Alternative splicing is associated with tissue differentiation, subgenome divergence, and agronomic trait regulation in hexaploid wheat

Hongyong Dai[1, 4], Pingchuan Deng[2, 4], Kai Wang[3], Baolian Lv[1], Zhe Yang[1], Chuizheng Kong[1], Zhencheng Xie[1], Jizeng Jia[1], Chuan Xia[1], Fei Du[1], Xu Liu[1*], Xiuying Kong[1*] and Lichao Zhang[1*]

[1]State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[2]College of Agronomy, Northwest A & F University, Yangling, Shaanxi 712100, China

[3]Xi'An Shansheng Biosciences Co., Ltd., Xi'an 710000, China

[4]The authors contribute equally to this work

*Corresponding authors: Lichao Zhang (zhanglichao@caas.cn), Xiuying Kong (kongxiuying@caas.cn), and Xu Liu (liuxu03@caas.cn).

Short title: Profile and function of AS in wheat

The authors responsible for the distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) are Lichao Zhang (zhanglichao@caas.cn), Xiuying Kong (kongxiuying@caas.cn), and Xu Liu (liuxu03@caas.cn).

ABSTRACT

Alternative splicing (AS) is a crucial post-transcriptional regulatory mechanism that enhances transcript and proteome diversity. However, AS in common wheat (*Triticum aestivum*) remains understudied due to the large and complex genome of this crop. Full-length transcriptome sequencing, which provides long, high-quality reads, offers a powerful tool for analyzing AS in wheat. In this study, we used the PacBio Sequel platform to sequence full-length transcripts from five wheat tissues (root, stem, leaf, spike, and grain) of the cultivar Aikang58 (AK58). We identified 560,631 isoforms from 86,073 genes, with 76.7% of genes producing multiple isoforms and 45.34% undergoing AS events (ASEs). Tissue-specific analysis revealed differences in the number and function of AS genes (ASGs), underscoring the potential role of AS in tissue differentiation. A comparison across the three wheat subgenomes showed similar numbers of ASGs and ASEs but distinct functional patterns, suggesting that AS is involved in subgenomic divergence. We also examined AS in genes linked to key agronomic traits, demonstrating association with trait regulation. These findings enhance our understanding of the adaptability and post-transcriptional gene regulation in wheat, offering insights for future research and breeding efforts.

KEYWORDS

wheat, full-length transcriptome, alternative splicing, tissue differentiation, subgenome divergence

INTRODUCTION

Alternative splicing (AS) is a process in which pre-mRNA is spliced in different ways to produce multiple isoforms of a gene (Wright et al., 2022). AS is widespread in higher eukaryotes, with AS genes (ASGs) found in up to 95% of *Homo sapiens*, 60.7% of *Drosophila melanogaster*, 61% of *Arabidopsis thaliana*, and 48% of *Oryza sativa* (Pan et al., 2008; Zhang et al., 2010; Graveley et al., 2011; Marquez et al., 2012). This widespread occurrence allows organisms to generate a diverse range of mRNA and protein isoforms from a limited number of genes,

significantly expanding transcriptome and proteome diversity. AS events (ASEs) are classified into seven types based on how exons are spliced: skipping exon (SE), mutually exclusive exons (MXE), alternative 5' splice site (A5), alternative 3' splice site (A3), retained intron (RI), alternative first exons (AF), and alternative last exons (AL) (Wright et al., 2022). The prevalence of these AS patterns varies among species—SE is more common in animals, whereas RI is the dominant type in plants (Ast, 2004; Marquez et al., 2012; Li et al., 2016). AS plays a pivotal role in regulating various biological processes throughout an organism's life, particularly under stress conditions such as salt, heat, and drought (Reddy et al., 2013). In plants, AS contributes to phenotypic variation and enhances adaptability to challenging environments. For example, in rice, AS of *GS3* fine-tunes grain size, influencing yield (Liu et al., 2022); in *Arabidopsis*, AS of *CONSTANS* regulates flowering in response to photoperiod changes (Wenkel et al., 2006); in eggplant, AS of *DFR* controls anthocyanin biosynthesis, which affects pigmentation and stress tolerance (Wang et al., 2022). Similarly, in wheat, AS of *Pm4* confers resistance to powdery mildew, while AS of *TaHSFA6e* helps plants adapt to heat stress, underscoring its importance in environmental adaptation (Sanchez-Martin et al., 2021; Wen et al., 2023).

Advancements in sequencing technology have continuously improved AS research. In the 1990s, researchers used expressed sequence tags (ESTs) from cDNA libraries to identify transcripts and gene expression levels, revealing the prevalence of AS in eukaryotes (Boguski et al., 1993; Modrek and Lee, 2002). In the early 2000s, DNA microarrays enabled large-scale AS studies by detecting ASEs across different species and conditions (Lee and Roy, 2004). Around the same time, RNA sequencing (RNA-seq) emerged as a powerful method for transcriptome-wide AS analysis (Pan et al., 2008; Dong et al., 2018; Yu et al., 2020). By generating high-throughput, short-read sequencing data, RNA-seq, combined with bioinformatics tools, allowed researchers to identify ASEs across various biological processes, tissue types, and physiological states. In the past decade, full-length transcriptome sequencing, also known as third-generation sequencing, has become an essential tool for AS research. Technologies like PacBio Single Molecule Real-Time (SMRT) sequencing and Oxford Nanopore sequencing directly read full-length RNA molecules, eliminating the need for amplification and fragment assembly (Tilgner et al., 2014; Cartolano et al., 2016). Compared to RNA-seq, these methods provide longer read lengths and greater sensitivity, allowing for a more detailed analysis of AS (Dong et al., 2015; Abdel-Ghany et al., 2016).

Bread wheat (*Triticum aestivum*) is an allohexaploid species with a large and complex genome (Xiao et al., 2022). Compared to model plants like *O. sativa* and *A. thaliana*, wheat genomic research has progressed more slowly. While the sequencing of the wheat genome has provided a strong foundation, AS studies in wheat remain limited (Yu et al., 2020; Gao et al., 2021; Sanchez-Martin et al., 2021; Wen et al., 2023). Recent advances in full-length transcriptome sequencing offer an opportunity to analyze AS in wheat more comprehensively. This study aims to generate a high-quality, full-length transcriptome of the wheat cultivar Aikang58 (AK58) and analyze its AS landscape. By comparing AS patterns across different tissues and subgenomes, we aim to explore how AS contributes to tissue differentiation or subgenome divergence. Our findings will help identify functional genes associated with key agronomic traits, such as yield, disease resistance, and stress tolerance. This knowledge will support wheat breeding strategies and contribute to the development of improved cultivars.

RESULTS

Construction and characterization of the full-length transcriptomes

To investigate AS in AK58, nine PacBio SMRT libraries were sequenced, covering leaf, stem, and root at the seedling stage, spike at young spike developmental stages, and grain tissues at various developmental stages (4, 10, 15, 20, and 25 days after pollination [DAP]). Each sample generated over 43 million subreads. The raw sequencing data were processed by removing barcode information, primer sequences, and chimeric reads. Next, clustering, collapsing, and filtering steps were applied to construct the full-length transcriptome. To generate a comprehensive grain-specific transcriptome, transcriptomes from different grain developmental stages were merged and deduplicated. Similarly, all tissue transcriptomes were combined and deduplicated to produce a consolidated full-length transcriptome. The final dataset includes transcriptomes for five distinct tissues, grain transcriptomes spanning five developmental stages, and an overall transcriptome (Supplementary Table S1).

To characterize the transcriptome, we analyzed its composition using the AK58 reference genome and its annotation. Isoforms were categorized based on their splicing junctions and structural variants relative to the reference transcripts. In total, 86,073 genes and 560,631 isoforms were identified. Each tissue expressed a diverse set of isoforms, with tens of thousands of unique isoforms and corresponding genes detected in the root, stem, leaf, spike, and grain

(Fig. 1; Supplementary Fig. S1). Previously unannotated isoforms from previously unannotated identified genes were categorized as antisense and intergenic, whereas previously unannotated isoforms of annotated genes were primarily classified as Novel in Catalog (NIC) and Novel Not in Catalog (NNC).

Functional annotation was performed to predict the potential roles of these isoforms. Of the 560,631 isoforms analyzed, 546,462 (97.47%) matched the NCBI-nr database, 416,501 (74.29%) aligned with Swiss-Prot, 436,981 (77.94%) were assigned to Pfam, 239,851 (42.78%) were associated with Gene Ontology (GO) terms, and 262,203 (46.77%) were linked to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. These annotations provided valuable insights into molecular functions, biological pathways, and protein associations. Additionally, transcription factor (TF) prediction identified 19,246 potential TF isoforms, which may play crucial roles in gene expression regulation.

Analysis of AS in the AK58 genome

Genes with multiple isoforms (MIGs) are not solely the result of true AS. Variations in transcription initiation and termination sites, along with other post-transcriptional modifications, contribute to isoform diversity. However, genes with a higher number of isoforms are more likely to undergo AS. Thus, analyzing MIGs within full-length transcriptomes provides insights into AS potential across the genome. In this study, 560,631 isoforms were identified from 86,073 expressed genes (EGs) in the AK58 genome, of which 66,033 (76.7%) were classified as MIGs (Fig. 2A). Among these, 21,432 (76.56%), 22,460 (76.59%), and 21,605 (77.30%) MIGs were found in the A, B, and D subgenomes, respectively. Based on previously defined homoeologous gene groups (Jia et al., 2023), three representative categories were selected for analysis: singleton (A:B:D configuration of 1:0:0, 0:1:0, or 0:0:1), dyad (1:1:0, 1:0:1, or 0:1:1), and triad (1:1:1). A total of 1,852 singleton, 5,184 dyad, and 15,752 triad gene sets were included, with each set comprising all homoeologous genes within the corresponding group. MIGs accounted for approximately 67%, 77%, and 85% of EGs in each category, respectively (Supplementary Table S2).

Splice sites play a critical role in isoform diversity, making the identification of splice junctions essential for understanding AS. Using the SQANTI3 pipeline, 3,236,161 splice

junctions were detected in the AK58 genome. Among these, 2,547,309 matched reference transcripts (known splice junctions), whereas 688,852 were unannotated (unannotated splice junctions). These junctions were further classified based on intron sequence type: canonical (GT-AG, GC-AG, and AT-AC intron sites) and non-canonical (all other combinations). All known splice junctions were canonical, whereas among unannotated splice junctions, 652,979 (94.8%) were canonical, and 35,873 (5.2%) were non-canonical. The distribution of known and unannotated splice junctions varied among isoform categories (Fig. 2B). All canonical splice junctions were present in Full Splice Match (FSM) and Incomplete Splice Match (ISM) isoforms, as these were derived entirely from known splicing events. Since NIC isoforms contain known splice sites in unannotated combinations, 98.3% of their splice junctions were canonical, with only 1.7% being unannotated canonical junctions. In contrast, 23.3% of NNC splicing junctions were unannotated, including both unannotated canonical and non-canonical types. Notably, unannotated non-canonical splice junctions were exclusive to NNC isoforms. Additionally, genic genomic, antisense, and intergenic isoforms—categorized as unannotated—contained 100% unannotated canonical splice junctions. Fusion isoforms exhibited a nearly equal distribution of known and unannotated canonical splice junctions, spanning two distinct annotated loci.

