



Communication

Long-Read cDNA Sequencing Revealed Novel Expressed Genes and Dynamic Transcriptome Landscape of Triticale (x *Triticosecale* Wittmack) Seed at Different Developing Stages

Ekaterina Polkhovskaya ¹, Anna Bolotina ^{1,2}, Pavel Merkulov ^{1,2}, Maxim Dudnikov ^{1,2}, Alexander Soloviev ^{1,3} and Ilya Kirov ^{1,2,*}

¹ All-Russia Research Institute of Agricultural Biotechnology, Moscow 127550, Russia
² Moscow Institute of Physics and Technology, Dolgoprudny 141701, Russia
³ N.V. Tsitsin Main Botanical Garden of the Russian Academy of Sciences, Botanicheskaya Str. 4, 127276 Moscow, Russia
* Correspondence: kirovez@gmail.com

Abstract: Developing seed is a unique stage of plant development with highly dynamic changes in transcriptome. Here, we aimed to detect the novel previously unannotated, genes of the triticale (x *Triticosecale* Wittmack, AABBRR genome constitution) genome that are expressed during different stages and at different parts of the developing seed. For this, we carried out the Oxford Nanopore sequencing of cDNA obtained for middle (15 days post-anthesis, dpa) and late (20 dpa) stages of seed development. The obtained data together with our previous direct RNA sequencing of early stage (10 dpa) of seed development revealed 39,914 expressed genes including 7128 (17.6%) genes that were not previously annotated in A, B, and R genomes. The bioinformatic analysis showed that the identified genes belonged to long non-coding RNAs (lncRNAs), protein-coding RNAs, and TE-derived RNAs. The gene set analysis revealed the transcriptome dynamics during seed development with distinct patterns of over-represented gene functions in early and middle/late stages. We performed analysis of the lncRNA genes polymorphism and showed that the genes of some of the tested lncRNAs are indeed polymorphic in the triticale collection. Altogether, our results provide information on thousands of novel loci expressed during seed development that can be used as new targets for GWAS analysis, the marker-assisted breeding of triticale, and functional elucidation.

Keywords: Nanopore sequencing; transcriptome; triticale; seed development; annotation



Citation: Polkhovskaya, E.; Bolotina, A.; Merkulov, P.; Dudnikov, M.; Soloviev, A.; Kirov, I. Long-Read cDNA Sequencing Revealed Novel Expressed Genes and Dynamic Transcriptome Landscape of Triticale (x *Triticosecale* Wittmack) Seed at Different Developing Stages.

Agronomy **2023**, *13*, 292. <https://doi.org/10.3390/agronomy13020292>

Academic Editors: Fang Bai and Kevin Begcy

Received: 9 December 2022

Revised: 13 January 2023

Accepted: 15 January 2023

Published: 18 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Seed development in grasses is a highly dynamic biological process that determines the final yield quantity and quality. Wheat seed development can be divided into three consecutive phases: cellularization, effective grain-filling, and maturation [1,2]. During these phases the transcriptome, proteome and metabolome programs change quickly [1–8]. This makes the identification of a full spectrum of genes expressed at distinct stages challenging. The spatiotemporal transcriptomic changes during seed development have been documented for several plant species, including *Arabidopsis* [9–11], maize [12,13], and wheat [14,15]. Despite such extensive investigation, there are a number of expressed genes that have not been detected as expressed [16]. Importantly, some of these genes can be key determinants of agronomically important traits. For example, the *wbm* gene that has been identified by the deep transcriptome sequencing of developing wheat seeds is important for endosperm maturation [17]. Some of its alleles have been found in wheat and triticale collections and were associated with improved bread making quality [17–20]. This example demonstrates that a search of previously unannotated genes involved in seed development may result in new genes with valuable functions for humans.

A grass seed contains two major tissues, germ and endosperm, which play distinct physiological roles. These two seeds parts also possess substantial differences in proteome and transcriptome levels during seed development as well as seed germination [4–7]. Proteomic analysis revealed that proteins differentially expressed in germ are enriched by stress-related functions [5,6]. In contrast, proteins differentially expressed in endosperm are mostly related to carbohydrate metabolism, starch and storage protein synthesis [5,6]. In addition to the listed proteomics and transcriptomic differences, genomic DNA of endosperm and germ parts have significantly different epigenetic landscapes. The DNA methylation reduction in endosperm part can be partially explained by specific action of miRNAs expressed during grain filling and targeted some DNA methyltransferases [21,22].

In addition to miRNAs, long non-coding RNAs (lncRNAs) were identified to be expressed during seed development. However, there are relatively few studies on lncRNAs that are associated with wheat yield, especially in the population transcriptome rather than at the individual level. More than two thousand differentially expressed lncRNAs were identified in six spike tissues containing inflorescence meristem, flower meristem, spikelet meristem, scale primordium, stamen primordium, and pistil, demonstrating regulatory differences in wheat spike development [23]. Of the identified lncRNAs, 170 were associated with spike-related traits [24]. Another example of an lncRNA with important implications for yield is lncRNA expressed from Flowering Locus C (FLC) [25]. lncRNAs were also shown to participate in grain quality turning [26]. In this study, the authors identified lncRNAs involved in the regulation of starch lysis resistance in the developing caryopsis (14 and 30 dpa) [26]. The analysis revealed 443 lncRNAs with a characteristic length of more than 200 bp and with little or no coding potential. Further analysis of expressed lncRNAs between 14 dpa and 30 dpa showed only six lncRNAs that showed contrasting results at 14 DPA and ten lncRNAs at 30 DPA, which have a role in highly resistant starch content.

Recent advances in long-read RNA sequencing have overcome many limitations of short-read sequencing for transcriptome characterization, establishing the concept of ‘one read—one transcript’ [27–30]. The application of the long-read sequencing of plant RNA (direct RNA sequencing) and cDNA significantly expanded the reference transcriptome by hundreds of even thousands of newly expressed genes and isoforms [16,27,28,30–34]. The long-read sequencing of the plant transcriptome has revealed that a small portion may consist of RNAs derived from transposable elements (TEs), particularly from LTR retrotransposons [16,35]. Each genome carries a different content of TEs and repeat elements that range from ~3% of the genome in *Utricularia gibba*, to 50% in *Arabidopsis thaliana*, 78% in *Helianthus annuus*, and up to 93% in cereals [36–40]. For a long time, it was believed that these elements did not range beyond transcriptional noise; however, a growing body of evidence shows that TEs are not just hidden genomic elements, but are also a major source of functional noncoding transcripts in eukaryotic genomes [41–43]. TE-derived lncRNAs are transcribed by RNA polymerase II (RNAPII) and have important functions in the regulation of gene expression [44,45]. Moreover, TEs can produce different transcripts encoding distinct TE proteins including GAG protein [16,35]. The function of these transcripts is poorly understood and some of them may encode proteins co-opted by the host. The direct RNA Nanopore sequencing of *Arabidopsis thaliana* discovered more than 38,500 previously unannotated transcript isoforms [30]. The identification of transcribed genes in (allo)polyploid plant genome such as wheat, triticale, and cotton are more challenging than in diploid species such as *Arabidopsis*. Due to long-read sequencing, the expressed gene annotation in (allo)polyploid species has been dramatically improved by long-read sequencing [26,31,32,34]. PacBio sequencing of the cotton transcriptome during salt stress revealed that more than 16% of the assembled transcripts were not previously annotated in the genome [31]. Using the Nanopore direct RNA sequencing of poly-A + RNA isolated from triticale seed at 10 days post-anthesis, we previously identified almost 281 previously unannotated long non-coding RNA (lncRNA) genes [16]. We also showed that this type

of lncRNA is more often located in the genomic regions with missing annotations than protein coding RNAs.

Here, we performed comprehensive transcriptome analysis using long-read cDNA and direct RNA sequencing of early (10 dpa), middle (15 dpa), and late (20 dpa) stages of triticale seed development. Our results showed that 17% (7128) of genes located expressed during seed development were in the regions of A, B, and R triticale genomes that were not annotated in wheat and rye genomes as transcriptionally active. We demonstrated that lncRNAs account for more than 10% of the transcriptome of triticale developing seed, suggesting their diverse roles in this biological process. The GO analysis of the expressed genes revealed transcriptome dynamics during seed development with a contrasting pattern of the enrichment of functional categories between early and middle/late stages.

