Research Square

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

Multi-omic analysis reveals the effects of interspecific hybridization on the synthesis of seed reserve polymers in a Triticum turgidum ssp. durum × Aegilops sharonensis amphidiploid

Qian Hu
Sichuan Agricultural University
Jing Liu
Sichuan Agricultural University
Xiaolei Chen
Sichuan Agricultural University
Carlos Guzmán
Universidad de Córdoba
Qiang Xu
Sichuan Agricultural University
Yazhou Zhang
Sichuan Agricultural University
Qian Chen
Sichuan Agricultural University
Huaping Tang
Sichuan Agricultural University
Pengfei Qi
Sichuan Agricultural University
Mei Deng
Sichuan Agricultural University
Jian Ma
Sichuan Agricultural University
Guoyue Chen
Sichuan Agricultural University
Yuming Wei
Sichuan Agricultural University
Jirui Wang
Sichuan Agricultural University
Youliang Zheng
Sichuan Agricultural University

Yong Tu
Xichang University
Qiantao Jiang
qiantaojiang@sicau.edu.cn

Sichuan Agricultural University

Research Article

Keywords: Aegilops sharonensis, Triticum turgidum ssp. durum, Amphidiploid, Multi-omics, Starch, Seed storage proteins

Posted Date: December 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3669976/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomics on June 20th, 2024. See the published version at https://doi.org/10.1186/s12864-024-10352-9.

Abstract

Background

Wheat grain endosperm is mainly composed of proteins and starch. The contents and the overall composition of seed storage proteins (SSP) markedly affect the processing quality of wheat flour. Polyploidization results in duplicated chromosomes, and the genomes are often unstable and may result in a large number of genes loss and gene rearrangements. However ,the instability of the genome itself, as well as the large number of duplicated genes generated during polyploidy, is an important driving force for genetic innovation This study compared the differences of starch and SSP, and analyse the transcriptomeand and metabolome conducted among *Ae. sharonensis*, durum wheat and amphidiploid to reveal the effects of polyploidization on the synthesis of seed reserve polymers.

Results

The total starch and amylose content of Z636×R7 was significantly higher than R7 and lower than Z636. The gliadin and glutenin contents of Z636×R7 were higher than those in Z636 and R7. Through transcriptome analysis, there were 21037,2197,15090 differentially expressed genes in the three comparison groups of R7 vs Z636, Z636 vs Z636 × R7, Z636 × R7 vs R7, which were mainly enriched in carbon metabolism and amino acid biosynthesis pathways. Transcriptome data and qRT-PCR were combined to analyze the expression levels of genes related to storage substances. It was found that the expression levels of AGP-L, AGP-S and GBSSI of some starch synthase genes in Z636×R7 were higher than R7 and among the 17 differentially expressed genes related to storage proteins, the expression levels of 14 genes in R7 were lower than those in Z636 and Z636 × R7. According to the classification analysis of all differential metabolites, most of the metabolites belong to carboxylic acids and derivatives, fattyacyls, and are enriched in the biosynthesis of unsaturated fatty acids, niacin and nicotinamide metabolism, one carbon pool by flate etc.

Conclusion

After allopolyploidization, it may be affected by the exogenous genome, the organic synthesis function is down-regulated, the seed development cycle is prolonged, and the starch accumulation period is delayed. Therefore, at the same development time point, the starch accumulation of Z636 × R7 lags behind that of Z636. In this study, the expression of GSe2 gene in Z636 × R7 was higher than that of the two parents, which was beneficial to protein synthesis and thus increased protein content. These results eventually led to the alteration of synthesis of seed reserve polymers. The current study provided a basis for a greater in-depth understanding of the mechanism of wheat allopolyploid formation and its stable preservation and promote the effective exploitation of high-value alleles

Background

Wheat grain is mainly composed of starch, protein and lipid, in which starch as the main energy storage material accounts for about 75% of the dry weight of the grain. Starch can be divided into amylose and amylopectin according to branching.and its synthesis involves at least four enzyme families: adenosine diphosphate glucose pyrophosphorylase (AGPase), starch synthases (SSs), starch branching enzymes (SBEs), and starch debranching enzymes (DBEs) [1]. Genetic evidence has demonstrated that DBEs, such as isoamylase, play a role in starch synthesis and are essential for the formation of crystalline amylopectin [2]. Therefore, the expression of starch synthase genes determines the content and composition of starch [3].

The content of protein in wheat grains only accounts for about 15% of its dry weight, but it plays an important role in determining the processing quality of wheat flour. Seed storage proteins are mainly composed of high-molecular-weight glutenin subunits (HMW-GSs), low-molecular-weight glutenin subunits (LMW-GSs), and gliadins (α, β, γ and ω)[4]. HMW-GSs have been considered to be the most important seed storage proteins, and are well known to play a key role in determining wheat flour quality [5]. In wheat and its related wild species, the HMW-GS-encoding genes were the *Glu-1* loci, located on the long arms of group 1 chromosomes, with each locus comprising a pair of tightly linked genes encoding one x-type subunit with a higher molecular mass and one y-type subunit with a lower molecular mass [6]. After long-term cultivation, the type and frequency of HMW-GS variation in common wheat are relatively low, while, there are abundant HMW-GS variation types in wheat related wild species such as *Ae. tauschii. Secale cereal, Ae.umbellulata, Elymus glaucus, Elytrigia elongata*[7–11].

Polyploidization or whole genome duplication (WGD), a phenomenon in which the genome in the nucleus is doubled and passed on in a heritable manner to its offspring, is the driving force of plant evolution, and all seed plants undergo this process throughout their evolutionary history[12, 13]. Polyploidy includes autopolyploidy and allopolyploidy. About 70% of flowering plants, including many important crops, are allopolyploids, such as common wheat [14], oilseed rape, *Brassica napus* [15], and cotton, *Gossypium hirsutum* [16]. Polyploidization results in duplicated chromosomes, and the genomes are often unstable and may result in a large number of genes loss and gene rearrangements [17]. Allopolyploids can act as important bridges by which to transfer valuable genes from wild crop ancestors into common wheat by hybridization of wheat and distantly related species to achieve wheat improvements. Therefore, studying the mechanism of allopolyploid formation and evolutionary variation is not only very important for basic research into plant evolution theory, but can also strengthen the directional selection of artificial allopolyploid evolution, which has broad production and utilization values in crops.

Aegilops sharonensis (Eig) Á. Löve. ($S^{sh}S^{sh}$, 2n = 2x = 14) is an annual diploid grass species (Sharon goatgrass), which belongs to *Aegilops* section Sitopsis, is distributed within the Near East, and may have contributed to the B-genome of common wheat. It exhibits many excellent traits, including abiotic stress tolerance and biotic resistance to pests and diseases, such as stripe rust, stem rust, leaf rust, and powdery mildew [18, 19], representing a rich gene pool for wheat improvement. In this species, the novel HMW-GS variants $1S^{sh}$x2.9 and $1S^{sh}$y2.3 were identified with high subunit molecular weights of 2.9 kb and 2.3 kb, respectively [20]. The *Glu-1S^{sh}* locus of *Ae. sharonensis* could be used for wheat quality

improvement because the glutenin subunits encoded possess longer repetitive domains than most of the known subunits.

As the foundation of this study, we successfully produced durum *T. turgidum* ssp. *durum* (durum wheat)-*Ae. Sharonensis* amphidiploids by conducting wide hybridizations [21]. In the present study, we analyze the regulatory mechanism of changes in starch and storage protein contents in developing grains in response to allopolyploidization by transcriptome and metabolome analyses. In this study, we analysis the transcriptomeand and metabolome conducted on *Ae. sharonensis*, durum wheat and amphidiploid, to explore the changes of DEGs and DEMs after polyploidization and reveal the effects of interspecific hybridization on the synthesis of seed reserve polymers.

Results

Identification of HMW-GS composition and ND-FISH observation of the amphidiploid and its parents

The amphidiploid (Z636×R7, AABB$S^{sh}S^{sh}$, 2n = 42) was generated from a wide cross between *Triticum turgidum* ssp. *durum* (Z636, AABB, 2n = 28) and *Ae. sharonensis* (R7, $S^{sh}S^{sh}$, 2n = 14). The amphidiploid seeds with *Ae. sharonensis*-specific HMW-GS (1S^{sh}x2.9 + 1S^{sh}y2.3) and durum wheat-specific HMW-GSs (1Bx6 + 1By8) were identified by SDS-PAGE (Fig. 1A). The stem nodes and ears of the amphidiploid plants were light purple after the jointing stage, with the plant morphology showing a phenotype intermediate between the two parental types (Fig. 1B, Fig. 1C). Cytological observation of the three genotypes showed that the number of chromosomes in the amphidiploid was 42, which was composed of 14 *Ae. sharonensis* chromosomes (Fig. 1D) and 28 durum wheat chromosomes (Fig. 1E), so that the genetic composition was consistent with that expected, indicating that the amphidiploid has a complete set of durum wheat and *Ae. sharonensis* chromosomes (Fig. 1F).

Content analysis of gluten proteins and starch

The contents of glutenins and gliadins of the three genotypes, R7, Z636, and Z636×R7, were analyzed by RP-HPLC (Fig. 2A, 2B). R7 and Z636 each exhibited two peaks in the HMW-GS region, namely 1S^{sh}x2.9 + 1S^{sh}y2.3 and 1Bx6 + 1By8, respectively, whereas Z636×R7 exhibited all four peaks in the HMW-GS region (1Bx6 + 1By8, 1S^{sh}x2.9 + 1S^{sh}y2.3). The above results were consistent with the SDS-PAGE results, which again proved that the HMW-GSs of R7 had been successfully introduced into the amphiploid. The content of the HMW-GSs of Z636×R7 (8,928.87 mAU*min) was significantly higher than that in Z636 (6,472.57 mAU*min), but not significantly different from the content of* HMW-GSs in R7 (8,935.36 mAU*min). The total LMW-GS content in Z636×R7 (16,590.25 mAU*min) was significantly higher than that in both R7 (12,482.67 mAU*min) and Z636 (12,009.62 mAU*min). Therefore, the total glutenin content in Z636×R7 (25,519.12 mAU*min) was significantly higher than that in R7 (21,418.03 mAU*min) and Z636 (18,482.19 mAU*min) (Fig. 2C).

Due to the small differences in molecular structure and molecular weight, the chromatographic peaks of α- and β-gliadins overlapped partially, so could not be separated from each other and could only be grouped as one class. There were differences in the peak shape and content of gliadins among the three genotypes; the contents of α-, β- and γ-gliadins of R7 were significantly lower than those of Z636 and Z636×R7, and the total gliadin content of R7 (45,527.80 mAU*min) was significantly lower than the contents in Z636 (60,665.47 mAU*min) and Z636×R7 (63,186.04 mAU*min) (Fig. 2D).

