

Genome-wide Identification, Expression, Characteristics of the Prolamin Superfamily in *Thinopyrum Elongatum* and Its Kneading Performance in Common Wheat

Yu Gao

Shandong Agricultural University

Shoushen Xu

Shandong Agricultural University

Xin Ma

Shandong Agricultural University

Hongwei Wang

Shandong Agricultural University

Lingrang Kong

Shandong Agricultural University

Silong Sun (✉ sunsl@sdau.edu.cn)

Shandong Agricultural University

Research Article

Keywords: *Thinopyrum elongatum*, genome, prolamins, expression, evolution, celiac disease

Posted Date: May 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-513138/v1>

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

[Read Full License](#)

Abstract

Background: Prolamins, unique to Gramineae (grasses), play a key role in the human diet. *Thinopyrum elongatum* (also known as tall wheatgrass, rush wheatgrass, or Eurasian quackgrass) of *Elytrigia* is genetically well-characterized, but little is known about its prolamin genes and the relationships with homologous loci in the *Triticum* genus.

Results: In this study, a total of 19 α -gliadin, 9 γ -gliadin, 19 ω -gliadin, 2 high-molecular-weight glutenin subunit (HMW-GS), and 5 low-molecular-weight glutenin subunit (LMW-GS) genes in the *Th. elongatum* genome were annotated. The transcriptome data of *Th. elongatum* exhibited differential expression in quantity and pattern in the same subfamily or different subfamilies. In addition, microsynteny and phylogenetic analysis revealed dynamic changes of prolamin gene region and genetic affinities among *Th. elongatum*, *T. aestivum*, *T. urartu*, and *Aegilops tauschii*. The E genome, like the B genome, only contained DQ8-glia- α 1/DQ8.5-glia- α 1, which provided a theoretical basis for the study of celiac disease (CD). Dough rheological properties of *T. aestivum*-*Th. elongatum* disomic substitution (DS) lines 1E(1A), 1E(1D), and 3E(3A) showed much higher peak height values than that of their parent.

Conclusions: Overall, this study provides a comprehensive overview of the prolamin gene superfamily in *Th. elongatum*, and suggests a promising use of this species in the generation of improved wheat breeds intended for the human diet.

Background

Prolamins (glutenins and gliadins), comprising 80% of wheat endosperm protein, are the main component of gluteins [1]. High- and low-molecular-weight glutenin subunits (HMW-GSs and LMW-GSs) form polymeric proteins by interchain disulfide bonds, imparting the elasticity of dough. Gliadins, divided into α -, β -, γ -, and ω -gliadins, are monomeric proteins of 30–78 kDa that determine the ductility and viscosity of dough [2–5]. These unique properties determine the quality of wheat flour in various technological processes and enable the manufacture of a wide range of products such as bread, pasta, noodles, cakes, and pastries [6, 7]. Prolamin genes were reported to be closely linked and mainly located on a few chromosomes in common wheat [8]. The HMW-GS genes are encoded by the *Glu-1* loci on the long arm of the first homologous group [9]. The γ -gliadin and ω -gliadin genes are encoded by the *Gli-1* loci closely linked to the *Glu-3* loci encoding LMW-GS genes on the short arm of the same chromosome, on which HMW-GS genes are located [10]. The α -gliadin genes are encoded by the *Gli-2* loci on the short arm of the sixth homologous group [11].

Prolamin genes belong to a large superfamily. Comprehensive understanding of these genes is challenging, but essential for improving the end-use quality of wheat flour [7]. Generally, *Triticum urartu* of the *Triticum* genus and *Aegilops tauschii* of the *Aegilops* genus are regarded as donors of the A and D genomes of common wheat [12]. *Thinopyrum elongatum* of *Elytrigia*, closely related to wheat, has excellent performance in biotic and abiotic stress resistance [13]. At present, genome data of these

gramineous crops have been published, but the relationship among their prolamin genes has not been well described, including gene features and molecular characteristics [13–16]. High pseudogene rate is associated with gliadin families and this was estimated to be 87% of 230 distinct α -gliadin gene sequences in several diploid wheat species [1]. Hence, it is necessary to understand the expression profile of these genes to clarify the mechanism of gene action. However, current research is still insufficient, especially regarding species other than common wheat [17].

Gene clusters of a gene family are often prone to genetic variations in copy number, sequence polymorphism, and gene expression [18]. Comparisons of these homologous regions between different genomes of related species will provide insights into gene differentiation, as well as local rearrangements [19]. For example, a comparison of the *Gli-2* loci between *T. dicoccoides* Korn and *T. aestivum* cv. Chinese Spring (CS) showed a large sequence difference between the two A genomes and a conservative region between the two B genomes [20]. To date, a detailed analysis of other genomes (except for the A, B, and D genomes) on the interval of these loci has not been reported. Prolamins can reflect the evolutionary relationship of Gramineae species to a certain extent [21].

Gluten is the most important source of proteins from wheat for human beings. Unfortunately, prolamins of gluten are also responsible for certain intolerances, among which celiac disease (CD) is one of the most common wheat-related disorders [22]. CD is a chronic intestinal immune-mediated bowel disease, which occurs in genetically susceptible people and is caused by the intake of gluten [23]. α -gliadins are the main substance causing CD [18]. A breeding effort had been proposed to develop wheat with reduced immunoreactive epitopes while retaining baking functions [24]. Prolamin genes from wild relatives of wheat were reported to have low gluten toxicity and could be used to improve wheat quality by distant hybridization [25].

In this study, prolamin genes were obtained and comparative analysis were carried out between *Th. elongatum* and three related species (*T. aestivum*, *T. urartu*, and *Ae. tauschii*) in terms of gene numbers, dynamic expression pattern, molecular characteristics, chromosome location, microsynteny, phylogenetic relationship, and CD epitope content. Mixogram tests of disomic substitution (DS) lines 1E(1A) and 1E(1D) were also performed. The results presented will also provide important evolution information in Triticeae and the application of prolamin genes of *Th. elongatum* in wheat breeding.

Results

Identification of the prolamin genes in *Th. elongatum* and error correction in the related species

A total of 19 α -gliadin genes, 9 γ -gliadin genes, 19 ω -gliadin genes, 2 HMW-GS genes, and 5 LMW-GS genes were identified and annotated in the *Th. elongatum* genome while no β -gliadin genes were detectable (Additional file 1: **Table S1**). In addition, some annotation errors were found in related species. For example, there were no sequences of *AET6Gv20125500.1* and *AET6Gv20126100.1* in the existing

CDS and protein files of *Ae. tauschii*. By checking their positions provided in that study, we found an incorrect coordinate layout [15]. All prolamin genes found in this study were manually checked and corrected according to their structure information (Additional file 1: **Table S2**) [8, 26–29]. Because of the lack of complete sequence of some ω -gliadins and HMW-GSs, detailed investigations of them were not performed.

Characteristics and sequence analysis of the prolamin gene superfamily

All prolamin genes of four species were first named based on abbreviations of species names and chromosomal locations. Then, the pl, MW, sequence length, and other characteristics of putative functional prolamin genes were calculated (**Additional file 1: Table S3**). No obvious differences in prolamin characteristics between *Th. elongatum* and other three species were found except for *Tel_LMW_glutenin_1E_1*, a special prolamin gene that has the highest MW and the longest protein sequence (**Additional file 1: Table S3**).

We describe the results of prolamin gene number changes among different ploidy species we studied (Fig. 1A). Only HMW-GS gene subfamilies, carrying one x- and one y-type in each diploid genome, follow a conservative pattern in the process of evolution [7]. In contrast, no specific pattern was observed among gliadin subfamilies across the species. What is particularly noticeable is that the numbers of *Th. elongatum* were the highest in three gliadin subfamilies of the three diploid species, and the number of ω -gliadin genes of *Th. elongatum* was about 3–5 times that of the other two diploid species. A common feature of gliadins of the four species was that the numbers of α -gliadin genes were the largest, followed by ω -gliadin genes, and finally γ -gliadin genes.

In order to determine the proportion of prolamin genes that were effectively expressed in different species, the pseudogene rates involved in related gene families were calculated according to the pseudogene number based on the inference from gene sequences to protein sequences (**Additional file 1: Table S3**). *Th. elongatum* again had the highest pseudogene rates in α -gliadin (84.2%) and γ -gliadin gene families (77.8%), while no pseudogenes were found in γ -gliadin and LMW-glutenin gene families in *T. urartu* (Fig. 1B). Comparisons of pseudogene rates revealed that few α -gliadin and γ -gliadin genes were effectively expressed in *Th. elongatum*, which may be related to their smaller seed.