2. Results

2.1. Reference Transcriptome Assembly Using cDNA and Direct RNA Long-Read Sequencing

To determine the genomic regions that were not previously annotated as transcriptionally active ('unannotated genes' further in the text) and that are expressed during seed development, we isolated RNA from middle (15 days post-anthesis (dpa)), and late (20 dpa) stages of triticale seed development. Three different samples were sequenced for each stage. One sample included RNA isolated from whole seed (WS15 and WS20). The second sample corresponded to RNA isolated from the embryo-less half-seed which included the dissected aleurone layer and starchy endosperm (EN15 and EN20). The third sample included the dissected embryo with a small endosperm part (EE15 and EE20). MinION Oxford Nanopore (ONT) sequencing of these samples resulted in 80,000–2,692,301 long-reads with an average N50 1.3 Kb. To establish the reference transcriptome, we also involved the previously obtained reads [16] from the direct ONT sequencing of RNA isolated from the early stage (10 dpa) of triticale seed development. In total, 6.9 million high-quality ONT reads (Qscore > 8) were used for this study. The pipeline used in this study is illustrated in Figure 1 and described in the Section 4. The transcripts were assembled for individual samples by reference-based transcriptome assembly approach using the artificial ABDR genome sequence (*Triticum aestivum* ABD genome plus *Secale cereale* R genome). Then, we combined the transcripts from all samples and obtained a non-redundant set of transcripts. Subsequently, only transcripts with an expression level TPM > 5 in at least one sample and transcript length < 5000 bp were kept for further analysis (Figure 1). The latter condition corroborates with N50 of the ONT read length and reduced artificially fused neighbor transcripts, as observed after manual curation of the assembly in JBrowse2.

This pipeline resulted in the reference transcriptome of developing triticale seed consisting of 44,149 transcripts expressed from 39,914 genes. The maximum number of isoforms per gene was eight. To confirm that our new datasets corresponded to the middle and late seed developing stages, we checked the expression of marker genes wbm, gliadin, and glutenin. These genes were among the most highly expressed genes in the 15 and 20 dpa datasets, whereas no expression of these genes was detected for 10 dpa [10].

2.2. Missed Transcripts of A, B, and R Triticale Genomes Expressed during Seed Development

We then determined the genes located in the intergenic space according to the latest genome annotation of A, B (IWGS54 [46]) and R genomes (Lo7 [47]). Notably, we applied very strict conditions and classified genes as previously unannotated if the genes exhibited no overlaps with annotated genes. With this approach, previously undetermined isoforms of annotated genes would not be classified as new genes. Using these criteria, we identified 7128 (17.6%) genes as previously unannotated ('novel') genes in A, B, and R genomes. Of them, 2552, 3195 and 2078 genes were located on A, B, and R genomes, respectively (Figure 2A, Supplementary Figure S1 and Table S1). The R genome contains the most genes expressed during triticale seed development and the lowest number of previously unannotated genes.

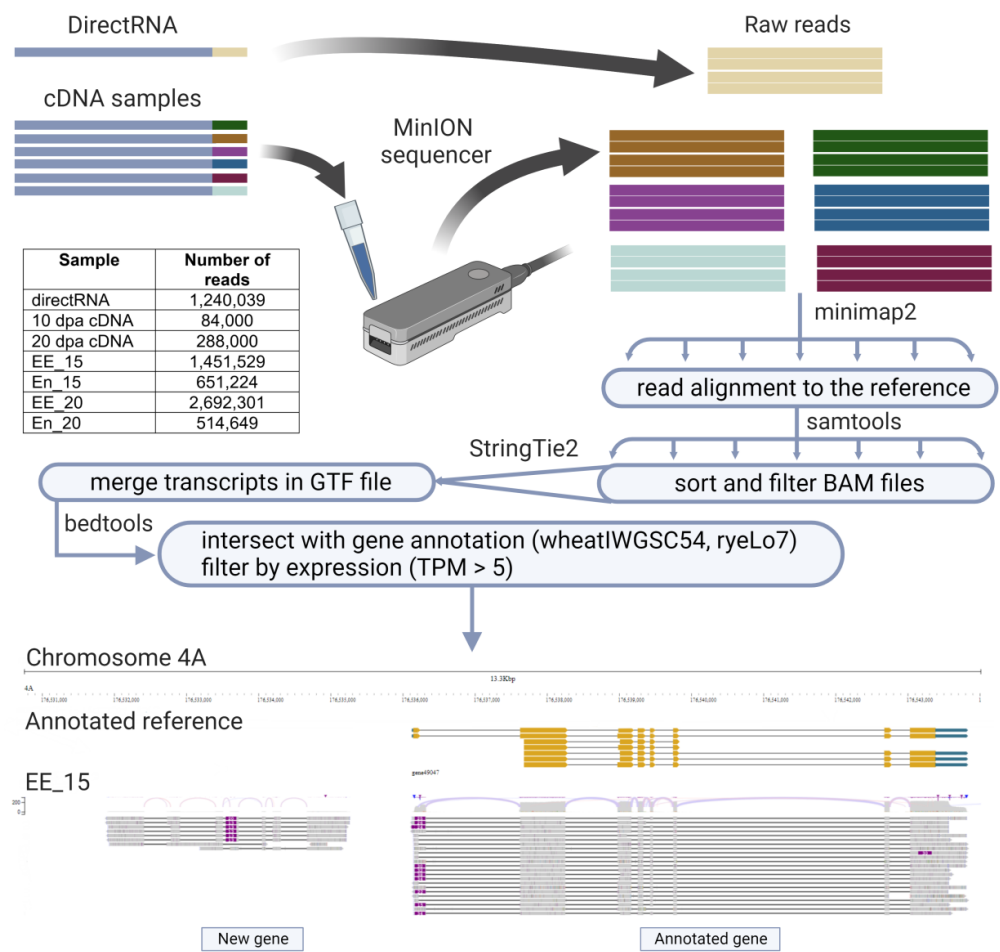


Figure 1. Experimental design and bioinformatic pipeline to find unannotated transcribed genes of the triticale. Six cDNA samples were sequenced, and the data were combined with the previous direct RNA sequencing. A snapshot of JBrowse2 of a region of Chromosome 4A with two previously unannotated genes (New gene 1 and New gene 2) is shown in the bottom.

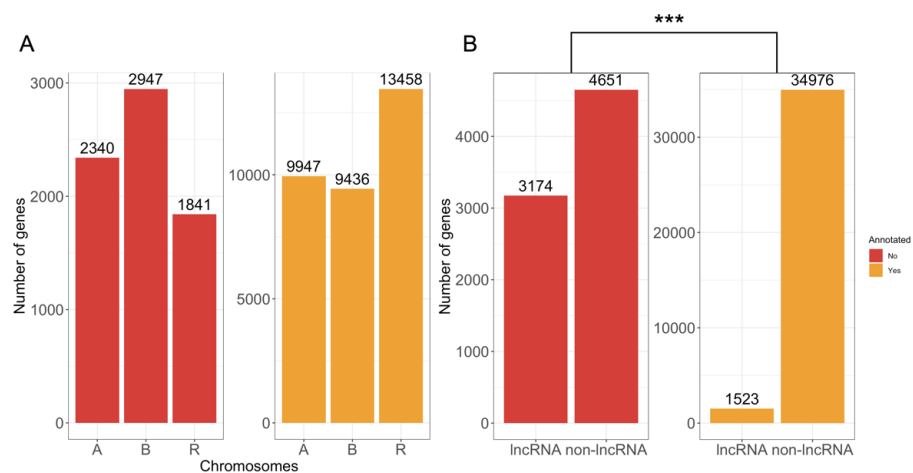


Figure 2. The characterization of previously annotated and unannotated transcripts by genome distribution and protein-coding potential. **(A)** Bar plot of the number of annotated and ‘novel’ genes located on three genomes of triticale. **(B)** The number of transcripts classified as lncRNAs expressed from annotated and novel genes. Three stars indicate the significance level (p -value $< 2.2 \times 10^{-16}$, Pearson’s chi-squared test with Yates’ continuity correction).

To estimate the impact of long non-coding RNAs (lncRNAs) to developing the seed transcriptome, we then classified the transcripts based on their protein-coding potential. lncRNAs were predicted by three tools (PLE, LncFinder, and CNCI, Supplementary Figure S2) and only the transcripts that were determined as non-coding by all three tools were classified as lncRNA. In total 4672 lncRNAs were identified. We then tested whether lncRNAs are equally presented in lncRNAs in annotated and unannotated transcript sets. We found that 3174 and 1523 lncRNAs were transcribed from annotated and unannotated genes, respectively. The portion of lncRNAs (40.5%) in unannotated transcripts was 10 times higher (Pearson's chi-squared test with Yates' continuity correction, p -value $< 2.2 \times 10^{-16}$) than in the annotated (4.1%) transcript set (Figure 2B). The results of this analysis generally dovetail well with our previous report [16], and further support the notion that some lncRNAs are underestimated in plants. Thus, we demonstrated that lncRNAs account for more than 10% of the transcriptome of triticale developing seed, suggesting their diverse roles in this biological process.

