genome, containing 294 annotated high-confidence genes (Supplementary Table 21). Among these, PI306540.6A02G00193400 emerged as a strong candidate, as it is orthologous to the wheat *BGI1* gene¹³ (Supplementary Fig. 8), which encodes cinnamyl alcohol dehydrogenase 1 (*TaCAD1*), a key enzyme in lignin biosynthesis. Mutations in this gene are known to alter pigmentation in lignified tissues¹³.

Screening the PI 306540 mutation database identified 37 independent mutations in PI306540.6A02G00193400 (Supplementary Table 22), including two splice-donor mutants (M1820 and M0560). Sanger sequencing confirmed that M1820 carries a G3846A transition at the splice donor site of the third intron, which causes aberrant splicing and premature truncation (Fig. 5b). The other mutant,

M0560, which carries a mutation at the splice donor site of the first intron, was used for genotyping and phenotyping. Plants homozygous for the mutant allele exhibited reddish-brown pigmentation (Fig. 5c). These results confirmed that *PI306540.6A02G00193400* is the causal gene.

To further validate this finding, we used CRISPR-Cas9 to knock out the corresponding orthologs in the bread wheat cultivar Fielder. Among 35 independent T₀ transgenic plants, 22 (62.86%) carried mutations at the target site (Fig. 5d). Two edited lines (KO1 and KO2) were homozygous knockout mutants in which all six homoeologous alleles were simultaneously modified (Fig. 5d). These frameshift mutations altered >97.2% of the protein sequences, abolishing the function of all three homoeologs. At the flowering stage, KO1 and KO2 exhibited reddish-brown pigmentation in multiple tissues (Fig. 5e), whereas single and double mutant plants displayed no phenotypic changes (Supplementary Fig. 9). Together, these results provide definitive evidence that *PI306540.6A02G00193400* is responsible for the brown-midrib phenotype.

Identification of the causal gene underlying a yellow-green leaf mutant

From an EMS-mutagenized population of *Aegilops tauschii* accession PI 511383, we identified a mutant A3437, which consistently exhibited yellow-green leaves throughout its entire growth period (Fig. 6a). Compared with WT PI 511383, the chlorophyll a and b contents in A3437 were significantly reduced by 69.38% and 66.37%, respectively (Fig. 6b). Genetic analysis of a cross between A3437 and another *Ae. tauschii* accession PI 511384 showed that all F₁ plants displayed the green-leaf phenotype. Segregation in the F₂ generation (91 green : 32 yellow-green) fit a 3:1 ratio ($\chi^2 = 0.07$, $P = 0.79$), indicating that the yellow-green trait is controlled by a single recessive gene, which we designated *yl1*.

BSR-Seq analysis of the A3437 × PI 511384 mapping population identified 396 phenotype-associated SNPs and InDels within a 101.66-Mb interval (211.47-313.13 Mb; AL8/78 Aet v6.0) on chromosome 1D (Supplementary Table 23). Using 695 F₂ plants and 11 cleaved amplified polymorphic sequence markers on chromosome 1D, we delimited the causal gene to a 0.14-cM interval flanked by markers *pkw247* and *pkw257* (Fig. 6c and Supplementary Table 24). This interval corresponds to a 10.47-Mb physical region in the AL8/78 genome (Aet v6.0), containing 60 annotated high-confidence genes (Supplementary Table 25). Within this candidate region, whole-genome resequencing of A3437 and WT PI 511383 identified only one SNP (G251157285A) in the coding region of a gene (Supplementary Table 26). This gene, *AET1D6G0155300*, encodes protochlorophyllide reductase B, and the mutation results in a Gly221Asp substitution. *AET1D6G0155300* spans 1,592 bp from start to stop codon, comprising four exons and three introns (Fig. 6c). Its complete coding sequence of 1,188 bp encodes a predicted protein of 395 amino acids. Phylogenetic analysis showed that *AET1D6G0155300* is orthologous to the *POR-B* gene (Supplementary Fig. 10), mutations of which are associated with light-green leaf blades²⁸. These results suggest that *AET1D6G0155300* is a strong candidate for *yl1*.

In PI 306540, the ortholog of *AET1D6G0155300* is *PI306540.1A02G00194000*, which shares 98.23% protein identity. Screening the PI 306540 mutant library identified 29 independent mutations in *PI306540.1A02G00194000* (Supplementary Table 27), including one premature stop codon (Q325*; line M4149) and one splice donor mutation (line M1126). Plants homozygous for either the Q325* allele or the splice donor allele exhibited the yellow-green leaf phenotype (Fig. 6d).

We then performed CRISPR/Cas9-mediated knockout of the corresponding orthologs in Fielder (Fig. 6e). Of 33 independent T₀ transgenic plants, 17 (51.52%) carried mutations at the target site and exhibited yellow or yellow-green leaves (Fig. 6f and Supplementary Fig. 11), with discoloration severity correlating with editing efficiency. One T₀ plant (T₀644-30), in which all six homoeologous alleles were simultaneously edited, displayed a completely chlorotic phenotype but was not viable (Figs. 6e and 6f). In the T₁ generation, plants with incomplete homozygous editing consistently exhibited yellow-green leaves (Fig. 6f). Collectively, these results confirm *AET1D6G0155300* and its orthologs as the causal genes for the yellow-green leaf phenotype.

DISCUSSION

Einkorn, the first domesticated wheat species, played a central role in the emergence of agriculture and the Neolithic Revolution in the Fertile Crescent ~10,000 years ago^{20,29}. Unlike *T. urartu*, einkorn has undergone prolonged cultivation and human selection across diverse environments. Approximately 1% of the bread wheat A subgenome can be traced to einkorn²¹. Natural and artificial introgressions from einkorn into bread wheat have been documented, many of which carry agriculturally important genes^{18-20,22,30}. Despite its agronomic value, genomic resources for einkorn remain limited; currently, only three draft genome assemblies are

available, each containing over 211 gaps^{20,21}. Here, we report a complete T2T assembly (5.11 Gb) for the cultivated einkorn accession PI 306540 (Fig. 1), representing the highest-quality A-genome reference for wheat to date. This assembly will serve as a fundamental resource for functional and evolutionary studies and accelerate the cloning of agronomically useful einkorn genes.

Using this assembly and the exome capture array developed for *T. monococcum*, we identified 1,338,432 EMS-induced mutations across 3,777 mutant lines, yielding a density of 28.30 mutations per kb. This density is comparable to those reported for major polyploid wheat mutant resources, such as the tetraploid Kronos (34.84 mutations/kb; 1,535 lines) and the hexaploid Candenza (39.54 mutations/kb; 1,200 lines) and Jing411 (47.10 mutations/kb; 2,162 lines) populations^{7,11}, and is notably higher than that of the hexaploid KN9204 library (19.94 mutations/kb; 2,090 lines)¹⁰. It far exceeds the 0.011 mutations/kb reported for a previously described einkorn mutant population derived from accession TA4342-96 (1,532 lines) based on Sanger sequencing of four genes³¹. The fact that our much larger diploid mutant population exhibits a relatively lower mutation density compared with some polyploid libraries may seem counterintuitive; however, this likely reflects the greater genomic buffering capacity of polyploids, which enables them to tolerate higher mutation loads. Indeed, hexaploid and tetraploid wheat can tolerate EMS concentrations up to 1.20% and 0.75%, respectively³², whereas only 32.23% of M₁ plants survived treatment with 0.45% EMS in our diploid population.

In hexaploid wheat, EMS-based functional analysis typically requires targeting all three homoeologs and pyramiding mutations through crossing, a process that can take years¹³⁻¹⁵. By contrast, the diploid system circumvents this complication, as most genes reside at single homoeologous loci. Consequently, the sequenced PI 306540 mutant library provides an efficient platform for studying traits and genes relevant to bread wheat, enabling direct phenotype-to-genotype associations. We demonstrated the utility of this diploid mutant system by rapidly validating *SR60*, *MLO*, *TmPDS*, and *TmGBSSI* (Fig. 4), and by identifying the causal genes underlying the brown-midrib and yellow-green leaf mutants (Figs. 5 and 6).