Expression pattern and electrophoretic maps of prolamin genes in *Th. elongatum*

The study of prolamin gene expression can expand our knowledge of their functions by using transcriptome data. Genes with an average TPM value greater than 5000 were consistent with putative functional genes, which further indicated that our inference is correct (Fig. 2A, **Additional file 1: Table S4**). Next, these highly expressed genes were further studied to discover their contribution to grain development. Comparative analysis revealed that the expression characteristics of prolamin genes in *Th.*

elongatum were different in the same subfamily, which was one of the factors influencing grain prolamin content. For example, the expression of *Tel_gamma_gliadin_1E_8* was about 5–6 times that of *Tel_gamma_gliadin_1E_2*. Slight differences in gene expression were found in the α -gliadin, LMW-GS and HMW-GS gene subfamilies. In addition, different prolamin gene subfamilies had their own specific expression patterns. In α -gliadin and LMW-GS gene subfamilies, the expression level of these genes in the semi-grain stage was higher than that in the grain stage, which indicated that the expression of these genes increased at first and then decreased [28]. However, the expression of γ -gliadin genes has been increasing and we speculate that they will decline rapidly or slowly in the future due to reduced gene activity with the maturity of grains. In the HMW-GS gene subfamily, the expression of x-type HMW-GS genes was different from that of y-type. The expression of *Tel_hmw_glutenin_1E_x* in the half-grain stage was higher than that in the grain stage, while the expression of *Tel_hmw_glutenin_1E_y* increased slightly from the half-grain to the grain stage. Among these expressed genes, the expression of *Tel_gamma_gliadin_1E_8* was the highest, followed by three genes of the α -gliadin gene subfamily. From these results, the expression patterns of prolamin gene subfamilies are complex and diverse.

Now, we turn to the experimental evidence for prolamin expression. In the electrophoretic map of gliadins, with CS as control, a total of 11 gliadin bands were obtained in *Th. elongatum* (Fig. 2B). In the ω -gliadin region, the number of protein bands of *Th. elongatum* was significantly more than that of CS, with about 7 bands. Two bands in γ -gliadin region were detected in *Th. elongatum*. Although two protein bands of *Th. elongatum* located at the β -gliadin region of CS, no β -gliadin genes in its genome were detected. We thought they should be α -gliadin genes. In the electrophoretic map of glutenin subunits, compared with CS, a total of 6 bands were obtained (Fig. 2C). A band (Ex) of HMW-GS was near the 8-subunit of CS and the other band was lower, which is consistent with the map of a previous study [29]. Both electrophoretic patterns show polymorphism of prolamins between *Th. elongatum* and bread wheat, indicating that *Th. elongatum* is rich in prolamin genes and can be utilized in manufacturing products for human consumption.

Chromosomal location and duplication of prolamin genes

Consistent with previous studies, the γ -gliadin, ω -gliadin, and LMW-GS genes of the four species are distributed on the short arm, and the HMW-GS genes are found on the long arm of the first homologous group [10, 30–31]. Most α -gliadins of the four species were mainly distributed on the sixth homologous group (Fig. 3) [11]. It is interesting to note that a new α -gliadin cluster was detected with four pseudogenes on the short arm of chromosome 7E (109202365–109312140 bp). Strangely, microsynteny analysis showed that these α -gliadin genes on chromosome 6E and 7E did not belong to paralogous genes. To date, the origin of the loci of chromosome 7E remains unknown.

Both tandem duplication and segmental duplication are associated with gene production. The proportions of tandem repeat genes of α -gliadins, γ -gliadins, and LMW-GSs in *Th. elongatum* were 73.7%, 88.9%, and 50%. This supports the hypothesis that the major expansion of prolamin genes in *Th. elongatum* was through tandem duplication of the above three families.

To better analyze the evolutionary mechanism of α -gliadin, γ -gliadin, and LMW-GS gene subfamilies, an interval (from 5417104 to 11909933 bp) covering the first to the last prolamin gene on the short arm of 1E chromosome was used for microsynteny analysis with the other genomes (Fig. 4). The results show that all these regions of analyzed chromosomes were located in collinear blocks and indicated two findings that an inversion occurred between *Glu-3* and *Gli-1* loci on chromosome 1D of bread wheat and an orthologous ω -gliadin gene far away from the other ω -gliadin genes was not detected in the A genome of *T. urartu*, the D genome of *Ae. Tauschii* and common wheat. Changes in these gene loci between homologous species provide evidence for the dynamic evolution of the prolamin gene superfamily.

The next section of research was concerned with the microsynteny of a single prolamin subfamily between other selected species and *Th. elongatum* (**Additional file 2**). Additional file 2A and 2B show a shared feature that most prolamin genes in *Th. elongatum* tend to exit synteny relationship with few gliadin genes of co-originated chromosomes in selected species. The numbers of α -gliadin genes in a syntenic relationship between *T. aestivum* and *Th. elongatum* were 12 (A genome: 7, B genome: 5), which may be due to the large number of alpha-gliadin gene subfamily (**Additional file 2C**). This phenomenon suggested that these few genes were evolutionarily conserved, and gene subfamily expansion occurred after genome differentiation. Without the limitation of using incomplete sequences, the HMW-GS genes of *Th. elongatum* showed a collinear relationship with those on co-originated chromosomes of selected species, which indicated that these genes existed before genome differentiation (**Additional file 2D**).

Evolutionary analysis of each subfamily of prolamins between *Th. elongatum* and three species selected

Based on the four selected species, we have constructed four phylogenetic trees of prolamin gene subfamilies to show the internal evolutionary relationship within Triticeae and the same prolamin gene subfamilies. The outgroups were *GQ139526.1* (*Psathyrostachys huashanica*), *X13508.1* (*Hordeum vulgare*), *HQ293220.1* (*Dasypyrum villosum*), and *FJ481574.2* (*Eremopyrum triticeum*), respectively.

All 66 α -gliadin genes were obviously divided into 8 clades, supporting the notion that genes from the same genome generally group together (Fig. 5) [18]. The distribution of genes with explicit chromosome location hinted that clade 1, clade 2, and clade 3 represented the A, B, and D genomes, respectively, and clade 6, with only 2 genes, represented the E genome, which supported the standpoint that *T. urartu* and *Ae. tauschii* contributed their genomes to common wheat. Among these clades, clade 1 and clade 2 were clustered together, representing a close relationship of the D genome of the *Aegilops* genus and the A genome of the *Triticum* genus. These results also clearly show that the E genome evolved earlier than the B genome, followed by the D genome and the A genome. Clade 4 and clade 5 were clustered in a larger branch, representing a collection of α -gliadin genes from the genomes of four species. Clade 5 – 2 contained most of the α -gliadin genes of the E genome while clade 5 – 1 consisted of parts of α -gliadin genes of the B and E genomes, indicating that the B and E genomes are, again, closely related. By comparison, clade 4 formed after the differentiation of clade 5 (A, B, D, and E genomes), therefore we inferred that it included some ancient genes of the four species. It is interesting to note that 4 genes of

the 7E chromosome formed a separate clade (clade 8) with a high bootstrap value and had a closer genetic distance with the root which indicates that these genes on chromosome 7E evolved earlier than those on chromosome 6E and formed as a result of different evolutionary trajectories in early stages. Importantly, the results consistent with the above evolutionary relationships of the A, B, D, and E genomes were also revealed by clade 2 of the HMW-GS gene subfamily that the HMW-GS genes of E and B genomes evolved first, followed by those of the D genome and the A genome (**Additional file 3C**). Also, the HMW-GS genes of the E genome were more closely related to those of the B genome of the *Triticum* genus, while those of the A genome showed a closer relationship within the *Triticum* genus with those of the D genome of the *Aegilops* genus (**Additional file 3C**). In HMW-GS gene subfamily, all 12 genes were well divided into 2 clades, representing y-type HMW-GSs (clade 1) and x-type HMW-GSs (clade 2) (**Additional file 3C**) [14]. The topological structure of y-type HMW-GSs had some differences with x-type HMW-GSs, which might be related to the sequence variation of the C-terminal region of y-type HMW-GSs (**Additional file 1: Table S6**).

The LMW-GS and γ -gliadin gene subfamilies did not exhibit similar results of genome evolution with the α -gliadin and y-type HMW-GS gene subfamilies due to the limited gene numbers, but a new classification method of LMW-GS gene subfamily has been proposed. All 28 LMW-GS genes were divided into four groups based on the type of the 21st–29th amino acids, corresponding to traditional LMW-m (clade 1 and clade 2), LMW-s (clade 3), and LMW-i (clade 4) (**Additional file 3B and Additional file 4**) [6, 32–33]. Interestingly, the distributions of the first and penultimate cysteine of LMW-GS sequence were found to be related to the classification of the LMW-m (clade 1 and clade 2) and LMW-s (**Additional file 4**). Although the penultimate cysteine of clade 1 was at the same site, clade 1 was divided into 3 subclades according to the different locations of their first cysteine residue (**Additional file 4**), and subclade 1–1 and subclade 1–2 showed a closer relationship (**Additional file 3B**). Clade 2 was defined as an intermediate type between subclade 1–3 and clade 3. The reason is that the sequence from the first to about 345th amino acids (including the first cysteine) were similar to that of clade 1–3, but the remaining amino acids sequence from 350th to about the last 483th amino acids (including the penultimate cysteine) were similar to that of clade 3 (**Additional file 4**). Whether based on the location of the third cysteine or the type of the 21st–29th amino acids, LMW-GSs of clade 4 belonged to the LMW-i type. We also found three special sequences (*Tel-LMW-glutenin-1E-2*, *-3* and *-4*) in the E genome with an extra serine in front of the stop codon (**Additional file 4**).

