agronomy

MDPI

Article

Genome-Wide Investigation of Spliceosomal SM/LSM Genes in Wheat (*Triticum aestivum* L.) and Its Progenitors

Ruiting Gao [1,†], Ning Su [1,†], Wenqiu Pan [1], Qiaoyu Bao [1], Zhen Li [1], Xiaojun Nie [1], Wei Tong [1,*] and Weining Song [1,2]

[1] State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy and Yangling Branch of China Wheat Improvement Center, Northwest A&F University, Yangling 712100, China; gruit@nwafu.edu.cn (R.G.); sunnying@nwafu.edu.cn (N.S.); wenqiu_pan@nwafu.edu.cn (W.P.); Baoqy@nwafu.edu.cn (Q.B.); lizhen.a@nwafu.edu.cn (Z.L.); small@nwafu.edu.cn (X.N.); sweining2002@yahoo.com (W.S.)

[2] Australia-China Joint Research Centre for Abiotic and Biotic Stress Management in Agriculture, Horticulture and Forestry, Yangling 712100, China

* Correspondence: tongw@nwsuaf.edu.cn

† These authors contributed equally to this work.

Abstract: The SSM/SLSM (spliceosomal Smith (SM)/SM-like (LSM)) genes are the central components of the spliceosome in eukaryotes, which play an important role in regulating RNA splicing, participating in diverse biological processes. Although it has been detected in *Arabidopsis* and rice etc. plants, the members and significance of the SSM/SLSM gene family in wheat are still not reported. In this study, we identified the SSM/SLSM genes in wheat and its progenitors at genome-scale, where 57 SSM/SLSM genes were identified in wheat, together with 41, 17and 19 found in *Triticum dicoccoides*, *Triticum urartu*, and *Aegilops tauschii*. Furthermore, their phylogenetic relationship, gene structures, conserved motifs, and cis-regulatory elements were systematically analyzed. By synteny analysis, good collinearity of SSM/SLSM genes was found among bread wheat and its progenitors' genomes, and the distribution of SMD2 genes in wheat chromosome 5A, 4B and 4D located in the 4AL-5AL-7BS chromosome model, due to the translocation. Then, the positively selected genes were further investigated based on the non-synonymous to synonymous (dN/dS) analysis of the orthologous pairs. Finally, the expression profiles of the SSM/SLSM genes were detected using RNA-seq datasets, and eight stress-responsive candidate genes were selected to validate their expression through qPCR (real-time quantitative polymerase chain reaction). According to the co-expression network analysis, the correlation between the LSM7-7A gene and related genes was illustrated through Gene Ontology (GO) enrichment analysis. Furthermore, the LSM7-7A gene was related to the *Arabidopsis* homologous salt tolerance gene RCY1. This investigation systematically identified the complete candidates of SSM/SLSM genes and their characters in wheat and its progenitors, and provided clues to a better understanding of their contribution during the wheat polyploidy process.

Keywords: SSM/SLSM gene; wheat; phylogenetic analysis; genome-scale; gene expression

Citation: Gao, R.; Su, N.; Pan, W.; Bao, Q.; Li, Z.; Nie, X.; Tong, W.; Song, W. Genome-Wide Investigation of Spliceosomal SM/LSM Genes in Wheat (*Triticum aestivum* L.) and Its Progenitors. *Agronomy* **2021**, *11*, 1429. https://doi.org/10.3390/agronomy11071429

Academic Editor: Daisuke Saisho

Received: 9 June 2021
Accepted: 6 July 2021
Published: 17 July 2021

1. Introduction

In eukaryotes, precursor mRNAs (pre-mRNAs) become mature mRNAs through fundamental processes of splicing, 5′ capping, 3′ polyadenylation, etc. Pre-mRNA splicing is catalyzed by the spliceosome, a highly dynamic ribonucleoprotein machinery containing small nuclear ribonucleo-proteins (snRNPs) and non-snRNP proteins [1,2]. The major spliceosomes (U2-dependent) contain five types of snRNP referred to as U1, U2, U4, U5, and U6 snRNPs.

The structure of those snRNPs contains seven core binding proteins, for U1, U2, U4, and U5 snRNPs, the seven binding proteins were belong to Smith (SM) proteins, which were first discovered in Miss Smith diagnosed in systemic lupus erythematosus [3], and

denoted as SMB/B′, SMD1, SMD2, SMD3, SME, SMF, and SMG (SSM gene family) [4–6]. For the U6 snRNP, the binding proteins were replaced with seven SM-like proteins (LSM) named as LSM2, LSM3, LSM4, LSM5, LSM6, LSM7, and LSM8 (SLSM gene family) [7–10]. These 14 genes are directly binding with the spliceosome, thus they are collectively referred to as SSM/SLSM (spliceosomal SM/LSM) gene family.

The 14 SSM/SLSM genes belong to an evolutionarily ancient family of small RNA-binding proteins involved in pre-mRNA splicing [9,11]. The protein of SM family is formed of two conserved fragments, SM1 and SM2, which are connected by a varying length sequence [11–14].

In plants, the SM gene family, including SSM/SLSM gene family, has been reported in *Arabidopsis thaliana* [15], longan [16], maize and rice [17]. In Arabidopsis thaliana [18], nine other plant species (*Glycine max*, *Lotus japonicus*, *Medicago truncatula*, *Oryza sativa*, *Physcomitrella patens*, *Populus trichocarpa*, *Sorghum bicolour*, *Vitis vinifera*, and *Zea mays*) [19], spliceosomal associated proteins were identified, including SSM/SLSM proteins, where the identification of SSM/SLMS members in maize and rice were mainly based on a bioinformatics search of related members in *Arabidopsis thaliana* [19]. However, the SM family in animals has not been specifically studied. While the SMB′ protein was found only in a small number of rodent cell types and suggest a role in the regulation of some cases of alternative RNA splicing [20]. Anne [21] l found that arginine methylation of SMB is required for *Drosophila* germ cell development. Scruggs [22] implicate SmD3 as a critical determinant in the processing of intronic non-coding RNAs in general and as an upstream mediator of metabolic stress response pathways through the regulation of snoRNA expression. SME and SMG proteins were associated with cancer [23,24].