Using SUPPA software, 39,026 ASGs were identified, accounting for 45.34% of EGs, along with 213,448 ASEs (Table 1, Supplementary Table S3). The frequency of AS varied by tissue, with 8,707, 12,939, 7,593, 10,591, and 29,344 ASGs in the root, stem, leaf, spike, and grain, respectively. Among these, 1,369, 1,810, 1,451, 1,261, and 12,338 ASGs were tissue-specific, with AS detected exclusively in a single tissue (Fig. 2C). Across the A, B, and D subgenomes, 12,674 (45.28%), 13,472 (45.94%), and 12,598 (45.08%) ASGs were identified, corresponding to 69,245, 75,868, and 66,843 ASEs, respectively (Supplementary Table S4). Meanwhile, singleton, dyad, and triad genes contained 722 (38.98%), 3,637 (41.86%), and 21,185 (50.37%) ASGs, respectively (Supplementary Table S2). Notably, 829, 875, and 846 triad genes were identified as subgenome-specific genes, exhibiting ASEs in only one subgenome (A, B, or D) (Fig. 2D).

The chromosomal distribution of MIGs, ASGs, and ASEs mirrored that of EGs, with fewer genes near centromeres and more genes towards chromosome ends (Fig. 2E). Although EG and

ASG counts were lower near centromeres, the ASGs/EGs ratio was higher in these regions, suggesting that genes near centromeres were more likely to undergo AS. RI events were the most prevalent ASE type, showing reduced occurrence near centromeres, whereas other ASEs (A3, A5, AF, AL, MX, SE) displayed no significant decrease in these regions. For example, AL events were more frequent near the centromeres of chromosomes 4A, 6A, and 6D. Despite EG counts varying across different tissues, MIG and ASG distributions, as well as ASE tissue preferences, remained largely consistent. Analysis across 21 chromosomes revealed a uniform distribution of MIGs and ASGs in root, stem, leaf, spike, and grain, with reduced occurrences near centromeres (Supplementary Fig. S2). The proportions of each ASE type remained stable across tissues, with A3, A5, AF, MX, RI, and SE accounting for approximately 20%, 15%, 8%, 3%, 0.2%, 40%, and 10% of ASEs, respectively (Table 1).

Protein-level analyses showed that 93.6% of ASEs retained coding potential, of which 58.8% altered amino acid sequences, 32.3% modified functional domain architecture, and 45.8% shifted translation termination sites (Supplementary Fig. S3). Among AS types, A3, A5, and SE events exerted the strongest effects on protein sequences, with alteration rates of 66.6%, 61.7%, and 66.4%, respectively.

AS facilitates tissue differentiation

Gene expression is intricately linked to tissue development and physiological functions; however, the relationship between AS and gene expression across different tissues remains poorly understood. To investigate the role of AS in tissue differentiation, AS patterns were compared across five tissues (root, stem, leaf, spike, and grain), and the association between tissue-specific ASGs and physiological functions was analyzed. Genes with a higher number of isoforms exhibited greater AS potential, suggesting a more complex regulatory landscape. The number of isoforms per gene varied significantly across tissues, with highly active ASGs displaying tissue-specific patterns (Fig. 3A). A substantial number of ASGs were identified, with a notable exhibiting tissue specificity (Fig. 2C; Table 1).

To understand functional differences, GO and KEGG enrichment analyses were performed on genes with a higher number of isoforms in each tissue on tissue-specific ASGs. The results aligned with those of tissue-specific EGs, with enriched pathways closely associated with tissue

development and physiological functions (Fig. 3B-C; Supplementary Fig. S4 and S5). In the root, ASGs were primarily associated with secondary metabolism and transport processes, consistent with their role in nutrient uptake and stress response. In the stem, enriched ASGs were linked to cell structure and transport, reflecting their function in structural support and nutrient translocation. In the leaf, pathways linked to photosynthesis and metabolism were highly enriched, supporting its primary role in energy production. In the spike, ASGs were associated with cell growth and hormone signaling, highlighting their role in reproductive development. In the grain, genes involved in starch and sugar metabolism were significantly enriched, underscoring its function in nutrient storage.

These findings suggest that AS may be involved in regulating tissue-specific gene expression and contribute to functional specialization. The observed variation in AS patterns across tissues reflects adaptive transcriptomic modulation in response to developmental or physiological demands. These results provide insights into the complexity of AS-mediated gene regulation and its potential involvement in tissue differentiation during wheat development.

AS is associated with subgenome divergence

A comparison of AS patterns across the A, B, and D subgenomes revealed a similar number of identified ASGs and ASEs (Supplementary Table S4), suggesting limited variation in AS among the three subgenomes. Moreover, the distribution of ASE types remained consistent across subgenomes, indicating no subgenome-specific preference for splicing types. These findings point to a similar AS landscape across the three subgenomes.

To further characterize subgenome-specific AS, we examined the conservation and divergence of AS types across 15,752 triad gene sets (Supplementary Table S5). Based on AS-type concordance among homoeologs, triads were classified into five categories (Fig. 4A): All_non-AS, in which none of the three homoeologs exhibited AS (6,114; 38.8%); Partially_non-AS, in which at least one homoeolog lacked AS or was not expressed (5,179; 32.9%); Partially_conserved, in which all three homoeologs showed AS but with only partially overlapping AS types (4,063; 25.8%); Conserved, in which all three homoeologs displayed identical AS types (358; 2.3%); and Divergent, in which all three homoeologs exhibited completely distinct AS types (38; 0.2%). This distribution mirrors the subgenome-based AS gene

patterns shown in Fig. 2D, indicating that many homoeologs have gained or lost splicing regulatory functions in specific subgenomes. Although overall numbers of ASGs and ASEs were comparable among subgenomes, homoeolog-level AS patterns showed substantial divergence. The rarity of fully conserved AS and the dominance of partially conserved or non-AS categories suggest extensive regulatory differentiation among homoeologs, likely contributing to subgenome functional diversification following polyploidization.

Then, we analyzed isoform composition of three subgenomes and subgenome-specific ASGs in triad genes. Among 15,752 triad gene sets, isoform numbers were quantified for each homologous gene, and the ratio of isoforms from each subgenome to the total isoforms in the set was calculated. Based on these ratios, triad gene sets were categorized as dominant, suppressed, or balanced. If two subgenomes had isoform ratios below 20%, the remaining subgenome was classified as dominant. Conversely, if two subgenomes had isoform ratios above 20% and the third subgenome's ratio was below 20%, the subgenome with the lower ratio was classified as suppressed. If none of these conditions were applied, the triad gene set was categorized as balanced. The results identified 720, 707, and 713 dominant triad gene sets in the A, B, and D subgenomes, respectively. Similarly, 2,241, 2,215, and 1,914 suppressed triad gene sets were identified in the A, B, and D subgenomes (Fig. 4B-C). Among the 15,752 triad gene sets with AS, 829, 875, and 846 contained ASEs unique to the A, B, or D subgenomes, respectively (Fig. 2D). Functional enrichment analysis of dominant, suppressed, and subgenome-specific ASGs revealed distinct pathway enrichment patterns (Fig. 4D-E; Supplementary Fig. S6). The A subgenome exhibited a high AS activity in nucleic acid regulation and functional development but lower AS activity in photosynthesis-related functions. The B subgenome showed increased AS activity in secondary metabolism and stress responses, while AS activity in nucleic acid metabolism was lower. The D subgenome demonstrated significant AS activity in cell cycle regulation and had reduced AS activity in metabolic regulation. These findings suggest that although AS patterns exhibit a similar landscape across the three subgenomes, functional divergence is evident across specific genes in each subgenome.

AS is involved in trait regulation

The role of AS in controlling agronomic traits in wheat has been partially explored, though only a small fraction of the identified ASEs have been investigated (Guo et al., 2020; Sanchez-Martin et al., 2021; Zhang et al., 2022; Wen et al., 2023). To further examine the relationship between AS and agronomic traits, AS patterns were analyzed during grain development, and their potential influence on grain traits was assessed.

The serine/arginine (SR) protein family, a key group of splicing factors involved in gene splicing regulation, plays a critical role in AS (Shepard and Hertel, 2009; Xie et al., 2022). A BLAST search of plant SR protein sequences against the AK58 protein database, followed by domain analysis, identified 59 candidate SR protein family genes in the AK58 genome. These genes contained RNA-binding domains (RBDs) at the N-terminus and arginine/serine-rich domains (RSs) at the C-terminus (Xie et al., 2022). RNA-seq data from different developmental stages of AK58 grain were used to quantify SR protein family gene expression. The results revealed a gradual decline in overall SR gene expression during grain development (Fig. 5A), indicating a reduced AS activity.

Analysis of the full-length transcriptome of grains at different developmental stages further demonstrated a progressive decrease in the number of MIGs (Fig. 5B). Genes with a higher number of isoforms were significantly more abundant during early stage (4DAP and 10DAP) compared to late stage (15DAP, 20DAP, and 25DAP) (Fig. 5C). Count and ratio analysis of ASGs from early and late stage confirmed significantly higher AS activity during early grain development (Fig. 5D; Supplementary Table S6). Functional enrichment analysis of early- and late-stage ASGs revealed distinct pathway enrichments (Fig. 5E-F). Early-stage ASGs were predominantly involved in growth, cell expansion, and photosynthesis, whereas late-stage ASGs were enriched in pathways related to macromolecule processing and transcriptional regulation. These findings suggest that AS plays a critical role in regulating developmental processes, with early-stage AS contributing to growth and late-stage AS facilitating transcriptional modifications associated with grain maturation.

To further understand the relationship between AS and grain development, the isoform profiles of grain trait-related genes were analyzed across developmental stages (Fig. 5G). Based on previously reported wheat grain trait-related genes, 98 homologous genes were identified in

the AK58 genome (Wei and Jiao, 2024). These genes exhibited higher AS potential during early grain development like the overall trend, suggesting that AS more actively influences key traits related to grain formation and maturation.

We next constructed a gene co-expression network using transcriptome data across multiple developmental stages, incorporating all ASGs, TF genes, and SR protein family members. In total, 20 co-expression modules comprising 25,271 expressed genes were identified (Supplementary Fig. S7). Four modules were significantly and positively correlated with AS activity, MEgreenyellow ($r = 0.96$, $p < 0.05$), MEred ($r = 0.85$, $p < 0.05$), MEblue ($r = 0.83$, $p < 0.05$), and MEturquoise ($r = 0.80$, $p < 0.05$). MEgreenyellow showed the strongest association with AS activity and contained 323 hub genes, including 25 TFs and 2 SR family members (Supplementary Fig. S7; Supplementary Table S7). A regulatory network centered on these two SR genes revealed that 87.2% (511/586) of module genes exhibited highly similar expression patterns with strong topological overlap (TOM > 0.2). These results indicate that SR proteins likely exert broad regulatory control within this module, acting as key splicing factors coordinating downstream AS events and potentially driving the dynamic AS changes observed during grain development.