2.3. Early vs. Late Stages and GO

Seed development is an extremely dynamic process involving several transcriptomic stages. We aimed to decipher the genes that were uniquely expressed in the early, middle, and late stages of seed development. For this, we compared the expression value of all identified genes in these periods using transcript count data from 10 dpa (early stage, 1 ONT direct RNA sequencing library, 1,240,039 reads), 15 dpa (middle stage, 3 ONT cDNA sequencing libraries, 2,186,753 reads) data, and 20 dpa (late stage, 3 ONT cDNA sequencing libraries, 3,494,950 reads). We filtered the genes by their expression values in the corresponding stages, with a cutoff TPM > 2 resulting in a set that consisted of 6057, 1575, and 3274 genes expressed specifically during early, middle, and late stages, respectively (Figure 3A). Among these genes, 878, 88, and 49 genes are expressed only during early, middle and late stages, respectively. GO analysis of the specifically expressed genes revealed remarkable differences between early and late/middle stages of seed development.

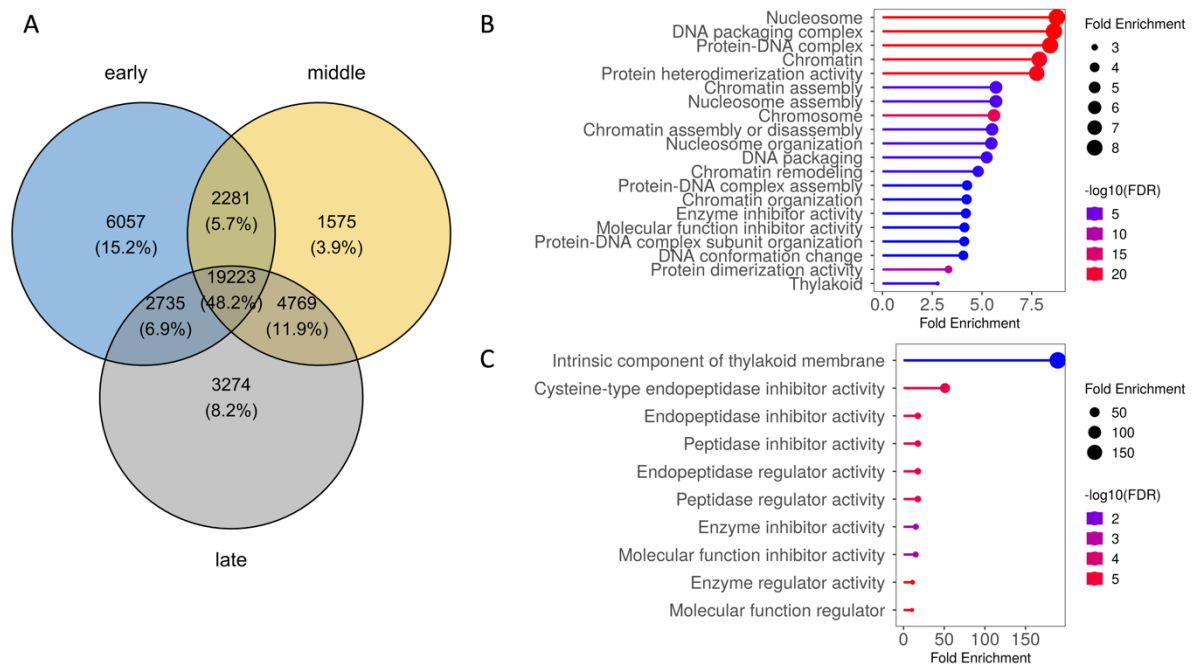


Figure 3. Characteristics of transcripts expressed during the early, middle, and late stages of seed development. (A) Venn diagram of genes expressed (TPM > 2) in specific stages of triticale seed development: early (10 dpa), middle (15 dpa), and late (20 dpa). (B) and (C) GO enrichment analysis of reference genes expressed during early and middle stages of seed development, respectively.

The genes specifically expressed during early stage are significantly (p -value < 0.0001) enriched by the genes involved in chromatin organization, packaging, and remodeling (Supplementary Table S2). These results reflect the intensive transcription; cell differentiation occurred during early stage of seed development when embryo differentiation occurred [48]. GO enrichment for the genes expressed during middle and late stages of seed development showed similar patterns: the most significant enrichment was detected for GOs linked with 'Peptidase inhibitor activity' and 'Glucose catabolic processes' (Supplementary Tables S3 and S4). Additionally, we found that genes encoding the proteins located in apoplast (late stage) and proteins, being components of thylakoid membrane (middle stage), are over-represented in middle/late gene sets (Figure 3). The GO enrichment analysis showed that transcriptomes of early and middle/late stages of triticale seed development are very different in terms of expressed genes and their functions.

Then, we aimed to understand the impact of lncRNA on the transcriptome of different stages. To check this, we compared the fraction of the lncRNAs among the transcripts of the genes expressed at specific stages of seed development. The results showed significant variation in the lncRNA fraction in the transcriptome between developmental stages. We found that lncRNAs accounted for 4.6% (879), 42.7% (756), and 31.6% (1739) transcripts expressed during early, middle, and late stages, respectively. To determine whether these differences were associated with a lower quality of transcriptome assembly for early and middle/late stages, we compared the fraction of lncRNAs in a set of assembled transcripts which have exon–intron structure that is identical to the reference transcripts (gffCompare class_code '='). We found significant differences in lncRNA fraction of the transcriptome when comparing middle vs. late stages (Pearson's chi-squared test with Yates' continuity correction, p -value = 1.6×10^{-7} , Figure 4 and Supplementary Figure S3).

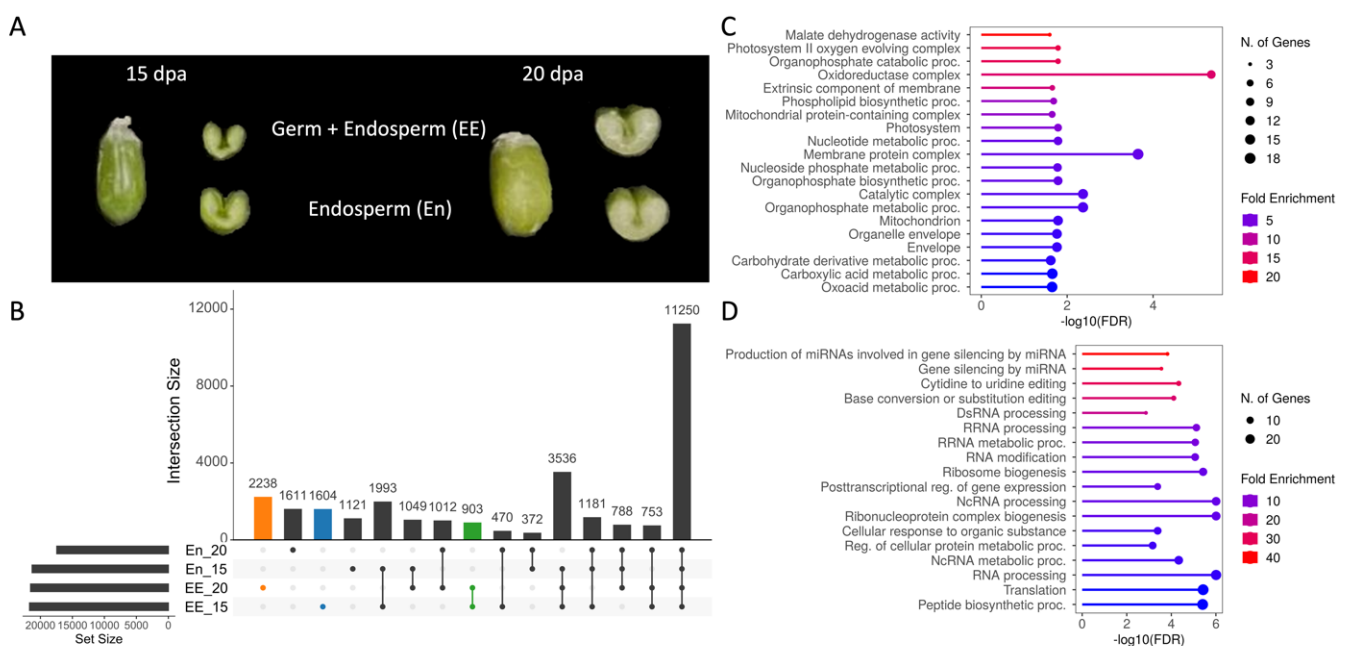


Figure 4. The identification of genes with expression in germ and endosperm parts of a seed. (A) The view of whole and dissected seeds at 15 and 20 dpa used for RNA isolation and ONT sequencing. (B) Upset plot showing intersection of genes expressed in four samples. The intersections for the genes expressed in EE_15, EE_20 and in both samples are highlighted. (C,D) GO enrichment analyses of genes expressed in EE_15 and EE_20, respectively.