The SSM/SLSM proteins in plant are not only associated with spliceosomes but also related to the circadian rhythm [25], mRNA degradation [10], and stress resistance [26]. So far, the distinct functions of the SSM/SLSM genes have not been extensively deciphered in plants. Only the LSM1-7 and LSM2-8 complexes and SME, SMD3, LSM5, and LSM4 genes have been studied in Arabidopsis [26–30]. However, the SSM/SLSM genes have not been identified and characterized in bread wheat up to now.

Bread wheat is an allohexaploid species (*Triticum aestivum* L., AABBDD) originating from two major allopolyploid events [31–33]. Firstly, diploid *Triticum urartu* (AA) hybridized with an unknown diploid grass (related to *Aegilops speltoides*, BB) to produce wild tetraploid wheat *Triticum dicoccoides* (AABB). Then, wild tetraploid wheat hybridized with the diploid goat grass *Aegilops tauschii* (DD) to form hexaploid bread wheat [34]. Allopolyploidy can result in the change of transcription and/or function in homologous genes [32]. Here, the genome-scale SSM/SLSM genes were systematically identified in bread wheat and its progenitors. Then, their genomic organization, phylogenetic relationships, gene structures, conserved motifs and gene expression patterns were comprehensively investigated.

2. Materials and Methods

2.1. Identification of Smith (SM)/SM-Like (SSM/SLSM) Genes in Wheat and Its Progenitors

The *Arabidopsis* SSM/SLSM protein sequences were retrieved and downloaded from the SRGD database (http://www.plantgdb.org/SRGD/, accessed on 1 June 2021), and then used as the queries to perform a BLASTP search against the local protein database of wheat (IWGSC_v1.1) and its progenitors, which was downloaded from Ensembl Plants (http://plants.ensembl.org/index.html, accessed on 1 June 2021, MBKbase (http://www.mbkbase.org, accessed on 1 June 2021 with the expected value (E-value) of 1×10^{-20}. Meantime, the SSM/SLSM domain (PF01423) was downloaded from the PFAM database (http://pfam.xfam.org/, accessed on 1 June 2021. Hmmsearch tool implemented in HMMER 3.3.1 [35] was used to search for the proteins with this domain in local wheat protein database with an E-value of 1×10^{-5} Furthermore, the protein sequences identified by both above methods were integrated and parsed by manual editing to remove the redundant. The remaining proteins were considered as candidate SSM/SLSM proteins. The candidates were finally submitted to the PfamScan database (https://www.ebi.ac.uk/Tools/pfa/

pfamscan/, accessed on 1 June 2021 and NCBI-CDD (https://www.ncbi.nlm.nih.gov/cdd/, accessed on 1 June 2021 to verify the SSM/SLSM conserved domain. Additionally, the same method was used to identify the SSM/SLSM genes in wild emmer wheat (*T. dicoccoides*), *A. tauschii*, and *T. urartu*. The relationship between gene name and gene ID is in Table 1.

2.2. Chromosome Location, Gene Structure, Collinearity Analyses and Non-Synonymous to Synonymous (dN/dS) of SSM/SLSM Genes

The physical properties of proteins were identified using ExPASy (https://web.expasy.org/protparam/, accessed on 1 June 2021, including the number of amino acids, molecular weight, theoretical pI, and grand average of hydropathicity (GRAVY). MEME online analysis (http://meme-suite.org/, accessed on 1 June 2021 of conservative motif of SSM/SLSM proteins was used, and TBtools was used to draw the gene structure and motif. The gene collinearity analysis of wheat and its relatives were predicted using the MCScanX [36] program and visualized with Circos [37]. In order to observe the gene changes during the process of wheat polyploidy, we calculated the dN/dS, which was displayed using ParaAT [38] and PAML [39].

2.3. Analysis of Cis-Acting Elements of SSM/SLSM Genes Promoter

The 1500 bp promoter sequences were processed through PERL script and used for the prediction of the plant cis-acting regulatory elements. The PlantCARE database was used for the identification of the elements in the promoters (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/ (accessed on 1 June 2021)). Then, PERL scripts were used to calculate the cis-acting elements of the promoters in Supplemental Table S1.

2.4. Analysis of the Specific Expression of SSM/SLSM Genes

We downloaded the original data from the NCBI database (detailed information is in Supplemental Table S2) and processed the data according to the transcriptome analysis process. In the first step, low-quality data were filtered through the software Trimmomatic, and in the second step, clean data was compared to the reference genome by Hisat2 [40]. The third step is to carry out quantitative calculation of reads to the reference genome by comparison with Stringtie software.

TPM values of transcripts in the five tissues (root, stem, leaf, spike and grain) under no pressure treatment were calculated by taking the mean value method. Similarly, TPM values in four stages (booting stage, heading stage, flowering stage and grain filling stage) were calculated by taking the mean value method. The same calculation was performed for the stress treatment including heat stress, drought stress and salt stress.

According to the RNA-seq analysis of drought and salt resistance, eight SM/LSM genes (TALSM5-1A, TALSM8-1A, TALSM5-1D, TASMD1-2A, TASMD1-2D, TASMD2-5A, TASME-6A, TASMF-6B, TASMD1-6D, and TALSM7-7A) were selected to perform real-time quantitative polymerase chain reaction (qPCR, detail in Figure S1).

The seeds of wheat genotype Chinese Spring were germinated in petri dishes and grown in a growth chamber at controlled conditions (23 ± 1 °C, 16-h light/8-h dark cycle). The three-leaf seedlings were used for stress treatments. The plants were incubated in 19% m/v polyethylene glycol (PEG6000) solution. Three plants with similar growth were selected for 6 h, 12 h, 24 h and 48 h under drought stress treatment. Meantime, the plants were incubated in 200 mM NaCl solution. Three plants with similar growth were selected for 6 h, 12 h, 24 h and 48 h under salt stress treatment. Seedlings under the normal condition were used as the control. Leaves of all these samples were collected with three biological replications. The total RNA of these samples was isolated by Plant RNA Kit reagent (Omega Bio-Tek, Norcross, GA, USA) according to the manufacturer's instructions. Eight primers for SM/LSM genes were designed using Primer 5.0 (Supplemental Table S3). The internal reference primer was the wheat's β-actin gene. qPCR were performed on the QuantStudioTM 7 Flex System (Thermo Fisher Scientifc, Waltham, MA, USA) with SYBR® Premix Ex Taq™ II (TaKaRa, Dalian, China) with the thermal cycling condition was 95 °C for 30 s followed by 40 cycles of 95 °C for 3 s, 60 °C for 30 s, then 95 °C for 15 s.