Beyond grain development, AS patterns were examined in genes associated with key agronomic traits, including tillering, grain shape, grain weight, and disease resistance (Supplementary Table S8 and S9). A significant proportion of these genes underwent AS, with certain genes, such as *ARF*, *Fhb1*, *IAA21*, *Q*, *Sus1*, and *Sus2*, exhibiting a high number of isoforms and diverse AS types. To validate the presence of these isoforms and ASEs, the *Q-5A* gene—a critical domestication gene regulating plant height, rachis length, and spike morphology—was analyzed. Three *Q* transcripts had previously annotated in the A subgenome (TraesCS5A02G473800.1, TraesCS5A02G473800.2, and TraesCS5A02G473800.3), with TraesCS5A02G473800.1 identified as the major transcript. In this study, 24 *Q-5A* isoforms were identified, undergoing ASEs of the A3, RI, and SE types. Visualization using Integrated Genomics Viewer (IGV) identified 10 *Q-5A* isoforms in stem tissue (Fig. 6A). To confirm these isoforms, primers spanning the first and last exons were designed (Fig. 6A; Supplementary Fig. S8). Among the PCR products, five distinct bands corresponding to target isoforms were observed and excised for Sanger sequencing (Fig. 6B; Supplementary Table S10). The resulting

sequences matched the PacBio full-length isoforms, confirming their presence. Aside from the major transcript, the remaining isoforms represented findings differing from previously annotated transcripts and potentially carrying uncharacterized functions. These results further underscore the regulatory role of AS in wheat agronomic traits and developmental processes.

To investigate the functional relevance of *Q-5A* isoforms, we focused on spike morphology and analyzed their transcriptional profiles. Using public RNA-seq data from 90 wheat accessions at the double ridge stage (Wang et al., 2017), we quantified expression of the *Q-5A* gene and all 24 *Q-5A* isoforms identified in this study. Each accession expressed 4–9 isoforms (Supplementary Table S11), and 15 isoforms were detected in spike tissues (Fig. 6C; Supplementary Table S12 and S13). These were designated *Q-5A*_isoform.1 to *Q-5A*_isoform.15, with *Q-5A*_isoform.1 corresponding to the reference transcript *TraesCS5A02G473800.1* and the remainder representing previously unidentified splice variants. Overall *Q-5A* expression was significantly correlated with spike-related traits (Fig. 6D). To evaluate isoform-specific effects, we assessed correlations between the relative abundance of each isoform and spike morphology traits (Fig. 6E; Supplementary Table S14). Using this proportion-based metric, *Q-5A*_isoform.6, *Q-5A*_isoform.11, and *Q-5A*_isoform.12 showed significant trait associations. By contrast, isoform number per accession was not significantly correlated with spike morphology (Supplementary Table S14), indicating that isoform composition, rather than AS frequency alone, underlies phenotypic variation. These results suggest that distinct *Q-5A* isoforms exert divergent regulatory roles during spike development, with highly expressed isoforms (e.g., *Q-5A*_isoform.1, 2, 3) likely fulfilling structural or constitutive functions, whereas low-abundance, trait-associated isoforms (e.g., *Q-5A*_isoform.6, 11, 12) may mediate fine-scale or genotype-specific regulatory effects.

To further evaluate the role of AS in trait regulation, we analyzed isoform–trait associations for four additional spike development genes (*DEP1-5D*, *FUL2-2A*, *Ppd-2D*, and *TaCol-B5-7B*). Using the same 90-accession RNA-seq dataset, we quantified isoform expression and calculated each isoform's proportion relative to total gene expression (Supplementary Table S12-S13). Spearman correlation analysis identified multiple significant associations between isoform proportions and spike traits (Supplementary Table S14; Supplementary Fig. S9). Specifically, *Ppd-2D*_isoform.13 was positively correlated with floret number ($r = 0.30$, $p = 0.005$), and *Ppd-*

*2D*_isoform.3 was positively associated with spikelet number (r = 0.28, $p = 0.010$). Conversely, *TaCol-B5-7B*_isoform.2 showed a negative correlation with floret number (r = −0.24, $p = 0.021$), and *FUL2-2A*_isoform.2 was negatively associated with spikelet number (r = −0.21, $p = 0.046$). Collectively, these findings demonstrate that AS-mediated isoform variation across multiple spike development genes is associated with phenotypic diversity in spike architecture, supporting a broad role for AS in the regulation of agronomic traits.

DISCUSSION

AS, a crucial post-transcriptional regulatory mechanism, plays a fundamental role in various biological processes (Wright et al., 2022). Recent advancements in sequencing technologies have greatly enhanced wheat genomic research, providing essential resources for studying AS in gene expression (Jia et al., 2023; Liu et al., 2024). In particular, full-length transcriptome sequencing offers substantial advantages in AS detection. Cross-platform validation in this study revealed that Iso-Seq identified approximately 7.1-fold more ASEs than short-read RNA-seq, while robustly capturing the majority of splicing events detectable by RNA-seq platforms (Supplementary Fig. S10). In wheat, AS research has primarily focused on stress responses, whereas studies examining AS differences across tissues and subgenomes remain limited (Sanchez-Martin et al., 2021; Zhang et al., 2022; Wen et al., 2023). In this study, full-length transcriptome analysis was conducted across different tissues and developmental stages of wheat, leading to the identification of numerous ASGs and ASEs. These findings provide a comprehensive overview of the AS landscape in wheat. Genome-wide analysis revealed the widespread and complex nature of AS, underscoring its crucial role in tissue and subgenome divergence. This study establishes a foundation for future research exploring the contribution of AS to agronomically important traits, such as yield and grain quality. Moreover, the resulting full-length transcriptome dataset represents a valuable resource for functional genomics and molecular breeding by enabling isoform-level analysis of gene regulation and trait-associated variants.

Universality and diversity of AS in wheat

The widespread occurrence of AS in wheat is evident from genome-wide analysis, which revealed that 76.7% of genes produce multiple isoforms, while 45.34% exhibit ASEs. Moreover, protein-level analysis revealed that a substantial proportion of ASEs led to changes in protein sequences or functional domains.These findings highlight AS as a pervasive mechanism that increases gene expression complexity and serves as a key strategy for meeting the developmental and physiological needs of the plant.

AS diversity in wheat is reflected in differences in AS patterns across tissues, subgenomes, and developmental stages. First, the number of ASGs and ASEs varies among tissues, with higher AS activity observed in young tissues exhibiting high transcriptional activity. The functions and biological pathways associated with ASGs differ between tissues, often aligning with their developmental and adaptive requirements. For instance, in the young stem during the seedling stage, where gene expression is particularly active, a greater number ratio of ASGs and ASEs were identified. In this stage, ASGs were enriched in pathways related to cellular structure and transport, which are crucial for stem development and its function. Second, despite overall landscape in ASG and ASE numbers across the three subgenomes, functional distribution biases were observed. AS patterns of specific genes exhibited clear differentiation across subgenomes, revealing functional diversity within the wheat subgenomes. Finally, AS diversity in wheat is also evident in its dynamic changes across developmental stages. The number and patterns of AS shift throughout grain development and maturation. AS activity was significantly higher in the early stages of grain development compared to later stages, with a greater number of ASGs and ASEs identified. Furthermore, functional differences were observed between early- and late-stage ASGs. Early-stage ASGs were primarily enriched in pathways related to growth, development, and cell expansion, whereas late-stage ASGs were enriched in pathways related to macromolecule processing pathways.

Together, these results establish AS as a central mechanism shaping transcriptomic complexity in wheat, characterized by pronounced spatial and temporal specificity. Tissue-specific ASGs are strongly enriched in pathways aligned with the core biological functions of their respective tissues, supporting a role for AS in tissue differentiation and physiological specialization. Moreover, several functional genes exhibit subgenome-specific splicing patterns,

suggesting that AS is associated with functional divergence among the A, B, and D subgenomes. AS also displays clear temporal dynamics during grain development, with elevated splicing activity at early developmental stages. This trend parallels the expression patterns of splicing factors, indicating coordinated regulation by stage-specific upstream regulators. Collectively, these findings highlight the multilayered regulatory architecture of AS and clarify its functional significance in wheat development.

Polyploidization enhances AS complexity in wheat

Wheat, a hexaploid species that has undergone two independent polyploidization events, exhibits a level of AS complexity that surpasses that of most diploid and tetraploid plants. Studies have shown that 42-61% of genes in plants undergo AS (Marquez et al., 2012; Reddy et al., 2013; Klepikova et al., 2016; Liu et al., 2024). However, AS in wheat is more complex, as evidenced by the identification of 560,631 isoforms from 86,073 EGs, exceeding the average observed in other plants. Additionally, 39,026 ASGs and 213,448 ASEs were detected in wheat, a number significantly higher than in diploid plants such as rice, corn, and soybean, as well as other polyploid species such as cotton, peanut, and zoysia (Supplementary Table S15). These findings indicate that wheat possesses a unique and highly complex transcriptome with distinct AS patterns. The greater AS complexity observed in wheat is likely caused by its larger genome size and higher gene redundancy, which provide additional opportunities for isoform generation and finer regulation of gene function.

The polyploid nature of wheat, characterized by its three subgenomes, plays a significant role in the flexibility of gene expression regulation. Among the three subgenomes, the number of ASEs and ASGs exhibits similar landscape; however, functional divergence is observed in ASGs, reflecting specialization across subgenomes. Furthermore, the distribution of AS across different homologous gene types (singleton, dyad, and triad) underscores this regulatory complexity. Triad genes, which possess homologs in all three subgenomes, exhibit the highest proportion of ASGs (50.38%), followed by dyads (41.86%) and singletons (38.98%). This pattern reflects the increased transcriptomic diversity contributed by the retention of multiple homologous copies in polyploid wheat, which may facilitate more flexible and robust regulation of gene expression. The observed AS bias among subgenomes and the differential ASG distribution among

homologous genes reveal intricate regulatory interactions within the wheat genome. These interactions enhance cooperative dynamics across subgenomes, allowing wheat to adapt flexibly to developmental and environmental challenges. Additionally, subgenome interplay in hexaploid wheat has reprogrammed AS patterns, particularly through the incorporation of the D subgenome, which significantly influences AS in the A and B subgenomes (Yu et al., 2020). This interaction further increases the complexity and flexibility of the gene regulatory network in polyploid wheat.

Polyploidization is crucial in plant evolution by providing abundant genetic resources for gene mutations, selection, and functional diversification. In wheat, the acquisition or loss of ASEs enables adaptation to diverse environmental conditions. However, AS is not uniformly distributed across the polyploid wheat genome but instead follows specific temporal, spatial, and functional biases. For example, during early grain development, ASGs are notably enriched in transport and energy metabolism pathways, whereas stress-responsive genes undergo specific ASEs under heat and drought stress (Liu et al., 2018). These findings suggest that AS optimizes key gene functions, enhancing wheat's evolutionary adaptability. Moreover, the divergence of AS patterns in triad gene sets in wheat further supports the notion that AS in polyploid plants strengthens the complexity of gene regulatory networks, providing a significant evolutionary advantage.