Thus, the analysis of transcriptome of early, middle, and late stages of triticale seed development revealed substantial differences in the sets of protein-coding and lncRNA genes expressed during these stages.

2.4. Genes Expressed in the Germ Parts of Developing Triticale Seed

We aimed to detect the differences between gene sets expressed in endosperm and germ parts. For this, we performed the ONT sequencing of cDNA prepared from RNA isolated from two parts of developing seeds—germ + endosperm (EE) and endosperm (En). We analyzed these data for seeds at 15 (EE_15 and En_15) and 20 dpa (EE_20 and En_20). At these stages, the seeds are sufficiently large to enable the isolation of sufficient amounts of RNA (Figure 4A).

We assumed that the gene was expressed in the sample if the expression value of the gene in the sample was TPM > 2. Using this cutoff, we identified that 11,250 genes were expressed in all samples (Figure 4B). Additionally, we found 1604, 2238, and 903 genes that expressed in EE_15, EE_20 and in both samples, respectively (Figure 4). These genes did not have sufficient expression in EE samples, suggesting that they may reflect the germ transcriptome at 15 and 20 dpa developmental stages. To understand the functional biases and differences between the genes that are expressed in the specific stages, we performed GO enrichment analysis. This analysis revealed the EE_15 gene set enrichment by the genes encoding the proteins of the oxidoreductase complex and membrane protein complex and which were involved in organophosphate metabolic processes. In the EE_20 gene set, the genes encoding the proteins of the ribonucleoprotein complex, nuclear-protein-containing complex, and ubiquitin ligase complex and were involved in gene silencing, ribosome biogenesis, RNA processing, and translation are the most over-represented (Enrichment FDR < 0.003).

The over-representation of gene silencing function in EE_20 may be explained by general hypomethylation of endosperm DNA, which could lead to TE activation and TE insertions in the germ line cells. The expression of gene silencing genes may prevent this scenario. To check this, we compared the number of expressed TE-related transcripts at EE-specific gene sets and found a significant (Pearson's chi-squared test with Yates' continuity correction, p -value = $2.482e^{-8}$) over-representation of TE-RNAs in EE_15 compared with EE_20 (37 of 1604 vs. 7 of 2238) gene sets. We also detected the expression of the shGAG isoform of the previously described active "MIG" LTR retrotransposon at EE_15 and En_15 samples, but not in 20 dpa samples. This implies that MIG is expressed in endosperm tissue rather than in the germ. The analysis of TE-RNA encoding genes demonstrated that TE-RNAs are present in the germ part of developing triticale seed, and they are more prevalent at the 15 dpa stage than at 20 dpa. This is strongly correlated with GO enrichment analysis and the over-representation of genes involved in gene silencing at 20 dpa, suggesting the protection role of these genes against TEs.

We also compared the fraction of lncRNA encoding genes expressed in the germ part at 15 and 20 dpa. The analysis demonstrated that lncRNA genes are over-represented among the genes expressed in the germ at 15 dpa compared with 20 dpa (462 of 1604 vs. 376 of 2238, Pearson's chi-squared test with Yates' continuity correction, p -value < 2.2×10^{-16} , Supplementary Figure S4).

2.5. Genomic Polymorphism of lncRNA Genes in the Triticale Collection

lncRNAs represent an interesting target for marker-assisted breeding. To determine whether the identified lncRNAs had different alleles, we exploited triticale germplasm collection consisting of 23 triticale genotypes of different origins (Table 1). For the analysis, we selected lncRNA genes located on A and B genomes. BLAST search against wheat genomes [41] was performed, lncRNA genes exhibiting InDel polymorphism were selected, and primers were designed for eight lncRNA genes (Table 2). The PCR analysis of the triticale collections with the designed primers revealed variation in PCR product size for seven primer pairs and the genomic DNA of different triticale genotypes (Figure 5). The most diverged genotype was line C95.

Table 1. The origin of the triticale collection lines used in this study.

No.	Variety Name	Origin
1	Dublet	Poland
2	Legalo	Poland
3	Sandro	Switzerland
4	Grebeshok	Russia
5	Lana	Belarus
6	Pamyati Merezhko	Russia–Belarus
7	Ukro	Ukraine–Russia
8	Ulyana	Belarus
9	Khlebodar Ukrainian	Ukraine
10	Yarilo	Russia
11	131/1656	Russia
12	131/7	Russia
13	6-35-5	Russia
14	C95	Russia
15	C238	Russia
16	C245	Russia
17	C259	Russia
18	V17/50	Russia
19	L8665	Russia
20	P2-16-20	Russia
21	P13-5-2	Russia
22	P13-5-13	Russia
23	PRAG551	Russia

Table 2. Primers used for lncRNA gene polymorphism analysis and the expected length of the PCR products.

Primer ID	Primer Sequences 5' - 3'	Expected Length of PCR Product, bp
1A2902	CCATGATTGAAGATGAATTAGATCAG GATATGCCGGGGTGTACTG	234
1A59	TGATTGAAGATGAATTAGATCAGAAGT TATGAGCGACGACATCTGCC	208
3B61	TTTGTTGTTGTTCGCACGAGC ACCTTGATTATGTGGGCCCG	302
3B75	CCGTGTGCTGCACAGAAAT CCAGAAAAGAAAAGGACAGGCA	162
3B82	GACCGACATTGTGACTCCGT ACACCCAAACAGAGAGGAGA	298
6A12	GCACATGTGACGTAGGGACA TGCAACTATCATAGGGTGTGTGT	168
7B70	GTGAAGGGGGTTGGAATCAC TGTTTTTCGTAGTTTGCACCCA	200