Distribution of CD epitopes of α -gliadins of four species

The most influential T cell epitopes in CD patients are PFPQPQLPY (DQ2.5-glia- α 1a), PYPQPQLPY (DQ2.5-glia- α 1b), PQPQLPYPQ (DQ2.5-glia- α 2), FRPQQQPYPQ (DQ2.5-glia- α 3), and QGSFQPSQQ (DQ8-glia- α 1/DQ8.5-glia- α 1), as well as the most toxic 33-mer peptide (LQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF) [34–35]. In our research, α -gliadin genes without clear chromosome coordinates were classified according to their evolutionary relationship (Fig. 5 and **Additional file 1: Table S5**). Based on previous conclusions, we also showed that genomic specificity also existed in the E genome [1]. Almost 86.7% of genes contained DQ2.5-glia- α 1a or DQ2.5-glia- α 3 in the A

genome, but only 28.6% of genes contained one type of epitopes, DQ8-glia- α 1/DQ8.5-glia- α 1 in the B genome. The epitope types of the D genome were the most abundant and all six types were included. The E genome, like the B genome, only contained one type of epitopes, DQ8-glia- α 1/DQ8.5-glia- α 1, with an existence rate of 66.7%. Another result similar to previous reports was that the 33-mer peptide was only detected in the D genome of bread wheat [1]. When it comes to the number of existence of every peptide, DQ2.5-glia- α 1b and DQ2.5-glia- α 2 were only found in the D genome and often existed in the form of multiple peptides, especially DQ2.5-glia- α 2. Conversely, other small peptides existed in the form of single peptides. These results provide a reference for the cultivation of wheat with low CD exacerbation.

Electrophoretic maps and kneading quality performance of DS1E(1A) and DS1E(1D)

Through chromosome engineering, prolamins of chromosome 1E of *Th. elongatum* were implanted into wheat by substituting chromosome 1A or 1D. In this part of the study, we aimed to detect the quality effect of chromosome 1E after replacing chromosome 1A or 1D of bread wheat. In order to ensure the correctness of the materials, we analyzed the cytological characteristics of substitution lines by fluorescence in situ hybridization (FISH). As expected, Fig. 6A and 6B show the chromosomal pattern of the chromosome DS1E(1D) and 3E(3A), which can be used to detect the quality effect of substitution lines. The DS1E(1A) we used has been verified in a previous study by Guo et al. from our lab [36]. Through SDS-PAGE and A-PAGE, we show that 1Ex subunits and some gliadins from the E genome were expressed normally in DS1E(1A) and DS1E(1D) (Fig. 2B, 2D) [53]. The 1Ey subunit of *Th. elongatum* was speculated not to be expressed on account of the deletion of its band in DS1E(1A) and DS1E(1D). Next, kneading parameters were detected to check the kneading performance of *Th. elongatum* with prolamins for common wheat. Width at 8 min and midline peak height of DS1E(1A) were much higher than those of CS, which is consistent with the results of Guo et al. (Fig. 6C and 6D) [53]. This is an interesting phenomenon that the protein content and water absorption of DS1E(1A) were similar to those of CS, but the protein content of DS1E(1D) was 3% higher than that of CS (Fig. 6C and 6E). There was no significant difference in width at 8 min between CS and DS1E(1D), but the peak height of DS1E(1A) and DS1E(1D) increased, indicating that the introduction of 1E chromosome enhanced the kneading resistance of dough (Fig. 6C and 6E). By examining the effects of other chromosomes of the E genome on bread wheat quality, we found that DS3E(3A) improved the 8-min band width and the peak height of midline in CS, which indicated that there may be regulatory sites related to flour mixing characteristics on bread wheat chromosome 3A (Fig. 6B and 6F).

Discussion

Prolamin gene families play important roles in flour viscoelasticity, nutritional quality, and CD epitope content. For research on these gene subfamily, the method of known genome-wide sequence searches are regarded as the most comprehensive method. The genome-wide identification of prolamin gene families have been widely carried out in *T. aestivum*, *T. urartu*, and *Ae. tauschii*, but lack systematic analysis [14–16, 37]. Currently, knowledge is still limited about prolamin gene families in *Th. elongatum*.

In this study, we identified 19 α -gliadins, 9 γ -gliadins, 19 ω -gliadins, 2 HMW-GSs, and 5 LMW-GSs from the *Th. elongatum* genome. Genes from the genomes of the above related species were summarized. Although there are limitations in genome assembly, these results at least provide a reference for the study of prolamin genes in a single germplasm [37].

Transcripts of grain development at different stages indicated the complexity of differential expression of prolamin genes. According to previous reports of hexaploid wheat, the expression of LMW-GS genes reached the peak at 10 days after anthesis, and then decreased with the maturity of seeds [28]. Another analysis also shows a similar result, in which LMW-GS genes began to express on the 5th day after flowering, reached the highest level at the 14th day, and then decreased gradually with seed ripening [17]. In *Th. elongatum*, the expression of three LMW-GS genes showed a consistent trend with the above studies (Fig. 2A). Compared with α/β -gliadin genes, their expression has genomic differences that genes of B and D genomes belong to early-expressed genes (highest level at 10 days after flowering), similar to the expression of LMW-GS genes, while those of the A genome are late-expressed genes (highest level at 20 days after flowering) [28]. The expression trend of α -gliadin genes is synchronous with that of LMW-GS genes in *Th. elongatum*, therefore the α -gliadin genes of E genome should be early-expressed genes (Fig. 2A). In CS, although the expression levels of all genes decreased to an exceptionally low level at 23–25 days post-anthesis (DPA), the two types of γ -gliadins showed different expression patterns; the expression of type I decreased rapidly at 10–15 DPA, while the expression of type II decreased slowly and gradually after reaching the highest level at 10 DPA [25]. In this study, the expression of γ -gliadin genes increased from the half grain to the grain stage and reached its peak later than that of α -gliadin genes and LMW-GS genes (Fig. 2A). The expression levels of different LMW-GS genes vary greatly (almost ten-fold) [17]. However, the largest difference was only found in two periods of γ -gliadin gene families in *Th. elongatum* (5–6 times) (Fig. 2A).

Previously, the collinearity of genes at different loci has been studied in genome fragments of rice, maize, sorghum, barley, and wheat [38]. Relieving limitations of DNA markers based on genetic maps, these studies and comparisons of the smaller regions will provide us with preliminary insights into the detailed composition and organization of many plant genomes [19]. Prolamins are concentrated into clusters on chromosomes, which is helpful in comparing the homologous regions of different species to elucidate their evolutionary characteristics. A comparison of a 307 kb physical contig was analyzed between the A and B genomes of durum wheat and the D genome of *Ae. tauschii*. It was realized that, although gene collinearity appears to be retained, four of six genes including the two paralogous HMW-GS genes are disrupted in the orthologous region of the A genome [39]. Another study inferred that considerable sequence changes caused rearrangements of prolamin genes in these genomic regions after the split of the two homoeologous wheat genomes [10]. In this study, the homology of the whole *Gli-1* and *Glu-3* intervals was shown between *Th. elongatum* and other selected species (Fig. 4). The order of ω -gliadin, γ -gliadin, and LMW-GS genes was maintained in the E genome of *Th. elongatum* and other genomes of common wheat (A and D subgenomes), *T. urartu* (A genome), and *Ae. tauschii* (D genome). However, an inversion occurred in the interval of *Gli-1* and *Glu-3* loci on chromosome 1B of common wheat, revealing a dynamic change in this region (Fig. 4). As reported, the homoeologous genomes of wheat are not as

well conserved as previously thought, owing largely to the differential insertion of transposable elements. In addition, a homologous ω -gliadin locus of *Th. elongatum* was detected only on chromosomes 1A and 1B of bread wheat (Fig. 4). We speculated that this locus was lost from chromosome 1D of bread wheat and *Ae. tauschii*, and chromosome 1A of *T. urartu* during the progress of evolution. These results lay the foundation for the further study of prolamin genes and the flanking regions of *Th. elongatum*.

The phylogenetic relationships of *Hystrix*, *Leymus*, and their relatives were investigated using the *Acc1* gene and obtained a result consistent with morphological and cytological studies, which indicates that the *Acc1* gene is a potentially valuable source for phylogenetic analysis in Triticeae [57]. In the past, the *Acc1* gene has also been successfully applied to the study of evolution of *Triticum/Aegilops*, as well as that of switchgrass (*Panicum virgatum* L.) [58,59]. However, the evolutionary relationship of species usually requires multiple groups of evidence. Prolamin genes are also considered as research resource for evolutionary relationship of Gramineae species. For example, the evolution of LMW-m genes indicated that there was a close relationship between the B genome and the Ss genome, which supported the view that the B genome originated from *Ae. apetala* [25]. Recently, a study indicated that the E genome of the *Elytrigia* genus was more related to the B genome of common wheat and the A genome was more related to the D genome of common wheat through the analysis of single copy genes of its genome [13]. In our study, this result is supported by phylogenetic trees of α -gliadin genes and x-type HMW-GS genes, respectively (Fig. 5 and S2C). In the α -gliadin subfamily, the genes on chromosome 7E were shown to diverge earlier than other genes (Fig. 5).