AS is associated with tissue differentiation and subgenome divergence

At the tissue level, distinct tissue-specific ASGs were identified across different wheat tissues. Further analysis revealed that these ASGs are closely associated with the physiological functions of specific tissues. For example, ASGs in leaves are predominantly involved in photosynthesis and metabolic pathways, whereas those in roots are enriched in secondary metabolism and transport functions. These findings indicate that AS generates diverse spliced isoforms to support the functional specialization of different tissues, enabling them to better meet their physiological demands and environmental conditions.

At the subgenome level, despite the overall similar landscape in ASG and ASE numbers among the three subgenomes, functional differences among ASGs reflected subgenome divergent evolution. For instance, genes involved in nucleic acid regulation exhibit higher AS activity in

the A subgenome, while genes related to secondary metabolism and stress response undergo more active AS in the B subgenome. In contrast, the D subgenome shows dominant AS activity in genes associated with cell cycle regulation. These AS differences among subgenomes suggest that AS facilitates functional complementation, promoting adaptive evolution under diverse environmental conditions. This underscores the role of AS in both subgenome coordination and the optimization of gene functionality and diversity, thereby enhancing the regulatory flexibility of gene expression network in wheat.

In summary, AS serves as a key factor contributing to both subgenome and tissue functional differentiation in wheat. By regulating diverse splice forms, AS enhances the flexibility and adaptability of wheat through gene regulatory mechanisms, allowing for optimized function and improved environmental resilience.

AS has potential regulatory roles in agronomic traits

AS plays a significant regulatory role in agronomic traits by generating diverse isoforms that enable precise modulation of gene expression and function. This process influences essential traits in wheat, such as plant architecture, stress resistance, and yield potential. In this study, AS of genes related to agronomic traits were analyzed, and the potential effects of AS on agronomic traits were revealed. For example, isoform–trait correlation analyses of spike-related genes, including *Q-5A*, *DEP1-5D*, *Ppd-2D*, *FUL2-2A*, and *TaCol-B5-7B*, revealed significant associations between the relative abundance of specific isoforms and spike morphology traits. These results indicate that AS-generated transcript diversity is closely associated with phenotypic variation in wheat spike architecture. Additionally, genes such as *IAA21* and *Fhb1*, which are associated with tillering and disease resistance, exhibit notable AS activity. The generation of specific isoforms in these genes suggests a role in enhancing the resilience to environmental stress and improving yield stability in wheat.

Furthermore, AS plays an important role in grain development, regulating physiological processes at different developmental stages. This study found that AS activity is significantly higher during early grain development than in later stages, closely correlating with the expression of genes involved in cell expansion and material accumulation. Such regulatory

patterns suggest that AS facilitates rapid grain growth and maturation by modulating functional gene expression over time. Additionally, key genes related to carbohydrate metabolism, such as *Sus1* and *Sus2*, exhibit distinct AS patterns that may directly influence starch biosynthesis and grain quality, thereby affecting wheat's agricultural value.

In conclusion, AS is likely involved in regulating key agronomic traits in wheat modulating the expression of essential genes, like *Q*, *IAA21*, and *Fhb1*, as well as genes involved in developmental pathways. This regulatory mechanism provides a molecular framework for improving traits such as stress resistance, tillering, and grain development. These insights offer directions for functional genomics research and establish a theoretical foundation for targeted breeding strategies in wheat improvement.

AS holds great potential for agricultural applications

A deeper understanding of the relationship between specific ASEs and agronomic traits offers opportunities for target genetic improvement in wheat. By identifying key genes and ASEs associated with growth, stress resistance, and yield, breeders can develop strategies to optimize desirable traits without altering the underlying gene structure. For instance, modifying the AS pattern of critical genes such as *Q* may enhance wheat morphology, improve resistance to pests and diseases, or increase grain yield while preserving the gene's natural sequence.

From a breeding perspective, AS represents a highly flexible and efficient tool for fine-tuning gene expression. Unlike traditional genetic modification or transgenic approaches, AS-based strategies allow precise regulation of gene function without introducing foreign DNA or inducing excessive genomic alterations. By selectively modulating AS, breeders can promote the expression of beneficial isoforms, leading to enhanced stress tolerance, improved grain quality, or increased yield potential.

MATERIALS AND METHODS

Sample selection and RNA extraction

Five tissues from the hexaploid wheat cultivar AK58 were selected, including root, stem, and leaf at the seedling stage, as well as spike at different developmental stages (W2, W3, W4, W5, W6, W7, W8), and grain at 4, 10, 15, 20, and 25 DAP (Waddington et al., 1983). Root, stem, leaf, and spike tissues were collected from greenhouse-grown plants, while grain was harvested from field-grown plants. For each tissue or developmental stage, samples were collected from at least five individual plants and pooled to form a composite biological sample, ensuring representation and reducing individual variation. Following collection, the samples were immediately flash-frozen in liquid nitrogen and stored at -80 °C. Total RNA was extracted from root, stem, leaf, spike, and grain samples using the TRIzol method (Chomczynski and Sacchi, 1987). Equal amounts of RNAs from spikes at different developmental stages were pooled to create a composite sample. Quality control assessment was performed on all RNA samples. Purity and concentration were evaluated using agarose gel electrophoresis and a Nanodrop spectrophotometer, while RNA integrity was determined using the Agilent 2100 Bioanalyzer (Schroeder et al., 2006; Desjardins and Conklin, 2010). The A260/280 ratios were approximately 1.9, and the RNA integrity number (RIN) values ranged from 7.5 to 9.4, confirming the high purity and integrity of the samples, which were suitable for library construction and sequencing.

cDNA library construction and sequencing

Polyadenylated mRNA was enriched using Oligo(dT) magnetic beads and reverse transcribed into cDNA using a SMARTer PCR cDNA Synthesis Kit. Fragment selection was performed with magnetic beads, followed by large-scale PCR amplification to generate sufficient total cDNA. The full-length cDNA underwent damage repair, end repair, and ligation of SMRT dumbbell-shaped adapters to construct the full-length transcriptome library. Exonuclease treatment was applied to remove sequences lacking ligated adapters at both ends of the cDNA. Subsequently, primers were added, and DNA polymerase was bound to the cDNA to form a complete SMRT

bell library. After quality control, full-length transcriptome sequencing was performed on the PacBio Sequel II/IIe platform.

Full-length transcriptome construction

The raw data from PacBio full-length transcriptome sequencing were processed using the IsoSeq pipeline in SMRT Link v13.0 (https://www.pacb.com/support/software-downloads). The pipeline included multiple steps, such as generating circular consensus sequences (CCS), removing primers, clustering, and collapsing redundant sequences. Default parameters were used, with key settings set as follows: minimum draft length before polishing = 10, minimum number of passes = 3, and minimum predicted accuracy = 0.99.

Subreads were extracted from each zero-mode waveguide (ZMW). If at least two full passes were detected, the subreads were corrected against each other to obtain CCS through self-correction from multiple passes of a single molecule. Sequences were then classified as full-length (FL) or non-full-length (non-FL) based on the presence of 5' primers, 3' primers, and a poly(A) tail at the 3' end. FL sequences were clustered using a hierarchical n*log(n) algorithm to generate both cluster consensus sequences and high-quality singleton consensus sequences. These were combined to obtain corrected consensus isoforms, which were subsequently aligned to the reference genome. Redundant isoforms were removed, yielding a set of non-redundant isoforms.

However, non-redundant isoforms generated by the IsoSeq pipeline may still have included false positives. To eliminate these, the SQANTI3 v5.2.1 tool was employed, applying the default filtering criteria in the *sqanti3_filter.py* pipeline under the *rules* option (Pardo-Palacios et al., 2024). The *sqanti3_rescue.py* pipeline was then executed in *full* mode with default parameters to restore isoforms that were incorrectly filtered, producing the final set of filtered isoforms.

Full-length transcriptome characterization and transcript classification

The SQANTI3 v5.2.1 tool was used to characterize the filtered isoforms (Pardo-Palacios et al., 2024). The *sqanti3_qc.py* pipeline was employed to analyze transcriptional and splicing characteristics using default parameters. This pipeline first aligned the isoforms to the AK58

reference genome and then compared their genomic positions with existing annotations (Jia et al., 2023). This comparison helped identify both annotated and previously unannotated genes and transcripts. In this study, genes and transcripts identified using the IsoSeq pipeline and SQANTI3 were labeled as "identified genes/transcripts." Those present in reference databases were classified as "annotated genes/transcripts," whereas those absent from the reference database were designated as "previously unannotated genes/transcripts." Isoforms aligned to unannotated genomic regions were considered unannotated gene loci. If multiple isoforms shared similar exon structures within the same genomic region, they were categorized under a single unannotated gene locus. Additionally, isoforms aligning to different exons of known gene regions were identified as unannotated transcripts.

SQANTI3 also classified isoforms based on splice junctions and splicing variant structures relative to reference transcripts. Isoforms were categorized into nine main types: FSM, ISM, NIC, NNC, Antisense, Genic Intron, Genic Genomic, Fusion, and Intergenic. FSM isoforms possessed the same exon number and internal splice junctions as reference transcripts. ISM isoforms lacked one or more 5' or 3' exons but maintained identical internal junctions. NIC isoforms used known splice junctions from annotated transcripts but did not match FSM or ISM criteria. NNC isoforms contained at least one unannotated exon or intron absent from the reference annotation and did not fit FSM or ISM categories. Antisense isoforms were transcribed in the opposite direction of an annotated gene. Genic Intron isoforms were entirely contained within an annotated intron, while Genic Genomic isoforms overlapped both exonic and intronic regions. Intergenic isoforms were located between annotated genes, and Fusion isoforms resulted from exon combinations of two or more distinct genes.

AS junction and events analysis

SQANTI3 analyzes AS junctions based on the splice donor and acceptor sites. Splice junctions present in reference transcripts were classified as "known splice junctions," while those absent from the reference were designated as "unannotated splice junctions." Furthermore, splice junctions were categorized as either canonical (e.g., GT-AG, GC-AG, AT-AC) or non-canonical based on the dinucleotide sequence at intron boundaries.

SQANTI3 generated an annotation file in GFF format for the filtered isoforms, which was subsequently used for the ASE identification via SUPPA v2.3. By examining exon and intron positions in both isoforms and gene annotations, SUPPA identifies seven main types of ASEs: SE, RI, MX, A5, A3, AF, and AL (Trincado et al., 2018).