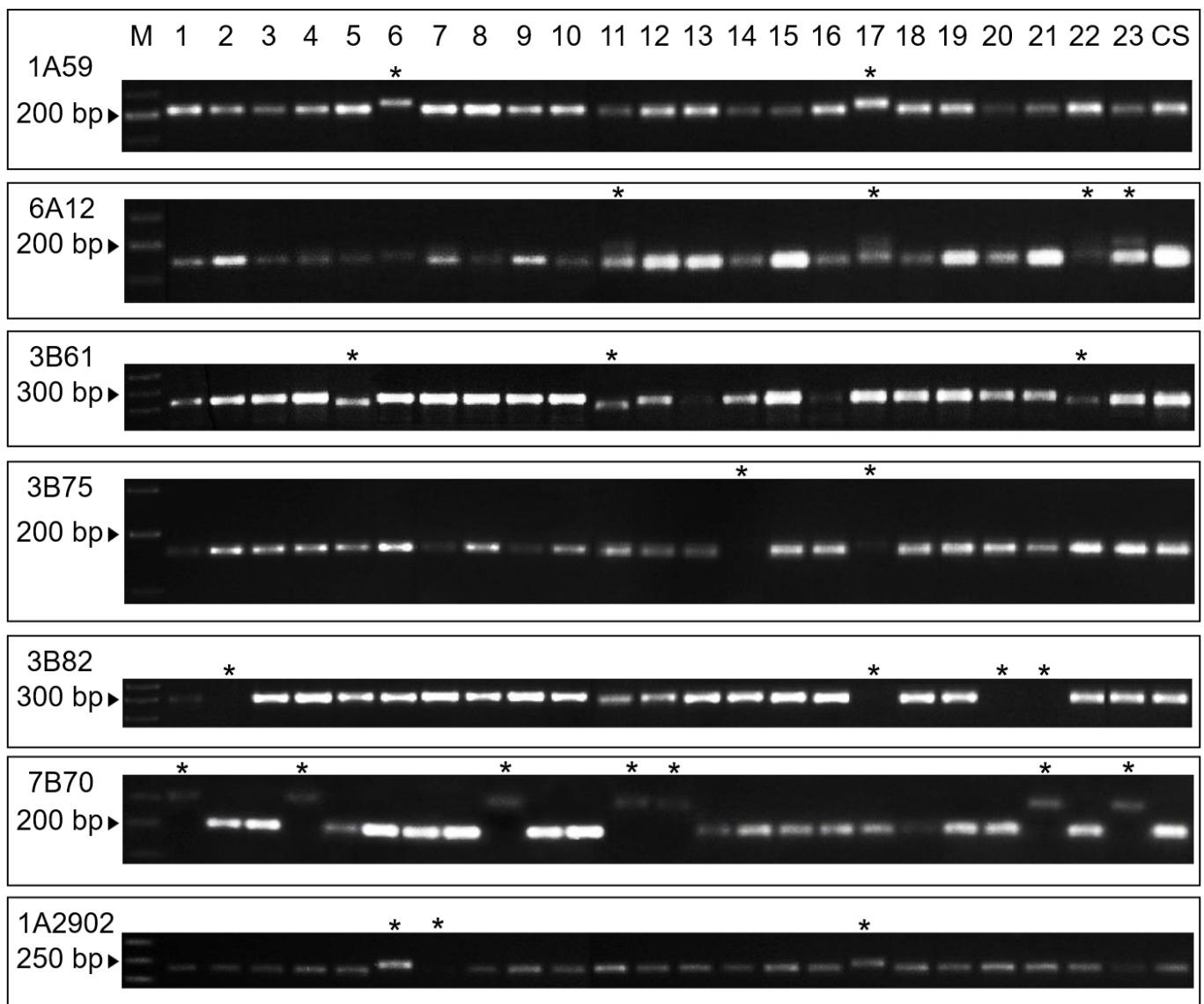


Figure 5. The analysis of lncRNA gene polymorphism in the collection of spring triticale. Individual triticale genotypes were: Dublet (1); legalo (2); Sandro (3); Scallop (4); Lana (5); In memory of Merezhko (6); Ukro (7); Ulyana (8); Khlebodar Ukrainian (9); Yarilo (10); 131/1656 (11); 131/7 (12); 6-35-5 (13); C238 (14); C245 (15); C259 (16); C95 (17); V17/50 (18); L8665 (19); P2-16-20 (20); P 13-5-2 (21); P13-5-13 (22); PRAG 551 (23); Chinese Spring (CS). DNA marker Step50 plus (Biolabmix, Novosibirsk, Russia) (M). Non-reference alleles are indicated by star (*).

In the future, it may be possible to use the genes analyzed here in collections of spring triticale to determine their role in the selection of grain crops, as well as the differences we detected in the process of caryopsis formation.

3. Discussion

In this study, we uncovered thousands of previously unannotated genes expressed at different stages of triticale seed development. We demonstrated that lncRNAs account for more than 10% of the transcriptome of triticale developing seeds, suggesting their possibly diverse roles in this biological process. Analysis of the transcriptome of early, middle, and late stages of triticale seed development revealed substantial differences in the sets of protein-coding and lncRNA genes expressed during these stages. We found that genes encoding the proteins located in the apoplast (late stage) and proteins being components of

the thylakoid membrane (middle stage) are over-represented in a set of genes expressed at middle/late stages (15–20 dpa) of seed development, whereas the genes involved in chromatin modeling were over-represented among the genes expressed at early stages (10 dpa). The comparison of genes expressed in the germ part with the genes expressed in the germ and endosperm parts demonstrated the over-representation of gene silencing function for the genes expressed in the germ at 20 dpa.

We found dramatic differences in transcriptome programs between the early and late stages of seed development, namely, during late and middle stages of seed development ('grain filling' [49]), we detected the expression of peptidase, whereas the genes involved in chromatin remodeling were expressed during the early stages. The expression of genes involved in the negative regulation of peptidase activity during the late (30 dpa) stage of seed development was also reported previously for wheat [14]. Plant cysteine proteases are important elements of the process of degradation and mobilization of storage proteins of cereals [46,47,49]. The most well-known plant protease inhibitors are phytocystatins [14]. Phytocystatins participate in the regulation of protein turnover during grain filling or the control of proteolysis during the development and/or germination and filling of cereal grains. A specific correlation was noted between the expression patterns of genes encoding barley, rice, wheat, and triticale cystatins in cluster A (Icy1, Icy2, Icy3, and Icy4) and their functional activity as protease inhibitors [46,47,49]. The analysis of cystatin genes revealed their highest expression in dry and germinating grains, and accordingly, this fact suggests that they may be specialized endogenous regulators of enzymes that are involved in the accumulation of storage proteins during their germination. The synthesis of storage proteins during the third week of seed formation is protected on the 10th–13th day after pollination by a high level of TrcC-4, and its expression in the first phase of development inhibits the expression of the main enzyme EP8 involved in the germination of triticale grains [50], because TrcC-4 can protect against the uncontrolled hydrolysis of storage proteins in germinating triticale seeds. In addition, WC5 has the function of protecting the aleurone layer or embryo from intense proteolytic activity occurring in the surrounding tissues. Thus, our transcriptomic survey corroborates previous reports and demonstrates the expression of protease inhibitors during late stages of triticale seed development.

Our enrichment analysis of gene sets expressed in embryo + endosperm and endosperm parts showed remarkable differences between 15 and 20 dpa. For example, we determined significant enrichment of EE_20 gene set by the genes involved in gene silencing, RNA processing and posttranscriptional regulation of gene expression and miRNA production. Previous reports on transcriptome and miRNAome analysis of maize, rice and wheat seed development demonstrated the key role of miRNAs in the modulation of transcriptional programs during seed development [21,51–53]. This includes the regulation of expression of transcription factor, e.g., MYB family in maize [53]. It is well known that besides gene expression regulation some miRNAs are involved in the control of transposon expression via post-transcriptional gene regulation and RNA-dependent DNA methylation [54]. The crosstalk between miRNAs, transposon silencing and DNA methylation during seed development was previously predicted for rice. Two miRNAs expressed during the filling stage of rice seed development target different transposons [51] and methyltransferases [21,22,51–54].

It is well known that endosperm is a hypomethylated tissue and DNA methylation is important for the accumulation of protein storage and endosperm biogenesis [55]. However, decreased DNA methylation promotes TE activation, creating a dangerous environment for the germ because TE activation may lead to lethal mutations and decreased seed viability. Indeed, we previously identified MIG Ty1/Copia LTR retrotransposon expressed during the early stage of seed development, although where MIG is expressed was unknown [16]. Using transcriptome profiling for the germ and endosperm, we showed that MIG is not exclusively expressed in the germ but is most probably expressed in the endosperm or simultaneously in endosperm and germ tissues. We identified the decreased representation of TE-RNAs in the germ part at the late stage of seed development compared with the

middle stage. At the same time, the set of genes expressed at this stage is enriched by the genes with transcriptional silencing functions. These results may suggest that the TE expression in the middle stage of the seed development triggers the activation of post-transcriptional gene silencing at the later stages of seed development and the accumulation of siRNAs. The latter may lead to TE methylation and protection of the germ from TE invasions. Thus, our data are consistent with the hypothesis that endosperm DNA hypomethylation has a TE-protective role for the embryo [55,56].

In summary, our study provides valuable information for genome annotation improvements, elucidation of the functional characterization of developing seed transcriptome, and searching for new targets for GWAS analysis. We believe that the obtained data of the spatiotemporal gene expression profile during seed development will be helpful for the genetic improvement of the triticale.

4. Material and Methods

4.1. Plant Material and RNA Isolation

For this study, the spring triticale line “L8665” obtained from the Department of Genetics, Russian State Agrarian University, was used.

Seeds of spring triticale line “L8665” were allowed to germinate in a dark place at room temperature on wet filter paper for 2 days. Then, the selected seedlings of the same size and strength were transferred to jars. The experiment was carried out in a greenhouse at a temperature range of 22–25 °C. Developing seeds at 10 days post-anthesis and 20 days post-anthesis were separated on the embryo and endosperm and placed into liquid nitrogen. We used three RNA samples for 15 and 20 dpa stages (whole seed, dissected endosperm and dissected germ part). RNA was isolated using the ExtractRNA kit (Evrogen, Moscow, Russia), following the manufacturer’s instructions. The RNA concentration and integrity were estimated by Nanodrop (Nanodrop Technologies, Wilmington, CA, USA) and gel electrophoresis using a 1.2% agarose gel with ethidium bromide staining. The results were detected using a Gel Doc XR + UV camera (Bio-Rad, USA). One microgram of the extracted RNA was DNase digested for 60 min at 37 °C in a final volume of 10 µL containing 1 µL of DNase I and 1 µL RDD Buffer (Qiagen, Germantown, United States). The reaction was blocked by adding 1 µL of 25 µM EDTA and heating at 75 °C for 10 min.

4.2. DNA Isolation

High-molecular-weight DNA was isolated from the triticale collection; the reference DNA was the wheat variety Chinese Spring (Table 1). Seeds were germinated in the dark at room temperature on wet filter paper disks. For extraction, 500 mg material was homogenized in liquid nitrogen. The DNA isolation process was performed according to the published protocol (<https://www.protocols.io/view/plant-dna-extraction-and-preparation-for389-ont-seque-bcvyiw7w>, accessed on 4 September 2021). The DNA concentration and integrity were estimated by Nanodrop (Nanodrop Technologies, Wilmington, CA, USA) and gel electrophoresis using a 1.2% agarose gel with ethidium bromide staining.