The total starch and amylose contents of the three genotypes, R7, Z636, and Z636×R7, grown in the greenhouse and field were determined (Fig. 2E, 2F), with the contents of total starch and amylose of Z636 × R7 in 2019 being significantly lower than those of Z636 (61.124%, 35.545%, respectively), but higher than those of R7 (54.822%, 5.995%, respectively). In 2021, the amylose content of Z636×R7 (61.163%) was significantly higher than that of either Z636 (58.272%) or R7 (42.455%).

RNA-Seq analysis

To compare the transcriptional differences among genotypes R7, Z636, and Z636×R7 in developing grains, three biological replicates were set up for each genotype, with a total of nine cDNA libraries being sequenced. After raw-read filtering and quality control, Q20 reached more than 95.29%, and Q30 (an error probability for base calling of 1% or 0.1%) reached more than 90.5% (Supplementary 5). The hierarchical clustering heatmap of Pearson's correlation coefficients showed the highly correlated expression profiles among the biological replicates for each genotype (Fig. 3A). The correlation within the three genotype groups reached more than 0.89. The high-quality sequencing results and the strong positive correlation between replicate samples showed that these transcriptomic data could be used for subsequent analyses.

Statistical and functional analyses of DEGs

In order to obtain reliable gene expression profiles, reads with log2 Fold Change ≥ 2 and FDR < 0.01 were selected to identify the DEGs. As shown in Fig. 3B, total (up-/down-regulated) DEGs in the three comparison groups of R7 vs Z636, Z636 vs Z636×R7, and Z636×R7 vs R7 were 21,037 (13,125/7,912), 2,197 (1,673/524),t and 15,090 (4,366/10,724), respectively.

DEGs were functionally annotated based on GO term enrichment analysis. The GO terms, namely biological process (BP), cellular component [22], and molecular function (MF), are shown in Supplementary 2. R7 vs Z636, Z636 vs Z636×R7, and Z636×R7 vs R7 comparisons were annotated and analyzed for GO function with respect to 16,651, 1,731, and 11,614 DEGs, respectively, and a total of 12,347, 1,250, and 8,547 DEGs were annotated, respectively. The DEGs in the three comparison groups were mainly concentrated in the term BP, in processes such as "metabolic process", "cellular process", "single organism process", and "biological process regulation", suggesting that genes involved in regulating physiological processes and cell metabolism play an important role in developing grains. Most DEGs in the term CC were enriched for "cell", "cell part", "membrane", "membrane part", and "organelles", suggesting that these DEGs may be enriched in wheat endosperm cells and the intercellular

cytoplasmic skeleton. With respect to the term MF, DEGs were mainly enriched in “binding”, “catalytically active”, and “transporter activity” processes, which affect the growth and development of wheat, enzymatic catalytic reactions, and ligand/receptor interactions.

In addition, KEGG pathway enrichment was analyzed for all DEGs in each inter-genotype comparison, selecting the top 20 pathways with significant enrichment to display (Supplementary 3). A total of 16,651 DEGs in the R7 vs Z636 comparison were annotated to 10,406 KEGG pathways, which were mainly enriched with respect to “carbon metabolism”, “biosynthesis of amino acids”, “ribosomes”, “photosynthesis-antenna protein”, and “RNA transport”. The Z636 vs Z636×R7 comparison contained 1,731 DEGs annotated to 1,059 KEGG pathways, mainly enriched in “starch and sucrose metabolism”, “amino sugar and nucleotide sugar metabolism”, “galactose metabolism”, and “pentose and glucuronate interconversion”. In the Z636×R7 vs R7 comparison, there were 11,614 DEGs annotated to 7,184 KEGG pathways, which were mainly enriched in “photosynthesis-antenna protein”, “ribosome biogenesis in eukaryotes”, “carbon metabolism”, and “ribosome”, indicating that genes related to starch and sucrose metabolism, protein synthesis and processing, and hormone regulation had important regulatory roles in developing wheat grains.

Differential expression of genes related to grain reserve polymers

The expression patterns of DEGs related to grain reserve polymers were investigated in developing grains of genotypes R7, Z636, and Z636×R7. Starch synthase genes, including *AGPase*, *SS*, *SBE*, and *DBE*, were identified in all differential cDNA libraries. The results indicated that the expression levels of starch synthase genes *AGP-L*, *AGP-S, SSII-2, SSIIIa, SBEII, SBEIII* and *GBSSI* in Z636×R7 were significantly higher than those in R7, but significantly lower than those in Z636×R7. Genes *SSI, SSII-3, ISAI, ISAII* and *PUL* exhibited significantly higher expression in Z636 than in R7 or Z636×R7, whereas the expression levels of *ISAIII, SBEI* and *SSIVa* genes in R7 were significantly higher than in Z636 or Z636×R7 (Fig. 4A, Supplementary 6).

When screening for genes encoding seed storage proteins, we found that gliadin and glutenin genes were differentially expressed between R7, Z636, and Z636×R7 (Fig. 4B, Supplementary 6). Seventeen glutenin- or gliadin-encoding DEGs were characterized in all libraries. The expression levels of genes encoding α/β-gliadin (TRITD0Uv1G108590), γ-gliadin (TRITD1Bv1G001870) and LMW-GS (TRITD1Bv1G008290) in R7 were significantly higher than those in Z636 or Z636×R7, whereas other seed protein-encoding DEGs showed significantly lower expression in Z636 or Z636×R7. The expression level of the LMW-GS (TRITD1Av1G002790) gene in Z636×R7 was significantly higher than that in R7 or Z636. The α/β-, ω- and γ -gliadin results indicated that the expression patterns in Z636×R7 were similar to those in Z636, whereas their expression levels were significantly higher than those in R7 (Fig. 3B, Supplementary 6). The glutenin profile of Z636×R7 is composed of the glutenins encoded by both R7 ($1S^{sh}$x2.9 + $1S^{sh}$y2.3) and Z636 (1Bx + 1By).

Twelve genes encoding grain reserve polymers were selected for qPCR validation of the RNA-Seq data. The IDs and primers of the selected genes are listed in Supplementary 6 and Supplementary 7 and the results of qPCR analysis are shown in Fig. 5. The qPCR results were consistent with the trends of the RNA-Seq expression patterns, indicating that the RNA-Seq data were reliable for expression pattern analysis.

Metabolome level and composition analysis

To construct a systematic profile of metabolic differences in R7, Z636, and Z636×R7, an untargeted metabolomic analysis was performed between R7 vs Z636, Z636 vs Z636×R7, and Z636×R7 vs R7. Principal component analysis (PCA) showed that the contribution rates of PC1 were 79.98% and 51.73%, and the contribution rates of PC2 were 7.88% and 31.05% in the positive and negative ion modes, respectively, of R7, Z636, and Z636×R7 (Fig. 6A, 6B). These results indicated that the differences between biological replicates during the experiment were small, and that there were differential metabolites between groups.

Differentially expressed metabolites (DEMs) were screened with a screening standard of VIP ≥ 1 and *P* value < 0.05. In the comparison R7 vs Z636, 501 and 370 DEMs were detected under positive and negative ion modes, respectively, of which 109 and 111 DEMs were upregulated, and 392 and 259 DEMs were downregulated, respectively. In the Z636 vs Z636×R7 comparison, 413 and 301 DEMs were detected under positive and negative ion modes, respectively, of which 297 and 198 were upregulated and 116 and 103 were downregulated, respectively. In the Z636×R7 vs R7 comparison, 480 and 356 DEMs were detected under positive and negative ion modes, respectively, of which 338 and 236 DEMs were upregulated and 142 and 120 DEMs were downregulated, respectively (Fig. 6C, 6D).

All DEMs identified in positive or negative ion modes were classified and analyzed using the Human Metabolome Database (HMBD) (Fig. 6E, Supplementary 8). When comparing R7 vs Z636, a total of 871 DEMs were detected, of which 319 were known metabolites, which, in turn, belonged to 50 classes. A total of 714 DEMs were detected in the Z636 vs Z636×R7 comparison, of which 263 were known metabolites, which were divided into 45 classes. A total of 836 DEMs were identified in the Z636×R7 vs R7 comparison, of which 293 were known metabolites, belonging to 50 classes. Of these known metabolites, most were carboxylic acids and their derivatives, fatty acyls, steroids, and steroid derivatives, organic oxygen compounds, benzene and its derivatives, prenol lipids, glycerophospholipids, and purine nucleosides. Among these, carboxylic acids and their derivatives accounted for the highest numbers of metabolites, with about 90 (R7 vs Z636), 80 (Z636 vs Z636×R7), and 84 (Z636×R7 vs R7), respectively, followed by fatty acyl groups, with 51 (R7 vs Z636), 30(Z636 vs Z636×R7), and 37 (Z636×R7 vs R7), respectively.

KEGG enrichment analysis of DEMs

KEGG enrichment analysis of DEMs was performed to identify significantly enriched metabolic pathways. In positive and negative ion modes, the differential metabolites between R7 and Z636 are mainly enriched in the biosynthesis of unsaturated fatty acids, niacin and nicotinamide metabolism, one-

carbon pool by folate, β-alanine, flavonoid and flavonol biosynthesis, carbamate metabolism, phenylalanine, tyrosine, and tryptophan biosynthesis, aminoyl-tRNA biosynthesis, alanine, aspartic acid, and glutamate metabolism, etc. (Fig. 7A). The differential metabolites between Z636 and Z636×R7 in positive and negative ion mode are mainly enriched in alanine metabolism, one-carbon pool by folate, glycine, serine and threonine metabolism, biosynthesis of unsaturated fatty acids, flavone and flavonol biosynthesis, aminoyl-tRNA biosynthesis, arginine biosynthesis, etc. (Fig. 7B). The differential metabolites between Z636×R7 and R7 are mainly enriched in the one-carbon pool by folate, niacin and nicotinamide metabolism, betaine biosynthesis, alanine metabolism, linoleic acid metabolism, phenylalanine, tyrosine and tryptophan biosynthesis, aminoyl-tRNA biosynthesis, etc. (Fig. 7C).