The qPCR for each primer was repeated three times in each different treatment. Three technological replications were applied and the expression level was calculated using the $2^{-\Delta\Delta CT}$ method [41].

2.5. Construction of the Co-Expression Network and Functional Search of Key Genes

The WGCNA package [42] was used to construct the co-expression network based on the wheat transcriptome data (Supplemental Table S2). Set the power to 26 to build a co-expression network. The genes associated with TaLSM7-7A with a weight of more than 0.3 from the co-expression network were extracted and displayed by Cytoscape3.6.0 software [43].

Then, the genes related to the LSM7 gene were extracted through the Perl script, and these genes were enriched in TBtools [44] and mapped by WEGO [45]. And we found the genes related to salt stress, and performed BLAST comparison with the Arabidopsis genome to find the homologous genes of the genes, which were related to salt stress by referring to the literature.

For the materials and methods part, we have prepared a diagram showing the whole steps in the Supplemental Figure S2.

3. Results

3.1. Identification of SSM/SLSM Gene Members and Their Physico-Chemical Properties in Wheat and Its Progenitors

By BLASTP and HMMER software searching, 57 SSM/SLSM genes were found in the wheat (TaSSM/SLSMs). In addition, 41, 17, and 19 SSM/SLSM genes were also identified in *Triticum dicoccoides*, *Triticum urartu*, and *Aegilops tauschii*, respectively. SSM/SLSM genes were named according to their chromosome location (Supplemental File 1). There are no significant sequence variation of the 19, 19, and 16 SSM/SLSM genes in wheat's A, B, and D subgenomes. The physico-chemical property features of identified SSM/SLSM proteins in wheat and its progenitors were listed in Table 1.

Table 1. The physico-chemical properties features of Smith (SM)/SM-Like (SSM/SLSM) genes.

Gene Name		AA	MW (Da)	PI	GRAVY	Gene Id
LSM2	TuLSM2-7	148	16,537.77	8.45	−0.511	TuG1812G0700005606.01.P01
	TdLSM2-7B	148	16,319.52	7.43	−0.326	TRIDC7BG033290.2
	TdLSM2-7A	93	10,737.36	6.82	−0.199	TRIDC7AG042160.1
	TaLSM2-7D	93	10,737.36	6.82	−0.199	TraesCS7D02G297400.1
	TaLSM2-7B	93	10,737.36	6.82	−0.199	TraesCS7B02G202200.1
	TaLSM2-7A	93	10,737.36	6.82	−0.199	TraesCS7A02G302100.1
	AetLSM2-7	93	10,737.36	6.82	−0.199	AET7Gv20754100.2
LSM3	TuLSM3-3	118	13,347.21	5.06	−0.311	TuG1812G0300004141.01.P01
	TdLSM3-3B	106	11,936.61	4.69	−0.105	TRIDC3BG060910.5
	TdLSM3-3A	104	12,178.97	5.14	−0.15	TRIDC3AG054250.1
	TdLSM3-1A	98	11,279.92	4.76	−0.182	TRIDC1AG040710.1
	TaLSM3-3D	82	96,07.99	4.57	−0.26	TraesCS3D02G371500.1
	TaLSM3-3A	97	11,196.83	4.76	−0.215	TraesCS3A02G378200.1
	TaLSM3-1D	97	11,196.83	4.76	−0.215	TraesCS1D02G273400.1
	TaLSM3-1B	97	11,196.83	4.76	−0.215	TraesCS1B02G283200.1
	TaLSM3-1A	98	11,279.92	4.76	−0.182	TraesCS1A02G273400.1
	AetLSM3-3	98	11,279.92	4.76	−0.182	AET3Gv20844000.2
	AetLSM3-1	119	13,735.82	4.8	−0.05	AET1Gv20658000.5
LSM4	TuLSM4-3	143	15,472.57	9.96	−0.914	TuG1812G0300001955.01.P01
	TdLSM4-3B	143	15,469.51	9.96	−0.952	TRIDC3BG028150.3
	TdLSM4-3A	143	15,472.57	9.96	−0.914	TRIDC3AG024440.1
	TaLSM4-3D	143	15,472.57	9.96	−0.914	TraesCS3D02G181600.1
	TaLSM4-3B	143	15,472.57	9.96	−0.914	TraesCS3B02G205600.2
	TaLSM4-3A	143	15,472.57	9.96	−0.914	TraesCS3A02G175900.1
	AetLSM4-3	167	18,150.39	10.27	−0.975	AET3Gv20400700.6