Based on amino acid sequences of N-terminal, six subgroups of LMW-GSs could be recognized, called LMW-s (SHIPGL-), LMW-m (METSHIPGL-, METSRIPGL-, METSCIPGL-), LMW-i (ISQQQ-), α -LMW (VRCPCP-), β -LMW (NMQVDP-), and ω -LMW (KELQSP-/ARQLNP-) [34–36]. The studies of α -LMW, β -LMW, and ω -LMW types are found in fewer numbers than others in general. In this study, multiple sequence alignment showed that LMW-m was no longer limited to three types, and 6 other types (METSHNPGL, METSHIPSL, METSRVPGL, METSCISGL, MDTSYIPGL, and METRCIPGL) have evolved (**Additional file 4**). Because only the M, T, and L sites of the nine amino acid residues at the beginning of the N-termini were conserved, we speculate that there will be more types of LMW-m with species diversification (**Additional file 4**). We also found that clustering of LMW-GSs was related to the position of the first and the penultimate cysteine. Therefore, an improved method of dividing LMW-m into four types was proposed (detailed analysis is shown in evolutionary analysis).

Based on the putative functional α -gliadin sequences of four species, our results showed that CD peptides were genome-specific, which is consistent with the results of a previous study (**Additional file 1: Table S6**) [1]. Importantly, α -gliadins of the E genome contain only one type CD peptide, which is beneficial to low-CD breeding. Although many germplasm resources of Triticeae have been used to test CD content based on protein sequences, they have not been applied to the cultivation of wheat varieties. DS6E(6D), which replaces the polypeptide D genome with the oligopeptide E genome and has been verified through cytological study (results not shown), will be used to detect the effect of low CD in the future.

Conclusions

Unlike the genome of *T. aestivum*, *T. urartu* and *Ae. tauschii*, prolamin genes have not been carried out in *Th. elongatum* genome artificially, and there is a lack of understanding of these genes. We completed this work, compared and characterized these genes with those we summarized in related crop genomes. We also explained the different expression pattern of prolamin gene subfamilies, located these genes on the chromosomes, found the dynamic changes of gene intervals, elaborated the evolutionary relationship of crop species. The pattern that large number prolamin genes existed but few were effectively expressed may be suitable for grain size in evolution. The E genome with low content of small CD peptides will be profitable genomic choice for the cultivation of wheat with low celiac disease. These results will promote the research of prolamin gene superfamily in wheat quality, and provide thinking for wild resources in improving wheat quality.

Methods

Identification of prolamin genes

Genome datasets and annotation files of four species (*Th. elongatum*, *T. aestivum*, *T. urartu*, and *Ae. tauschii*) were downloaded from the Genome Warehouse (<https://bigd.big.ac.cn/gwh/Assembly/965/show>), IWGSC (https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.1/), MBKBASE (<http://mbkbase.org/Tu/>), and <http://aegilops.wheat.ucdavis.edu/ATGSP/annotation/>, respectively. Prolamin genes in *T. aestivum*, *T. urartu*, and *Ae. tauschii* were obtained from annotation files by extracting gene sequences and protein sequences from coding sequence (CDS) and protein files [14–16, 43]. The common wheat prolamin gene sequences, including HMW-GS, LMW-GS, α -, β -, γ -, and ω -gliadin genes, were used as queries in blastn searches against the *Th. elongatum* genome sequence with an e-value cutoff of 1×10^{-10} , and then matched genes with full-length sequences (including stop codon) were extracted and manually annotated [16]. All CDSs and protein sequences were regenerated using the gffread program [44].

Feature analysis

In this study, putative functional genes were deduced according to whether the amino acid sequences of prolamins could produce proteins with complete structure. The pseudogene rate of the gene subfamily was calculated with the number of pseudogenes divided by the total number of genes. Two genes were deleted because the sequences were not completed. All putative functional prolamin sequences were submitted to ExPasy (https://web.expasy.org/compute_pi/) for calculation of molecular weight (MW) and theoretical isoelectric point (pI).

Transcriptome analysis

In order to analyze the expression patterns of prolamin genes in *Th. elongatum*, we sampled tissues at six stages with three biological replicates: young spike, pre-flowering, flowering, ovary expansion, half-grain,

and grain stages. Total RNA was extracted for each sample using TRIzol® Reagent (Invitrogen) according to the manufacturer's instructions. A total of 18 samples were paired-end sequenced on the Illumina HiSeq 4000 platform. The quality of all samples was checked by fastqc (v0.11.9) and controlled by fastp (v0.20.0) with default parameters [45–46]. Hisat2 (v2.2.1) was used to map clean reads to the genome of *Th. elongatum* [47]. The transcripts per kilobase of exon model per million mapped reads (TPM) value for each gene was calculated by StringTie (v2.1.2) and the heatmap was constructed by TBtools (v1.047) [48–49]. The transcriptome data used in this study can be downloaded from the NCBI (<https://www.ncbi.nlm.nih.gov/>) with accession numbers PRJNA540081 (<https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA540081>).

Protein extraction and gel electrophoresis

The method for glutenins

Single milled seed was suspended in 0.5 mL mixture extraction liquid and then incubated at 65 °C for 2 hours. The components of the extracted liquid mixture were as follows: 25 mL Tris-HCl (0.5 M, pH 6.8), 20 mL glycerol, 4 g sodium dodecyl sulfate (SDS), 30 g bromophenol blue, and 1 g dithiothreitol (DTT). The total volume of above extracted liquid mixture was then made to 100 mL with deionized water. The glutenin subunits were separated by a discontinuous sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) system. The 100 mL 10% separating gel included 40.7 mL deionized water, 33.3 mL 30% Acryl/Bis solution (29:1), 25 mL 4X separation gel buffer, 1 mL 10% (m/v) ammonium persulfate solution and 800 µL TEMED. The 15% separating gel were calculated according to the 10% separating gel. The 80 mL 5% stacking gel included 45.6 ml deionized water, 13.6 mL 30% Acryl/Bis solution (29:1), 20 mL concentrated rubber buffer, 800 µL 10% (m/v) ammonium persulfate solution and 65 µL TEMED. Gels were run at a constant current (12 mA) for 20 hours.

The method for gliadins

Single milled seed was suspended in 0.3 mL 70% ethanol and then incubated at 37 °C for 2 hours. Acid gel composition was as follows: 0.5 g ascorbic acid, 0.01 g FeSO₄, 50 g acrylamide, and 2.5 g Bis-Tris. The total volume of each sample was then made to 500 mL with deionized water. The separating gel included 40 mL acid gel liquid, 0.3 mL formic acid, and 30 µL 0.6% H₂O₂. The stacking gel included 10 mL acid gel liquid and 0.1 µL 0.6% H₂O₂. Gels were run at a constant voltage in two stages (150 V for 30 min, then 350 V for 4 hours).

Coomassie brilliant blue dye liquor included 1 g coomassie brilliant blue R-250, 100 mL glacial acetic acid, 250 mL isopropanol and 650 mL deionized water. The decolorizing solution included 50 mL anhydrous ethanol, 100 mL glacial acetic acid and 850 mL deionized water. The gels were stained with a Coomassie brilliant blue dye liquor for 4 hours and destained overnight.

Distribution and duplication of prolamin genes

The distribution of prolamin genes on chromosomes was determined by the MapChart software [50]. Three criteria were used to determine whether they were tandem repeat genes [51–52]: (a) The distance between two genes on the same chromosomal fragment was less than 100 kb; (b) the shorter aligned sequence covered > 70% of the longer sequence; (c) the similarity of aligned sequences was > 70%. If AB and BC are two pairs of tandem repeat genes (A, B, and C are three adjacent genes) but AC does not meet the tandem replication criteria, A and C are considered to be tandem repeat genes. The microsynteny analysis was conducted using the JCVI software with appropriate parameters (cscore = 0.8) (<https://github.com/tanghaibao/jcvi/wiki>).

Phylogenetic analysis and classification of each prolamin subfamily

Four evolutionary trees were constructed using CDSs of the conservative domains derived from α -gliadin, γ -gliadin, LMW-GS, and HMW-GS genes (**Additional file 1: Table S6**). Multiple sequence alignments were first performed using the MUSCLE program with default parameters [53]. Phylogenetic trees were constructed using the MEGA X software with the maximum likelihood method [54]. Bootstrap was set as 1000 replicates. Based on model tests, the best models of four phylogenetic trees were set to be TN93 + G, K2 + G, K2 + I, and T92 + I, respectively [55].

CD content detection

The key peptides involved in the pathogenesis of CD are DQ2.5-glia- α 1a (PFPQPQLPY), DQ2.5-glia- α 1b (PYPQPQLPY), DQ2.5-glia- α 2 (PQPQLPYPQ), DQ2.5-glia- α 3 (FRPQQPYPQ), and a 33-mer peptide (LQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF). These small peptide sequences were retrieved from the deduced functional protein sequences, and the contents of these small peptides of single prolamin subfamily in the genomes of the four species were determined.