The selection of the mutant generation is a critical consideration in constructing indexed EMS libraries. We performed WES on 3,777 mutant lines (1,381 M₂ and 2,396 M₃) and systematically compared mutation load, homozygosity ratio, and validation efficiency. M₂ lines carried significantly more mutations per line (470.48) than M₃ lines (287.44), consistent with mutation loss over generations due to segregation in single-plant progenies. However, M₂ mutants were predominantly heterozygous (63.64%), with only 36.36% homozygous (Supplementary Table 10), which may mask recessive phenotypes in direct screening. In contrast, M₃ lines showed a substantially higher homozygous proportion (61.51%), while heterozygotes dropped to 38.49% (Supplementary Table 10), reflecting Mendelian segregation after selfing. Thus, M₃ and later generations are more suitable for phenotypic screening, as recessive traits become penetrant.

The sequenced diploid wheat mutant database generated in this study can be complemented by CRISPR, which enables precise genome editing but remains technically demanding and genotype-dependent in wheat. The process of designing gene-specific guide RNAs, constructing transformation vectors, and executing stable genetic transformation remains challenging and labor-intensive, meaning that CRISPR is not readily accessible to every laboratory. By contrast, our EMS mutant library provides an immediately accessible and cost-effective platform for functional analysis. Researchers can perform *in silico* searches to identify mutant alleles for any gene of interest. Seeds for the desired line can then be directly requested and grown for phenotypic evaluation, bypassing the lengthy transgenic pipeline.

In summary, this work integrates a complete T2T reference genome of diploid wheat with a gene-indexed EMS mutant library comprising 3,777 lines. This integrated platform overcomes the challenges of homoeologous gene redundancy by enabling a "diploidized" approach to functional genomics, thereby accelerating gene discovery and trait dissection to support the genetic improvement of modern wheat.

METHODS

Plant materials, EMS mutagenesis, and mapping populations

Einkorn wheat accession PI 306540 was obtained from the USDA National Small Grains Collection and self-pollinated for ten generations under greenhouse conditions. Approximately 15,000 high-quality seeds were mutagenized with 0.45% EMS (Sigma, St. Louis, MO) for 18 h. Treated seeds were rinsed under running water for 3 h, germinated on moistened germination paper at 4°C, and transplanted to the field. A single M₂ plant derived from each M₁ plant was grown in the greenhouse. Genomic DNA was extracted

from each M₂ or M₃ plant for whole-exome sequencing (one plant per line), and seeds were harvested from the sequenced individuals. For phenotypic screening, 15-20 seeds from each sequenced line were sown in single-row plots. Seeds from each row were harvested and archived for distribution. Subsequent generations were continuously evaluated in the field for phenotypic variation.

Three F₂ mapping populations were developed. To validate the stem rust resistance gene *SR60*, the mutant line M1461 was crossed with the cultivated *T. monococcum* accession PI 272557²⁴. The brown-midrib mutant M1820 was crossed with its WT PI 306540 to map the causal gene. Additionally, the yellow-green leaf mutant A3437, identified in an EMS-mutagenized population of *Ae. tauschii* accession PI 511383¹³, was crossed with another *Ae. tauschii* accession PI 511384 for map-based cloning.

Genomic DNA sequencing and assembly

Fresh leaf tissue was collected from 10-day-old seedlings. Genomic DNA extraction, library construction, and sequencing were performed by Novogene Bioinformatics Technology Co., Ltd. (Tianjin, China) and the Single Cell Omics/Single Molecule Analysis Platform at the Peking University Institute of Advanced Agricultural Sciences (Weifang, China). Sequencing was conducted using PacBio Sequel II (HiFi), Oxford Nanopore Technologies (ultra-long), and Illumina HiSeq (Hi-C).

We used hifiasm v0.19.9³³ with the -UL option to integrate PacBio HiFi reads with ONT ultra-long reads, generating a contig-level assembly. Hi-C paired-end reads were mapped to contigs using 3D-DNA³⁴ and Juicer v1.6³⁵, followed by manual refinement with Juicebox³⁶. The final four gaps were closed using QuartetT v1.2.1³⁷ with default settings. We then mapped PacBio HiFi reads (with -x map-hifi) and ONT reads (with -x map-ont) to the gap regions with Minimap2 v2.28³⁸ and checked their consistency with the flanking assembly. The contiguous structure was visually inspected for each gap in IGV v12.19.7³⁹.

Gene annotation

The PI 306540 T2T genome was annotated by integrating homology-based predictions, PacBio Iso-Seq, and RNA-seq data. RNA was extracted from the leaves, roots, stems, seeds, and spikes at six growth stages (seedling, vernalization, elongation, booting, heading, and filling; Supplementary Table 1). Library construction and sequencing were performed by Novogene Bioinformatics Technology Co., Ltd. (Tianjin, China). High-confidence annotations from PI 306540 v1.0, TA10622, and TA299 were lifted over to the PI 306540 T2T assembly using Liftoff v1.6.3⁴⁰. RNA-seq reads were trimmed with fastp v0.23.2⁴¹, aligned using HISAT2 v2.2.1⁴², and assembled with StringTie v2.2.1⁴³. Iso-Seq reads were processed with the Iso-Seq3 pipeline, mapped to the genome with pbmm2 (<https://github.com/PacificBiosciences/pbmm2>), and collapsed using cDNA_Cupcake (https://github.com/Magdoll/cDNA_Cupcake). Transcript assemblies from RNA-seq and Iso-Seq were merged with StringTie v2.2.1⁴³, and coding sequences were predicted using TransDecoder (<https://github.com/TransDecoder/TransDecoder>). All annotation files were consolidated into a single GFF3 file using the AGAT toolkit⁴⁴. Transcript isoforms were then clustered at the gene level using CD-HIT v4.8.1⁴⁵. High-confidence gene models were retained if they met either of the following criteria: 1) significant homology to the UniProtKB Poaceae reference proteome (BLASTP e-value ≤ 1e-5, amino acid identity ≥ 80%, and query coverage ≥ 50%), or 2) supporting evidence from RNA-seq data.

Assessment of genome assembly and annotation quality

Genome completeness was evaluated using BUSCO v5.8.0 (poales_odb10)⁴⁶. Assembly quality was assessed with the LTR Assembly Index calculated by LTR_retriever⁴⁷ based on intact LTR retrotransposons identified using LTR_Finder⁴⁸. The quality value was estimated using Merqury v1.3 (k = 21)⁴⁹. BUSCO⁴⁶ was used for gene annotation quality assessment.

Annotation of TEs

A custom repeat library was constructed by combining a *de novo* library generated with RepeatModeler2 v2.0.3⁵⁰ and curated TE sequences from ClariTeRep (<https://github.com/jdaron/CLARI-TE>). Repetitive elements were annotated using RepeatMasker v4.1.5⁵¹.

Centromere analysis

ChIP-seq libraries were prepared using an antibody against CENH3, with input DNA as a control. Raw sequencing reads were trimmed using fastp v0.23.2⁴¹ and aligned to the PI 306540 genome using Bowtie2 v2.5.1⁵². SAMtools v1.10⁵³ was used for format

conversion, sorting, and duplicate removal (mapping quality ≥ 30). Peaks were called using MACS2 v2.2.9.1⁵⁴. Enrichment profiles ($\log_2$ fold-change over input) were generated with deepTools2 v3.5.5⁵⁵ and visualized in IGV v12.19.7³⁹.

Comparative analysis of *T. monococcum* genomes

Pairwise whole-genome alignments between PI 306540 and two published assemblies (TA10622 and TA299) were performed using MUMmer4⁵⁶. Alignments were filtered to retain collinear, high-identity ($\geq 95\%$) blocks of at least 2 kb using delta-filter, and coordinate information was extracted with show-coords. Structural variants were identified with SyRI⁵⁷ and visualized using Plotsr v1.1.0⁵⁸.