4.3. cDNA Synthesis

The reverse transcription required 2 µg of total RNA with double-stranded complementary DNA (ds-cDNA) and was performed using poly-A specific primers and the MINT cDNA kit (Evrogen, Moscow, Russia), according to the manufacturer’s instructions. During synthesis, the optimal number of PCR cycles (22 cycles) was adapted to reach the exponential phase of amplification. The ds-cDNA concentration and integrity were estimated by Nanodrop (Nanodrop Technologies, Wilmington, CA, USA) and by Qubit (Qubit dsDNA BR Assay Kits, ThermoFisher Scientific, Waltham, Massachusetts, USA), and checked by gel electrophoresis. Synthesized ds-cDNA was purified by 1.8× Agencourt AMPure XP Beads (Beckman Coulter, Pasadena, CA, USA), in accordance with the manufacturer’s instructions.

4.4. cDNA Nanopore Sequencing

For Nanopore sequencing, a library was prepared from ds-cDNA using the nanopore native barcoding genomic DNA SQK-NBD110-24 (Oxford Nanopore Technologies, Oxford, UK), with some modification in the process of using the NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (New England Biolabs, MA, USA). Approximately 500 ng of ds-cDNA in 24 µL was mixed with 1.75 µL NEBNext FFPE DNA Repair Buffer, 1.75 µL Ultra II End-prep reaction buffer, 1.5 µL Ultra II End-prep enzyme mix, 1 µL NEBNext FFPE DNA Repair Mix, and the mixture was incubated in a thermal cycler. The DNA samples were transferred to a new tube and an extra 30 µL H₂O was added to reach the required volume for the next preparations of the DNA library, according to the manufacturer's instructions. Each end-prepped sample was barcoded with 2.5 µL native barcodes. After purification, 4 barcoded DNA samples were pooled in one tube per 16.25 µL with the further ligation of adapters for sequencing using Quick T4 DNA ligase (New England Biolabs, MA, USA). Sequencing was carried out using MinION and flow cell SQK-LSK109. Basecalling was performed with Guppy (Version 5.0.11).

4.5. Identification of Novel Loci Expressed during Triticale Seed Development

Before data were obtained after the analysis, we combined genome sequences of *Triticum aestivum*, which were downloaded *Triticum aestivum* from the Ensembl Plants server (https://plants.ensembl.org/Triticum_aestivum/Info/Index, accessed on 1 October 2022), with *Secale cereale* genome sequences into a single fasta file. For this study, a hexaploid triticale possessing only ABR genomes was used; however, the D genome sequence was also included for bioinformatic analysis because introgression of the D genome into ABR genomes is possible during triticale breeding [49]. To decrease the artefacts, the transcripts assembled for D genome were not analyzed in this study.

Nanopore reads, obtained during sequencing, according to each barcode were aligned to merged fasta genomes of wheat and rye using minimap2 with the '-ax splice' argument. We added headers to mapped reads and converted sam to bam followed by bam file sorting using SAMtools. Bam files were filtered by mapping a quality score > 40 with the samtools view -q 40 parameter. The transcript assembly and estimation of the transcript abundance were performed using StringTie2 software. The individual gtf files were merged using the command 'stringtie merge', by which the gtf files for each barcode were obtained and merged into one. From each sorted and filtered file, abundance tables were created and also merged into one table. After converting the gtf files to the bed format using gtf2bed, the bedtools intersect were applied for the identification of unannotated genes (parameter -v) and annotated genes (-wa parameter). Thus, the unannotated genes were those without any overlaps with annotated genes. The annotated and unannotated genes assembled from our data were further filtered by expression value determined by StringTie2, and the genes with TPM > 5 were kept for further analysis. The applied procedure resulted in 36,525 annotated and 7610 unannotated genes expressed during triticale seed development.

4.6. GO Enrichment

For GO enrichment analysis, the gtf file with a merged and non-redundant transcript set was compared with the reference wheat annotation gtf file using the GffCompare tool [57]. Then, the refmap file was parsed and transcripts with class_code '=' were retained. The corresponding reference transcript IDs were used for GO enrichment analysis by ShinyGo [58] using the "All available gene sets" setting. To dissect GO enrichments that were specific for the developing stage and embryo part, we created a random set of transcripts. The GOs enriched in the random sets were not considered for further analysis.

4.7. PCR Analysis of lncRNA

PCR amplification was conducted using specific primers (Table 2), which were designed using Primer 3.0 software (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) (accessed on 1 September 2022).

PCR was performed in a 25 μ L mixture containing 2 \times PCR, 70 mM Tris-HCl; 16 mM $(\text{NH}_4)_2\text{SO}_4$, 1.5 mM MgSO_4 , 0.25 mM of each dNTP, 10 pmol of each primer, 2 units of Taq polymerase (Sintol, Russia), and 20 ng of the high-molecular-weight genomic DNA. The PCR included 35 cycles of DNA denaturation at 95 $^\circ\text{C}$ for 15 s, annealing of the primers at 60 $^\circ\text{C}$ for 15 s, and DNA synthesis at 72 $^\circ\text{C}$ for 15 s, followed by the final synthesis at 72 $^\circ\text{C}$ for 5 min. Visualization of the obtained PCR results was performed via gel electrophoresis using a 2.5% agarose gel with ethidium bromide staining.

4.8. Data Visualization and Statistical Analysis

Statistical analysis was carried out in RStudio Version 2021.09.1 (<http://www.rstudio.com/>) (accessed on 1 September 2021) with R version 4.2.0. Bar plots were drawn with ggplot2 [59]. Upset plots were drawn using the UpSetR R package [60].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy13020292/s1>, Tables S1–S4: The information about the expressed genes and enrichment analysis; Figure S1. Number of previously unannotated genes for each A, B and R chromosomes; Supplementary Figure S2. Number of lncRNAs identified by three tools; Supplementary Figure S3. Bar plot showing the fractions of lncRNAs among reference transcripts expressed during early, middle and late stages of triticale seed development; Supplementary Figure S4. Bar plot showing the fractions of lncRNAs among the transcripts expressed specifically in EE samples (Germ + Endosperm).

Author Contributions: Conceptualization, I.K.; methodology, I.K. and E.P.; software, I.K. and E.P.; validation, E.P., A.B., M.D., and P.M.; formal analysis, A.S.; resources, A.S.; writing—original draft preparation, I.K. and E.P.; writing—review and editing, I.K.; visualization, P.M. and E.P.; supervision, I.K.; project administration, A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Education and Science of Russian Federation (goszadanie No. FGUM-2022-0005).

Data Availability Statement: Nanopore data produced for this study are available in Sequence Read Archive (SRA) NCBI under Bioproject Accession PRJNA 924567.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nadaud, I.; Girousse, C.; Debiton, C.; Chambon, C.; Bouzidi, M.F.; Martre, P.; Branlard, G. Proteomic and morphological analysis of early stages of wheat grain development. *Proteomics* **2010**, *10*, 2901–2910. [[CrossRef](#)] [[PubMed](#)]
2. Shewry, P.R.; Mitchell, R.A.; Tosi, P.; Wan, Y.; Underwood, C.; Lovegrove, A.; Freeman, J.; Toole, G.A.; Mills, E.C.; Ward, J.L. An integrated study of grain development of wheat (cv. Hereward). *J. Cereal Sci.* **2012**, *56*, 21–30. [[CrossRef](#)]
3. Yang, M.; Gao, X.; Dong, J.; Gandhi, N.; Cai, H.; von Wettstein, D.H.; Rustgi, S.; Wen, S. Pattern of Protein Expression in Developing Wheat Grains Identified through Proteomic Analysis. *Front. Plant Sci.* **2017**, *8*, 962. [[CrossRef](#)]
4. Zhang, S.; Ghatak, A.; Bazargani, M.M.; Bajaj, P.; Varshney, R.K.; Chaturvedi, P.; Jiang, D.; Weckwerth, W. Spatial distribution of proteins and metabolites in developing wheat grain and their differential regulatory response during the grain filling process. *Plant J.* **2021**, *107*, 669–687. [[CrossRef](#)] [[PubMed](#)]
5. Gu, A.; Hao, P.; Lv, D.; Zhen, S.; Bian, Y.; Ma, C.; Xu, Y.; Zhang, W.; Yan, Y. Integrated Proteome Analysis of the Wheat Embryo and Endosperm Reveals Central Metabolic Changes Involved in the Water Deficit Response during Grain Development. *J. Agr. Food Chem.* **2015**, *63*, 8478–8487. [[CrossRef](#)]
6. Cao, H.; He, M.; Zhu, C.; Yuan, L.; Dong, L.; Bian, Y.; Zhang, W.; Yan, Y. Distinct metabolic changes between wheat embryo and endosperm during grain development revealed by 2D-DIGE-based integrative proteome analysis. *Proteomics* **2016**, *16*, 1515–1536. [[CrossRef](#)] [[PubMed](#)]
7. He, M.; Zhu, C.; Dong, K.; Zhang, T.; Cheng, Z.; Li, J.; Yan, Y. Comparative proteome analysis of embryo and endosperm reveals central differential expression proteins involved in wheat seed germination. *BMC Plant Biol.* **2015**, *15*, 97. [[CrossRef](#)]
8. Yang, M.; Liu, Y.; Dong, J.; Zhao, W.; Kashyap, S.; Gao, X.; Rustgi, S.; Wen, S. Probing early wheat grain development via transcriptomic and proteomic approaches. *Funct. Integr. Genom.* **2020**, *20*, 63–74. [[CrossRef](#)]
9. Palovaara, J.; Saiga, S.; Wendrich, J.R.; van 't Wout Hofland, N.; van Schayck, J.P.; Hater, F.; Mutte, S.; Sjollem, J.; Boekschoten, M.; Hooiveld, G.J.; et al. Transcriptome Dynamics Revealed by a Gene Expression Atlas of the Early Arabidopsis Embryo. *Nat. Plants* **2017**, *3*, 894–904. [[CrossRef](#)]