FISH

FISH was performed with the probes oligo-psc119.2-1 from *Secale cereal* and oligo-pta535-1 from *T. aestivum*. Hybridization solution was prepared according to the number of samples prepared. The two probes were mixed at a ratio of 1:1 before hybridization. The specific methods used were performed according to that of Han et al. [56].

Dough rheological properties test

Mature grains were milled into flour using a mill for further testing. In this study, the protein and water contents of flour were determined by a DA7200 multi-function near infrared analyzer. By referring to the formula of the "AACC54-40A" method, the main mixing parameters, such as the mixing time, middle peak height, middle peak time, middle peak at 8 min, and width at 8 min, were determined with a 10 g mixograph.

Abbreviations

HMW-GS: High-molecular-weight glutenin subunit

LMW-GS: Low-molecular-weight glutenin subunit

DS: Disomic substitution

CD: Celiac disease

CS: Chinese Spring

FISH: Fluorescence in situ hybridization

bp: Base pair

kb: Kilobase

pI: Isoelectric point

MW: Molecular Weight

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

A-PAGE: Acid polyacrylamide gel electrophoresis

Declarations

Acknowledgements

I thank Mingzhu Li for the help with FISH experiment and Xuye Du for the guidance of SDS-PAGE and A-PAGE experiments.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 31901492) (data collection), the Shandong Provincial Foundation (Grant No. 2019LZGC016) (Manuscript polishing).

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files. The datasets used during the current study are available from NCBI (<https://www.ncbi.nlm.nih.gov/>) with accession numbers PRJNA540081 (<https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA540081>).

Author Contribution

YG arranged and analyzed the data, performed the experiments, and wrote the manuscript. SSX managed and provided materials. SLS, HWW, XM and LRK provided technical support in data analysis, critical reading, and suggestions regarding the manuscript. SLS designed and supervised this study. All of the authors discussed the results and commented on the manuscript.

Ethics approval and consent to participate

Th. elongatum, commonly known as wild wheat grass, is widely distributed in Central Asia, Siberia, Mongolia, China and other countries. *T. urartu*, *Ae. tauschii* and *T. aestivum* are common breeding materials. None of the above plant materials are listed in the appendices I, II and III of the Convention on the Trade in Endangered Species of Wild Fauna and Flora, that has been valid from 4 October 2017 (<https://cites.org/eng/app/appendices.php>). The plant samples used in this study were collected from State Key Laboratory of Crop Biology, College of Agronomy, Shandong Agricultural University, Taian, China. Collection of plant materials complied with the institutional, national and international guidelines. No specific permits were required.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

References

1. van Herpen TV, Goryunova S, van der Schoot J, Mitreva M, Salentij E, Vorst O, Schenk M, van Veelen PA, Koning F, van Soest LJM, et al. alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. *BMC Genomics*. 2006;7: 1.
2. Shewry PR, Halford NG, Lafiandra D. Genetics of wheat gluten proteins. *Adv Genet*. 2003;49:111-184.
3. Wieser H. Chemistry of gluten proteins. *Food Microbiol*. 2007;24:115-119.
4. Payne PI, Holt LM, Lawrence GJ, Law CN. The genetics of gliadin and glutenin, the major storage proteins of the wheat endosperm. *Plant Foods Hum Nutr*. 1982;31:229-241.
5. Peter I, Payne Kathryn G, Corfield. Subunit composition of wheat glutenin proteins, isolated by gel filtration in a dissociating medium. *Planta*. 1979;145:83-88.
6. Sissons MJ, Bekes F, Skerritt JH. Isolation and functionality testing of Low molelght glutenin subunits. *Cereal Chem*. 1998;75:380-382.

7. Shewry PR, Halford NG, Belton PS, Tatham AS. The structure and properties of gluten: an elastic protein from wheat grain. *Phil Trans R Soc Lond B-Biol Sci.* 2002;357:133-142.
8. Anderson OD, Hsia CC, Torres V. The wheat γ -gliadin genes: characterization of ten new sequences and further understanding of γ -gliadin gene family structure. *Theor Appl Genet.* 2001;103:323-330.
9. Payne PI. Genetics of wheat storage proteins and the effect of allelic variation on Bread-Making quality. *Ann Rev Plant Physiol.* 1987;38:141-53.
10. Gao SC, Gu QY, Wu JJ, Devin CD, Huo NX. Rapid evolution and complex structural organization in genomic regions harboring multiple prolamin genes in the polyploid wheat genome. *Plant Mol Biol* 2007;65:189-203.
11. Anderson OD, Litts JC, Gautier MF, Greene FC. Nucleic acid sequence and chromosome assignment of a wheat storage protein gene. *Nucleic Acids Res.* 1984;12:8129-8144.
12. Petersen G, Seberg O, Yde M, Berthelsen K. Phylogenetic relationships of Triticum and Aegilops and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Mol Phylogenet Evol.* 2006;39:70-82.
13. Wang H, Sun S, Ge W, Zhao L, Hou B, Wang K, Lyu Z, Chen L, Xu S, Guo J, et al. Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. *Science.* 2020;368:eaba5435.
14. The International Wheat Genome Sequencing Consortium (IWGSC), Rudi A, Kellye E, Nils S, Catherine , Beat K, Jane R, Curtis JP, Frédéric C, Assaf D, Jesse P, Gil R, Andrew GS, Omer B, Kobi B, Gabriel KG, Martin M, Gil B, Ambre AJ, Axel H, François B, Juan GG, Matthew H, ChuShin K, Gary M, Raj KP, Pozniak, CJ, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science.* 2018;361:661+.
15. Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature.* 2017;551:498-502.
16. Ling HQ, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, et al. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* 2018;557:424-428.
17. Gorwa MF, Croux C, [Soucaille P](#). Molecular characterization and transcriptional analysis of the putative hydrogenase gene of *Clostridium acetobutylicum* ATCC 824. *J Bacteriol.* 1996;178:2668-2668.
18. Huo N, Zhu T, Altenbach S, Dong L, Wang Y, Mohr T, Liu Z, Dvorak J, Luo MC, Gu YQ. Dynamic evolution of alpha-gliadin prolamin gene family in homeologous genomes of hexaploid wheat. *Sci Rep.* 2018;8:5181.
19. Bennetzen JL. Comparative Sequence Analysis of Plant Nuclear Genomes. Microcolinearity and Its Many Exceptions. *Plant Cell.* 2000;12:1021-9.
20. Kong XY, Gu YQ, You FM, Dubcovsky J, Anderson OD. Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat. *Plant Mol Biol.* 2004;54:55-69.

21. Xu JH, Messing J. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc Natl Acad Sci.* 2008;105:14330-14335.
22. Stamnaes J, Sollid LM: Celiac disease. Autoimmunity in response to food antigen. *Semin Immunol.* 2015;27:343-352.
23. Sharma N, Ch V, Sharma S, Kapoor P, Garg M. Pathogenesis of celiac disease and other gluten related disorders in wheat and strategies for mitigating them. *Front Nutr.* 2020; doi:10.3389/fnut.2020.00006.
24. Shewry PR, Tatham AS. Improving wheat to remove coeliac epitopes but retain functionality. *J Cereal Sci.* 2016;67:12-21.
25. Wang S, Shen X, Ge P, Li J, Subburaj S, Li X, Zeller FJ, Hsam SLK, Yan Y. Molecular characterization and dynamic expression patterns of two types of γ -gliadin genes from *Aegilops* and *Triticum* species. *Theor Appl Genet.* 2012;125:1371-1384.
26. Tatham AS, Shewry PR. The S-poor prolamins of wheat, barley and rye: Revisited. *J Cereal Sci.* 2012;55:79-99.
27. Anderson OD, Litts JC, Greene FC. The α -gliadin gene family. I. Characterization of ten new wheat α -gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and Southern analysis of the gene family. *Theor Appl Genet.* 1997;95:50-58.
28. Liu CY, Shepherd KW. Inheritance of B subunits of glutenin and omega-and gamma-gliadins in tetraploid wheats. *Theor Appl Genet.* 1995;90:1149-1157.
29. Masci S, D'Ovidio R, Lafiandra D, Kasarda DD. A 1B-coded low-molecular-weight glutenin subunit associated with quality in durum wheats shows strong similarity to a subunit present in some bread wheat cultivars. *Theor Appl Genet.* 2000;100:396-400.
30. Jackson EA, Holt LM, Payne PI. Characterisation of high molecular weight gliadin and low-molecular-weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localisation of their controlling genes. *Theor Appl Genet.* 1983;66:29-37.
31. Payne PI, Holt LM, Worland AJ, Law CN. Structural and genetical studies on the high-molecular-weight subunits of wheat glutenin : Part 3. Telocentric mapping of the subunit genes on the long arms of the homoeologous group 1 chromosomes. *Theor. Appl. Genet.* 1982;63:129-138.
32. Gianibelli MC, Larroque OR, Macritchie F, Wrigley CW. Biochemical, genetic and molecular Ccharacterization of wheat glutenin and its component subunits. *Cereal Chem.* 2001;78:635-646.
33. D'Ovidio R, Masci S. The low-molecular-weight glutenin subunits of wheat gluten. *J Cereal Sci.* 2004;39: 321-339.
34. Kasarda DD, Ovidio RD. Deduced amino acid sequence of an α -gliadin gene from Spelt wheat (*Spelta*) includes sequences active in celiac disease. *Cereal Chem.* 1999;76:548-551.
35. Stern M, Ciclitira PJ, van Eckert R, Feighery C, Janssen FW, Mendez E, Mothes T, Troncone R, Wieser H. Analysis and clinical effects of gluten in coeliac disease. *Eur J Gastroenterol Hepatol.* 2001;13:741-747.