Phylogenetic and comparative genomic analysis

Orthologous gene families were identified across 13 *Triticeae* species plus two outgroups (*Oryza sativa* and *Brachypodium distachyon*) using GeneTribe2 v1.2.1²³. Single-copy orthologs were aligned with MUSCLE v5.1⁵⁹ and trimmed with trimAl⁶⁰. A maximum likelihood phylogenetic tree was constructed with IQ-TREE v2.1.4⁶¹ using ModelFinder and 1,000 ultrafast bootstrap replicates. Divergence times were estimated with MCMCTREE in PAML v4.9⁶². Whole-genome-wide synteny was detected using MCScanX⁶³ and visualized with the JCVI⁶⁴. Protein sequences of PI306540.6A02G00193400 and AET1D6G0155300 and their orthologs/homologs from representative *Triticeae* and outgroup species were aligned and used to construct neighbor-joining phylogenetic trees in MEGA7⁶⁵. Multiple sequence alignments were visualized using ggmsa⁶⁶.

Exome capture design

Exonic sequences (including both coding exons and UTRs) from high-confidence genes in the PI 306540 v1.0 annotation were targeted for probe design by Tcuni Technology Co., Ltd. (Chengdu, Sichuan, China). The v1.0 annotation was selected because it was available prior to the T2T assembly. Exons were divided into 200-bp non-overlapping windows using BEDTools v2.30.0⁶⁷. For each window, a 40-bp candidate region was centered at the midpoint, and nine overlapping 120-bp probe candidates were generated with a 5-bp sliding step. The optimal one or two probes per window were selected based on the following priority criteria: GC content within 0.30-0.65, minimal mononucleotide repeats, fewest genomic alignments (determined by Minimap2 v2.28)³⁸, and highest local sequence complexity.

Exome library construction, capture, and sequencing

Library preparation and exome capture sequencing were conducted by Tcuni Technology Co., Ltd. (Chengdu, China). Genomic DNA was extracted from a single plant per family. Whole-exome libraries were prepared for 1,381 M₂ and 2,396 M₃ individuals using the BGI DNA Library Prep Kit (BGI, Shenzhen, China). For each sample, ~2 μ g of genomic DNA was fragmented to 200-300 bp using a Bioruptor UCD-200 sonicator (Diagenode, Denville, NJ, USA). Fragmented DNA was end-repaired and deoxyadenosine-tailed. Size selection was performed with magnetic beads (Twist Bioscience, CA, USA), followed by 4-5 cycles of pre-capture PCR. Equal amounts of library products were pooled for exome capture using a custom-designed probe panel targeting ~47.29 Mb of coding sequence. Hybridization products were captured using the M1 Beads Capture Kit (Thermo Fisher, MA, USA), washed, and amplified via ligation-mediated PCR. Library quality was assessed on an Agilent 4200 TapeStation system (Agilent, CA, USA), and final libraries were sequenced on the BGI DNBSEQ-T7 platform.

Mutation calling and annotation

A total of 3,777 plants were subjected to exome sequencing, generating ~36.38 Tb of raw data (mean 9.63 Gb per sample). Raw reads were trimmed to remove adapters and low-quality bases using fastp v0.23.2⁴¹ and aligned to the PI 306540 T2T genome using the GTX-One platform (Genetalks, China)⁶⁸. Sequence alignment was performed with the gtx align module, and variant calling was conducted with the gtx vc module using the parameters `-pcr-indel-model HOSTILE` and `-min-base-quality-score 20`. To ensure data quality, raw variants were filtered to retain only those with allele depth ≥ 6 and quality-by-depth ≥ 10 . The resulting VCF files were merged using BCFtools v1.14⁶⁹ to generate a multi-sample VCF. The merged VCF was further filtered to remove SNPs present in more than three mutant lines and InDels present in more than one mutant line. Functional annotation of the remaining variants was performed using snpEff v5.2⁷⁰ based on the PI 306540 genome annotations (v1.0 and T2T).

Validation of EMS-induced mutations

Mutations identified by WES were validated by PCR and Sanger sequencing. A total of 43 genes spanning all seven chromosomes, comprising 196 SNPs and five InDels, were randomly selected for validation (Supplementary Table 14). For each mutant line, genomic DNA was extracted from 8-16 progeny and pooled in equimolar amounts to serve as PCR templates. Gene-specific primers were designed to amplify the genomic regions flanking each mutation. The resulting amplicons were purified and subjected to Sanger sequencing.

Identification of mutant lines via online BLASTN search

The sequenced EMS-mutagenized population of PI 306540 can be queried online at <http://223.85.181.157:8182/>. BLASTN searches using the sequences of target genes (e.g., *SR60*, *MLO*, *TmPDS*, or *TmGBSSI*) readily identify mutant lines harboring mutations in those genes. Candidate mutations were confirmed using genome-specific primer pairs (Supplementary Table 14), and plants homozygous for the mutant alleles were subsequently used for phenotypic analysis.

Stem rust and powdery mildew assays

For stem rust assays, PI 306540 and F₂ plants derived from the cross between PI 272557 and M1461 were inoculated with *Pgt* race 34PKUSC²². Seedlings at the three-leaf stage were inoculated with *Pgt* urediniospores mixed with talcum powder at a 1:20 ratio using the shaking-off method¹⁸. Inoculated plants were placed in a dark dew chamber at 22°C for 24 hours and then transferred to a growth chamber (22-24 °C, 16-h photoperiod). Disease symptoms were scored ~12 days post-inoculation.

For powdery mildew assays, PI 306540 was initially tested against 25 isolates of *Bgt* at Yantai University to identify virulent isolates (Supplementary Table 18). Subsequently, one-leaf-stage seedlings of PI 306540 and mutant line M1416 were inoculated with a mixture of virulent *Bgt* isolates.

Iodine staining

Amylose content was assessed using iodine staining. Longitudinal and transverse hand-cut sections of grains from PI 306540 and M3261 were immersed in a 1:128 dilution of Lugol's iodine solution (5% I₂, 10% KI) for 15 min and imaged under bright-field microscopy.

Measurement of chlorophyll content

Chlorophyll content in PI 511383 and A3437 was determined using a plant Chlorophyll Content Assay Kit (Catalog No. C930493; Macklin, Shanghai, China). Absorbances were measured at 663 nm and 645 nm using a microplate reader. Chlorophyll concentrations were calculated using the equations provided by the manufacturer.

BSR-seq analysis and marker development

Leaf samples were collected from two F₂ populations: 20 brown-midrib and 20 WT individuals from the M1820 × PI 306540 population, and 20 yellow-green and 20 WT individuals from the A3437 × PI 511384 population. RNA extraction, library construction, and sequencing were performed by Novogene Bioinformatics Technology Co., Ltd (Tianjin, China). Raw reads were trimmed and filtered using fastp v0.23.2⁴¹ and aligned to PI 306540 or AL8/78 Aet v6.0 (*Ae. tauschii*) reference genomes using STAR v2.7.10a⁷¹. Variants were called from each bulk using FreeBayes v1.3.6 (<https://github.com/freebayes/freebayes>) and filtered with BCFtools v1.14⁶⁹. SNP-index and Δ(SNP-index) were calculated to identify candidate genomic regions associated with each phenotype. Primers targeting polymorphisms were designed using Primer3 v4.1.0 (Supplementary Tables 20 and 24). PCR products were Sanger-sequenced to validate polymorphisms.

CRISPR/Cas9-based gene editing

CRISPR/Cas9 gene-editing system was used to validate the candidate genes *bm1* (PI306540.6A02G00193400) and *yl1* (AET1D6G0155300) in the wheat cultivar Fielder. gRNA design, vector construction, and application of Cas9-Trex2 and GRF4-GIF1 fusion proteins were performed as described^{72,73}. The gRNA expression cassette (TaU3p:gRNA) was generated by two rounds of PCR using the pOPGR-TS1 plasmid as template⁷⁴ and cloned into the modified pCas9T vector at the *StuI* site using the In-Fusion® HD Cloning Kit (Clontech, CA, USA). The final construct was introduced into Fielder via *A. tumefaciens*-mediated transformation. Mutations in transgenic plants were identified using genome-specific primer pairs (Supplementary Tables 20 and 24).