10. Day, R.C.; Herridge, R.P.; Ambrose, B.A.; Macknight, R.C. Transcriptome Analysis of Proliferating Arabidopsis Endosperm Reveals Biological Implications for the Control of Syncytial Division, Cytokinin Signaling, and Gene Expression Regulation. *Plant Physiol.* **2008**, *148*, 1964–1984. [[CrossRef](#)]
11. Mizzotti, C.; Rotasperi, L.; Moretto, M.; Tadini, L.; Resentini, F.; Galliani, B.M.; Galbiati, M.; Engelen, K.; Pesaresi, P.; Masiero, S. Time-Course Transcriptome Analysis of Arabidopsis Siliques Discloses Genes Essential for Fruit Development and Maturation. *Plant Physiol.* **2018**, *178*, 1249–1268. [[CrossRef](#)] [[PubMed](#)]
12. Yi, F.; Gu, W.; Chen, J.; Song, N.; Gao, X.; Zhang, X.; Zhou, Y.; Ma, X.; Song, W.; Zhao, H.; et al. High Temporal-Resolution Transcriptome Landscape of Early Maize Seed Development. *Plant Cell* **2019**, *31*, 974–992. [[CrossRef](#)] [[PubMed](#)]
13. Chen, J.; Zeng, B.; Zhang, M.; Xie, S.; Wang, G.; Hauck, A.; Lai, J. Dynamic Transcriptome Landscape of Maize Embryo and Endosperm Development. *Plant Physiol.* **2014**, *166*, 252–264. [[CrossRef](#)]
14. Rangan, P.; Furtado, A.; Henry, R.J. The Transcriptome of the Developing Grain: A Resource for Understanding Seed Development and the Molecular Control of the Functional and Nutritional Properties of Wheat. *BMC Genom.* **2017**, *18*, 766. [[CrossRef](#)] [[PubMed](#)]
15. Yu, Y.; Zhu, D.; Ma, C.; Cao, H.; Wang, Y.; Xu, Y.; Zhang, W.; Yan, Y. Transcriptome Analysis Reveals Key Differentially Expressed Genes Involved in Wheat Grain Development. *Crop J.* **2016**, *4*, 92–106. [[CrossRef](#)]
16. Kirov, I.; Dudnikov, M.; Merkulov, P.; Shingaliev, A.; Omarov, M.; Kolganova, E.; Sigaeva, A.; Karlov, G.; Soloviev, A. Nanopore RNA Sequencing Revealed Long Non-Coding and LTR Retrotransposon-Related RNAs Expressed at Early Stages of Triticale SEED Development. *Plants* **2020**, *9*, 1794. [[CrossRef](#)]
17. Furtado, A.; Bundock, P.C.; Banks, P.M.; Fox, G.; Yin, X.; Henry, R.J. A Novel Highly Differentially Expressed Gene in Wheat Endosperm Associated with Bread Quality. *Sci. Rep.* **2015**, *5*, 10446. [[CrossRef](#)]
18. Kirov, I.; Pirsikov, A.; Milyukova, N.; Dudnikov, M.; Kolenkov, M.; Gruzdev, I.; Siksin, S.; Khrustaleva, L.; Karlov, G.; Soloviev, A. Analysis of Wheat Bread-Making Gene (Wbm) Evolution and Occurrence in Triticale Collection Reveal Origin via Interspecific Introgression into Chromosome 7AL. *Agronomy* **2019**, *9*, 854. [[CrossRef](#)]
19. Guzmán, C.; Xiao, Y.; Crossa, J.; González-Santoyo, H.; Huerta, J.; Singh, R.; Dreisigacker, S. Sources of the Highly Expressed Wheat Bread Making (Wbm) Gene in CIMMYT Spring Wheat Germplasm and Its Effect on Processing and Bread-Making Quality. *Euphytica* **2016**, *209*, 689–692. [[CrossRef](#)]
20. Henry, R.J.; Furtado, A.; Rangan, P. Wheat Seed Transcriptome Reveals Genes Controlling Key Traits for Human Preference and Crop Adaptation. *Curr. Opin. Plant Biol.* **2018**, *45*, 231–236. [[CrossRef](#)]
21. Jin, X.; Fu, Z.; Lv, P.; Peng, Q.; Ding, D.; Li, W.; Tang, J. Identification and Characterization of microRNAs during Maize Grain Filling. *PLoS One* **2015**, *10*, e0125800. [[CrossRef](#)] [[PubMed](#)]
22. Lu, C.; Jeong, D.H.; Kulkarni, K.; Pillay, M.; Nobuta, K.; German, R.; Thatcher, S.R.; Maher, C.; Zhang, L.; Ware, D.; et al. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4951–4956. [[CrossRef](#)] [[PubMed](#)]
23. Yang, G.; Deng, P.; Guo, Q.; Shi, T.; Pan, W.; Cui, L.; Liu, X.; Nie, X. Population Transcriptomic Analysis Identifies the Comprehensive lncRNAs Landscape of Spike in Wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2022**, *22*, 450. [[CrossRef](#)] [[PubMed](#)]
24. Cao, P.; Fan, W.; Li, P.; Hu, Y. Genome-Wide Profiling of Long Noncoding RNAs Involved in Wheat Spike Development. *BMC Genom.* **2021**, *22*, 493. [[CrossRef](#)]
25. Heo, J.B.; Sung, S. Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. *Science* **2011**, *331*, 76–79. [[CrossRef](#)]
26. Madhawan, A.; Sharma, A.; Bhandawat, A.; Rahim, M.S.; Kumar, P.; Mishra, A.; Parveen, A.; Sharma, H.; Verma, S.K.; Roy, J. Identification and Characterization of Long Non-Coding RNAs Regulating Resistant Starch Biosynthesis in Bread Wheat (*Triticum aestivum* L.). *Genomics* **2020**, *112*, 3065–3074. [[CrossRef](#)] [[PubMed](#)]
27. Parker, M.T.; Knop, K.; Sherwood, A.V.; Schurch, N.J.; Mackinnon, K.; Gould, P.D.; Hall, A.J.; Barton, G.J.; Simpson, G.G. Nanopore Direct RNA Sequencing Maps the Complexity of Arabidopsis mRNA Processing and m6A Modification. *eLife* **2020**, *9*, e49658. [[CrossRef](#)]
28. Byrne, A.; Cole, C.; Volden, R.; Vollmers, C. Realizing the Potential of Full-Length Transcriptome Sequencing. *Philos. Trans. R. Soc. B* **2019**, *374*, 20190097. [[CrossRef](#)]
29. Zhao, L.; Zhang, H.; Kohnen, M.V.; Prasad, K.V.S.K.; Gu, L.; Reddy, A.S.N. Analysis of Transcriptome and Epitranscriptome in Plants Using PacBio Iso-Seq and Nanopore-Based Direct RNA Sequencing. *Front. Genet.* **2019**, *10*, 253. [[CrossRef](#)]
30. Zhang, S.; Li, R.; Zhang, L.; Chen, S.; Xie, M.; Yang, L.; Xia, Y.; Foyer, C.H.; Zhao, Z.; Lam, H.-M. New Insights into Arabidopsis Transcriptome Complexity Revealed by Direct Sequencing of Native RNAs. *Nucleic Acids Res.* **2020**, *48*, gkaa588. [[CrossRef](#)]
31. Wang, D.; Lu, X.; Chen, X.; Wang, S.; Wang, J.; Guo, L.; Yin, Z.; Chen, Q.; Ye, W. Temporal Salt Stress-Induced Transcriptome Alterations and Regulatory Mechanisms Revealed by PacBio Long-Reads RNA Sequencing in *Gossypium hirsutum*. *BMC Genom.* **2020**, *21*, 838. [[CrossRef](#)]
32. Wang, M.; Wang, P.; Liang, F.; Ye, Z.; Li, J.; Shen, C.; Pei, L.; Wang, F.; Hu, J.; Tu, L.; et al. A Global Survey of Alternative Splicing in Allopolyploid Cotton: Landscape, Complexity and Regulation. *New Phytol.* **2018**, *217*, 163–178. [[CrossRef](#)] [[PubMed](#)]
33. Lyu, J.I.; Ramekar, R.; Kim, J.M.; Hung, N.N.; Seo, J.S.; Kim, J.-B.; Choi, I.-Y.; Park, K.-C.; Kwon, S.-J. Unraveling the Complexity of Faba Bean (*Vicia faba* L.) Transcriptome to Reveal Cold-Stress-Responsive Genes Using Long-Read Isoform Sequencing Technology. *Sci. Rep.* **2021**, *11*, 21094. [[CrossRef](#)]