Table 1. *Cont.*

Gene Name		AA	MW (Da)	PI	GRAVY	Gene Id
LSM5	TuLSM5-1	148	16,232.73	9.33	−0.264	TuG1812G0100002831.01.P01
	TdLSM5-1B	150	16,187.58	8.74	−0.29	TRIDC1BG041610.1
	TdLSM5-1A	118	12,956.97	7.93	−0.098	TRIDC1AG036620.3
	TaLSM5-1D	88	9590.03	4.42	0	TraesCS1D02G242900.1
	TaLSM5-1B	89	9753.21	4.42	−0.015	TraesCS1B02G254400.1
	TaLSM5-1A	89	9753.21	4.42	−0.015	TraesCS1A02G242900.1
	AetLSM5-1	89	9753.21	4.42	−0.015	AET1Gv20591900.1
LSM6	TuLSM6-2	128	13,818.85	9.72	−0.241	TuG1812G0200003237.01.P02
	TdLSM6-2B	139	14,806.67	9.86	−0.623	TRIDC2BG044080.1
	TdLSM6-2A	99	10,360.7	9.13	−0.362	TRIDC2AG041140.2
	TaLSM6-2D	99	10,350.66	9.13	−0.354	TraesCS2D02G283200.1
	TaLSM6-2B	99	10,350.66	9.13	−0.354	TraesCS2B02G301300.1
	TaLSM6-2A	99	10,374.72	9.13	−0.361	TraesCS2A02G284300.1
	AetLSM6-2	99	10,350.66	9.13	−0.354	AET2Gv20643200.1
LSM7	TuLSM7-7	104	11,432.98	4.97	−0.421	TuG1812G0700004021.01.P01
	TuLSM7-2	100	10,675.18	4.87	−0.183	TuG1812G0200005747.01.P01
	TdLSM7-7B	110	12,203.98	4.74	−0.201	TRIDC7BG042390.2
	TdLSM7-7A	140	15,455.53	8.84	−0.652	TRIDC7AG051520.1
	TdLSM7-2B	85	9209.38	4.62	−0.184	TRIDC2BG080800.1
	TdLSM7-2A	72	7675.66	4.27	−0.101	TRIDC2AG074130.1
	TaLSM7-U	147	15,822.08	5.61	−0.196	TraesCSU02G104500.2
	TaLSM7-7B	104	undefined	undefined	−0.286	TraesCS7B02G259600.1
	TaLSM7-7A	104	11,447.01	4.97	−0.418	TraesCS7A02G368900.1
	TaLSM7-2D	100	10,675.18	4.87	−0.183	TraesCS2D02G531700.1
	TaLSM7-2B	100	10,675.18	4.87	−0.183	TraesCS2B02G559200.1
	TaLSM7-2A	100	10,675.18	4.87	−0.183	TraesCS2A02G528800.1
	TaLSM7-1D	97	10,593.4	8.31	0.218	TraesCS1D02G062600.1
	AetLSM7-7	132	14,456.32	7.16	−0.572	AET7Gv20882600.3
	AetLSM7-2	100	10,675.18	4.87	−0.183	AET2Gv21167900.2
LSM8	TuLSM8-1	99	10,756.24	4.55	0.044	TuG1812G0100004839.01.P01
	TdLSM8-1B	98	10,642.14	4.55	0.077	TRIDC1BG074340.1
	TdLSM8-1A	97	10,770.39	5.63	−0.115	TRIDC1AG065090.2
	TaLSM8-1D	99	10,756.24	4.55	0.044	TraesCS1D02G451800.1
	TaLSM8-1B	99	10,756.24	4.55	0.044	TraesCS1B02G478200.1
	TaLSM8-1A	99	10,770.27	4.55	0.047	TraesCS1A02G443700.1
	AetLSM8-1	99	11,084.79	4.66	0.205	AET1Gv21045900.3
SMB	TuSMB-2	265	27,936.58	11.3	−0.689	TuG1812G0200002708.01.P02
	TdSMB-2B	265	27,965.58	11.3	−0.708	TRIDC2BG038770.1
	TdSMB-2A	265	27,936.58	11.3	−0.689	TRIDC2AG034250.1
	TaSMB-2D	265	28,038.72	11.3	−0.665	TraesCS2D02G248300.1.cds1
	TaSMB-2B	265	27,965.58	11.3	−0.708	TraesCS2B02G270200.1.cds1
	TaSMB-2A	265	27,936.58	11.3	−0.689	TraesCS2A02G243500.1.cds1
	AetSMB-2	265	28,038.72	11.3	−0.665	AET2Gv20552800.1
SMD1	TuSMD1-6	114	12,727.99	11.23	−0.538	TuG1812G0600002404.01.P01
	TuSMD1-2	172	18,940.35	8.55	0.271	TuG1812G0200003790.01.P03
	TdSMD1-6B	146	16,050.8	11.05	−0.471	TRIDC6BG036850.1
	TdSMD1-6A	184	20,193.36	11.29	−0.507	TRIDC6AG032030.1
	TdSMD1-2B	136	15,176.65	10.72	−0.59	TRIDC2BG050890.3
	TdSMD1-2A	175	19,362.8	9.8	0.126	TRIDC2AG048420.3
	TaSMD1-6D	107	11,809.78	11.02	−0.645	TraesCS6D02G188600.1
	TaSMD1-6B	107	11,809.78	11.02	−0.645	TraesCS6B02G230500.1
	TaSMD1-6A	114	12,727.99	11.23	−0.538	TraesCS6A02G207000.1
	TaSMD1-2D	114	12,722.1	11.23	−0.642	TraesCS2D02G331500.1
	TaSMD1-2B	107	11,821.72	11.02	−0.721	TraesCS2B02G350800.1
	TaSMD1-2A	114	12,703.42	11.21	−0.641	TraesCS2A02G331400.1
	AetSMD1-2	140	15,650.33	10.74	−0.487	AET2Gv20748800.2