36. Guo J, Lu M, Wu Z, et al. Effects of *Thinopyrum elongatum* chromosome 1Ee on wheat agronomical and quality related Traits. *Journal of Plant Genetic Resources*. 2019;20:854-860.
37. Anderson OD, Huo N, Gu YQ. The gene space in wheat: the complete γ -gliadin gene family from the wheat cultivar Chinese Spring. *Funct Integr Genomics* 2013;13:261-273.
38. Bennetzen JL, Ramakrishna W. Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. *Plant Mol* 2002;48:821-827.
39. Gu YQ, Coleman D D, Kong X, Anderson OD. Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four triticeae genomes. *Plant Physiol*. 2004;135:459-470.
40. Fan X, Zhang HQ, Sha LN, Zhang L, Yang RW, Ding CB, Zhou YH: Phylogenetic analysis among *Hystrix*, *Leymus* and its affinitive genera (Poaceae: Triticeae) based on the sequences of a gene encoding plastid acetyl-CoA carboxylase. *Plant Sci*. 2007;172:701-707.
41. Huang SX, Su XJ, Haselkorn R, Gornicki P: Evolution of switchgrass (*Panicum virgatum* L.) based on sequences of the nuclear gene encoding plastid acetyl-CoA carboxylase. *Plant Sci*. 2003;64:43-49.
42. Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc. Natl Acad* 2002;99:8133-8138.
43. Juhasz A, Belova T, Florides CG, Maulis C, Fischer I, Gell G, Birinyi Z, Ong J, Keeble GG, Maharajan A, et al. Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Sci Adv*. 2018;4:eaar8602.
44. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. *F1000Research*. 2020;9:ISCB Comm J-304.
45. Chen S, Zhou Y, Chen Y, Jia G. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884-i890.
46. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000 Research*. 2018;7:1338.
47. Kim D, Langmead B, Salzberg SL: HISAT: A fast spliced aligner with low memory requirements. *Nat Methods* 2015;12:357–360.
48. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*. 2019; doi:10.1186/s13059-019-1910-1.
49. Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R: TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant*. 2020;13:1194-1202.
50. Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol Biol Evol*. 2002;3:256-262.
51. Yang SH, Zhang XH, Yue JX, Tian DC, Chen JC. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Genet Genomics* 2008;280:187-198.

52. Sun H, Fan HJ, Ling HQ. Genome-wide identification and characterization of the bHLH gene family in tomato. *BMC Genomics*. 2015;16:9.
53. Lozano R, Hamblin MT, Prochnik S, Jannink JL. Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genomics*. 2015;16:360.
54. Sudhir K, Glen S, Michael L, Christina K, Koichiro T. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547-1549.
55. Thomas RH. Molecular Evolution and Phylogenetics. *Heredity*. 2000;86:385-385.
56. Han F, Liu B, Fedak G, Liu Z. Genomic constitution and variation in five partial amphiploids of wheat–*Thinopyrum intermedium* as revealed by GISH, multicolor GISH and seed storage protein analysis. *Theor Appl Genet*. 2004;109:1070-1076.

Figures

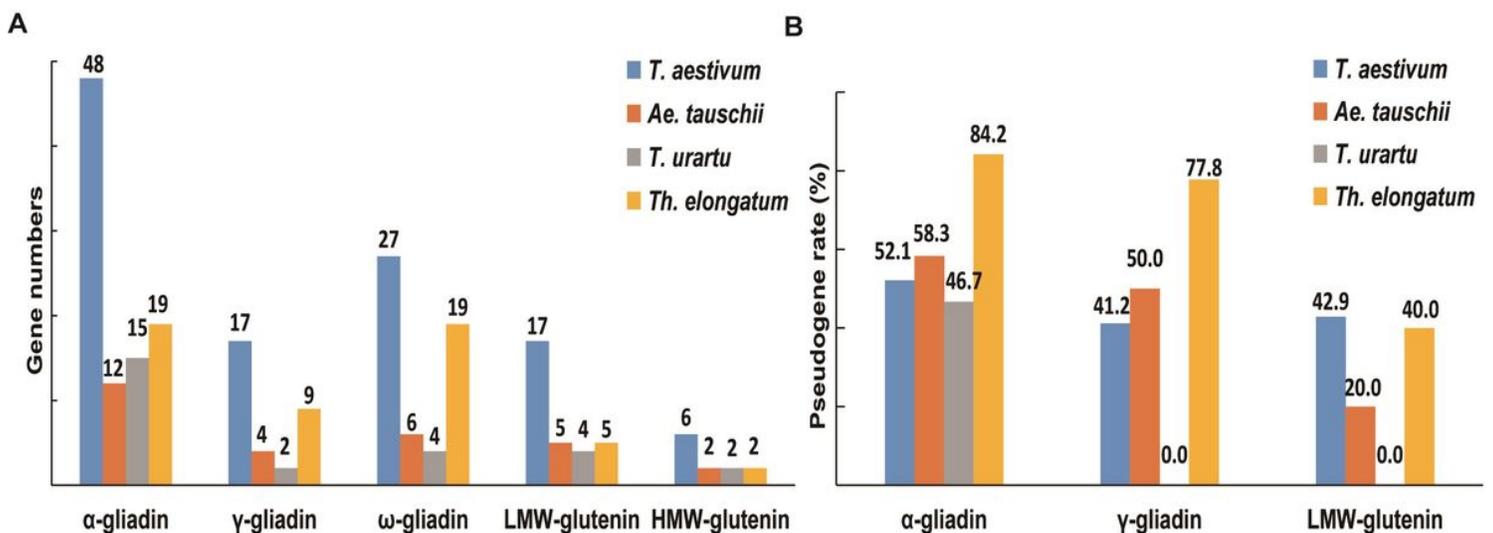


Figure 1

Characteristic comparisons of prolamins genes. (A) The number of each prolamins gene subfamily among *Th. elongatum*, *T. aestivum*, *T. urartu* and *Ae. tauschii*. (B) Pseudogene rates of α -gliadin, γ -gliadin and LMW-GS gene families among four selected species.