Declarations

Data availability

All sequencing data generated from wheat accession PI 306540 and its mutant lines, along with transcriptome data, genome assembly, and gene annotation, have been deposited at the National Genomics Data Center (Beijing Institute of Genomics, Chinese Academy of Sciences) under BioProject accession number PRJCA060576 (<https://ngdc.cnbc.ac.cn/gwh/Assembly/reviewer/upkhhfNbmCMBFJxBjMHJVambmIIOADYjGmXBYGPwGljAjlufpoelkOtNEgAEvLdV>). The PI 306540 genome assembly and gene annotation have also been deposited at NCBI under BioProject number PRJNA1465037 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA1465037>) and in Figshare (<https://figshare.com/s/434cf9f5cbf1bd2df2f2>). All mutant lines and variant annotation data are freely available to the research community at <http://223.85.181.157:8182/>. The re-sequencing data of PI 511383 were obtained from the National Genomics Data Center (BioProject accession PRJCA061522). The raw sequencing data associated with mapping of *bm1* and *y11* have also been deposited at the National Genomics Data Center under the same BioProject accession number PRJCA060576.

Code availability

All software packages used in the study are publicly available and as described in Methods and Reporting Summary.

Acknowledgements

We thank Dr. Baoxing Song (Peking University Institute of Advanced Agricultural Sciences) for providing the TaCENH3 antibody. This research was supported by the Key R&D Program of Shandong Province (2024CXPT072, 2024LZGC034, and 2023LZGC022), the National Natural Science Foundation of China (32472159), the Key Research and Development Program of Shaanxi (2024NC2-GJHX-35), the Shandong Provincial Natural Science Foundation (SYS202206), the National Key Research and Development Program of China (2022YFD1201300), the Shandong Provincial Natural Science Fund for Excellent Young Scientists Fund Program (Overseas) (2026HWYQ-078), and the Taishan Scholars Program.

Author contributions

S.C. designed and supervised the project; Hong.L., S.L., Y.C., W.H., and P.A. performed the experiments and analyzed the data; Y.C. supervised the genome assembly; T.S., M.L., S.R., He.L., Sheng.C., and L.S. contributed to the EMS mutant population; L.H. and K.W. contributed to vector construction; Z.C. assisted with sequence data analysis and database construction; B.L. contributed to ONT sequencing; L.G., Hao.L., and X.W.D provided scientific support; S.C., Hong.L., S.L., and Y.C. wrote the initial draft of the manuscript; S.C. obtained the funding and generated the final manuscript. All authors reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Feldman, M. & Levy, A.A. Genome evolution due to allopolyploidization in wheat. *Genetics* **192**, 763-774 (2012).
2. 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 (2025).
3. Shiferaw, B. et al. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Secur* **5**, 291-317 (2013).
4. Walkowiak, S. et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**, 277-283 (2020).
5. Rasheed, A. & Xia, X. From markers to genome-based breeding in wheat. *Theor Appl Genet* **132**, 767-784 (2019).
6. Stadler, L.J. Some genitic effects of X-rays in plants. *J Hered* **21**, 3-20 (1930).
7. Krasileva, K.V. et al. Uncovering hidden variation in polyploid wheat. *Proc Natl Acad Sci USA* **114**, E913-E921 (2017).
8. Henikoff, S., Till, B.J. & Comai, L. TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol* **135**, 630-636 (2004).

9. Kaur, P., and Gaikwad, K. From genomes to GENE-omes: exome sequencing concept and applications in crop improvement. *Front Plant Sci* **8**, 2164 (2017).
10. Wang, D. et al. Boosting wheat functional genomics via an indexed EMS mutant library of KN9204. *Plant Commun* **4**, 100593 (2023).
11. Xiong, H. et al. A large-scale whole-exome sequencing mutant resource for functional genomics in wheat. *Plant Biotechnol J* **21**, 2047-2056 (2023).
12. Uauy, C., Wulff, B.B. & Dubcovsky, J. Combining traditional mutagenesis with new high-throughput sequencing and genome editing to reveal hidden variation in polyploid wheat. *Annu Rev Genet* **51**, 435-454 (2017).
13. Hua, L. et al. Manipulation of the *brown glume and internode 1* gene leads to alterations in the colouration of lignified tissues, lignin content and pathogen resistance in wheat. *Plant Biotechnol J* **23**, 1548-1564 (2025).
14. Schönhofen, A., Hazard, B., Zhang, X. & Dubcovsky, J. Registration of common wheat germplasm with mutations in *SBEII* genes conferring increased grain amylose and resistant starch content. *J Plant Regist* **10**, 200-205 (2016).
15. Li, K. et al. Interactions between SQUAMOSA and SHORT VEGETATIVE PHASE MADS-box proteins regulate meristem transitions during wheat spike development. *Plant Cell* **33**, 3621-3644 (2021).
16. Dvorak, J., McGuire, P.E. & Cassidy, B. Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome* **30**, 680-689 (1988).
17. Li, H. et al. High-resolution mapping of *SrTm4*, a recessive resistance gene to wheat stem rust. *Theor Appl Genet* **136**, 120 (2023).
18. Chen, S. et al. Wheat gene *Sr60* encodes a protein with two putative kinase domains that confers resistance to stem rust. *New Phytol* **225**, 948-959 (2020).
19. Chen, S. et al. Stripe rust resistance gene *Yr34* (synonym *Yr48*) is located within a distal translocation of *Triticum monococcum* chromosome 5A^mL into common wheat. *Theor Appl Genet* **134**, 2197–2211 (2021).
20. Wang, X. et al. A near-complete genome sequence of einkorn wheat provides insight into the evolution of wheat A subgenomes. *Plant Commun* **5**, 100768 (2024).
21. Ahmed, H.I. et al. Einkorn genomics sheds light on history of the oldest domesticated wheat. *Nature* **620**, 830-838 (2023).
22. Luo, J. et al. Identification and characterization of *Sr22b*, a new allele of the wheat stem rust resistance gene *Sr22* effective against the Ug99 race group. *Plant Biotechnol J* **20**, 554-563 (2022).
23. Chen, Y. et al. A Collinearity-incorporating homology inference strategy for connecting emerging assemblies in the triticeae tribe as a pilot practice in the plant pangenomic era. *Mol Plant* **13**, 1694-1708 (2020).
24. Chen, S. et al. Fine mapping and characterization of *Sr21*, a temperature-sensitive diploid wheat resistance gene effective against the *Puccinia graminis* f. sp. *tritici* Ug99 race group. *Theor Appl Genet* **128**, 645-656 (2015).
25. Kumar, S. et al. The impact of GBSSI inactivation on starch structure and functionality in EMS-induced mutant lines of wheat. *BMC Genom* **26**, 501 (2025).
26. Wang, Y. et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat biotechnol* **32**, 947 (2014).
27. Li, T. et al. Highly efficient heritable genome editing in wheat using an RNA virus and bypassing tissue culture. *Mol Plant* **14**, 1787-1798 (2021).
28. Sakuraba, Y. et al. The rice faded green leaf locus encodes protochlorophyllide oxidoreductase B and is essential for chlorophyll synthesis under high light conditions. *Plant J* **74**, 122-133 (2013).
29. Heun, M. et al. Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* **278**, 1312-1314 (1997).
30. Chen, S., Zhang, W., Bolus, S., Rouse, M.N. & Dubcovsky, J. Identification and characterization of wheat stem rust resistance gene *Sr21* effective against the Ug99 race group at high temperature. *PLoS Genet* **14**, e1007287 (2018).
31. Rawat, N. et al. A diploid wheat TILLING resource for wheat functional genomics. *BMC Plant Biol* **12**, 205 (2012).
32. Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N. & Facciotti, D. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat Biotechnol* **23**, 75-81 (2005).
33. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170-175 (2021).

34. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).
35. Durand, N.C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**, 95-98 (2016).
36. Durand, N.C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* **3**, 99-101 (2016).
37. Lin, Y. et al. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic Res* **10**, uhad127 (2023).
38. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
39. Thorvaldsdóttir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192 (2013).
40. Shumate, A. & Salzberg, S.L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639-1643 (2021).
41. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
42. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015).
43. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).
44. Dainat, J., Hereñú, D. & Pucholt, P. AGAT: another gff analysis toolkit to handle annotations in any GTF/GFF format. *Zenodo*, <https://www.doi.org/10.5281/zenodo.3552717> (2021).
45. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).
46. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
47. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol* **176**, 1410-1422 (2018).
48. Ou, S. & Jiang, N. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mobile DNA* **10**, 48 (2019).
49. Rhie, A., Walenz, B.P., Koren, S. & Phillippy, A.M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).
50. Flynn, J.M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* **117**, 9451-9457 (2020).
51. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinform* **5**, 4-10 (2004).
52. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
53. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
54. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X.S. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**, 1728-1740 (2012).
55. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160 (2016).
56. Marçais, G. et al. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol* **14**, e1005944 (2018).
57. Goel, M., Sun, H., Jiao, W.-B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* **20**, 277 (2019).
58. Goel, M. & Schneeberger, K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922-2926 (2022).
59. Edgar, R.C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun* **13**, 6968 (2022).
60. Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

61. Minh, B.Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37**, 1530-1534 (2020).
62. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007).
63. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012).
64. Tang, H. et al. JCVI: A versatile toolkit for comparative genomics analysis. *Imeta* **3**, e211 (2024).
65. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**, 1870-1874 (2016).
66. Zhou, L. et al. ggmsa: A visual exploration tool for multiple sequence alignment and associated data. *Brief Bioinform* **23**, bbac222 (2022).
67. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
68. Bu, L. et al. Improving read alignment through the generation of alternative reference via iterative strategy. *Sci Rep* **10**, 18712 (2020).
69. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
70. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly* **6**, 80-92 (2012).
71. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
72. Yang, J. et al. Genome-assisted identification of wheat leaf rust resistance gene *Lr. ace-4A/Lr30*. *Nat Commun* **16**, 9339 (2025).
73. Shen, T. et al. The wheat *Sr8155B1* gene encodes a typical NLR protein that confers resistance to the Ug99 stem rust race group. *Plant Commun* **6**, 101296 (2025).
74. Li, J., Wang, Z., He, G., Ma, L. & Deng, X.W. CRISPR/Cas9-mediated disruption of *TaNP1* genes results in complete male sterility in bread wheat. *J Genet Genomics* **47**, 263-272 (2020).

Figures

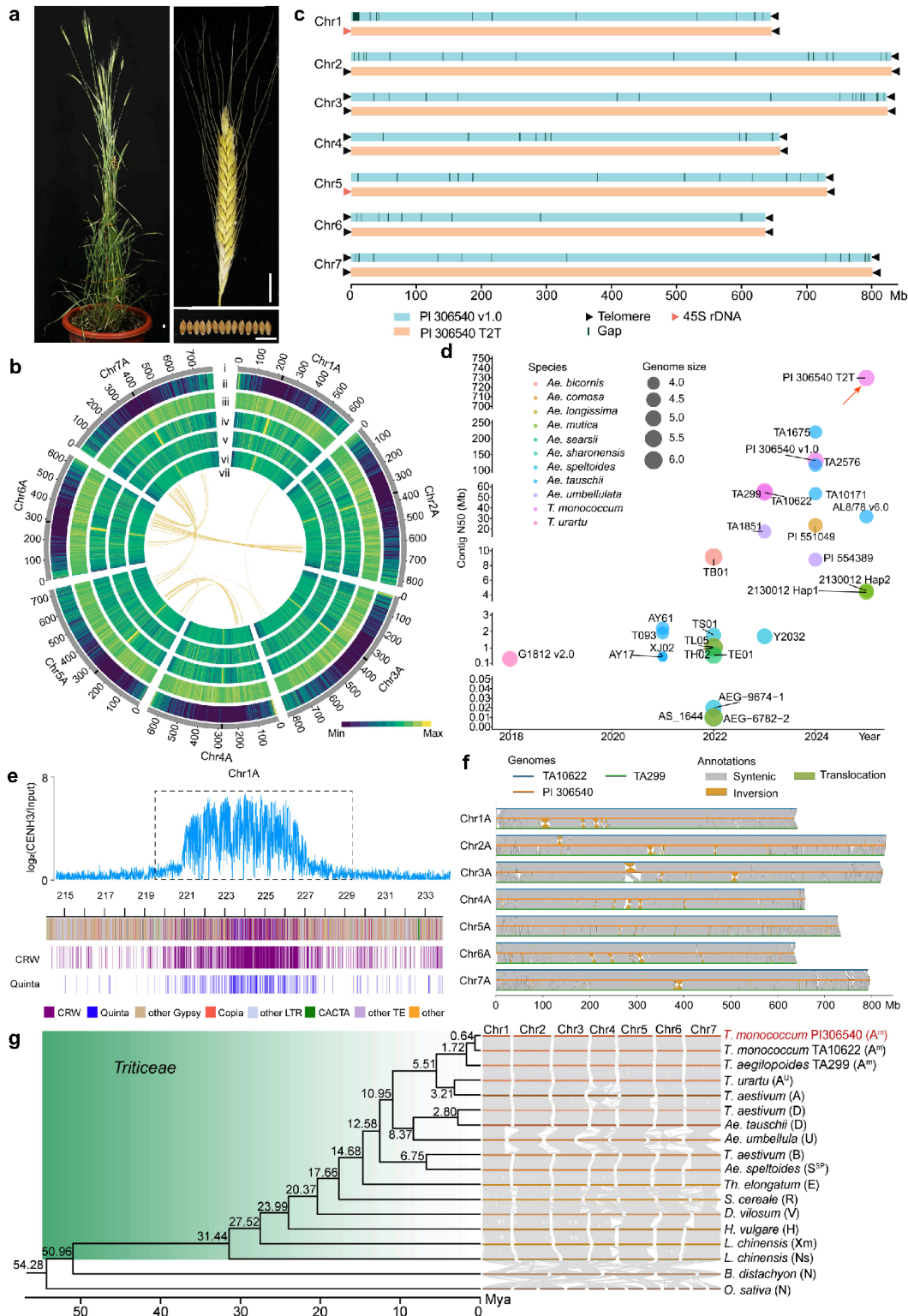


Figure 1

Telomere-to-telomere genome assembly of *T. monococcum* accession PI 306540. **a**, Morphological feature of PI 306540. Scale bar, 1cm. **b**, Circos plot of the PI 306540 genome assembly. Rings from outside to inside: (i) chromosomes, (ii) gene density, (iii) GC content, (iv) repetitive sequence density, (v) *Copia* density, (vi) *Gypsy* density, (vii) syntenic links. **c**, Comparison of genome continuity between the PI 306540 v1.0 assembly (top) and the T2T assembly (bottom). Vertical black lines mark resolved gaps; triangles indicate the telomeres and 45S rDNA regions. **d**, Contig N50 comparison of the PI 306540 T2T assembly with other published diploid wild wheat relatives. **e**, Centromere identification on chromosome 1A via CENH3 ChIP-seq. The track shows the CENH3 $\log_2(\text{ChIP}/\text{Input})$ signals (blue peaks) in 50-bp windows. Distributions of the centromeric retrotransposons CRW and Quinta are shown below. **f**,

Schematic diagram of major structural variants (>100 kb) among *T. monococcum* accessions TA10622 (blue), PI 306540 (orange), and TA299 (green). **g**, Phylogenetic and syntenic analysis of Triticeae species. The Triticeae clade is shaded in green. Syntenic relationships across genomes are displayed on the right.