Correlation analysis of DEGs and DEMs associated with starch and amino acid synthesis

Based on GO and KEGG enrichment analyses, many DEMs and DEGs of R7, Z636 and Z636×R7 associated with starch and sucrose metabolism and with amino acid synthesis metabolism were identified (Fig. 8, Supplementary 6). The correlations between the content of cellobiose and the expression levels of genes *AGP-L*, *AGP-S*, *SS*, *SDE*, *DBE*, *GBSS*, α-1, 4-glucanase (*GBE*), and β-glucosidase [22] in Z636×R7 were higher than those in R7 and lower than those in Z636×R7. The expression levels of genes associated with starch synthesis were positively correlated with contents of amylose and total starch. GBE is a key enzyme in glycogen anabolism, catalyzing the formation of glycogen containing α-1,4-glycosidic and α-1,6-glycosidic bonds from ADP-glucose. In the current study, the correlation between *GBE* gene expression level and glycogen content in Z636×R7 was positive, but the correlations were opposite in R7 and Z636.

In amino acid synthesis, the contents of amino acids, such as homoserine, proline, arginine, and isoleucine, in Z636 were significantly downregulated, relative to R7, to varying degrees, whereas the contents of aspartic acid and glutamic acid were significantly upregulated. Aspartate and glutamate are also important reserves for nitrogen supply. The contents of proline, arginine, tryptophan, and phenylalanine in Z636×R7 were significantly higher, but the contents of asparagine and glutamate were significantly downregulated. Compared with Z636×R7, the content of many differential metabolites in R7, such as isoleucine, homoserine, phenylalanine, arginine, lysine, etc., showed different degrees of increase, whereas the significantly down-regulated metabolites were mainly aspartic acid, asparagine, etc. These results indicate that, in developing grains, low-molecular-weight molecules are gradually converted into macromolecular storage compounds, such as starch, proteins, lipids, and polysaccharides.

In addition, DEGs associated with amino acid biosynthesis, such as phosphoserine transaminase (*SER*), tryptophan synthase (*TSB*), phospho-2-dehydrogen-3-deoxy heptanoate aldolase (*DAHPS*), alanine aminotransferase (*ALT*), asparagine synthetase (*ASNS*), glutamate synthase (*GOGAT*), and glutamate dehydrogenase (GDH1), were highly expressed in R7. It was also found that, for two DEGs responsible for encoding glutamate synthase (GOGAT), *GSr1* and *GSe2*, genotypes Z636 and Z636×R7 had significantly

higher expression levels of *GSr1* than R7, whereas Z636×R7 had significantly higher gene expression levels of *GSe2* than either of its parents.

Discussion

Wheat grain dry biomass contains 65–80% carbohydrates, mainly composed of starch [23]. Biosynthesis of starch is accomplished by two types of enzymes, grain-binding starch synthase (GBSS) and starch synthase (SS) [24, 25]. Different expression levels of these starch synthase genes may affect starch synthesis and accumulation during grain filling [26]. Considerable progress has been made to date in the study of genes related to starch synthesis in tetraploid and hexaploid wheat [27], but little is known about the changes in the expression of genes involved in starch synthesis during polyploidization.

The present omics and phenotypic study of Z636×R7, obtained from the cross between Z636 and R7, allowed analysis of the effect of polyploidization on the synthesis of seed storage polymers in wheat. Through display of KEGG enrichment of DEGs, carbon metabolism, amino acid biosynthesis, RNA transport, starch and sucrose metabolism, amino and nucleotide sugar metabolism, galactose metabolism, pentose and glucuronic acid interconversion were the main metabolic pathways associated with differential gene enrichment. Some studies have shown that the starch content and the expression level of starch synthase genes in synthetic hexaploid wheat are higher than those of their parents, with synthetic hexaploid wheat inheriting the starch granule development and starch synthase gene expression pattern of the tetraploid parents [28]. In the current study, the expression levels of the starch biosynthesis genes *SSII*, *SSIIIa*, *SBEII*, *SBEIII*, and *GBSSI* in Z636×R7 were significantly higher than those in R7, but significantly lower than those in Z636. Z636×R7 starch synthase gene expression and starch content were more similar to Z636 than to R7, supporting the hypothesis that amphidiploid starch synthase gene expression is determined more by the tetraploid than by the diploid parent. Many studies have shown that allopolyploidy induces targeted sequence elimination, random structural changes, epigenetics and DNA methylation, and gene silencing of gene expression [29, 30]. The ADPase activity in starch biosynthesis of Z636×R7 may be restricted by the addition of the additional gene. The expression level of the corresponding starch biosynthesis gene in Z636×R7 being downregulated compared with Z636, which, in turn, affects the synthesis of amylose and amylopectin. And this finally causes the starch content of Z636×R7 being significantly lower than that of Z636.

Synthesis of seed storage proteins requires an adequate supply of amino acids, most of which are transported through the phloem, with the supply capacity of amino acids being strongly positively correlated with storage protein content [31]. Storage proteins are composed of storage protein particles known as protein body I and protein body II. Among these storage proteins, gluteninis mainly stored in protein body II, are rich in amino acids such as lysine, arginine, and glycine, and are readily decomposed by protease activity, whereas gliadins are mainly stored in protein body I, contain less lysine and are not readily decomposed by trypsin. Compared with the parent lines R7 and Z636, the content of seed storage proteins in Z636×R7 changed to different degrees, with the accumulation of the differential amino acid arginine decreasing in the amphiploid. Glutamine synthase (GS) is a key enzyme in plant nitrogen

metabolism, responsible for the first step of ammonium assimilation and its conversion to glutamine [32, 33]. Glutamate can be converted into other amino acids by various transaminases and transported to wheat grains. In addition, the *GS* gene can be regulated to enhance the nitrogen reuptake capacity of grains. Some studies have found that *GS1* can regulate grain size through quantitative trait locus (QTL) analysis [34], and that *GS2* expression is related to grain protein content [33]. In the current study, the *GSe2* expression level of Z636×R7 was higher than that of either of the two parents, an observation which is conducive to increased protein synthesis and thus increased grain protein content in the amphidiploid.

Recent studies have shown that polyploidy can increase glutenin content by upregulating glutenin biosynthesis, transport, and storage, thus affecting the content of related essential amino acids[35] Compared with R7, the expression levels of genes encoding HMW-GSs, LMW-GSs, α/β, γ- and ω-gliadins in Z636×R7 were significantly upregulated. RP-HPLC results showed that the total glutenin content of Z636×R7 was significantly higher than that of either of the two parents, whereas the total gliadin content was significantly higher than that of R7, but not significantly different from Z636. We speculate that the higher glutenin content of Z636×R7 than either parent might be due to the amphidiploid consists of four HWM-GSs ($1S^{sh}$x2.9 + $1S^{sh}$y2.3, 1Bx + 1By), so that the glutenin content of the polyploid is increased. Following hybridization of Z636 with R7 and subsequent spontaneous polyploidization, the alien *Aegilops* HMW-GSs contentcould be easily introduced into common wheat and improve the processing quality. The research results provide a basis for a greater in-depth understanding of the mechanism of wheat allopolyploid formation and its stable preservation and promote the effective exploitation of high-value alleles from related species in wheat genetic improvement. It also provides new materials for the transfer of good genes to cultivated wheat.

Conclusion

The mechanism of allopolyploid formation and evolutionary variation is very important for basic research into plant evolution theory. This study indicated that the carbohydrate, amino acid biosynthesis pathway and the expression of enzymes related to starch and protein was changed, which eventually led to the alteration of synthesis of seed reserve polymers after polyploidization. The current study provided a basis for a greater in-depth understanding of the mechanism of wheat allopolyploid formation and its stable preservation and promote the effective exploitation of high-value alleles from related species in wheat genetic improvement. It also provides new materials for the transfer of good genes to cultivated wheat.

Materials and Methods

Plant materials

Seed of *Aegilops sharonensis* (R7, $S^{sh}S^{sh}$, 2n = 14) PI584388 were kindly provided by the United States Department of Agriculture-Agricultural Research Service (USDA-ARS) (https://www.ars-grin.gov/),

Triticum turgidum ssp. *durum* (Z636, AABB, 2n = 28) was provided by the Triticeae Research Institute, Sichuan Agricultural University, China. And the amphidiploid ($AABBS^{sh}S^{sh}$, 2n = 42) of *Triticum turgidum* subsp. *durum*× *Ae. sharonensis* were generated in our previous work [21]. All the plant materials were grown in greenhouses in 2019 and fields in 2021.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Storage proteins were extracted from half of the endosperm of individual mature seeds as described previously [36]. The extracts were heated at 95℃ for 5 min and centrifuged at 10621×g for 10 min. The supernatant was loaded onto a 10% (w/v) SDS-PAGE gel and electrophoresis was carried out [6, 37].

Non-denaturing fluorescent in-situ hybridization (ND-FISH)

The methods described by [38] were used to prepare mitotic metaphase chromosome spreads. The probes Oligo-pSc119.2-1, Oligo-pTa535-1 and $(AAG)_6$ (Shanggon, shanghai) (Supplementary 4) were used in ND-FISH experiments to characterize alien chromosomes of *Ae. sharonensis.* The probes of Oligo-pSc119.2-1 and Oligo-pTa535-1 were developed based on repeat sequences available in NCBI(https://www.ncbi.nlm.nih.gov/), and can give a signal to distinguish chromosomes from wheat and other wild species [39]. Oligo-pSc119.2-1, Oligo-pTa535-1 and $(AAG)_6$ were 5' end-labeled with 6-carboxyfluorescein (6-FAM), 6-carboxytetramethylrhodamine (TAMRA), or Cy5, respectively. The ND-FISH analysis was performed according to the procedure described by [40]. Images were obtained using an epifluorescence microscope (BX51, Olympus, Tokyo, Japan).

Reverse-phase high-performance liquid chromatography (RP-HPLC) analysis of gluten proteins

RP-HPLC was used to separate and compare the contents of glutenins (HMW-GS, LMW-GS) and gliadins (ω, α/β, γ) of R7, Z636, and Z636×R7. The extraction method for glutenins and gliadins was carried out as described by [41] and [42]. The gluten proteins from each genotype were extracted from three independent flour samples. The glutenins (HMW-GS and LMW-GS) were extracted from 45 mg of flour from each genotype at room temperature by adding 1 mL of 0.3 mol/L NaI and 7.5% 1-propanol solution and carrying out gentle shaking for 10 min. The mixture was centrifuged at 17949×g for 10 min to recover the pellet. The pellet was suspended in 1 mL of 70% ethanol with gentle agitation at room temperature for 30 min. The suspension was then centrifuged at 17949×g for 10 min to collect the pellet for suspension in 1 mL of 55% isopropanol. The suspension was then incubated at 65℃ for 30 min with periodic agitation every 5 min, followed by centrifugation at 17949×g for 10 min. This step was repeated once to recover the pellet, with the pellet being suspended in 500 µL of 50% isopropanol, 80 mmol/L Tris-HCl (pH 8.0), and 1% (w/v) dithiothreitol and incubated at 65℃ for 1 h, after which 1% (w/v) 4-vinylpyridine was added to the mixture, which was incubated at 65℃ for 30 min, following which the suspension was centrifuged at 17949×g for 10 min to collect the supernatant, which represented the glutenin fraction. The gliadins were extracted from 45 mg flour in 1 mL of 70% ethanol, with periodic

agitation, for 1 h. The suspension was then centrifuged at 17949×g for 10 min, and the supernatant was retained as the gliadin fraction.