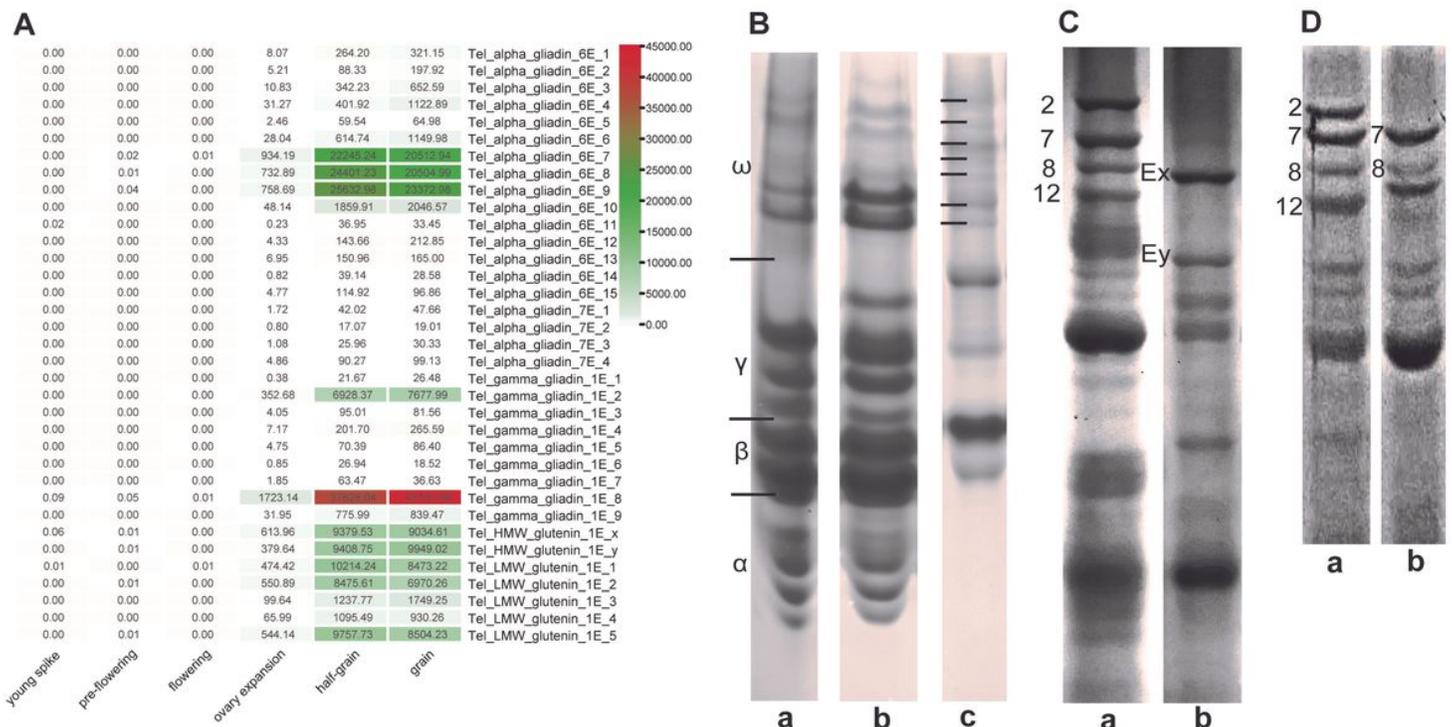


Figure 2

Expression profiles of prolamin genes. (A) Expression heatmap in different stages of α -gliadin, γ -gliadin, LMW-GS, and HMW-GS genes of *Th. elongatum*. The marked TPM values are from the average of three samples at each stage in the figure and the heatmap is constructed by TBtools. (B) A-PAGE electrophoretic patterns of gliadins of *Th. elongatum* and DS1E(1D). a-CS, b-DS1E(1D), c-*Th. elongatum*. (C) 10% SDS-PAGE electrophoretic patterns of glutenins from *Th. elongatum*. a-CS, b-*Th. elongatum*. (D) 15% SDS-PAGE electrophoretic patterns of glutenin subunits from DS1E(1D). a-CS, b-DS1E(1D). The lanes in B, C and D come from their respective gels. B, C and D gels were cropped parallel to their respective proportions to show clear and concise.

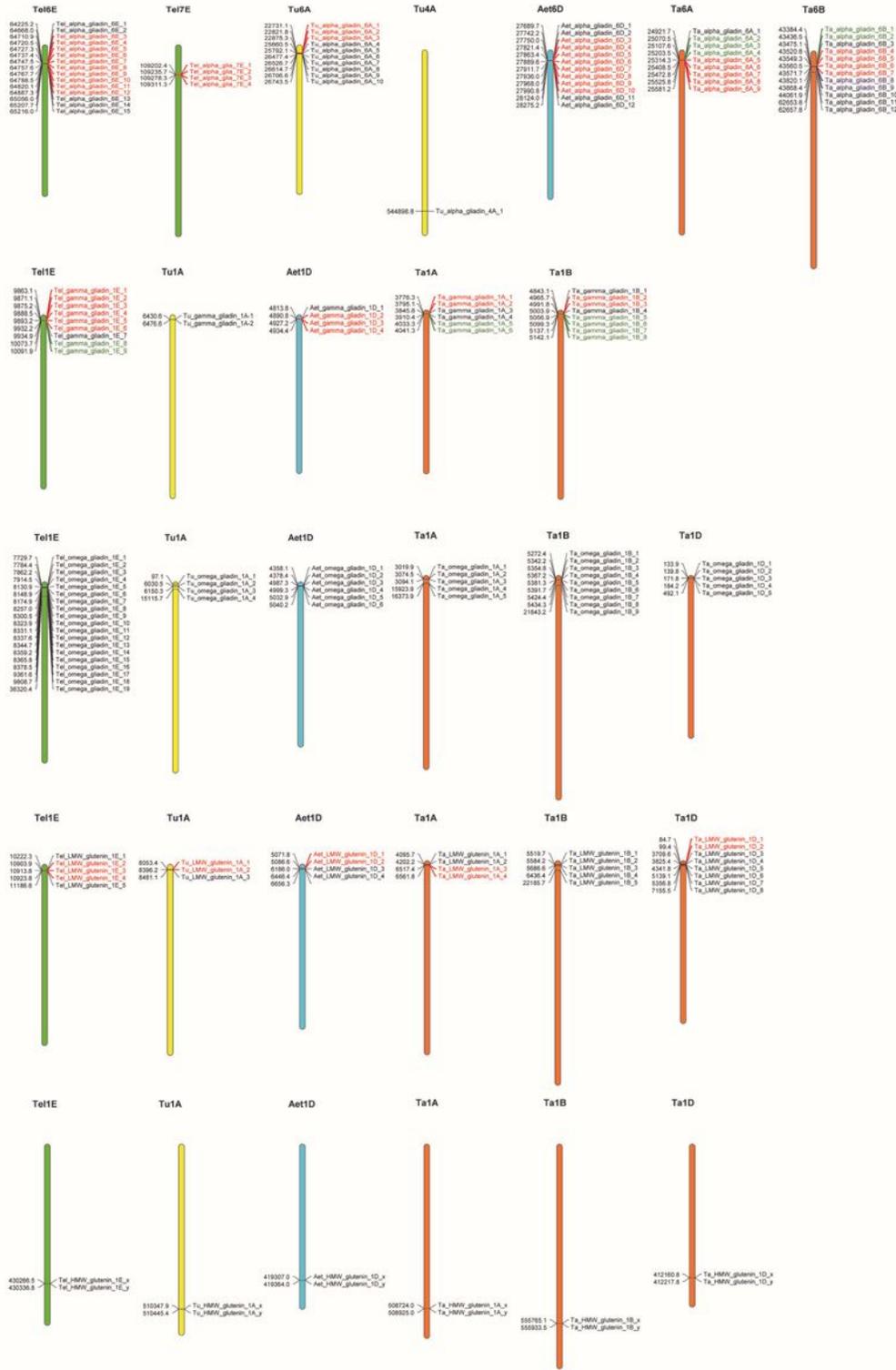


Figure 3

Chromosome location and gene duplication of prolamin genes. Tandem duplicated genes were marked by red or green.

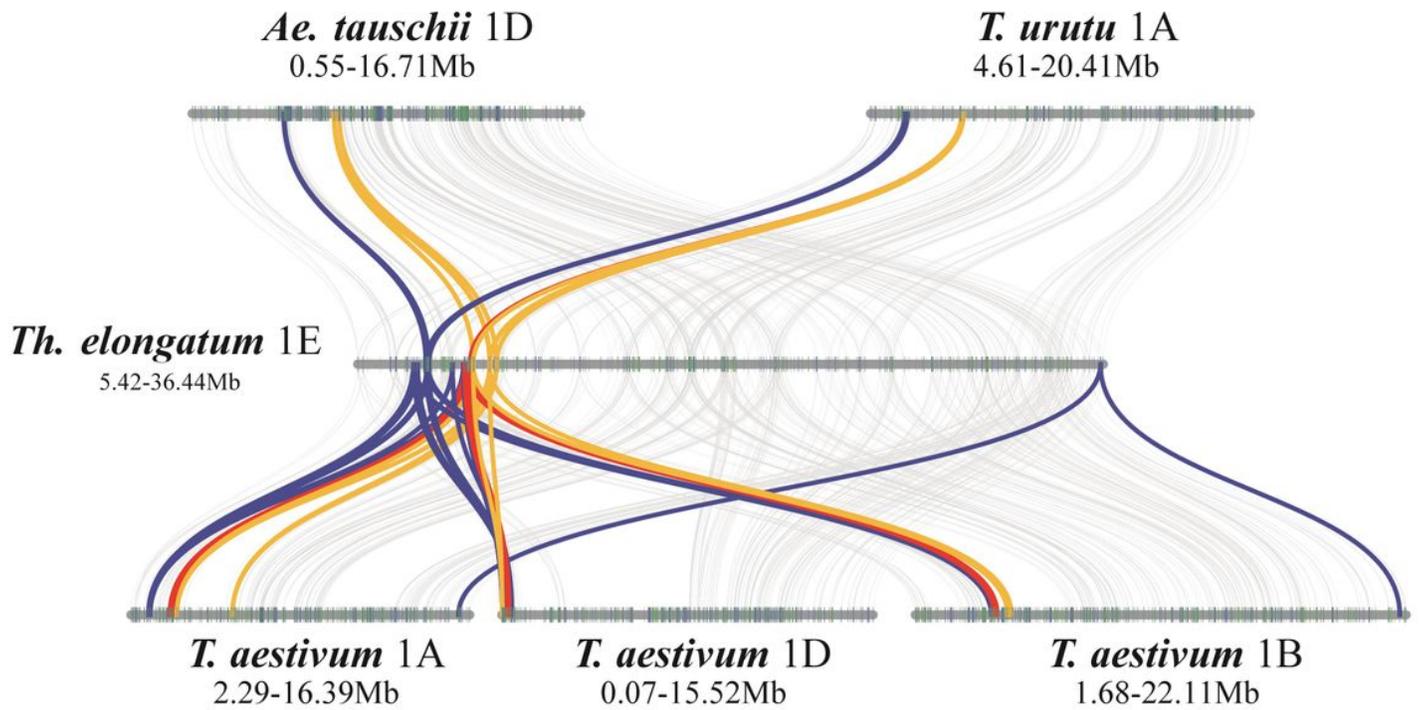


Figure 4

Microsynteny analysis between the E genome of *Th. elongatum* and other studied genomes. The relationship of these families is indicated by blue (ω -gliadin subfamily), red (γ -gliadin subfamily), and orange lines (LMW-GS subfamily), respectively.