Functional annotation of isoforms

The filtered isoforms were systematically annotated using multiple major databases, including NCBI-nr, Swiss-Prot, Pfam, GO, and KEGG. These databases provided complementary insights into isoform functions. The NCBI-nr and Swiss-Prot protein sequence databases facilitated functional prediction through sequence homology analysis (Boutet et al., 2007). For this purpose, a high-throughput sequence aligner, Diamond (v2.1.8), was used to perform BLAST searches against these databases (Buchfink et al., 2015). GO annotation categorized isoform functions into three domains: Cellular Component (CC), Molecular Function (MF), and Biological Process (BP), providing a structured framework for functional characterization (Gene Ontology, 2015). KEGG pathway analysis was performed to determine isoform involvement in biological processes and signaling pathways, offering insights into potential regulatory roles (Kanehisa et al., 2016). To classify protein families, the Pfam database was employed to identify conserved domains and structural features (Punta et al., 2012). Functional annotation was further enhanced using the eggNOG-mapper platform, which integrated GO, KEGG, and Pfam data (Huerta-Cepas et al., 2019; Cantalapiedra et al., 2021). Functional enrichment analysis of target gene sets was conducted using TBtools to identify significantly enriched biological categories (Chen et al., 2023). P-values adjusted using the Benjamini-Hochberg method in TBtools, represented as "p-adjust". Additionally, transcription factor prediction was performed using PlantTFDB (v5.0), a specialized database for plant transcription factors, offering insights into the regulatory functions of identified isoforms (Tian et al., 2020).

Identification of homoeologous gene groups among subgenomes

Homoeologous gene groups across the A, B, and D subgenomes were defined based on the dataset reported by Jia (Jia et al., 2023), who established homoeologous relationships using the AK58 reference genome. Only gene groups with unambiguous homoeologous assignments were

retained, including singletons (1 : 0 : 0, 0 : 1 : 0, or 0 : 0 : 1), dyads (1 : 1 : 0, 1 : 0 : 1, or 0 : 1 : 1), and triads (1 : 1 : 1). Groups with complex or ambiguous configurations (e.g., 1 : 1 : n, 0 : 0 : n, or n : n : n) were excluded because copy number variation in multi-copy groups may confound AS pattern analyses. Genes unanchored to chromosomes were also removed from subgenome-level splicing analyses.

Co-expression network analysis

Gene co-expression networks were constructed using the WGCNA R package (Langfelder and Horvath, 2008). RNA-seq data from AK58 grains at multiple developmental stages were normalized as fragments per kilobase of exon model per million mapped fragments (FPKM) and used as input. To improve network robustness, genes with negligible expression (FPKM < 1 in all samples or total expression across samples < 3) were filtered out. A total of 25,271 expressed genes, including ASGs, TF genes, and SR family genes, were retained for network construction.

The pickSoftThreshold function identified a soft-thresholding power of 6, satisfying the scale-free topology criterion ($R^2 \geq 0.85$). Networks and modules were generated using blockwiseModules with the following parameters: power = 6, maxBlockSize = 25271, TOMType = "unsigned", minModuleSize = 100, pamRespectsDendro = FALSE, reassignThreshold = 0, mergeCutHeight = 0.1, numericLabels = TRUE, corType = "pearson", and verbose = 3. This analysis yielded 20 co-expression modules. To identify modules associated with AS activity during grain development, correlations between module eigengenes and phenotypic traits were calculated. Hub genes were defined by high module membership (MM > 0.9) and strong trait correlation (GS > 0.6). The resulting networks were visualized using Cytoscape.

Characterization of isoforms with agronomic traits

Isoforms related to key agronomic traits were identified based on SQANTI3 output data. As a representative example, the well-characterized *Q* gene, which influences agronomic traits in wheat, was analyzed. Structural isoforms of the *Q* gene were visualized using IGV (v 2.16.2), enabling detailed examination of isoform architecture and splicing patterns (Robinson et al., 2011). Additionally, the isoforms were experimentally validated through PCR amplification of

their corresponding cDNA sequences, followed by gel extraction and Sanger sequencing to confirm isoform identity.

Trait association analysis of *Q-5A* isoforms

To evaluate the functional relevance of *Q-5A* isoforms in spike development, we analyzed transcript abundance and phenotypic associations using RNA-seq data from 90 wheat accessions at the double ridge stage (Wang et al., 2017). Expression levels of all 24 *Q-5A* isoforms identified in this study were quantified with Salmon (v1.9.0) (Patro et al., 2017) and expressed as transcripts per million (TPM). Isoforms with TPM > 1 in at least one accession were considered expressed and retained for analysis.

Spike morphology traits, including spikelet number, floret number, and seed number, were used to assess isoform–trait relationships. Spearman correlation coefficients were calculated to determine the strength and direction of associations between isoform abundance and individual traits, with corresponding *p*-values used to evaluate statistical significance.

Accession Numbers

The accession numbers (gene IDs) and sequences of the major genes mentioned in this study are provided in Supplementary Tables.

LIST OF ABBREVIATIONS

A3: Alternative 3’ Splice Site

A5: Alternative 5’ Splice Site

AF: Alternative First Exons

AK58: Aikang58

AL: Alternative Last Exons

AS: Alternative Splicing

ASE: Alternative Splicing Event

ASG: Alternative Splicing Gene

BP: Biological Process

CC: Cellular Component

CCS: Circular Consensus Sequence

DAP: Days After Pollination

EG: Expressed Gene

EST: Expressed Sequence Tag

FL: full-length

FPKM: Fragments Per Kilobase of exon model per Million mapped fragments

FSM: Full Splice Match

GO: Gene Ontology

ISM: Incomplete Splice Match

KEGG: Kyoto Encyclopedia of Genes and Genomes

MF: Molecular Function

MIG: Multiple Isoform Gene

MXE: Mutually Exclusive Exons

NIC: Novel in Catalog

NNC: Novel Not in Catalog

RBDs: RNA-binding domains

RI: Retained Intron

RIN: RNA Integrity Number

RNA-seq: RNA sequencing

SE: Skipping Exon

SMRT: Single Molecule Real-Time

SR: Serine/Arginine

TPM: Transcripts Per Million

ZMW: Zero-Mode Waveguide

DATA AVAILABILITY STATEMENT

The full-length transcriptome sequencing data generated in this study were deposited in NCBI (PRJNA1245628).

FUNDING

This research was supported by the National Key R&D Program (2022YFF1003501, 2024YFF1001200), the National Natural Science Foundation of China (32201844, 31991213), and the Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-CSCB-202401, CAAS-BRC-202609, CAAS-ZDRW202403).

ACKNOWLEDGMENTS

We thank Novogene for the support in sequencing, Lingli Zheng and Wuman Xu for excellent technical support, and Yijing Zhang from Fudan University for valuable suggestions.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

L.Z., X.K., and X.L. conceived and designed the project. H.D. and P.D. performed the experiments and analyzed the data. B.L., Z.Y., K.W., C.K., Z.X., C.X., and F.D. provided suggestions on experiments and analysis. H.D., L.Z., and X.K. wrote the manuscript. J.J. provided suggestions for manuscript improvement. All authors accepted the final version of the manuscript.

TABLES

Table 1. Summary of identified ASEs and ASGs

Sample	ASGs	ASGs/EGs	ASEs	A3	A5	AF	AL	MX	RI	SE
Root	8,707	21.2%	18,881	4,641	3,077	1,596	266	50	7,351	1,900
Stem	12,939	25.8%	30,609	7,005	4,546	2,729	560	71	12,717	2,981
Leaf	7,593	23.5%	19,521	4,427	2,670	1,073	545	37	9,090	1,679
Spike	10,591	24.2%	22,502	5,394	3,821	1,380	318	45	9,048	2,496
Grain	29,344	40.5%	126,687	23,794	17,395	14,799	5,300	393	53,907	11,099
All	39,026	45.3%	213,448	38,413	27,741	29,022	10,570	741	88,780	18,181

ASEs: alternative splicing events; ASGs: alternative splicing genes

SE: skipping exon; A5/A3: alternative 5′/3′ splice sites; MX: mutually exclusive exons; RI: retained intron; AF/AL: alternative first/last exons

FIGURE LEGENDS

Figure 1. Construction and characterization of the full-length transcriptomes.

The number of genes and isoforms included in the full-length transcriptomes across different tissues (root, stem, leaf, spike and grain), with isoform classification. Isoform classification categories include Full Splice Match (FSM), Incomplete Splice Match (ISM), Novel in Catalog (NIC), Novel Not in Catalog (NNC), Antisense, Genic Intron, Genic Genomic, and Intergenic. Identified genes/transcripts refer to those detected through the IsoSeq pipeline and SQANTI3, annotated genes/transcripts are those present in reference databases, and previously unannotated genes/transcripts are those absent from the reference database.

Figure 2. Genome-wide alternative splicing (AS) landscape in AK58.

(A) Distribution of genes with varying numbers of isoforms in the full-length transcriptome. **(B)** Distribution of splicing junction categories across different isoform classifications. **(C)** Statistics of AS genes (ASGs) across different tissues. **(D)** Statistics of ASGs in triad gene sets within the A, B, and D subgenomes. **(E)** Distribution of expressed genes (EGs), genes with multiple isoforms (MIGs), ASGs, and AS events (ASEs) on chromosomes: a. chromosome ideograms (red boxes indicate centromere positions), b. EG density, c. MIG density, d. ASG density, e. ASGs/EGs proportion density, f. distribution of different AS types across chromosomes, and g. ASE density.

Figure 3. AS facilitates tissue differentiation.

(A) Comparison of isoform numbers per gene across tissues. The heatmap using $\log_2(n+1)$. **(B)** KEGG enrichment of genes with higher isoform numbers in each tissue. **(C)** KEGG enrichment of tissue-specific ASGs.

Figure 4. AS is associated with subgenome divergence.

(A) Conservation and divergence of AS types among triad homoeologous genes across the A, B, and D subgenomes. **(B)** Proportion of isoform numbers for triad genes across the A, B, and D subgenomes. **(C)** Gene number in the ternary plot. **(D)** KEGG enrichment analysis of dominant and suppressed genes. **(E)** KEGG enrichment of subgenome-specific ASGs.

Figure 5. AS is involved in grain development and maturation.

(A) Expression levels of serine/arginine (SR) protein family candidate genes in AK58 grain across different developmental stages. The heatmap using $\log_2(FPKM+1)$. **(B)** Comparison of isoform numbers per gene in grain at different stages. **(C)** Distribution of genes with varying isoform numbers in grain full-length transcriptomes of different stages. The heatmap using $\log_2(n+1)$. **(D)** Statistics of ASGs across different grain developmental stages. Bars represent the number of ASGs, and black dots connected by lines indicate the trend across stages. **(E)** GO enrichment of genes with higher AS activity in early (4 DAP, 10 DAP) and late (15 DAP, 20

DAP, 25 DAP) grain development. **(F)** KEGG enrichment results of genes with higher AS activity in early and late grain development. **(G)** Isoform numbers of grain trait-related genes in the AK58 genome at different developmental stages. The heatmap using $\log_2$ (n+1).

Figure 6. AS is involved in trait regulation.