The gluten protein fractions were then filtered through a 0.45-μm pore size nylon filter before separating on an RP-HPLC, after loading 20 μL of supernatant onto the C18 column (Cat. 880995–902, ZORBAX300SB-C18; 4.6 mm×250 mm, 5 μm particle size; Agilent, Palo Alto, CA, USA). The RP-HPLC was conducted with an Agilent 1260 Infinity LC instrument (Agilent Technologies, Palo Alto, CA, USA). The separation conditions for gliadins began with 79% of mobile phase A (0.06% trifluoroacetic acid in ultra-pure water) and 21% of mobile phase B (0.06% trifluoroacetic acid in acetonitrile) at 0 min, followed by 52% A and 48% B at 55 min, and 79% A and 21% B at 60 min to achieve the separation of all gliadins. The isolation of glutenins begin with 75% A and 25% B at 0 min, followed by 52% A and 48% B at 45 min, and 75% A and 25% B at 50 min until all the glutenins were separated. The separation conditions were set to 60℃ for column temperature and 1 mL/min for flow velocity using a diode array detector (DAD) at a wavelength of 210 nm. There was a 10-min pause between runs to allow the column to be flushed by methyl alcohol. The contents of individual glutenins and gliadins were calculated by combining the related peak areas in the chromatograms.

Starch analysis

The total starch, amylopectin, and amylose contents of R7, Z636 and Z636×R7 were measured using the Total Starch Assay Kit and Amylose/Amylopectin Assay Kit (Megazyme, Bray, Ireland) according to the manufacturer's protocols. The final contents of total starch, amylopectin, or amylose in each genotype were determined using the data from three independent flour sample replicates.

Total RNA isolation and transcriptome sequencing

The developing grains of R7, Z636, and Z636×R7 were harvested at 15 days post-anthesis (DPA), transferred immediately into liquid nitrogen, and stored at − 80℃ until RNA was to be extracted. Total RNA was isolated using the Plant RNA Kit (Biofit, Chengdu, China), in accordance with the manufacturer's protocol. The RNA concentration was measured using the NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). For RNA-Seq (RNA-sequencing)/transcriptome analysis and metabolome analysis of the three biological replicates of each sample, analyses were sub-contracted to BioMarker Technologies (Beijing, China), with RNA-Seq being conducted on the Illumina HiSeq high-throughput sequencing platform (HiSeq 2000, San Diego, CA, USA).

Transcriptome assembly, annotation, and differential expression analysis of the genes

Raw data in the FastQ format were first processed through perl script developed in-house. In this step, clean data were obtained by removing from the raw data reads containing adapters, poly-N sequences, and low-quality sequence reads. Simultaneously, Q20, Q30, GC-content, and sequence duplication levels

of the clean data were calculated. All the downstream analyses were based on clean data of high quality. The clean reads were then mapped to the reference genome sequence of *Triticum_turgidum*_ssp._*durum* (https://www.interomics.eu/durum-wheat-genome); only reads with a perfect match or with one mismatch were further analyzed and annotated based on the reference genome, using the HISAT2 V2.1.0 tool (daehwankimlab.github.io).

Gene function was annotated based on the following databases: Pfam (Protein family) (https://pfam.xfam.org/); KO [Kyoto Encyclopedia of Genes and Genomes (KEGG) Ortholog database (https://www.genome.jp/kegg/ko.html); GO (Gene Ontology) (http://www.geneontology.org); and Swiss-Prot (a manually annotated and reviewed protein sequence database) (https://www.uniprot.org/). Quantification of the gene expression levels was achieved by fragments per kilobase of transcript per million fragments mapped (FPKM).

The differential expression analysis between two genotype groups was completed using the R package 'DEseq'. The resulting *p*-values were adjusted using the approach of Benjamini and Hochberg for controlling the false discovery rate (FDR). Genes with log2 Fold Change ≥ 2 and FDR < 0.01 were assigned as DEGs.

Metabolite extraction, detection, and bioinformatic analysis

The developing grain samples from each of the three genotypes were extracted for metabolome analysis. Briefly, 50 mg of each sample (with three replicates of each sample) were placed in an Eppendorf tube, to which was added 1 mL extraction solvent (methanol: acetonitrile: water = 2:2:1 [v/v/v]), containing an internal standard (2 μg/ml). Each sample was then homogenized in a ball mill for 4 min at 45 Hz, then ultrasonically extracted for 5 min, the container being cooled by incubation in ice water. After three rounds of homogenization, the sample was incubated for 1 h at − 20℃ to precipitate proteins. The sample was then centrifuged at 15294 ×g for 15 min at 4℃, after which the supernatant (500 μL) was transferred into a fresh Eppendorf tube. The supernatant was evaporated to dryness in a vacuum concentrator without heating and 100 μL extraction solvent (acetonitrile: water = 1:1 [v/v]) were added to re-suspend it. The sample was then vortexed for 30 s, sonicated for 10 min (in a 4℃ water bath), and centrifuged for 15 min at 15294 ×g at 4℃. Afterward, the supernatant (60 μL) was transferred into a fresh 2-mL liquid chromatography/mass spectrometry (LC/MS) glass vial. Lastly, 10 μL from each of the three independent replicates of each sample was then taken and pooled as quality control (QC) samples and 60 μL of supernatant were taken for an ultra-high-performance liquid tandem chromatography-quadrupole time of flight-mass spectrometry (UHPLC-QTOF-MS) analysis.

The LC/MS system for metabolomic separation and detection was composed of ultra-high performance liquid chromatography (UPLC Acquity I-Class PLUS, Waters, USA) and high resolution mass spectrum (UPLC Xevo G2-XS QTof, Waters, USA). The column used was Acquity UPLC HSS T3 (1.8 μm particle size, 2.1×100 mm) (Waters, USA). For the positive and negative ion modes, the mobile phases were A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile) and the injection volumes were 1.5 μL (positive) or 1 μL (negative). The TripleTOF mass spectrometer (Waters, USA)was used for its ability to

acquire MS/MS spectra on an information-dependent basis during an LC/MS experiment. In this mode, the acquisition software (Analyst TF 1.7, AB Sciex) continuously evaluates the full-scan survey MS data as it collects and triggers the acquisition of MS/MS spectra, depending on preselected criteria. In each cycle, 12 precursor ions with intensities < 100 were chosen for fragmentation at a collision energy (CE) of 30V (15 MS/MS events with product ion accumulation time of 50 ms each). The electrospray ionization source conditions were set as follows: ion source gas 1 at 60 psi, ion source gas 2 at 60 psi, curtain gas at 35 psi, and Ion Spray Voltage Floating (ISVF) at 5,000V value of OPLS-DA and the univariate statistical analysis of the t-test P-value to screen differentially expressed metabolites (DEMs) among different genotype groups for comparison. The screening criteria log2 Fold Change ≥ 1, P-value < 0.05, and VIP > 1 were taken into account to identify the DEMs between groups. Gene Ontology enrichment analysis of the differentially expressed genes (DEGs) was executed by the 'goseq' R package, based on the Wallenius non-central hyper-geometric distribution, which can adjust for gene length bias in DEGs [43]. The KEGG pathway enrichment analysis of the DEGs and DEMs was carried out using the KOBAS software [44]. The heatmap was created using the TBtool V0.66839 software [45].

Real-time quantitative PCR (RT-qPCR) validation

The expression levels of 12 genes encoding components of HMW-GSs, gliadins, LMW-GSs, and starch were quantified by qPCR to verify the RNA-Seq data. First-strand cDNA was synthesized using the PrimeScript™ RT Reagent Kit with gDNA Eraser (Takara, Dalian, China), and qPCR was carried out with SYBR® Premix Ex *Taq™* II (Takara, Beijng) on a CFX 96 Real-Time System (Bio-Rad, Hercules, CA, USA). The CFX Manager software (Bio-Rad) was used to analyze the qPCR data and to calculate relative expression with the $2^{-\triangle\triangle Ct}$ method. The wheat actin and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) genes were used as internal reference genes to normalize the relative expression of candidate genes.

Statistical analysis

Statistical analysis was conducted using ANOVA.

Declarations

Statistical analysis

Statistical analysis was conducted using ANOVA.

Acknowledgements

We are grateful for the funding support.

Author Contributions

Q. H: Validation, Formal analysis, Investigation, Data curation, Writing-original draft, Visualization. J. L: Methodology, Resources. XL. C: Software, Methodology. C. G, Y. T, Q. X and YZ. Z: Software,

Methodology. Q. C, HP. T: Resources, Supervision. PF. Q, M. D and J. M: Visualization, Writing - review & editing. GY. C, JR. W: Writing - review & editing. YM. W and YL. Z: Supervision, Funding acquisition. QT. J: Conceptualization, Supervision, Funding acquisition, Project administration, Writing - review & editing. All authors read and approved the final manuscript.

Fundings

This work was supported by the Sichuan Science and Technology Program, PR China, (2023YFH0041), and the Science & Technology project of Chengdu, Sichuan Province, PR China (2022-YF05- 00022-SN).

Availability of data and materials

All data generated or analysed during this study are available in Supplementary Materials.

Ethics approval and consent to participate

The authors declare that all study complies with relevant institutional, national, and international guidelines and legislation for plant ethics in the methods section.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Hannah LC, James M: **The complexities of starch biosynthesis in cereal endosperms**. *Current Opinion in Biotechnology* 2008, **19**(2):160-165.
2. Wang Z, Li W, Qi J, Shi P, Yin Y: **Starch accumulation, activities of key enzyme and gene expression in starch synthesis of wheat endosperm with different starch contents**. *Journal of Food Science Technology* 2014, **51**(3):419-429.
3. Nakamura T, Yamamori M, Hirano H, Hidaka S: **Identification of three Wx proteins in wheat (Triticum aestivum L.)**. *Biochemical genetics* 1993, **31**:75-86.
4. Jenner C, Ugalde T, Aspinall D: **The physiology of starch and protein deposition in the endosperm of wheat**. *Functional Plant Biology* 1991, **18**(3):211-226.
5. !!! INVALID CITATION !!! [5-8].
6. Payne PI: **Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality**. *Annual Review of Plant Physiology* 1987, **38**:141-153.
7. Yan Z, Wan Y, Liu K, Zheng Y, Wang D: **Identification of a novel HMW glutenin subunit and comparison of its amino acid sequence with those of homologous subunits**. *Chinese Science*

Bulletin 2002, **47**: 222-226.