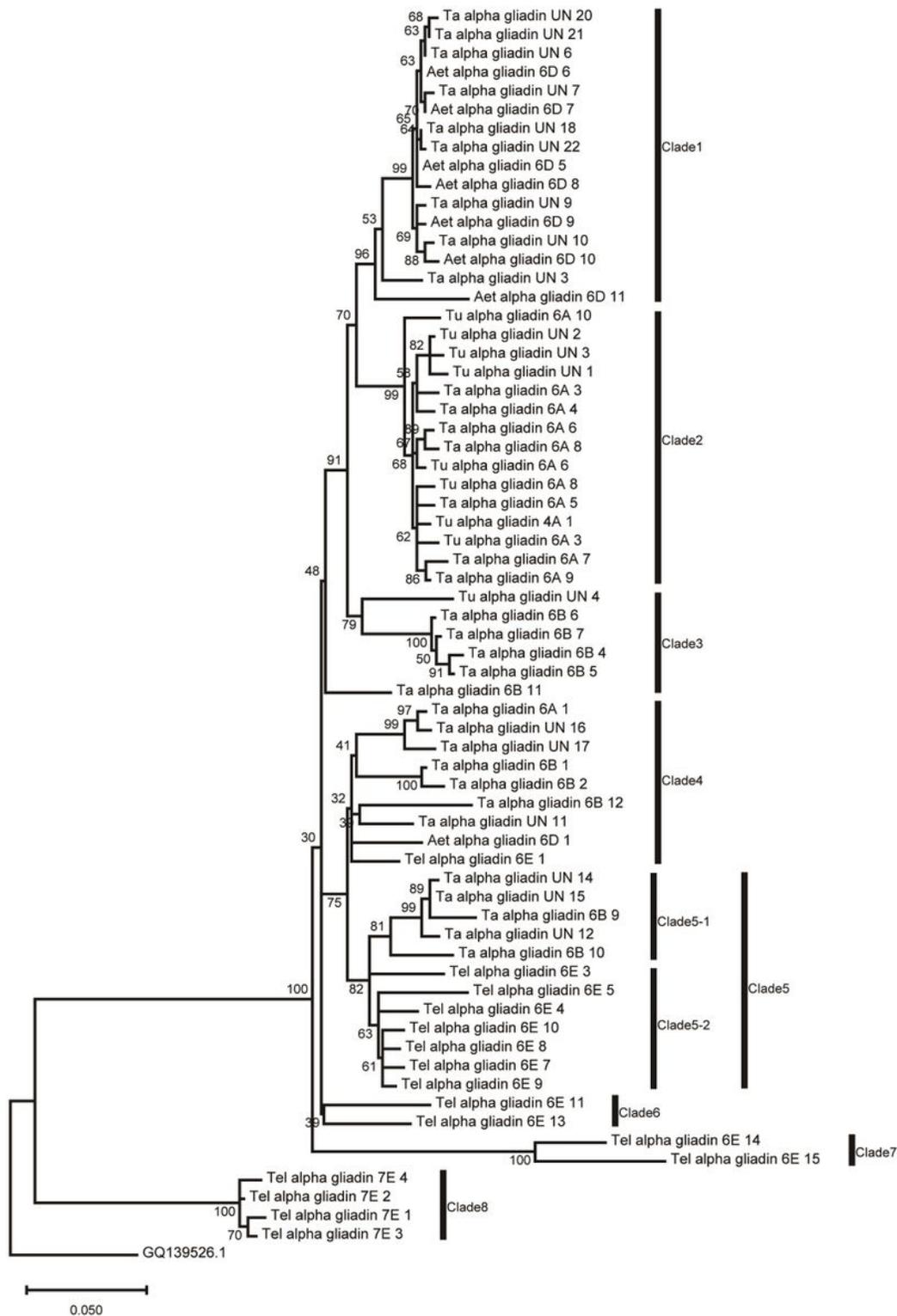


Figure 5

Phylogenetic tree of α -gliadins of *Th. elongatum*, bread wheat, *T. urartu*, and *Ae. tauchii*. The phylogenetic tree was constructed by MEGA X with the maximum likelihood (ML) method and 1000 bootstrap replications. GQ139528.1 was set as the outgroup. Different clades are marked with vertical bars and the genes of different species can be judged by their names.

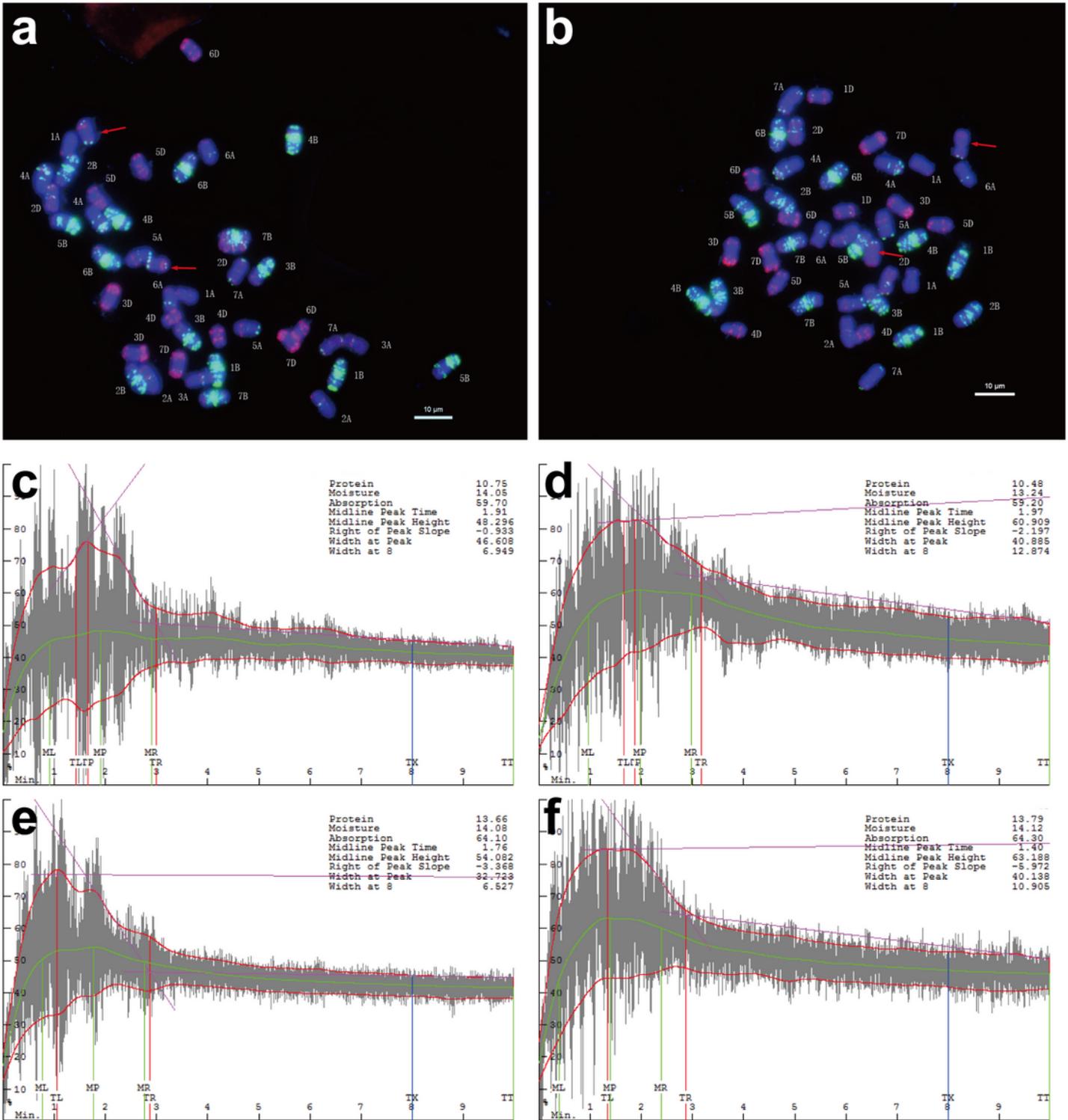


Figure 6

Cytological identification and rheological properties. (A-B) The FISH pattern of DS1E(1D) and DS3E(3A), respectively. Chromosome 1E and 3E was indicated with arrows respectively. (C-F) Rheological properties of CS, DS1E (1A), DS1E (1D) and DS3E (3A) respectively.

Supplementary Files

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

- [Additionalfile1.xls](#)
- [Additionalfile2.tif](#)
- [Additionalfile3.tif](#)
- [Additionalfile4.tif](#)