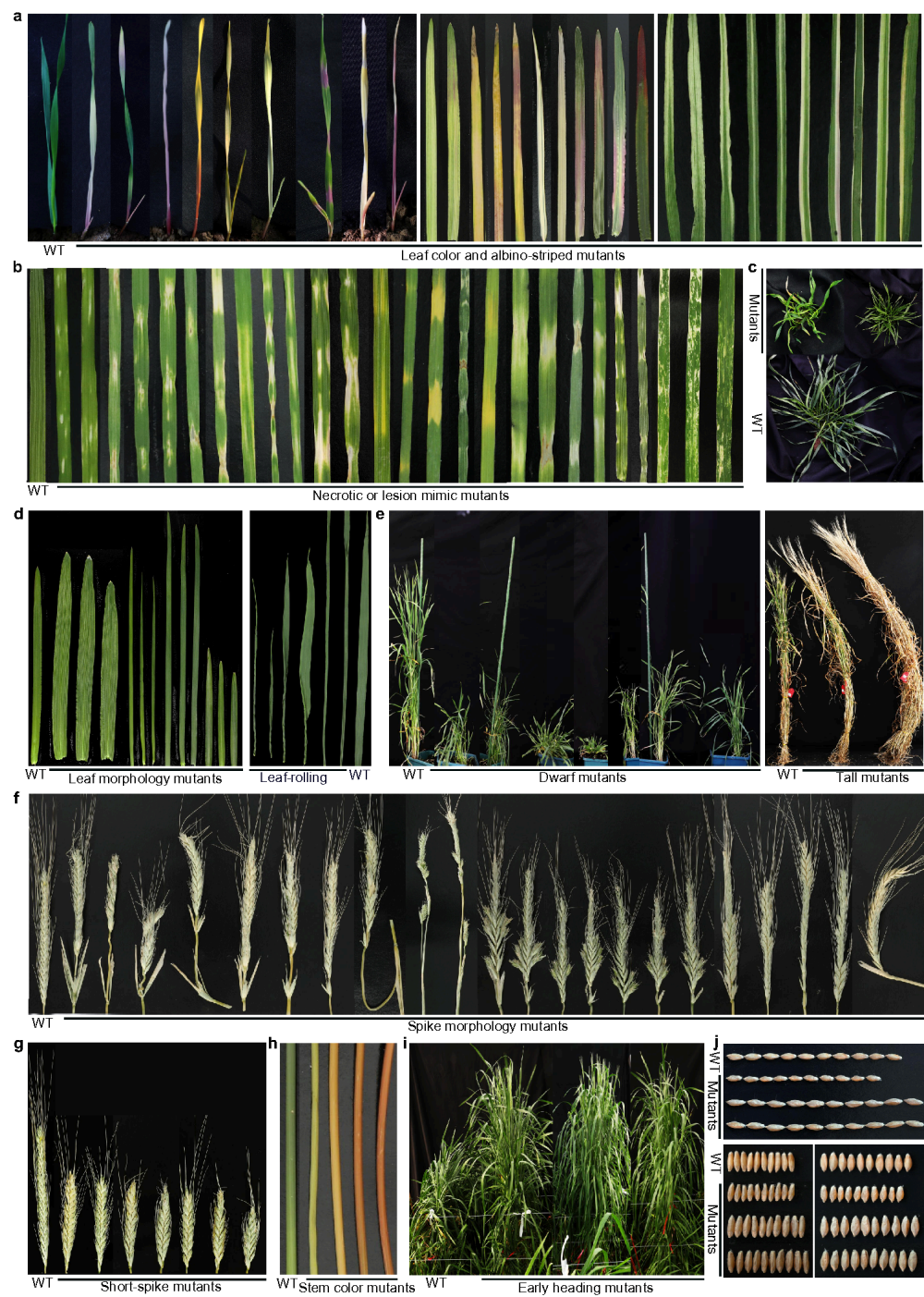


Figure 2

Representative mutant phenotypes from the EMS-mutagenized library of PI306540. **a**, Leaf color and albino-striped mutants. **b**, Necrotic and lesion mimic mutants. **c**, Plant architecture mutants. **d**, Leaf morphology and leaf-rolling mutants. **e**, Plant height

mutants. **f**, Spike morphology mutants. **g**, Short-spike mutants. **h**, Stem color mutants. **i**, Heading date mutants. **j**, Grain size and morphology mutants.

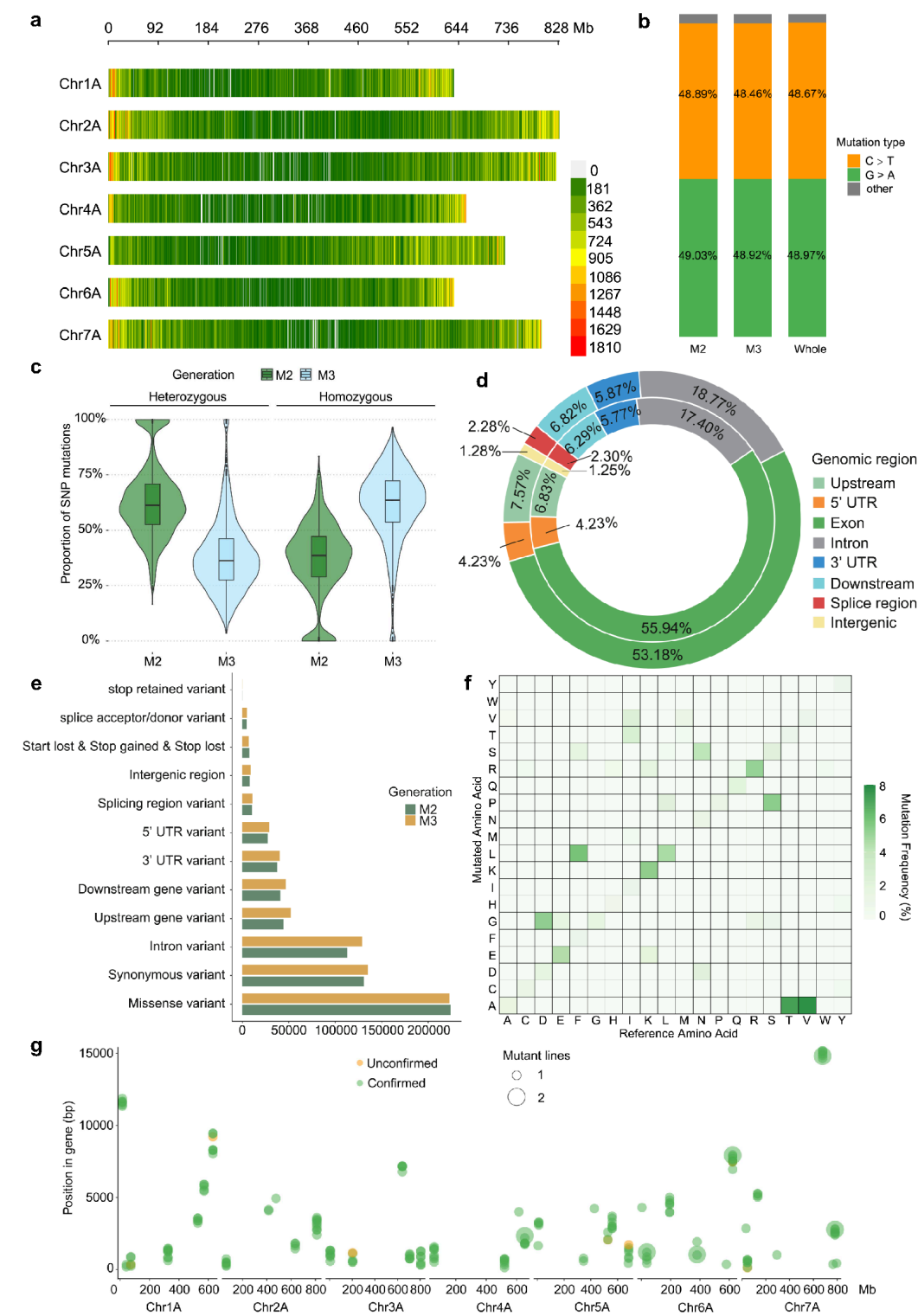


Figure 3

Characterization of mutations in the diploid wheat PI306540 mutant library. **a**, Genome-wide distribution of identified mutations across the PI306540 T2T genome assembly. **b**, Proportions of nucleotide substitution types in the M₂, M₃, and combined mutant populations. Orange and green represent C>T and G>A, respectively; gray indicates other mutation types. **c**, Distribution of homozygous and heterozygous SNP mutations across mutant lines in the M₂ (green) and M₃ (light blue) populations. In the violin