8. De Bustos A, Jouve N: **Characterisation and analysis of new HMW-glutenin alleles encoded by the Glu-R1 locus of Secale cereale**. *Theor Appl Genet* 2003, **107**(1):74-83.
9. Liu Z, Yan Z, Wan Y, Liu K, Zheng Y, Wang D: **Analysis of HMW glutenin subunits and their coding sequences in two diploid Aegilops species**. *Theor Appl Genet* 2003, **106**(8):1368-1378.
10. Jiang QT, Wei YM, Lu ZX, Pu ZE, Lan XJ, Zheng YL: **Structural variation and evolutionary relationship of novel HMW glutenin subunits from Elymus glaucus**. *Hereditas* 2010, **147**(3):136-141.
11. Wang J, Yan Z, Wei Y, Zheng Y: **A novel high-molecular-weight glutenin subunit gene Ee1.5 from Elytrigia elongata (Host) Nevski**. *Journal of Cereal Science* 2004, **40**(3):289-294.
12. Chen ZJ: **Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids**. *Annual Review of Plant Biology* 2007, **58**(1):377-406.
13. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS *et al*: **Ancestral polyploidy in seed plants and angiosperms**. *Nature* 2011, **473**(7345):97-100.
14. Hao M, Li A, Shi T, Luo J, Zhang L, Zhang X, Ning S, Yuan Z, Zeng D, Kong XJBg: **The abundance of homoeologue transcripts is disrupted by hybridization and is partially restored by genome doubling in synthetic hexaploid wheat**. 2017, **18**:1-12.
15. Martinez Palacios P, Jacquemot M-P, Tapie M, Rousselet A, Diop M, Remoue C, Falque M, Lloyd A, Jenczewski E, Lassalle GJMb *et al*: **Assessing the response of small RNA populations to allopolyploidy using resynthesized Brassica napus allotetraploids**. 2019, **36**(4):709-726.
16. Yoo M, Szadkowski E, Wendel JJH: **Homoeolog expression bias and expression level dominance in allopolyploid cotton**. 2013, **110**(2):171-180.
17. Wang X, Shi X, Hao B, Ge S, Luo JJNP: **Duplication and DNA segmental loss in the rice genome: implications for diploidization**. 2005, **165**(3):937-946.
18. Olivera PD, Steffenson BJ: **Aegilops sharonensis: Origin, genetics, diversity, and potential for wheat improvement**. *Botany* 2009, **87**(8):740-756.
19. Yu G, Champouret N, Steuernagel B, Olivera PD, Simmons J, Williams C, Johnson R, Moscou MJ, Hernández-Pinzón I, Green P: **Discovery and characterization of two new stem rust resistance genes in Aegilops sharonensis**. *Theoretical Applied Genetics* 2017, **130**(6):1207-1222.
20. Jiang Q-T, Ma J, Wei Y-M, Liu Y-X, Lan X-J, Dai S-F, Lu Z-X, Zhao S, Zhao Q-Z, Zheng Y-L: **Novel variants of HMW glutenin subunits from Aegilops section Sitopsis species in relation to evolution and wheat breeding**. *BMC plant biology* 2012, **12**(1):1-13.
21. Zhao QZ, Jiang QT, Yang Q, Ma J, Wang JR, Chen GY, Liu YX, Lu ZX, Wei YM, Zheng YL: **Characterization of intergeneric hybrid between common wheat and Aegilops sharonensis (Eig) and transfer of alien high molecular weight glutenin subunits into wheat**. *Cereal Research Communications* 2014, **42**(4):640-647.

22. Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG *et al*: **Durum wheat genome highlights past domestication signatures and future improvement targets**. *Nat Genet* 2019, **51**(5):885-895.
23. Kucek LK, Veenstra LD, Amnuaycheewa P, Sorrells ME: **A Grounded Guide to Gluten: How Modern Genotypes and Processing Impact Wheat Sensitivity**. *Comprehensive Reviews in Food Science Food Safety* 2015, **14**(3).
24. Mcmaugh SJ, Thistleton JL, Emma A, Jixun L, Christine KR, Hong W, Min H, Oscar L, Ahmed R, Jobling SA: **Suppression of starch synthase I expression affects the granule morphology and granule size and fine structure of starch in wheat endosperm**. *Journal of Experimental Botany* 2014(8):2189.
25. Cabrera, Antonio, Griffey, Carl, Sneller, Clay, McKendry, Anne, Hoffstetter, Amber: **Genomic selection for wheat traits and trait stability**. *Theoretical Applied Genetics* 2016, **129**(9):1697-1710.
26. Yu K, Liu D, Chen Y, Wang D, Yang W, Yang W, Yin L, Zhang C, Zhao S, Sun J: **Unraveling the genetic architecture of grain size in einkorn wheat through linkage and homology mapping and transcriptomic profiling**. *Journal of Experimental Botany* 2019, **70**(18).
27. Regina A, Berbezy P, Kosar-Hashemi B, Li S, Cmiel, Mark, Larroque: **A genetic strategy generating wheat with very high amylose content**. *Plant biotechnology journal* 2016, **13**(9):1276-1286.
28. Li W, Chen Z, Li Z, Zhao X, Pu Z, Chen G, Jiang Q, Wei Y, Zheng YJCRC: **Characterization of starch synthetic genes and starch granule during seeds development between synthetic hexaploid wheat and its parents**. 2018, **46**(2):275-286.
29. Liu B, Vega J, Feldman MJG: **Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences**. 1998, **41**(4):535-542.
30. Ozkan H, Levy AA, Feldman MJTPC: **Allopolyploidy-induced rapid genome evolution in the wheat (Aegilops–Triticum) group**. 2001, **13**(8):1735-1747.
31. Veliz CG, Criado MV, Roberts IN, Echeverria M, Prystupa P, Prieto P, Boem FG, Caputo C: **Phloem sugars and amino acids as potential regulators of hordein expression in field grown malting barley (Hordeum vulgare L.)**. *Journal of Cereal Science* 2014, **60**(2):433-439.
32. Good AG, Shrawat AK, Muench DG: **Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production?** *Trends in Plant Science* 2004, **9**(12):597-605.
33. Gadaleta A, Nigro D, Giancaspro A, Blanco A: **The glutamine synthetase (GS2) genes in relation to grain protein content of durum wheat**. *Functional Integrative Genomics* 2011, **11**(4):665-670.
34. Ying G, Sun J, Zhang G, Wang Y, Li S: **Haplotype, molecular marker and phenotype effects associated with mineral nutrient and grain size traits of TaGS1a in wheat**. *Field Crops Research* 2013, **154**(3):119-125.
35. Gan L, Huang B, Song Z, Zhang Y, Zhang Y, Chen S, Tong L, Wei Z, Yu L, Luo X: **Unique glutelin expression patterns and seed endosperm structure facilitate glutelin accumulation in polyploid rice seed**. *Rice* 2021, **14**:1-19.

36. Mackie AF, S. LE, J. SP, D L: **Molecular and Biochemical Characterisation of HMW Glutenin Subunits from T. tauschii and the D Genome of Hexaploid Wheat**. *Journal of Cereal Science* 1996, **23**(3):213-225.
37. Payne P, Holt LM, Krattiger A, Carrillo J: **Relationships between seed quality characteristics and HMW glutenin subunit composition determined using wheats grown in Spain**. *Journal of Cereal Science* 1988, **7**(3):229-235.
38. Han F, Lamb JC, Birchler JA: **High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize**. *Proceedings of the National Academy of Sciences* 2006, **103**(9):3238-3243.
39. Tang Z, Yang Z, Fu S: **Oligonucleotides replacing the roles of repetitive sequences pAs1, pSc119.2, pTa-535, pTa 71, CCS1, and PA WRC.1 for FISH analysis**. *Journal of Applied Genetics* 2014, **55**(3).
40. Fu S, Lei C, Wang Y, Meng L, Tang Z: **Oligonucleotide Probes for ND-FISH Analysis to Identify Rye and Wheat Chromosomes**. *Scientific Reports* 2015, **5**(1):10552.
41. Marchylo BA, Kruger JE, Hatcher DW: **Quantitative reversed-phase high-performance liquid chromatography analysis of wheat storage proteins as a potential quality prediction tool**. *Journal of Cereal Science* 1989, **9**(2):113-130.
42. Yan Y, Hsam S, Yu JZ, Jiang Y, Ohtsuka I, Zeller FJ: **HMW and LMW glutenin alleles among putative tetraploid and hexaploid European spelt wheat (Triticum spelta L.) progenitors**. *Theoretical Applied Genetics* 2003, **107**(7):1321-1330.
43. Anders S, Huber W: **Differential expression analysis for sequence count data**. *Genome Biology* 2010, **11**(10):R106.
44. Mao X, Tao C, Olyarchuk JG, Wei L: **Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary**. *Bioinformatics* 2005, **21**(19):3787-3793.
45. Chen C, Chen H, Zhang Y, Thomas HR, Xia R: **TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data**. *Molecular Plant* 2020, **13**(8).