34. Athiyannan, N.; Abrouk, M.; Boshoff, W.H.P.; Cauet, S.; Rodde, N.; Kudrna, D.; Mohammed, N.; Bettgenhaeuser, J.; Botha, K.S.; Derman, S.S.; et al. Long-Read Genome Sequencing of Bread Wheat Facilitates Disease Resistance Gene Cloning. *Nat. Genet.* **2022**, *54*, 227–231. [[CrossRef](#)]
35. Kirov, I.; Omarov, M.; Merkulov, P.; Dudnikov, M.; Gvaramiya, S.; Kolganova, E.; Komakhin, R.; Karlov, G.; Soloviev, A. Genomic and Transcriptomic Survey Provides New Insight into the Organization and Transposition Activity of Highly Expressed LTR Retrotransposons of Sunflower (*Helianthus annuus* L.). *Int. J. Mol. Sci.* **2020**, *21*, 9331. [[CrossRef](#)]
36. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A.; et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* **2009**, *326*, 1112–1115. [[CrossRef](#)] [[PubMed](#)]
37. Staton, S.E.; Bakken, B.H.; Blackman, B.K.; Chapman, M.A.; Kane, N.C.; Tang, S.; Ungerer, M.C.; Knapp, S.J.; Rieseberg, L.H.; Burke, J.M. The Sunflower (*Helianthus annuus* L.) Genome Reflects a Recent History of Biased Accumulation of Transposable Elements. *Plant J.* **2012**, *72*, 142–153. [[CrossRef](#)] [[PubMed](#)]
38. Ibarra-Laclette, E.; Lyons, E.; Hernández-Guzmán, G.; Pérez-Torres, C.A.; Carretero-Paulet, L.; Chang, T.-H.; Lan, T.; Welch, A.J.; Juárez, M.J.A.; Simpson, J.; et al. Architecture and Evolution of a Minute Plant Genome. *Nature* **2013**, *498*, 94–98. [[CrossRef](#)] [[PubMed](#)]
39. Bennetzen, J.L.; Wang, H. The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Plant Biol.* **2014**, *65*, 505–530. [[CrossRef](#)]
40. Baud, A.; Wan, M.; Nouaud, D.; Francillonne, N.; Anxolabéhère, D.; Quesneville, H. Traces of Transposable Elements in Genome Dark Matter Co-Opted by Flowering Gene Regulation Networks. *Peer Community J.* **2022**, *2*, e14. [[CrossRef](#)]
41. Kapusta, A.; Kronenberg, Z.; Lynch, V.J.; Zhuo, X.; Ramsay, L.; Bourque, G.; Yandell, M.; Feschotte, C. Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genet.* **2013**, *9*, e1003470. [[CrossRef](#)]
42. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory Activities of Transposable Elements: From Conflicts to Benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [[CrossRef](#)] [[PubMed](#)]
43. Cho, J.H.; Choi, M.N.; Yoon, K.H.; Kim, K.-N. Ectopic Expression of SjCBL1, Calcineurin B-Like 1 Gene from *Sedirea Japonica*, Rescues the Salt and Osmotic Stress Hypersensitivity in *Arabidopsis* Cbl1 Mutant. *Front. Plant Sci.* **2018**, *9*, 1188. [[CrossRef](#)] [[PubMed](#)]
44. The International Wheat Genome Sequencing Consortium (IWGSC); Appels, R.; Eversole, K.; Stein, N.; Feuillet, C.; Keller, B.; Rogers, J.; Pozniak, C.J.; Choulet, F.; Distelfeld, A.; et al. Shifting the Limits in Wheat Research and Breeding Using a Fully Annotated Reference Genome. *Science* **2018**, *361*, eaar7191. [[CrossRef](#)]
45. Rabanus-Wallace, M.T.; Hackauf, B.; Mascher, M.; Lux, T.; Wicker, T.; Gundlach, H.; Baez, M.; Houben, A.; Mayer, K.F.X.; Guo, L.; et al. Chromosome-Scale Genome Assembly Provides Insights into Rye Biology, Evolution and Agronomic Potential. *Nat. Genet.* **2021**, *53*, 564–573. [[CrossRef](#)] [[PubMed](#)]
46. Lukaszewski, A.J.; Curtis, C.A. Transfer of the Glu-D1 Gene from Chromosome 1D of Breadwheat to Chromosome 1R in Hexaploid Triticale. *Plant Breed.* **1992**, *109*, 203–210. [[CrossRef](#)]
47. Ma, X.; Wang, Q.; Wang, Y.; Ma, J.; Wu, N.; Ni, S.; Luo, T.; Zhuang, L.; Chu, C.; Cho, S.-W.; et al. Chromosome Aberrations Induced by Zebularine in Triticale. *Genome* **2016**, *59*, 485–492. [[CrossRef](#)]
48. Shi, C.; Xu, L. Characters of Cysteine Endopeptidases in Wheat Endosperm during Seed Germination and Subsequent Seedling Growth. *J. Integr. Plant Biol.* **2009**, *51*, 52–57. [[CrossRef](#)]
49. Tottman, D.R. The Decimal Code for the Growth Stages of Cereals, with Illustrations. *Ann. Appl. Biol.* **1987**, *110*, 441–454. [[CrossRef](#)]
50. Szewińska, J.; Simińska, J.; Bielawski, W. The Roles of Cysteine Proteases and Phytocystatins in Development and Germination of Cereal Seeds. *J. Plant Physiol.* **2016**, *207*, 10–21. [[CrossRef](#)]
51. Yi, R.; Zhu, Z.; Hu, J.; Qian, Q.; Dai, J.; Ding, Y. Identification and Expression Analysis of microRNAs at the Grain Filling Stage in Rice (*Oryza sativa* L.) via Deep Sequencing. *PLoS One* **2013**, *8*, e57863. [[CrossRef](#)] [[PubMed](#)]
52. Meng, F.; Liu, H.; Wang, K.; Liu, L.; Wang, S.; Zhao, Y.; Yin, J.; Li, Y. Development-associated microRNAs in grains of wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2013**, *13*, 140. [[CrossRef](#)] [[PubMed](#)]
53. Hu, Y.; Li, Y.; Weng, J.; Liu, H.; Yu, G.; Liu, Y.; Xiao, Q.; Huang, H.; Wang, Y.; Wei, B.; et al. Coordinated regulation of starch synthesis in maize endosperm by microRNAs and DNA methylation. *Plant J.* **2021**, *105*, 108–123. [[CrossRef](#)] [[PubMed](#)]
54. Hung, Y.-H.; Slotkin, R.K. The initiation of RNA interference (RNAi) in plants. *Curr. Opin. Plant Biol.* **2021**, *61*, 102014. [[CrossRef](#)]
55. Zemach, A.; Kim, M.Y.; Silva, P.; Rodrigues, J.A.; Dotson, B.; Brooks, M.D.; Zilberman, D. Local DNA Hypomethylation Activates Genes in Rice Endosperm. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 18729–18734. [[CrossRef](#)]
56. Hsieh, T.-F.; Ibarra, C.A.; Silva, P.; Zemach, A.; Eshed-Williams, L.; Fischer, R.L.; Zilberman, D. Genome-Wide Demethylation of *Arabidopsis* Endosperm. *Science* **2009**, *324*, 1451–1454. [[CrossRef](#)]
57. Pertea, G.; Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **2020**, *9*, 304. [[CrossRef](#)]
58. Ge, S.X.; Jung, D.; Yao, R. ShinyGO: A Graphical Enrichment Tool for Animals and Plants. *Bioinformatics* **2019**, *36*, 2628–2629. [[CrossRef](#)]

-
59. Wickham, H. Ggplot2. *WIREs Comput. Stat.* **2011**, *3*, 180–185. [[CrossRef](#)]
 60. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties. *Bioinformatics* **2017**, *33*, 2938–2940. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.