Table 1. *Cont.*

Gene Name		AA	MW (Da)	PI	GRAVY	Gene Id
SMD2	TuSMD2-5	119	13,325.73	9.81	−0.354	TuG1812G0500005377.01.P02
	TuSMD2-3	106	12,085.23	9.95	−0.477	TuG1812G0300005644.01.P01
	TdSMD2-5A	107	12,224.42	9.95	−0.457	TRIDC5AG075140.1
	TdSMD2-4B	89	10,219.24	10.82	−0.343	TRIDC4BG060470.1
	TdSMD2-3B	107	12,156.31	9.95	−0.456	TRIDC3BG084500.1
	TdSMD2-3A	129	14,372.68	10.24	−0.729	TRIDC3AG073200.5
	TaSMD2-5A	107	12,224.42	9.95	−0.457	TraesCS5A02G529800.2
	TaSMD2-4D	107	12,224.42	9.95	−0.457	TraesCS4D02G354800.2
	TaSMD2-4B	107	12,224.42	9.95	−0.457	TraesCS4B02G361800.1
	TaSMD2-3D	105	11,985.11	9.88	−0.422	TraesCS3D02G524800.1
	TaSMD2-3B	115	13,068.47	10.07	−0.365	TraesCS3B02G584900.1
	TaSMD2-3A	106	12,085.23	9.95	−0.477	TraesCS3A02G517300.1
	AetSMD2-4	107	12,224.42	9.95	−0.457	AET4Gv20832900.1
	AetSMD2-3	106	12,113.28	9.95	−0.455	AET3Gv21214400.2
SMD3	TuSMD3-5	135	14,505.94	11.06	−0.23	TuG1812G0500005669.01.P01
	TdSMD3-7A	139	15,484.09	9.96	−0.266	TRIDC7AG005000.1
	TdSMD3-5A	158	17,171.04	11.5	−0.385	TRIDC5AG078210.2
	TdSMD3-4B	196	22,029.53	10.43	−0.37	TRIDC4BG064290.3
	TdSMD3-4A	150	Undefined	undefined	−0.135	TRIDC4AG065300.1
	TaSMD3-U	135	14,505.94	11.06	−0.23	TraesCSU02G033900.1
	TaSMD3-5A	135	14,505.94	11.06	−0.23	TraesCS5A02G554700.1
	TaSMD3-4B	135	14,505.94	11.06	−0.23	TraesCS4B02G392700.1
	AetSMD3-4	166	18,806.14	8.73	−0.654	AET4Gv20876200.2
SME	TuSME-7	146	16,460.43	10.74	−0.116	TuG1812G0700003308.01.P01
	TdSME-7B	88	10,359.28	9.89	−0.265	TRIDC7BG033630.1
	TdSME-7A	143	16,103.99	10.75	−0.093	TRIDC7AG042500.1
	TdSME-6B	63	7409.69	8.3	−0.335	TRIDC6BG015620.1
	TdSME-6A	95	11,020.96	9.58	−0.052	TRIDC6AG011490.2
	TaSME-U	86	10,185.06	9.52	−0.141	TraesCSU02G017200.1
	TaSME-7D	88	10,359.28	9.89	−0.265	TraesCS7D02G299900.1
	TaSME-7B	88	10,359.28	9.89	−0.265	TraesCS7B02G204700.1
	TaSME-7A	88	10,359.28	9.89	−0.265	TraesCS7A02G304400.1
	TaSME-6B	97	11,483.56	5.37	−0.069	TraesCS6B02G111500.1
	TaSME-6A	117	13,453.83	10.26	−0.306	TraesCS6A02G089100.2
	AetSME-7	103	12,280.32	9.2	−0.403	AET7Gv20759600.3
	AetSME-6	143	16,406.26	11.53	−0.49	AET6Gv20214400.2
SMF	TuSMF-7	143	15,245.44	7.23	−0.272	TuG1812G0700005332.01.P01
	TdSMF-7B	97	10,571.1	4.69	−0.205	TRIDC7BG063710.1
	TdSMF-7A	94	10,270.78	4.5	−0.151	TRIDC7AG069290.2
	TdSMF-6B	86	9664.09	4.42	−0.127	TRIDC6BG026950.1
	TdSMF-6A	66	7491.65	8.24	−0.191	TRIDC6AG021120.1
	TaSMF-7B	87	9673.16	4.5	−0.116	TraesCS7B02G401200.1
	TaSMF-7A	87	9673.16	4.5	−0.116	TraesCS7A02G496000.1
	TaSMF-6D	113	12,580.39	4.43	−0.068	TraesCS6D02G141900.1
	TaSMF-6B	113	12,612.47	4.56	−0.05	TraesCS6B02G180100.1
	AetSMF-7	87	9768.23	4.69	−0.167	AET7Gv21208500.4
	AetSMF-6	86	9632.97	4.42	−0.217	AET6Gv20390000.6
SMG	TuSMG-2	80	8933.39	8.06	−0.226	TuG1812G0200001854.01.P01
	TdSMG-2B	83	9151.47	4.37	0.2	TRIDC2BG026270.2
	TdSMG-2A	80	8933.39	8.06	−0.226	TRIDC2AG021760.1
	TaSMG-2B	81	9121.44	6.55	−0.249	TraesCS2B02G198200.1
	TaSMG-2A	79	8818.08	5.67	−0.234	TraesCS2A02G171800.1
	AetSMG-2	80	8933.39	8.06	−0.226	AET2Gv20357300.1

Note: AA, amino acid sequence length; MW, molecular weight; pI, isoelectric point; Splicing, alternative splicing number; GRAVY—grand average of hydropathicity.

The average length of the SSM/SLSM protein of wheat was 113 amino acids, while that of wild emmer wheat was 126.56, 126.76 in *T. urartu*, and 119.53 in *A. tauschii*. Accordingly, the molecular weight (Mw) ranged from 8.82 to 28.04 kDa, 8.93 to 27.9 kDa, 7.41 to 27.97 kDa, and 8.93 to 28.04 kDa for wheat, *T. urartu*, *T. dicoccoides*, and *A. tauschii*, respectively. Compared to *T. dicoccoides* (from 4.27 to 11.5 with an average of 8.29), *A. tauschii* (4.42 to 11.53 with an average of 7.82), and *T. urartu* (4.42 to 11.3 with an average of 8.12), the predicted isoelectric points (pI) of the wheat SSM/SLSM proteins varied from 4.42 to 11.3 with an average of 7.87, suggesting no significant difference occurred among them.

In wheat subgenomes, the similarity of nucleic acid sequence in the CDS region of LSM2 gene was between 98.6% and 99.3%, the similarity of nucleic acid sequence in CDS region of the LSM3 gene was between 83.3% and 99.7%, the sequence similarity of LSM4 was between 96.1% and 97.5%, the sequence similarity of LSM5 was between 97.4% and 99.3%, the sequence similarity of LSM6 was between 94.7% and 97.37%. Similarly, the sequence similarity distribution range of LSM7, LSM8, SMB, SMD1, SMD2, SMD3, SME, SMF and SMG genes was between 77.5% and 99.01%; between 96.3% and 99.3%; between 97.6% and 98.2%; between 87.1% and 99.4%; between 87.1% and 100%; between 97.30% and 97.8%, between 85.7% and 100%; between 97.7% and 97.77%; between 95.9% and 95.9% (Table 2).

Table 2. The similarity of nucleic acid sequence in CDS region of SSM/SLSM genes in wheat.

Gene Name	Wheat Minimum Similarity (%)	Wheat Maximum Similarity (%)
LSM2	98.582	99.291
LSM3	83.274	99.66
LSM4	96.065	97.454
LSM5	97.407	99.259
LSM6	94.667	97.333
LSM7	77.5	99.01
LSM8	96.333	99.333
SMB	97.619	98.246
SMD1	87.126	99.383
SMD2	90.034	99.074
SMD3	97.304	97.794
SME	85.992	100
SMF	97.661	97.727
SMG	95.935	95.935

Wheat minimum similarity (%): The minimum similarity between different nucleic acid sequences in CDS region of the same gene in wheat by MEGAX. Wheat maximum similarity (%): The maximum similarity between different nucleic acid sequences in CDS region of the same gene in wheat by MEGAX.