(A) Visualization of 10 isoforms identified in the stem. Thick blue boxes represent coding regions, thin blue boxes represent untranslated regions (UTRs), thin lines connecting boxes represent spliced junctions, and arrowheads indicate the direction of transcription. Red boxes indicate primer locations. Genomic coordinates are shown at the top. **(B)** PCR amplification results of *Q-5A*. **(C)** Expression profiles of *Q-5A* isoforms across 90 wheat accessions. The heatmap using $\log_2$ (TPM+1). **(D)** Correlation analysis between *Q-5A* gene expression and spike-related traits (the number of spikelets, seeds and florets per main spike). Blue dots represent individual wheat accessions. Dashed red lines represent LOESS regression curves, and gray shaded areas represent 95% confidence intervals to indicate data trends. Spearman correlation coefficients (r) and p-values (p) are shown in each panel. **(E)** Correlation coefficients between *Q-5A* isoforms' proportion and spike morphology traits (the number of spikelets, seeds and florets per main spike). The heatmap using -$\log_{10}$ (p-value).

REFERENCES

Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS (2016) A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun **7:** 11706

Ast G (2004) How did alternative splicing evolve? Nat Rev Genet **5:** 773-782

Boguski MS, Lowe TM, Tolstoshev CM (1993) dbEST--database for "expressed sequence tags". Nat Genet **4:** 332-333

Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A (2007) UniProtKB/Swiss-Prot. Methods Mol Biol **406:** 89-112

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nat Methods **12:** 59-60

Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol **38:** 5825-5829

Cartolano M, Huettel B, Hartwig B, Reinhardt R, Schneeberger K (2016) cDNA library enrichment of full length transcripts for SMRT long read sequencing. PLoS One **11:** e0157779

Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, Xia R (2023) TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol Plant **16:** 1733-1742

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem **162:** 156-159

Desjardins P, Conklin D (2010) NanoDrop microvolume quantitation of nucleic acids. J Vis Exp

Dong C, He F, Berkowitz O, Liu J, Cao P, Tang M, Shi H, Wang W, Li Q, Shen Z, Whelan J, Zheng L (2018) Alternative splicing plays a critical role in maintaining mineral nutrient homeostasis in rice (*Oryza sativa*). Plant Cell **30:** 2267-2285

Dong L, Liu H, Zhang J, Yang S, Kong G, Chu JS, Chen N, Wang D (2015) Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. BMC Genomics **16:** 1039

Gao P, Quilichini TD, Zhai C, Qin L, Nilsen KT, Li Q, Sharpe AG, Kochian LV, Zou J, Reddy ASN, Wei Y, Pozniak C, Patterson N, Gillmor CS, Datla R, Xiang D (2021) Alternative splicing dynamics and evolutionary divergence during embryogenesis in wheat species. Plant Biotechnol J **19:** 1624-1643

Gene Ontology C (2015) Gene Ontology Consortium: going forward. Nucleic Acids Res **43:** D1049-1056

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis CA, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, Sturgill D, Tuch BB, Zaleski C, Zhang D, Blanchette M, Dudoit S, Eads B, Green RE, Hammonds A, Jiang L, Kapranov P, Langton L, Perrimon N, Sandler JE, Wan KH, Willingham A, Zhang Y, Zou Y, Andrews J, Bickel PJ, Brenner SE, Brent MR, Cherbas P, Gingeras TR, Hoskins RA, Kaufman TC, Oliver B, Celniker SE (2011) The developmental transcriptome of *Drosophila melanogaster*. Nature **471:** 473-479

Guo W, Yu K, Han L, Li X, Wang H, Liu Y, Zhang Y (2020) Global profiling of alternative splicing landscape responsive to salt stress in wheat (*Triticum aestivum L.*). Plant Growth Regulation **92:** 107-116

Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P (2019) eggNOG 5.0: a hierarchical, functionally and

phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res **47:** D309-D314

Jia J, Zhao G, Li D, Wang K, Kong C, Deng P, Yan X, Zhang X, Lu Z, Xu S, Jiao Y, Chong K, Liu X, Cui D, Li G, Zhang Y, Du C, Wu L, Li T, Yan D, Zhan K, Chen F, Wang Z, Zhang L, Kong X, Ru Z, Wang D, Gao L (2023) Genome resources for the elite bread wheat cultivar Aikang 58 and mining of elite homeologous haplotypes for accelerating wheat improvement. Mol Plant **16:** 1893-1910

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res **44:** D457-462

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. Plant J **88:** 1058-1070

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9**: 559

Lee C, Roy M (2004) Analysis of alternative splicing with microarrays: successes and challenges. Genome Biol **5:** 231

Li S, Yamada M, Han X, Ohler U, Benfey PN (2016) High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. Dev Cell **39:** 508-522

Liu L, Zhou Y, Mao F, Gu Y, Tang Z, Xin Y, Liu F, Tang T, Gao H, Zhao X (2022) Fine-tuning of the grain size by alternative splicing of *GS3* in rice. Rice (N Y) **15:** 4

Liu Y, Do S, Huynh H, Li J-X, Liu Y-G, Du Z-Y, Chen M-X (2024) Importance of pre-mRNA splicing and its study tools in plants. Advanced Biotechnology **2:** 4

Liu Z, Qin J, Tian X, Xu S, Wang Y, Li H, Wang X, Peng H, Yao Y, Hu Z (2018) Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (*Triticum aestivum L.*). Plant biotechnology journal **16:** 714-726

Liu Z, Yang F, Deng C, Wan H, Tang H, Feng J, Wang Q, Yang N, Li J, Yang W (2024) Chromosome-level assembly of the synthetic hexaploid wheat-derived cultivar Chuanmai 104. Sci Data **11:** 670

Marquez Y, Brown JW, Simpson C, Barta A, Kalyna M (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. Genome Res **22:** 1184-1195

Modrek B, Lee C (2002) A genomic view of alternative splicing. Nat Genet **30:** 13-19

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet **40:** 1413-1415

Pardo-Palacios FJ, Arzalluz-Luque A, Kondratova L, Salguero P, Mestre-Tomas J, Amorin R, Estevan-Morio E, Liu T, Nanni A, McIntyre L, Tseng E, Conesa A (2024) SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. Nat Methods **21:** 793-797

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods **14:** 417-419

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. Nucleic Acids Res **40:** D290-301

Reddy AS, Marquez Y, Kalyna M, Barta A (2013) Complexity of the alternative splicing landscape in plants. Plant Cell **25:** 3657-3683

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol **29:** 24-26

Sanchez-Martin J, Widrig V, Herren G, Wicker T, Zbinden H, Gronnier J, Sporri L, Praz CR, Heuberger M, Kolodziej MC, Isaksson J, Steuernagel B, Karafiatova M, Dolezel J, Zipfel C, Keller B (2021) Wheat *Pm4* resistance to powdery mildew is controlled by alternative splice variants encoding chimeric proteins. Nat Plants **7:** 327-341

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol **7:** 3

Shepard PJ, Hertel KJ (2009) The SR protein family. Genome Biol **10:** 242

Tian F, Yang DC, Meng YQ, Jin J, Gao G (2020) PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res **48:** D1104-D1113

Tilgner H, Grubert F, Sharon D, Snyder MP (2014) Defining a personal, allele-specific, and single-molecule long-read transcriptome. PNAS **111:** 9869-9874

Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyras E (2018) SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol **19:** 40

Waddington SR, Cartwright PM, Wall PC (1983) A quantitative scale of spike initial and pistil development in barley and wheat. Annals of Botany**:** 119-130

Wang X, Chen X, Luo S, Ma W, Li N, Zhang W, Tikunov Y, Xuan S, Zhao J, Wang Y, Zheng G, Yu P, Bai Y, Bovy A, Shen S (2022) Discovery of a DFR gene that controls anthocyanin accumulation in the spiny *Solanum* group: roles of a natural promoter variant and alternative splicing. Plant J **111:** 1096-1109

Wang Y, Yu H, Tian C, Sajjad M, Gao C, Tong Y, Wang X, Jiao Y (2017) Transcriptome Association Identifies Regulators of Wheat Spike Architecture. Plant Physiol **175**: 746-757

Wei B, Jiao Y (2024) Grain size control in wheat: toward a molecular understanding. Seed Biology **3**

Wen J, Qin Z, Sun L, Zhang Y, Wang D, Peng H, Yao Y, Hu Z, Ni Z, Sun Q, Xin M (2023) Alternative splicing of *TaHSFA6e* modulates heat shock protein-mediated translational regulation in response to heat stress in wheat. New Phytol **239:** 2235-2247

Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. Plant Cell **18:** 2971-2984

Wright CJ, Smith CWJ, Jiggins CD (2022) Alternative splicing as a source of phenotypic diversity. Nature Reviews Genetics **23:** 697-710

Xiao J, Liu B, Yao Y, Guo Z, Jia H, Kong L, Zhang A, Ma W, Ni Z, Xu S, Lu F, Jiao Y, Yang W, Lin X, Sun S, Lu Z, Gao L, Zhao G, Cao S, Chen Q, Zhang K, Wang M, Wang M, Hu Z, Guo W, Li G, Ma X, Li J, Han F, Fu X, Ma Z, Wang D, Zhang X, Ling HQ, Xia G, Tong Y, Liu Z, He Z, Jia J, Chong K (2022) Wheat genomic study for genetic improvement of traits in China. Sci China Life Sci **65:** 1718-1775

Xie M, Zuo R, Bai Z, Yang L, Zhao C, Gao F, Cheng X, Huang J, Liu Y, Li Y, Tong C, Liu S (2022) Genome-wide characterization of serine/arginine-rich gene family and its genetic effects on agronomic traits of *Brassica napus*. Front Plant Sci **13:** 829668

Yu K, Feng M, Yang G, Sun L, Qin Z, Cao J, Wen J, Li H, Zhou Y, Chen X, Peng H, Yao Y, Hu Z, Guo W, Sun Q, Ni Z, Adams K, Xin M (2020) Changes in alternative splicing in response to domestication and polyploidization in wheat. Plant Physiol **184:** 1955-1968

Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu Z, He Z, Fang X, Chen L, Tian W, Tao Y, Kristiansen K, Zhang X, Li S, Yang H, Wang J, Wang J (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res **20:** 646-654

Zhang L, Liu Y, Wang Q, Wang C, Lv S, Wang Y, Wang J, Wang Y, Yuan J, Zhang H, Kang Z, Ji W (2022) An alternative splicing isoform of wheat TaYRG1 resistance protein activates immunity by interacting with dynamin-related proteins. J Exp Bot **73:** 5474-5489