plots, the width indicates the data density. The overlaid box plots show the median (horizontal line), the 25th and 75th percentiles (box edges), and whiskers that extend to the most extreme data points within 1.5× the interquartile range (IQR). **d**, Distribution of EMS-induced SNPs across genomic regions. Mutations identified in the M₂ population are plotted on the inner circle, and those in the M₃ population on the outer circle. **e**, Predicted functional effects of SNP mutations. Mutations from the M₂ and M₃ populations are shown in green and orange, respectively. **f**, Amino acid substitutions resulting from EMS-induced mutations. The matrix displays the substitution patterns from reference amino acids (x-axis) to mutated amino acids (y-axis). The color intensity in each cell represents the frequency of mutations leading to a specific amino acid change. **g**, Validation of randomly selected mutations from the mutant library by Sanger sequencing. Confirmed mutations are shown in green; unconfirmed mutations are shown in orange.

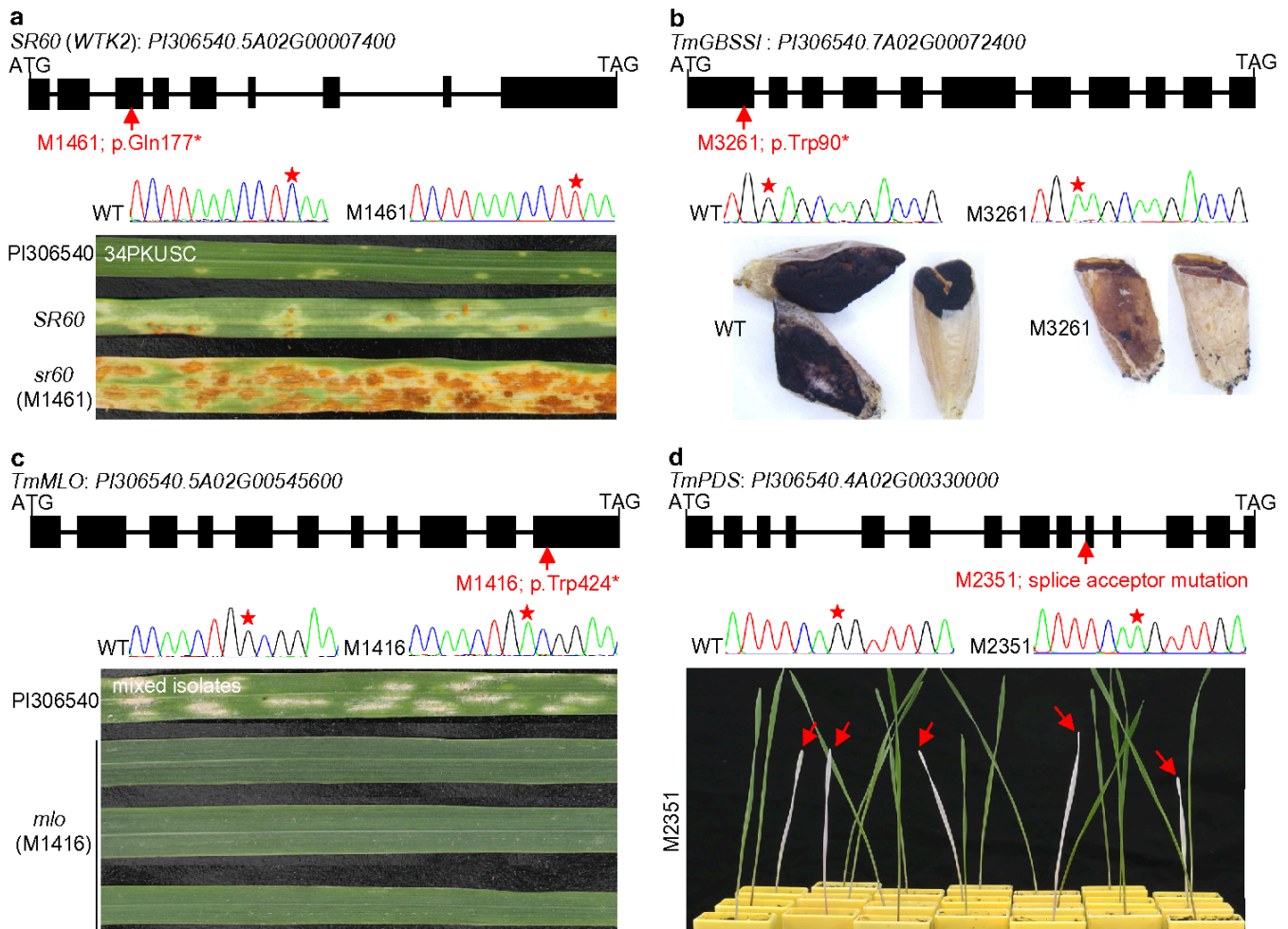


Figure 4

Functional validation of *SR60*, *TmGBSSI*, *TmMLO*, and *TmPDS* using the PI 306540 mutant library. **a**, Functional validation of the stem rust resistance gene *SR60*. Plants homozygous for the mutant allele (Q177*; line M1461) were inoculated with *Pgt* race 34PKUSC. **b**, Functional validation of *TmGBSSI*. Endosperms homozygous for the mutant allele (W90*; line M3261) exhibited reddish-brown staining with iodine. **c**, Functional validation of the powdery mildew susceptibility gene *TmMLO*. Plants homozygous for the mutant allele (W424*; line M1416) displayed strong resistance against *Bgt*. **d**, Functional validation of *TmPDS*. Plants homozygous for the frameshift mutation (G>A; line M2351) exhibited a complete albino phenotype. The mutated sites are highlighted in red and marked with red arrows and stars.