3.2. Genome Distribution and Synthetic Analysis of SSM/SLSM Genes among Bread Wheat and Its Progenitors

In wheat and its progenitors, the LSM2 genes were distributed on chromosome 7, the LSM3 genes were distributed on chromosomes 3 and 1, the LSM4 genes were distributed on chromosome 3, the LSM5 genes were distributed on chromosome 1, the LSM6 genes were distributed on chromosome 2, the LSM7 genes were distributed on chromosomes 7 and 2, and the LSM8 genes were distributed on chromosome 1. SMBs were distributed on chromosome 2; SMD1s were on chromosomes 2 and 6; SMD2s were on chromosomes 5, 3, and 4; SMD3s were on chromosomes 5, 7, and 4; SMEs were on chromosomes 6 and 7; SMFs were on chromosomes 6 and 7; and SMGs were on chromosome 2. We revealed that the chromosome distribution of the LSM gene was mainly concentrated on single chromosome (except for LSM7 and LSM3), while the chromosome distribution of the SM gene was relatively scattered, with only SMB and SMG genes distributed on single chromosome, and the rest were distributed on two or three chromosomes.

The synteny analysis of the SSM/SLSM genes among bread wheat and its progenitors was performed (Figure 1). A total of 94 pairs of orthologous genes were identified in bread wheat and wild emmer wheat. Similarly, pairs of genes in bread wheat and *A. tauschii*,

bread wheat and *T. urartu*, and *T. dicoccoides* and *T. urartu*, respectively, were 42 pairs, 38 pairs, and 48 pairs, and both were identified in the above.

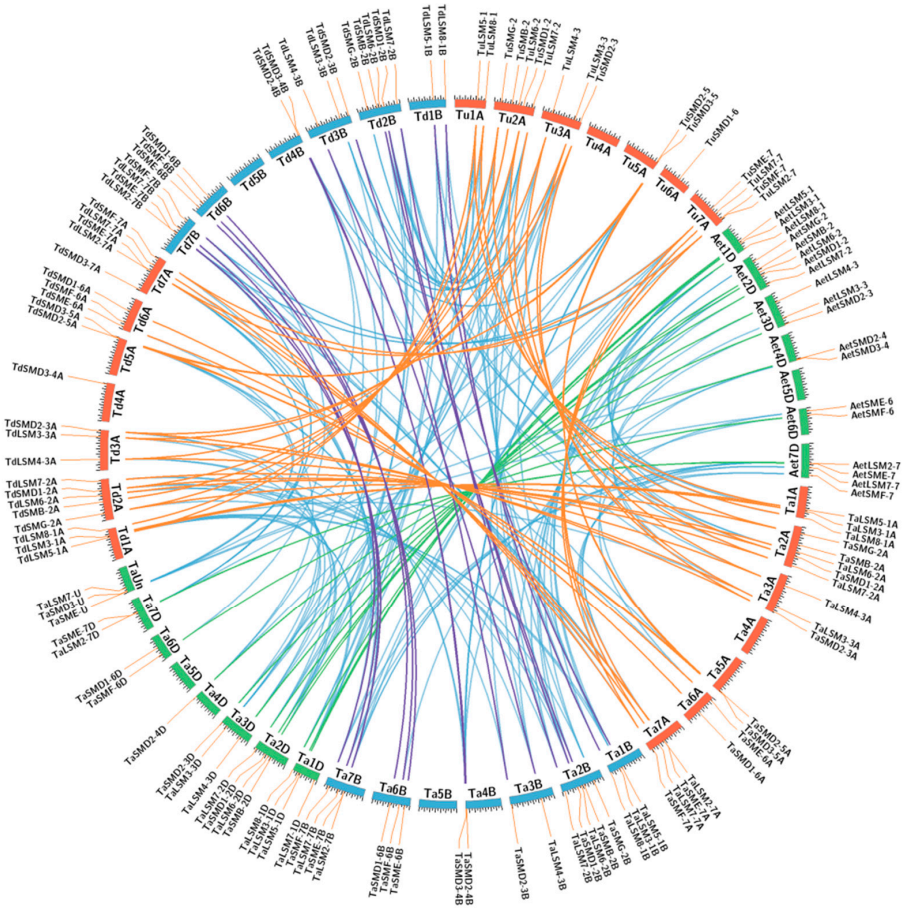

Figure 1. The chromosomal distribution and collinearity analysis of SSM/SLSM genes among bread wheat and relatives. The three subgenomes of bread wheat and its relatives chromosomes are represented by different colors. The SSM/SLSM genes are labeled according to their positions on the chromosomes. The A subgenome chromosome is shown in red, and homeoalleles of the SSM/SLSM gene links are shown in orange. The B subgenome chromosome is shown in blue, and homeoalleles of the SSM/SLSM gene links are shown in purple. The D subgenome chromosome is shown in green, and homeoalleles of the SSM/SLSM gene links are shown in green. Homeoalleles of the SSM/SLSM gene are linked by lines in blue, which represents the correspondence between different subgenomes.

Most characterized domestication events are associated with primitive extreme genetic mutations and selection pressures. These factors are predicted to increase the relative rate of nonsynonymous to synonymous (dN/dS) substitution, potentially resulting in the fixation of deleterious alleles. Therefore, in order to observe the gene changes during the process of wheat polyploidy, we calculated the dN/dS rate, which is displayed in Figure 2. We found that, in bread wheat and wild emmer wheat, the dN/dS of LSM2, LSM4, LSM6, LSM7, LSM8, and SMB was much greater than 1, so these genes were positively selected by environmental pressures, while other genes were purified and selected (dN/dS << 1). In bread wheat and *T. urartu*, for the LSM7 genes dN/dS >> 1, and for all the other genes dN/dS << 1. In bread wheat and *A. tauschii*, the LSM4 and LSM6 were positively selected, while others were purified and selected. In *T. dicoccoides* and *T. urartu*, the LSM7 and SMD1 were positively selected, others were purified and selected. In general, the SMD1 gene was positively selected only in the process of polyploidy from *T. urartuto* wild emmer wheat under the environmental influence. The LSM7 gene was positively selected only in the process of polyploidy from *T. urartu* to wild emmer wheat and then to bread

wheat. From the result of dN/dS, we found that most SM genes were purely selected during polyploidization except in the first stage of polyploidization SMD1 were positively selected and in the second stage of polyploidization SMB were positively selected, while spliceosomal LSM genes mostly undergo positive selection except for LSM3 and LSM7 genes. The evolutionary tree constructed for SMD2 gene is shown in Figure 3. Compared with TaSMD2-3D, TaSMD2-3B and TaSMD2-3A, TaSMD2-5A, the evolutionary relationship between TaSMD2-4D and TaSMD2-4B is closer.