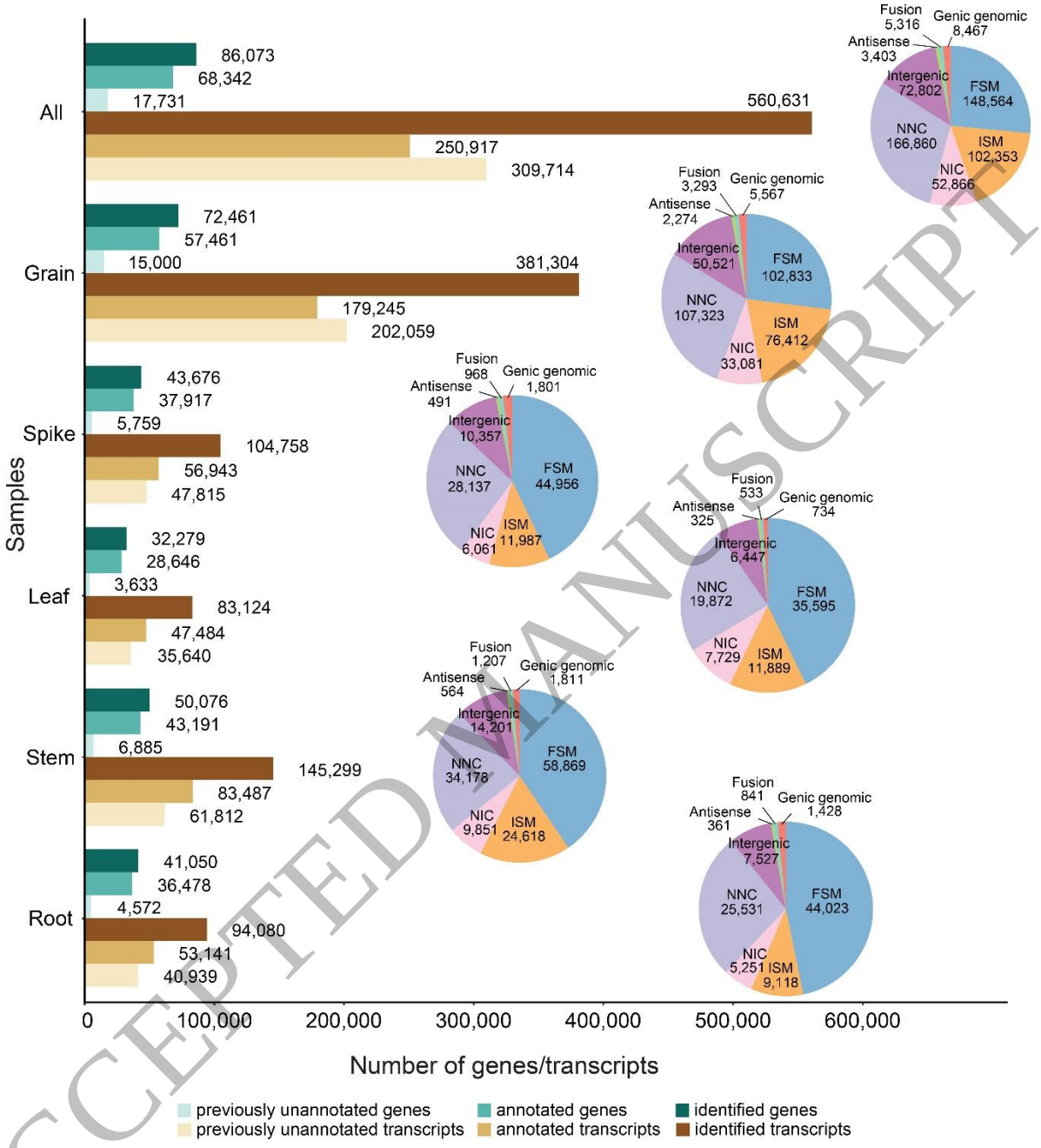

Figure 1
159x174 mm (x DPI)

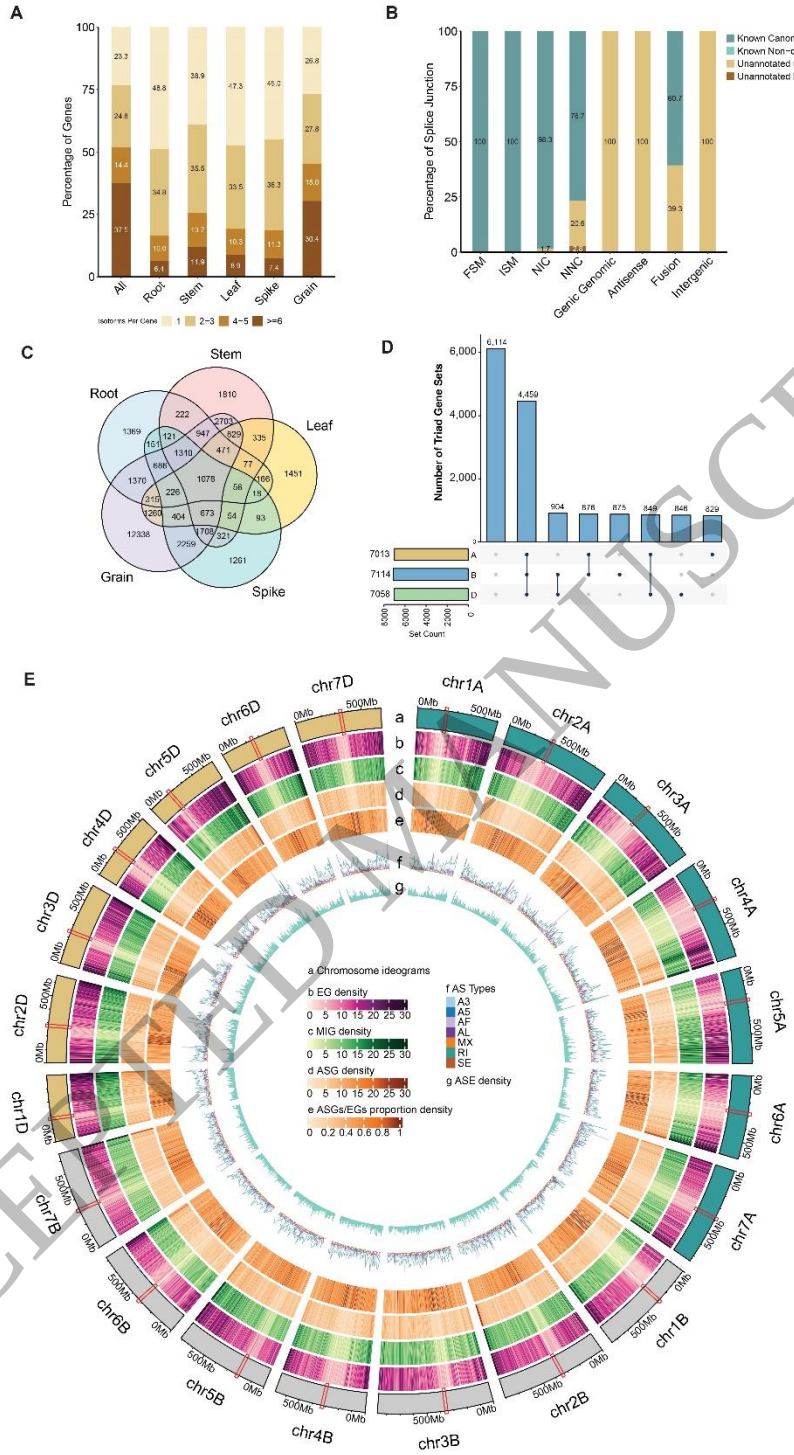

Figure 2
149x246 mm (x DPI)

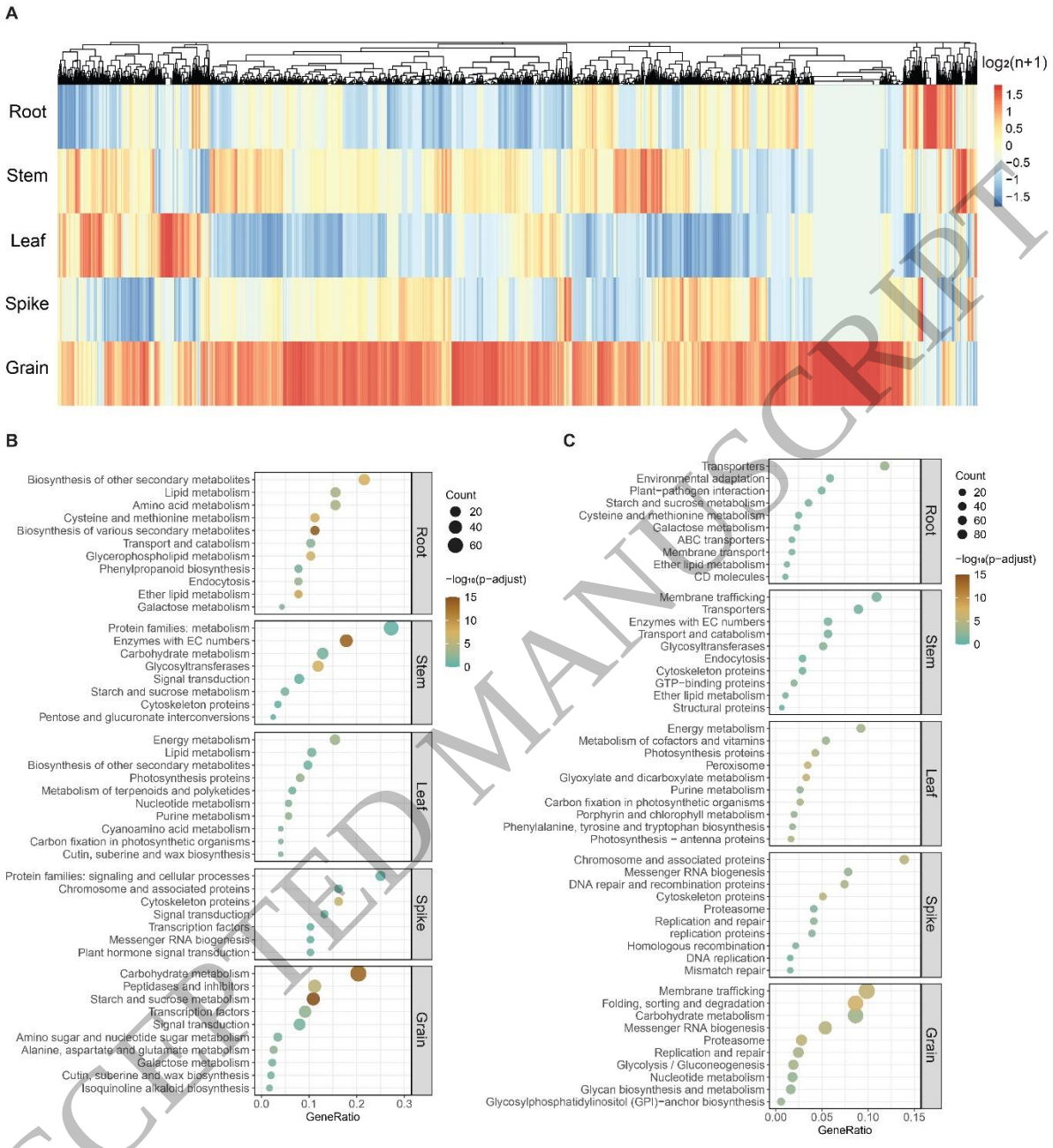

Figure 3
159x171 mm (x DPI)