Figures

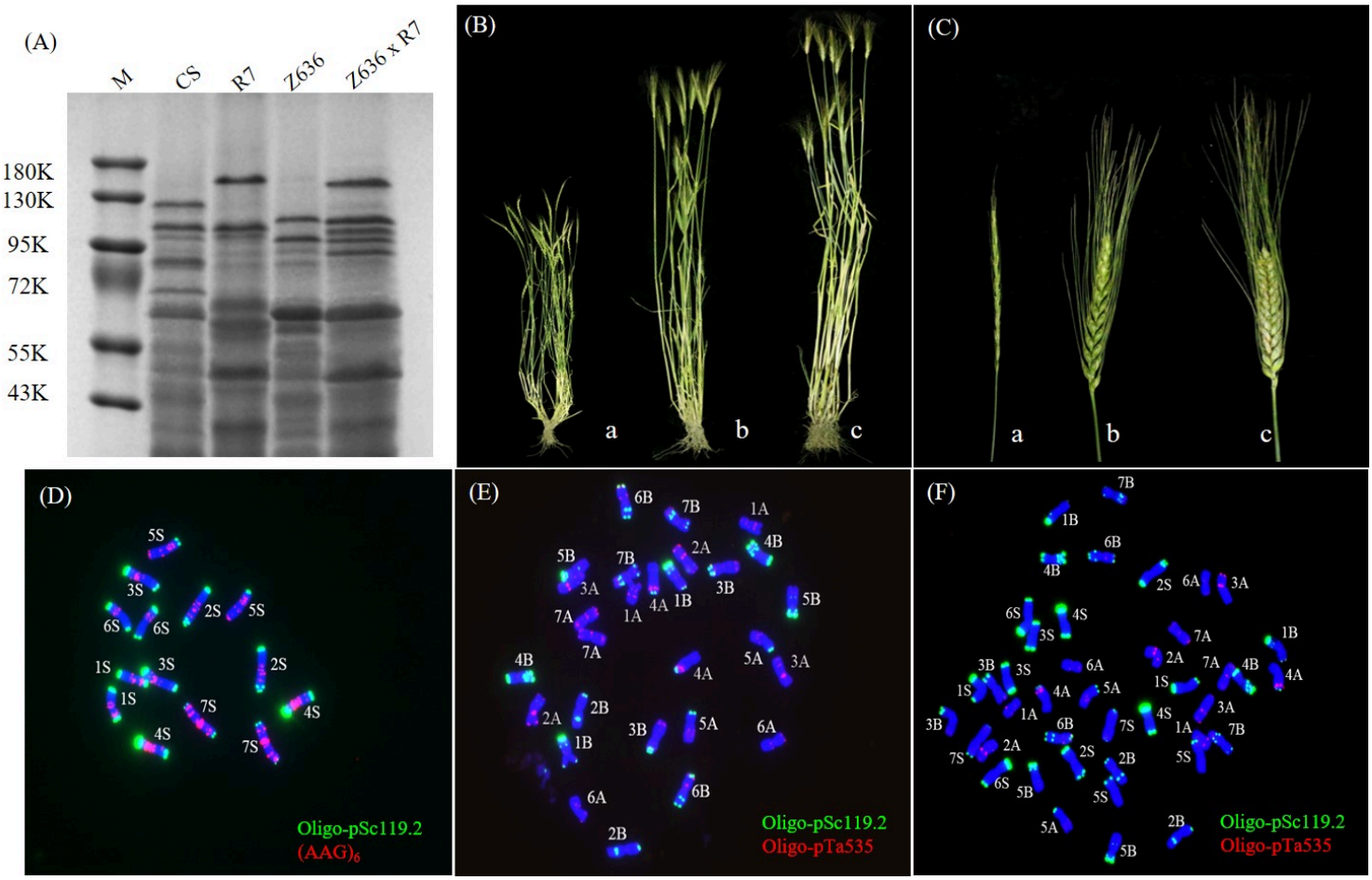


Figure 1

The morphology and cytology of R7, Z636 and Z636×R7. **A**, HMW-GS composition of R7, Z636 and Z636×R7 determined by SDS-PAGE (the full-length gel is included in Supplementary 1). **B**, **C**, Comparison of mature plant and panicle morphology. Number a-c, R7; Z636; Z636×R7, respectively. **D**, **E**, **F**, Sequential multi-colour FISH of R7、Z636、Z636×R7, showing that chromosomes of the A, B and S genome was represented on basis of probes pSc119.2 (green) and pTa535.

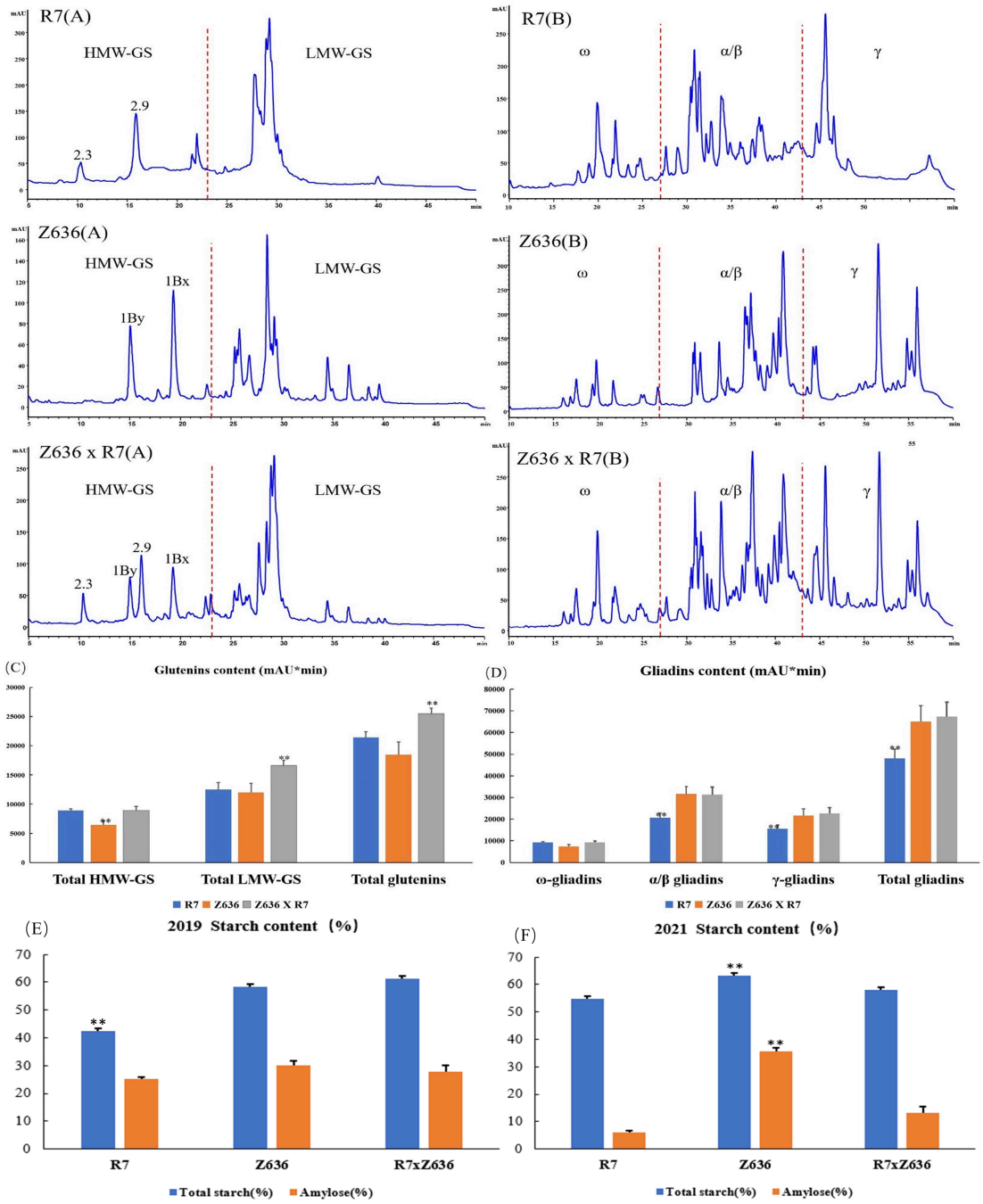


Figure 2

Contents of gluten protein and starch in the grains of R7, Z636 and Z636×R7. **A**, Contents of glutenin subunits by RP-HPLC analysis. **B**, Contents of gliadins by RP-HPLC analysis. **C**, Comparison of relative contents of different glutenins. **D**, Comparison of relative contents of different gliadins. **E**, **F**, Comparison of starch contents of R7, Z636 and Z636×R7 in 2019 and 2021.

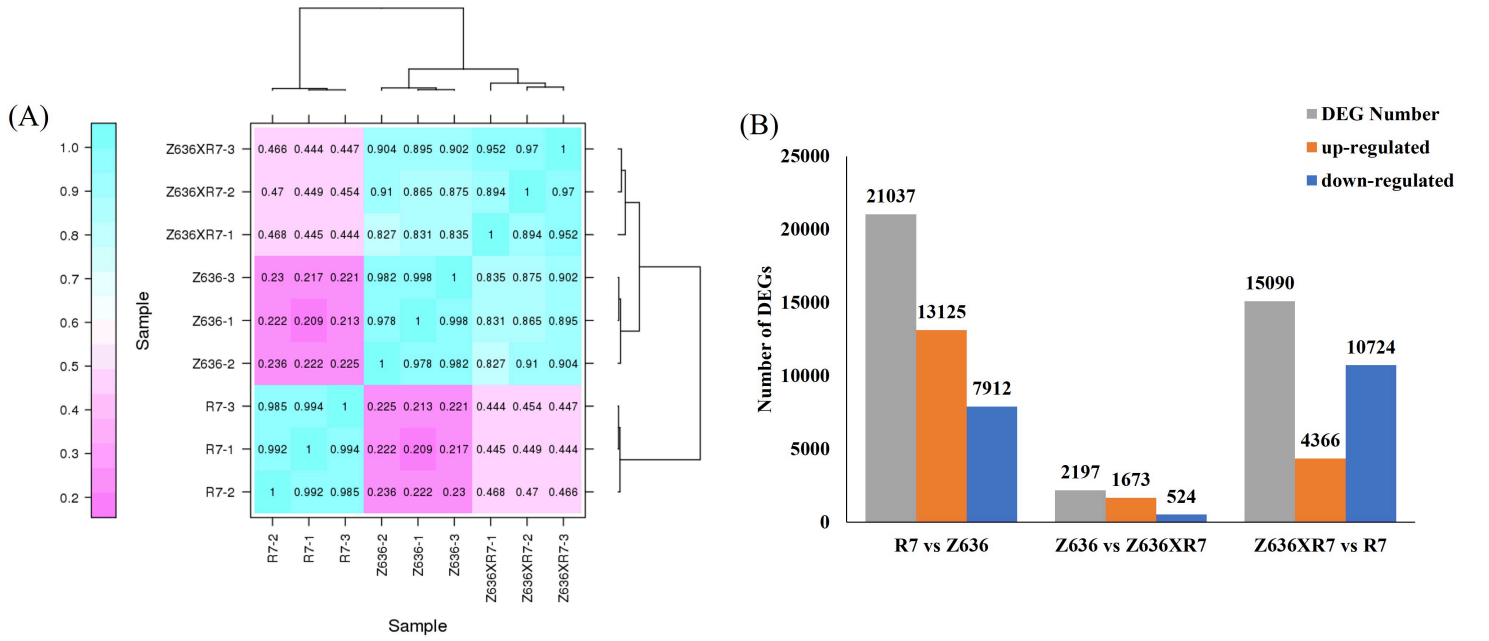


Figure 3

RNA-seq analysis in R7, Z636 and Z636×R7. **A**, Analysis of Pearson's correlation. **B**, Diferentially expressed genes (DEGs). Gray is DEGs number, orange is up-regulated expression, blue is down-regulated expression.