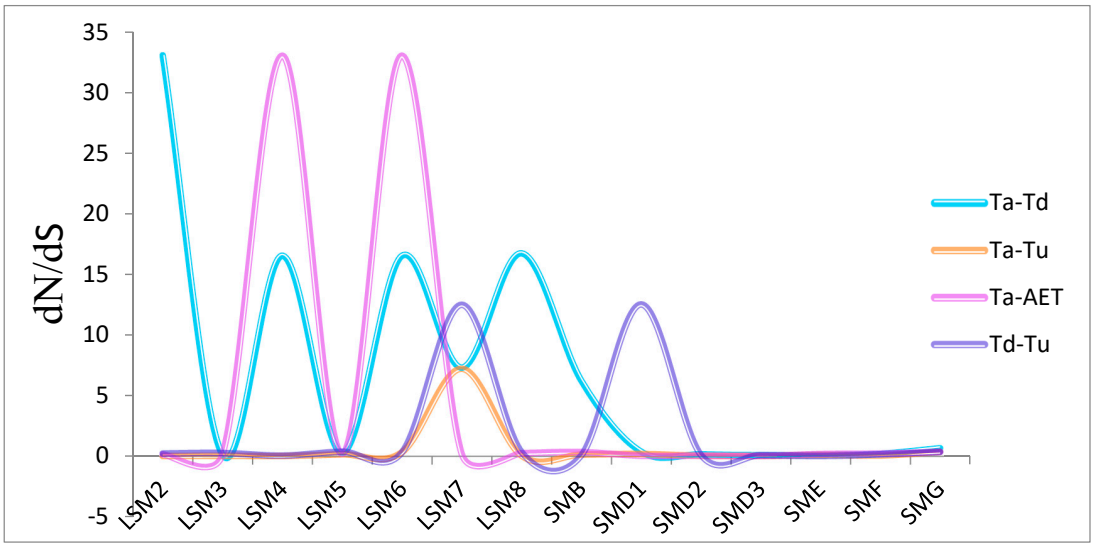

Figure 2. Non-synonymous to synonymous (dN/dS). The dN/dS ratios for orthologous SSM/SLSM genes between wheat and its relatives. The blue line represents the dN/dS ratios between SSM/SLSM genes of wheat and *Triticum dicoccoides*. The orange line represents the dN/dS ratios between SSM/SLSM genes of wheat and *Triticum urartu*. The pink line represents the dN/dS ratios between SSM/SLSM genes of wheat and *Aegilops tauschii*. The purple line represents the dN/dS ratios between SSM/SLSM genes of *Triticum dicoccoides* and *Aegilops tauschii*.

3.3. The Phylogenetic and Gene Structure of SSM/SLSM Proteins in Wheat and Its Progenitors

To further understand the evolutionary relationships of SSM/SLSM genes, their structural features and phylogenetic characters were analyzed. The phylogenetic tree was constructed using the full-length protein sequence alignments of the identified 57 TaSSM/SLSM, 41 TdSSM/SLSM, 17 TuSSM/SLSM, and 19 AetSSM/SLSM (Figure 4). The neighbor-joining (NJ) tree of SSM/SLSM genes can be clearly divided into 14 known groups. By contrast, within each SSM/SLSM gene, a strong amino acid sequence conservation was found, suggesting strong evolutionary relationships among all the members.

Moreover, other evidence, such as motif compositions and gene structure as described below, additionally support the truth. Form the NJ tree, we found that the pair-wise relationships between LSM and SM genes are as follows: LSM3-SMD2, LSM4-SMD3, LSM6-SMF, LSM7-SMG, and LSM8-SMB.

The SSM/SLSM structure was analyzed based on the arrangement of their exons (Figure 5). The number and distribution of exons in each SSM/SLSM gene between wheat and its ancestor genes were very similar. The SMB genes have only one exon in wheat and its progenitors, whereas the SMD1 genes have three or four exons in wheat, it has four or five exons in *T. urartu*, it has two or four exons in *T. dicoccoides*, and it has three exons in *A. tauschii*. SMD2 genes in wheat and its ancestors have four exons except TdSMD2-4B, TdSMD2-3A and TaSMD2-3B. In wheat and *T. urartu* the SMD3 genes have four exons, it has three, four or five exons in *T. dicoccoides*, and it has three exons in *A. tauschii*. In wheat the SME genes have four or six exons, it has six exons in *T. urartu*, it has three or four or six exons in *T. dicoccoides*, and it has five or six exons in *A. tauschii*. SMF genes in wheat and its ancestors have five exons except TdSMF-6A, AetSMF-6, TaSMF-6B and TaSMF-6D. SMG genes in wheat and its ancestors have four exons except TdSMG-2B.

Figure 3. Unrooted phylogenetic tree representing the relationships among the SMD2 gene in wheat and its progenitors. Purple represents the SMD2 gene on chromosomes 5 and 4, and orange represents the SMD2 gene on chromosome 3.