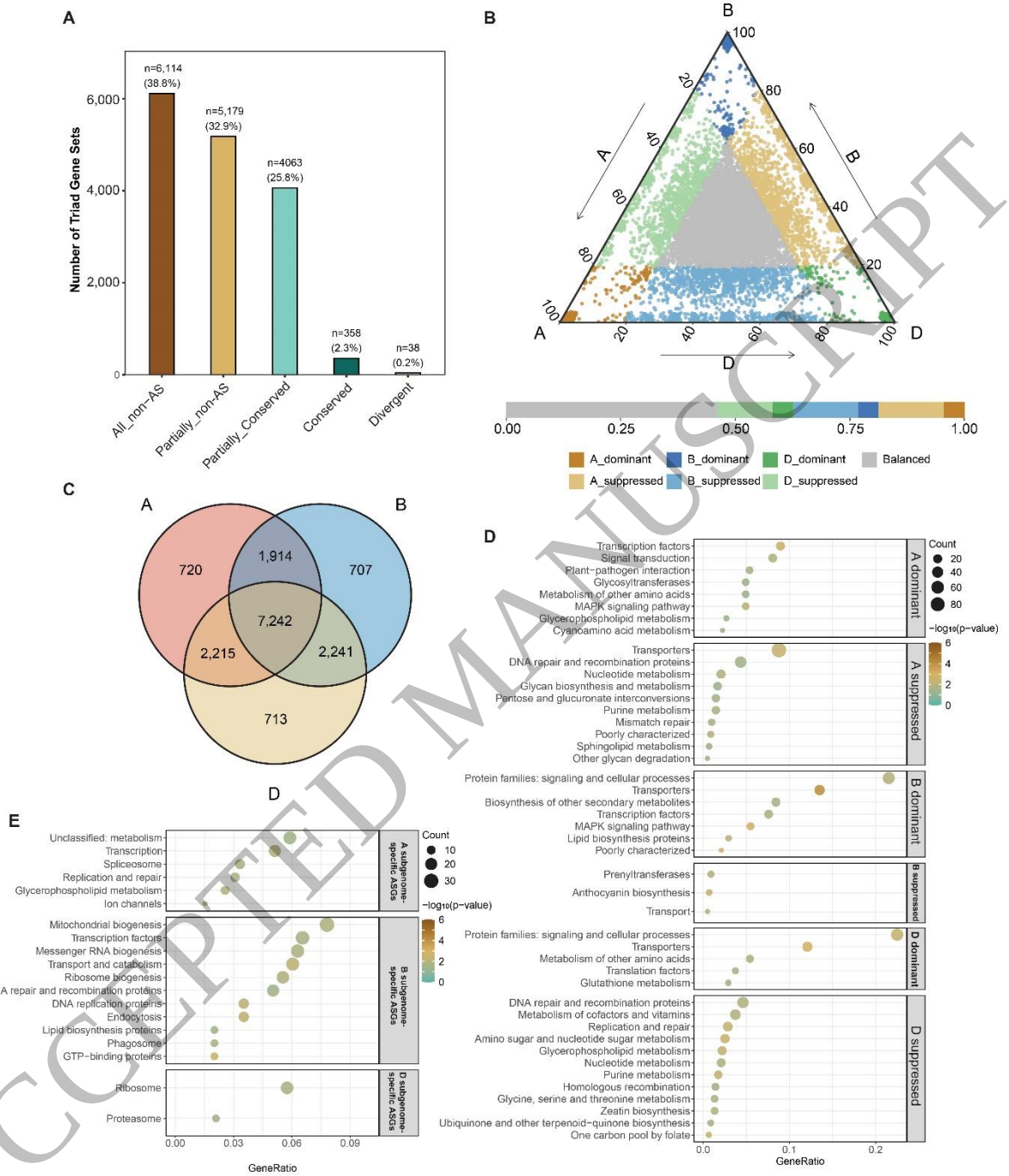

Figure 4
159x185 mm (x DPI)

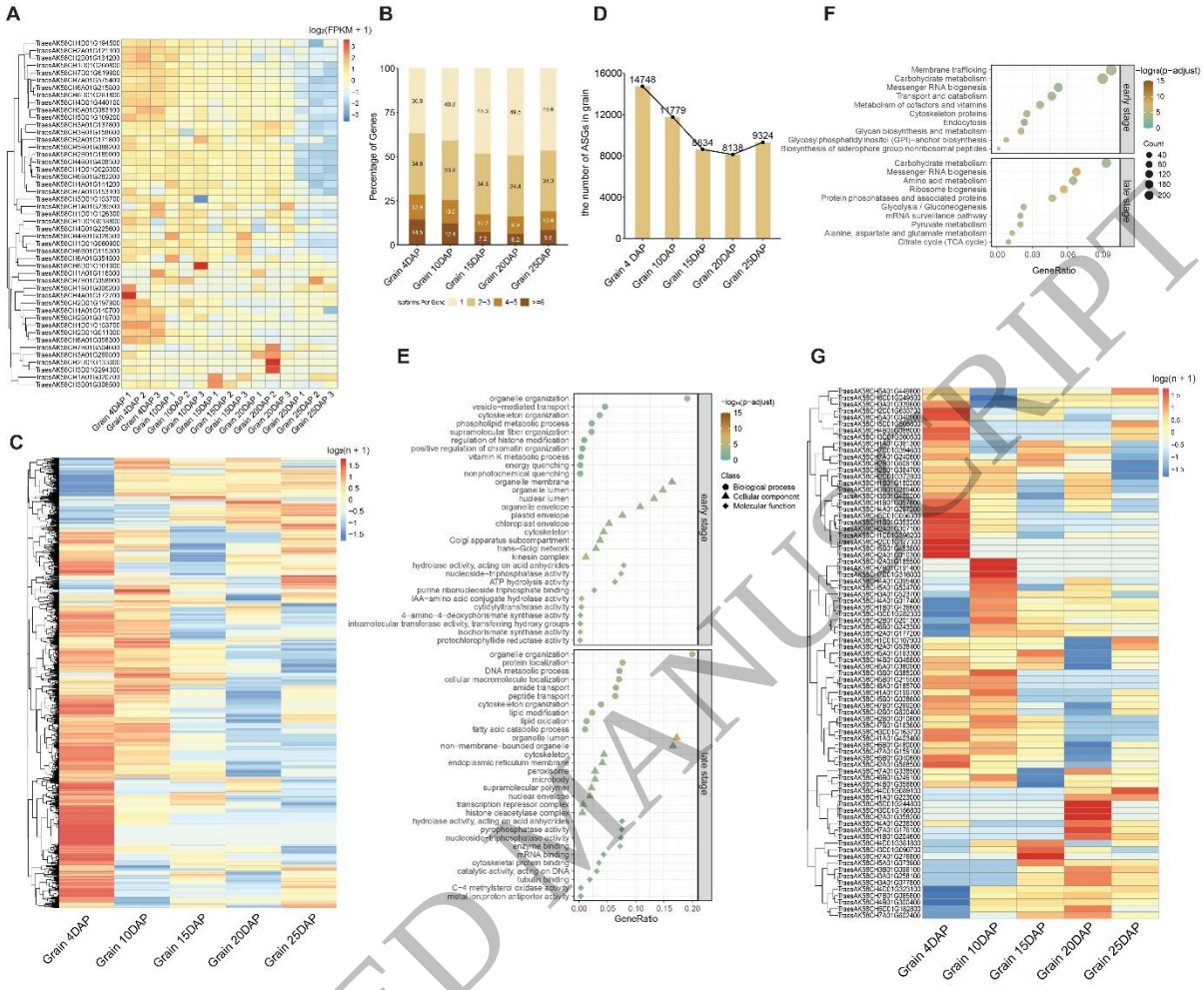

Figure 5
159x132 mm (x DPI)

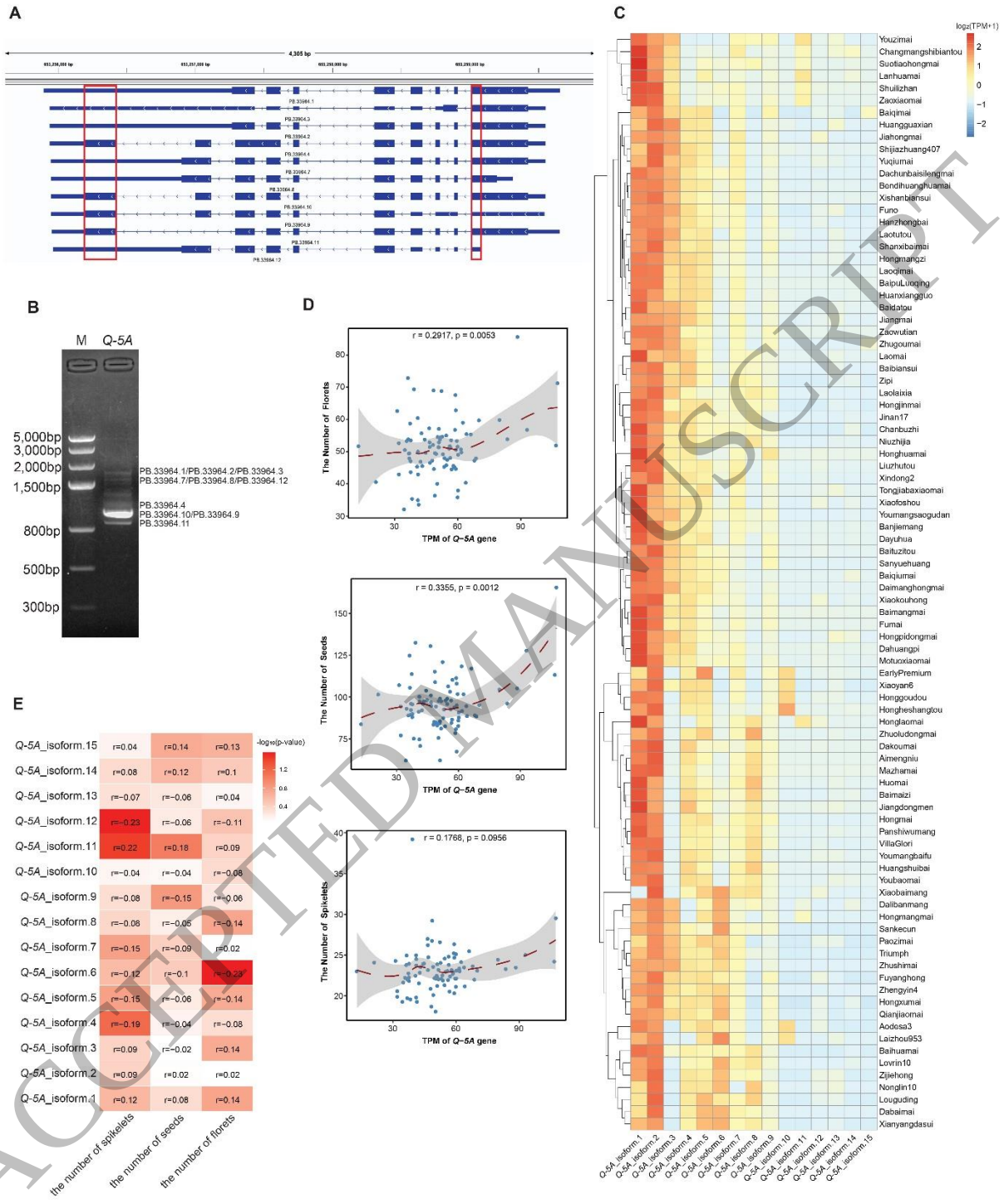

Figure 6
159x193 mm (x DPI)