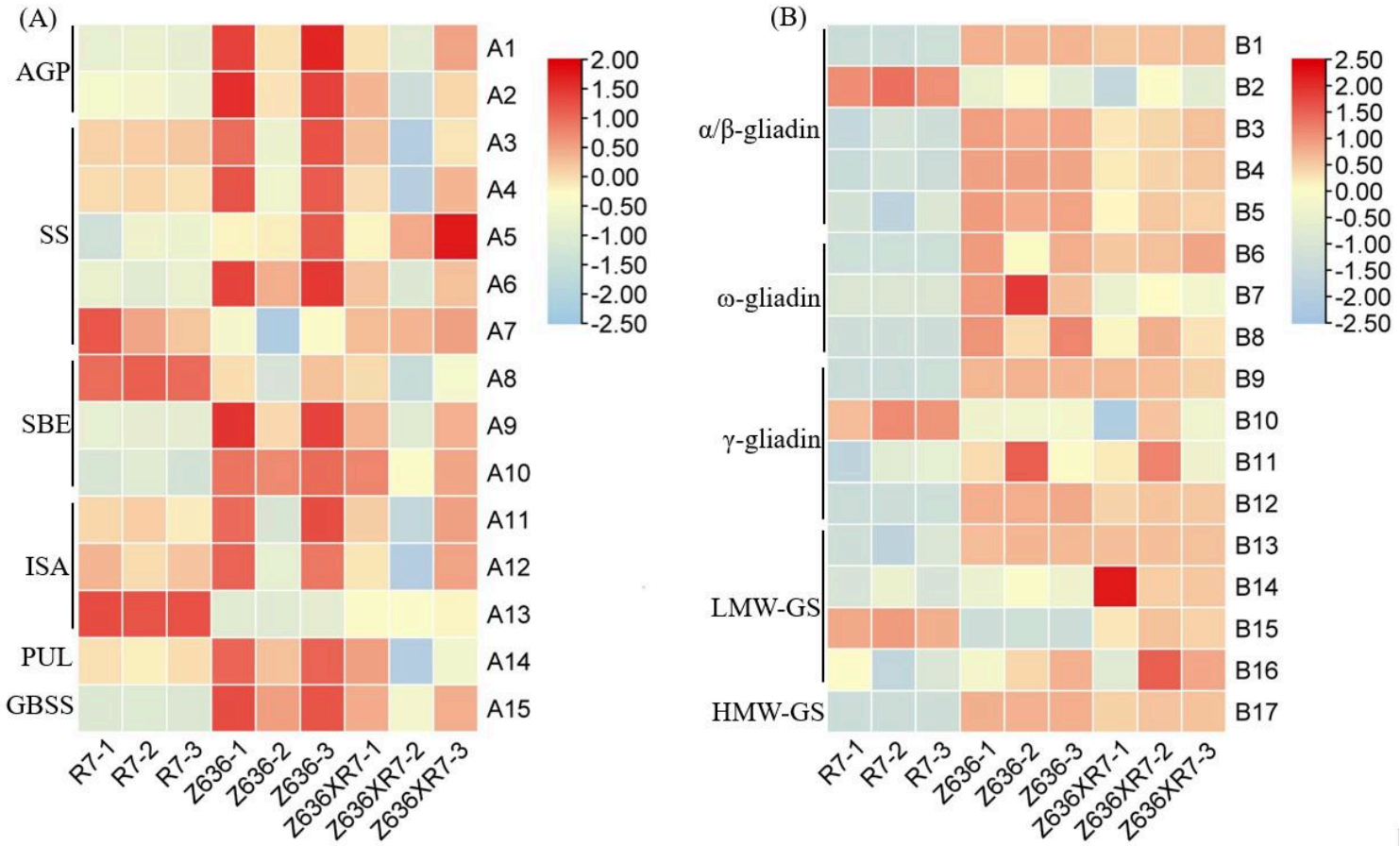


Figure 4

Expression patterns of starch synthase gene and storage protein gene. **A**, starch synthase genes. **B**, storage protein genes.

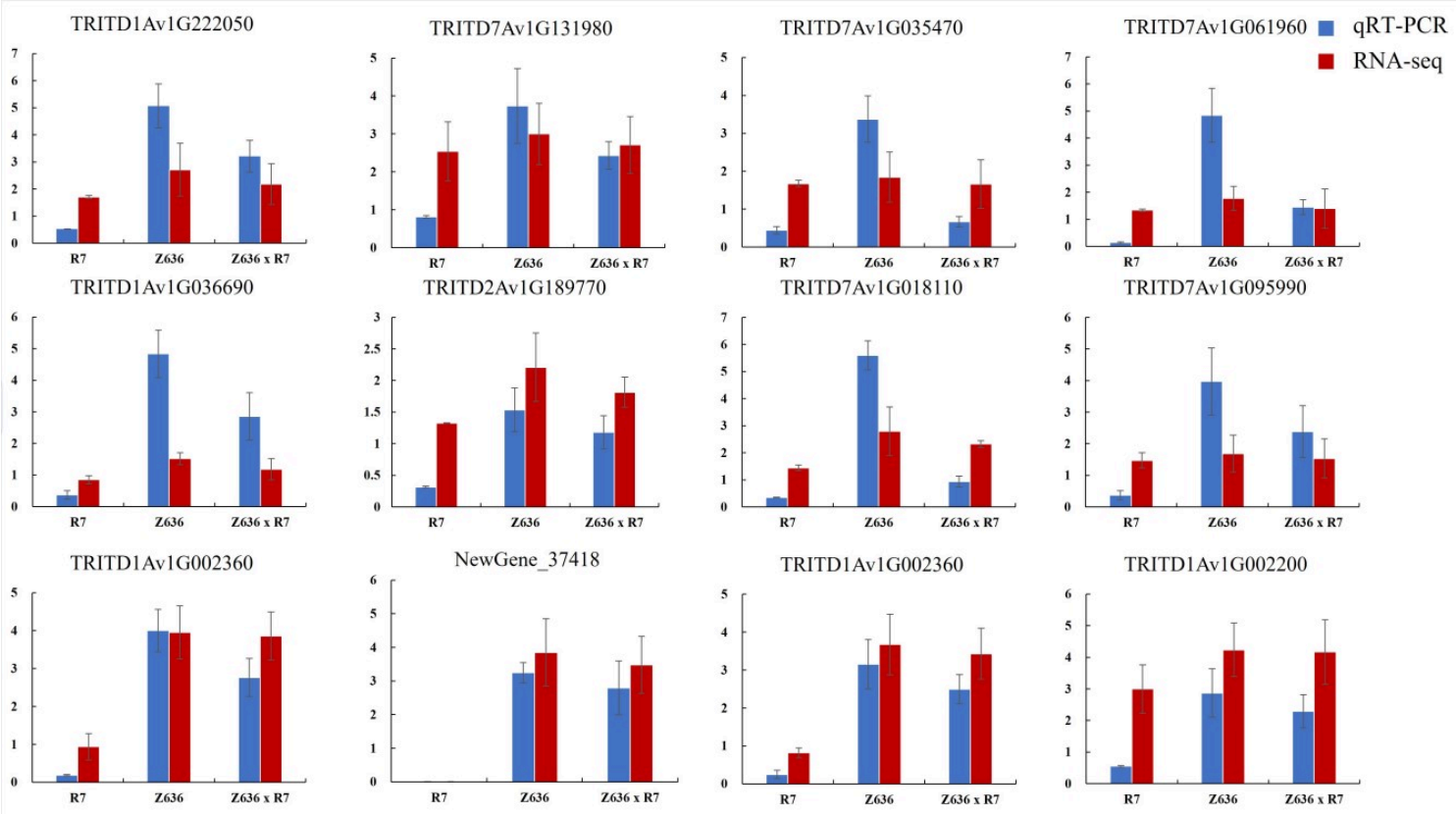


Figure 5

RT-qPCR validation of DEGs identified by RNA-Seq. The blue and red column represent RT-qPCR and RNA-Seq data.