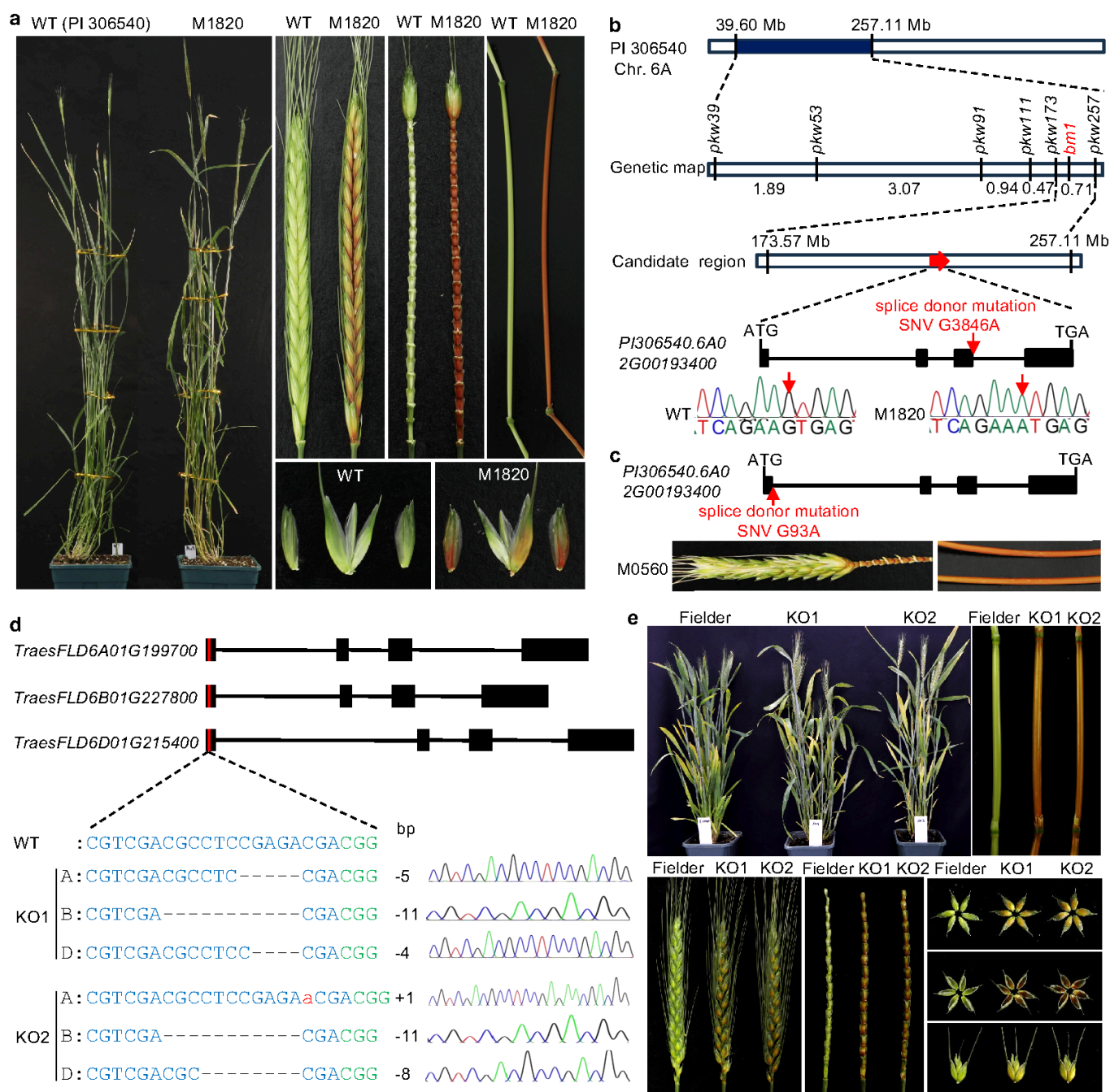


Figure 5

Cloning and functional characterization of the brown-midrib mutant in *T. monococcum*. **a**, Phenotypic comparison between PI 306540 (WT) and the brown-midrib mutant M1820. **b**, Genetic mapping of the causal gene using 212 F_2 plants derived from the M1820 $\times$ PI306540 cross. Gene models and physical positions are based on the PI 306540 T2T reference genome. The mutated sites are highlighted with red arrows. **c**, Validation of the candidate gene using an independent EMS-induced mutant line M0560, which carries a splice donor mutation (G93A). **d**, CRISPR/Cas9-mediated functional validation. The single gRNA target site is highlighted in red; sequencing chromatograms show insertion/deletion mutations at the targeted locus. **e**, Phenotypic comparison between Fielder and CRISPR/Cas9-generated knockout (KO) mutants. Photographs were taken at the flowering stage.

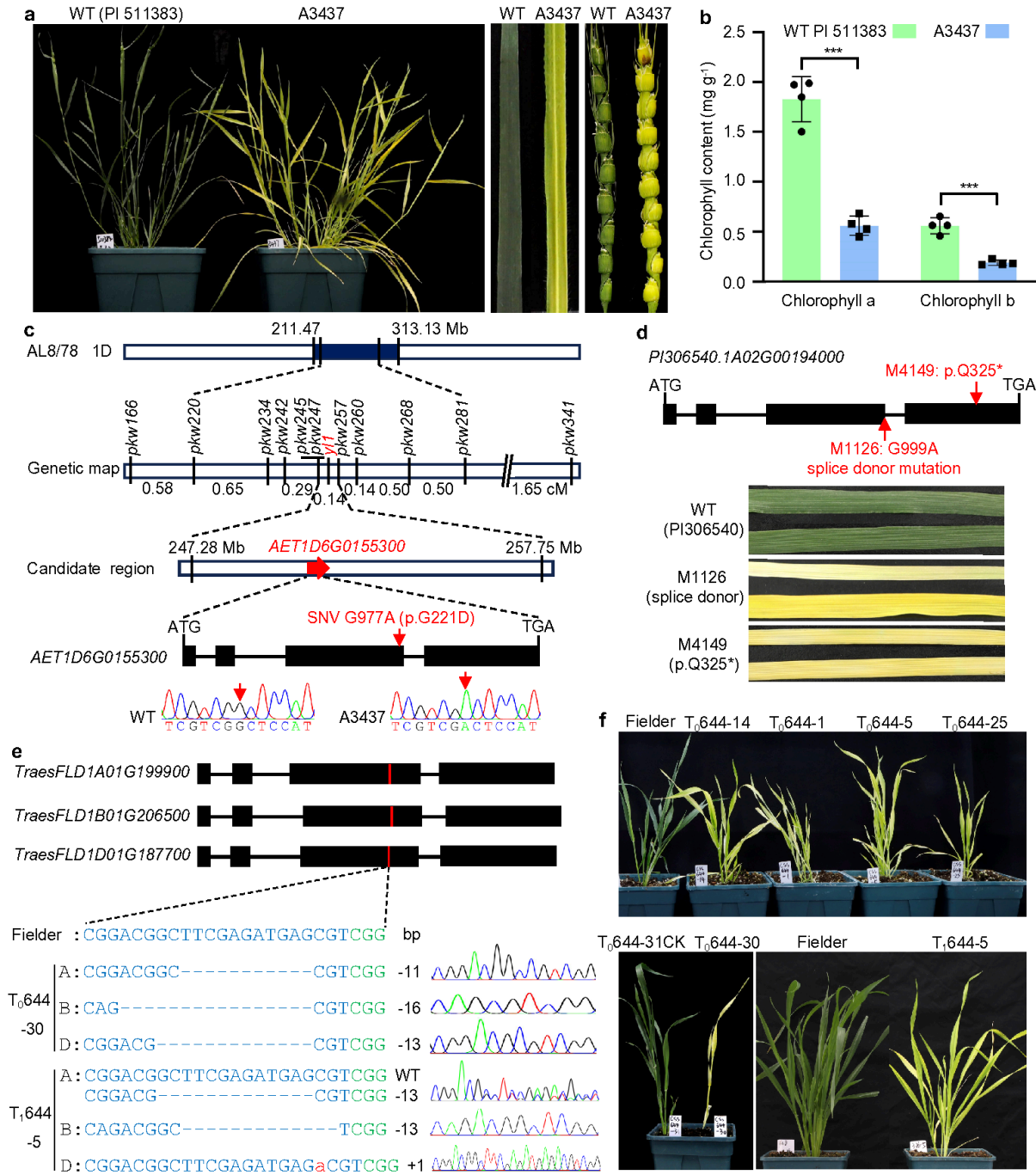


Figure 6

Cloning and functional characterization of the yellow-green leaf mutant in *Ae. tauschii*. **a**, Phenotypes of *Ae. tauschii* accession PI 511383 (WT) and the yellow-green leaf mutant A3437. **b**, Chlorophyll a and chlorophyll b contents in PI 511383 and its mutant A3437. ***, $P < 0.001$. Statistical significance was determined using a two-sided unpaired t -test ($n = 4$). Error bars represent standard errors of the means. **c**, Genetic mapping of the causal gene using 695 F₂ plants derived from the A3437 × PI 511384 cross. Gene models and physical positions are based on the AL8/78 Aet v6.0 reference genome. The mutated sites are indicated by red arrows. **d**, Validation of the candidate gene using PI 306540 EMS mutants. Mutant lines M1126 (splice donor mutation) and M4149 (Q325*) are highlighted

in red. **e**, CRISPR/Cas9-mediated functional validation. The single gRNA target site is marked in red; sequencing chromatograms show insertion/deletion mutations at the targeted locus. **f**, Phenotypic comparison between Fielder and CRISPR/Cas9-edited mutant plants.

Supplementary Files

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

- [SupplementaryData1.xlsx](#)
- [SupplementaryTables.xlsx](#)
- [SupplementaryFigures.pdf](#)