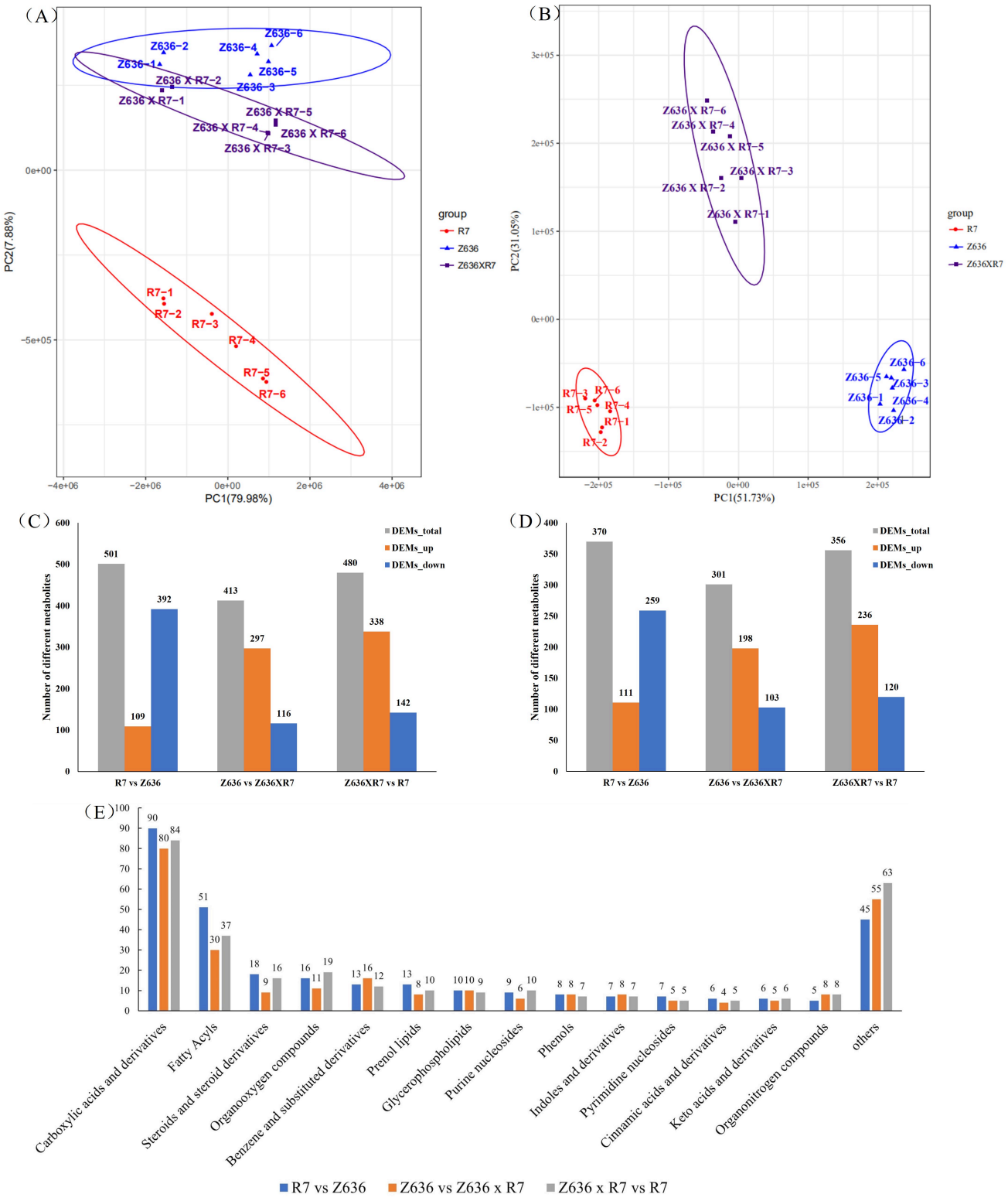


Figure 6

Metabolic variation in R7, Z636, Z636×R7. **A**, **B**, Principal component analysis in positive ion mode and negative ion mode,the X-axis represents the first principle component (PC1), and the Y-axis represents the second principal component (PC2). **C**, **D**, Statistical analysis of the number of different metabolites in positive and negative ion modes. Grey is DEMs-total, blue is DEMs-up, yellow is DEMs-down. **E**, Category of differential metabolites.

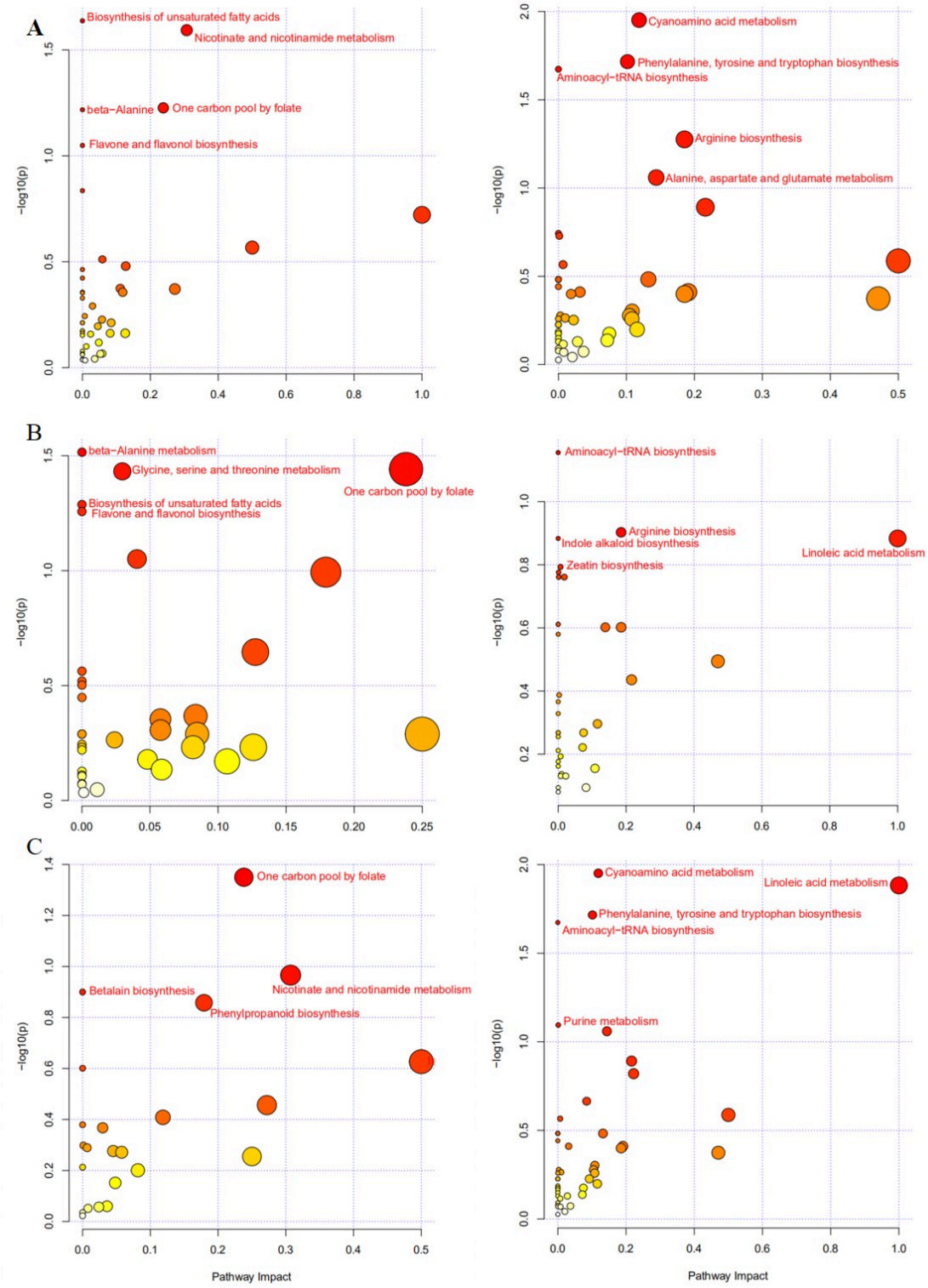


Figure 7

Pathway enrichment analysis of metabolites in positive ion modes and negative ion modes. **A**, R7 vs Z636. **B**, Z636 vs Z636×R7. **C**, Z636 vs Z636×R7. The left is positive ion modes and the right is negative ion modes.

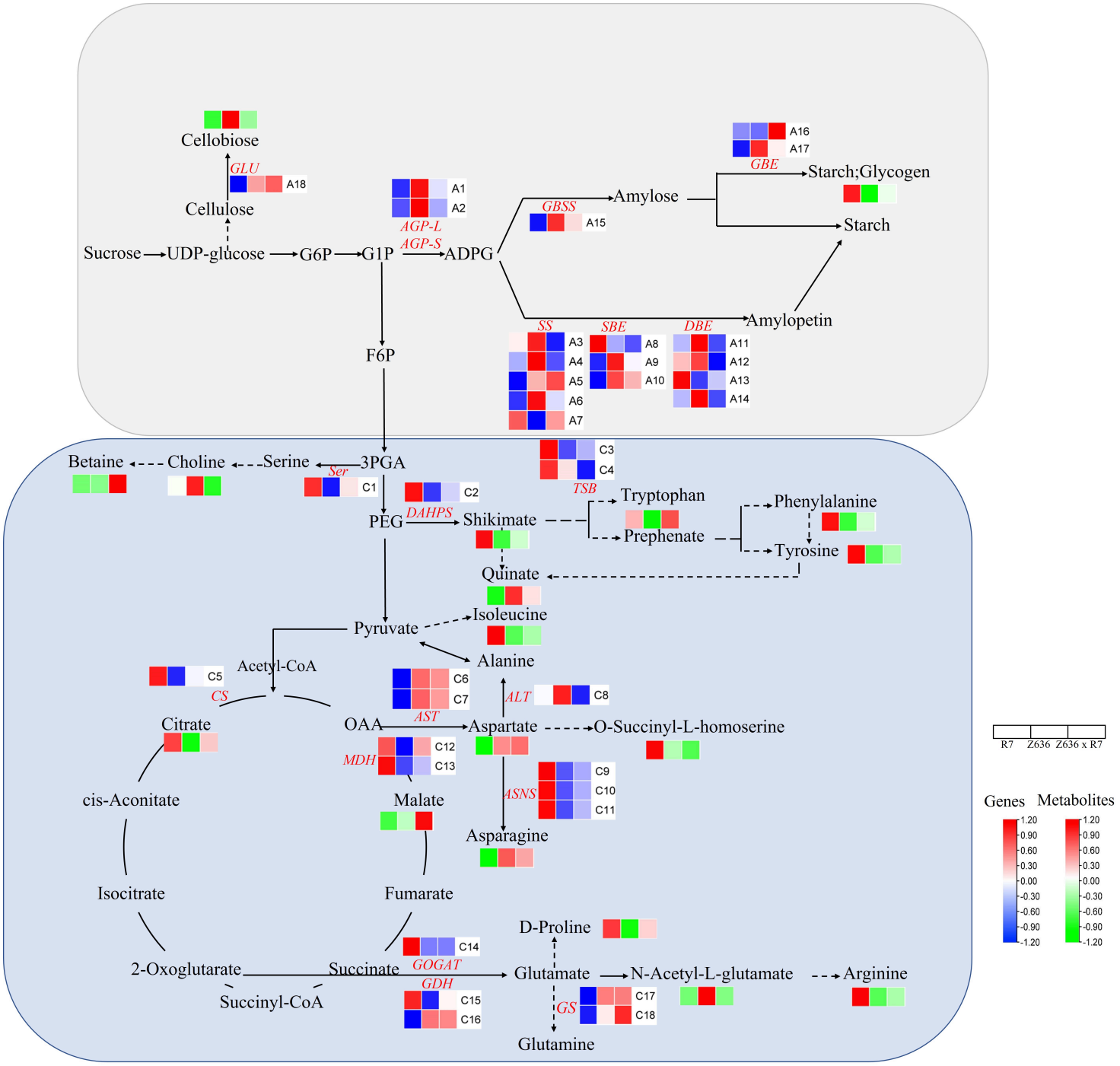


Figure 8

Metabolic pathways of R7, Z636 and Z636×R7. Black characters and red-green heat maps represent the names and heat maps of DEMs. Red characters and red-blue heat maps represent the abbreviations and heat maps of DEGs annotation function, and three single cells represent R7, Z636 and Z636×R7 (metabolites and genes are in the same order). Solid arrows indicate that there are no intermediates between the two metabolites, and dotted arrows indicate that some intermediates are omitted from the pathway between the two metabolites.

Supplementary Files

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

- SupplementaryMaterial1.docx
- SupplementaryMaterial2.docx
- SupplementaryMaterial3.docx
- SupplementaryMaterial4.xlsx
- SupplementaryMaterial5.xlsx
- SupplementaryMaterial6.xlsx
- SupplementaryMaterial7.xlsx
- SupplementaryMaterial8.xlsx