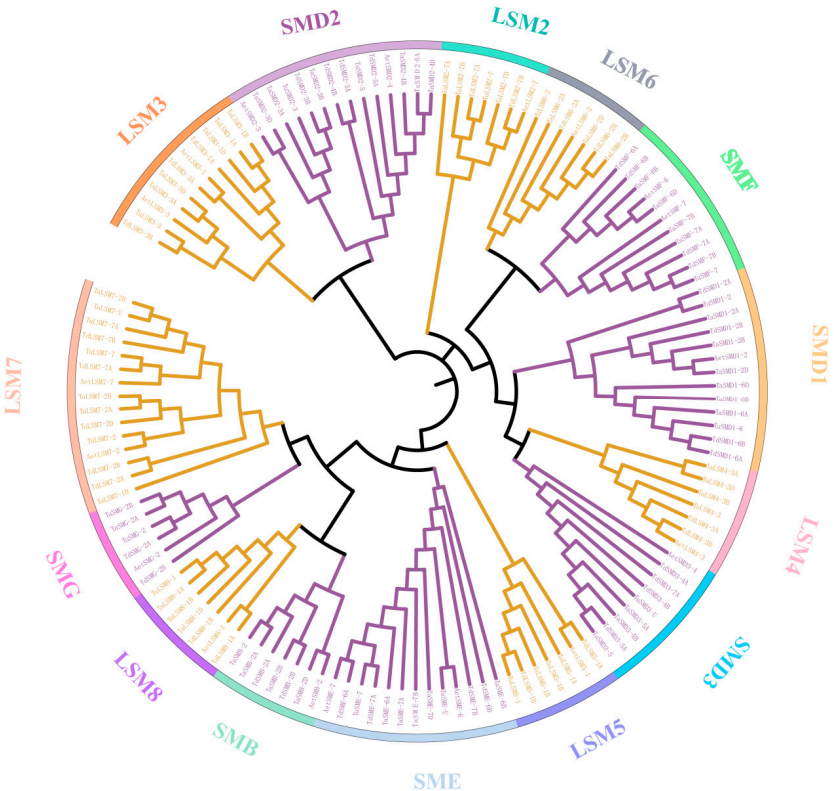

Figure 4. Phylogenetic analysis of SSM/SLSM homolog proteins from wheat and its relatives. The tree was generated with MEGAX using the neighbor-joining (NJ) algorithm. The SLSM and SSM genes are shown in different colors. The SLSM gene branch is shown in purple, and the SSM gene is shown in yellow. Two major genes are distinguished with 14 subcategories.

Figure 5. Phylogenetic relationships (**A**), motif compositions (**B**), gene structure (**C**). The maximum likelihood (ML) tree was constructed using MEGAX based on the full-length protein sequence. The exon–intron structures of these genes were graphically displayed by the Gene Structure Display Server using the CDS and genome sequence of SSM/LSM genes. The protein sequences of SSM/LSM genes were used to predict the conserved motifs using the MEME Suite web server.

In wheat, *A. tauschii* and *T. urartu* the LSM2 gene has three exons, it has four exons in *T. dicoccoides*. LSM3 gene in wheat and its ancestors has three exons except TdLSM3-3A.

LSM4 gene in wheat and its ancestors has seven exons except TdLSM4-3B. LSM5 gene in wheat and its ancestors has four exons except TdLSM5-1A and AetLSM5-1. The LSM6 gene only has three exons in wheat and its progenitors. LSM7 gene in wheat and its ancestors has five exons except TdLSM7-7B, TdLSM7-2B, TdLSM7-2A and TaLSM7-U. LSM8 gene in wheat and its ancestors has five exons except AetLSM8-1.

3.4. Analysis of Cis-Regulatory Elements in the Promoter Regions of TaSSM/SLSMs

To better understand the transcriptional regulation and potential function of the SSM/SLSM genes in wheat and its ancestors, the cis-regulatory elements were investigated using the PlantCARE database [46] (Supplemental Table S2). We found light-response elements Sp1 (GGGCGG), a meristem expression CAT-box (GCCACT), a root regulation cis-element G-box (CACGTC), jasmonic acid-responsive elements (TGACG-motif and CGTCA-motif), gibberellin responsive elements GARE-motif (TCTGTTG) and P-box (CCTTTTG), abscisic acid-responsive elements (ABRE) (ACGTG), auxin-responsive cis-elements AuxRR-core (GGTCCAT) and TGA-element (CCATCTTTTT), salicylic acid-responsive TCA-element (CCATCTTTTT and TCAGAAGAGG) heat-responsive elements (AAAAAATTTC), anoxic specific inducibility element GC-motif (A/CGCCGCGCA), drought-inducibility cis-element MBS (CAACTG), and low-temperature responsive cis-element LTR (CCGAAA). Among these cis-elements, the G-box, CGTCA-motif, TGACG-motif, and ABRE were found in most cases [47]. The presence of a significant number of cis-regulatory elements suggests strong participation of the SSM/SLSM genes in wheat and its progenitors' development and hormone-response.

3.5. Expression Profile Analysis of SSM/SLSM Genes in Wheat

We used high-throughput RNA-seq data to analyze the expression patterns in different wheat tissues, developing periods and various stress. We detected the spatial and temporal specific expression patterns of SSM/SLSM genes in wheat, and in general, most members of wheat SSM/SLSM genes were expressed in lower levels in the leaf and higher levels in the spike, implying that SSM/SLSM genes may prefer to be involved in spike development of the spike in bread wheat. The calculation of wheat germination after a week of SSM/SLSM gene expression in different tissues found that the LSM2 and LSM4 expression in the stem was generally high. For genes in the spike tissue, the SMD1, SMD2, and SME expression were high, the other SMG expression was relatively high in the root. In the grain, the LSM5 and LSM7 expression were relatively high and the expression of these genes in the leaf were relatively low.

In general, the expression of the LSM4, SMD2, SME, SMF, and SMG genes was relatively high one week after germination among abiotic stress treatments. The expression levels of these genes in leaves at different stages were calculated, and we found that the expression levels of SME, SMD1, SMD2, SMD3, and LSM5 were generally high at the grain-filling stage, and the expression levels of LSM4 and SMB were high at the flowering stage. In addition, the expression levels of LSM6, LSM7, and SMG were relatively high at the booting stage, while the expression levels of these genes were relatively low at the heading stage.

The RNA-seq data confirmed previous indications that abiotic stress markedly alters alternative splicing in plants [48–55], and altered alternative splicing is regulated by the spliceosome, and thus, the expression pattern of the core protein of the spliceosome may change under the control of stress. Therefore, we also analyzed the expression of these genes under different abiotic stresses (Figure 6). TaSME-6B and TaLSM7-1D displayed no expression levels when exposed to any level of any treatment. In general, compared with the contrast, most SSM/SLSM genes were expressed at lower levels.

Base on the results, we selected eight genes (SMD1-6D, SMD1-2D, SMD1-2A, SMD2-5A, SMF-6B, LSM5-1D, LSM7-7A, LSM8-1A) to further validate their expression profiles by qPCR (Figure 7). We found the SMF-6B, LSM7-7A, SMD1-2A, and LSM8-1A in salt and

drought stresses were highly expressed, which indicated that such genes may play more important roles in wheat drought and salt resistance.

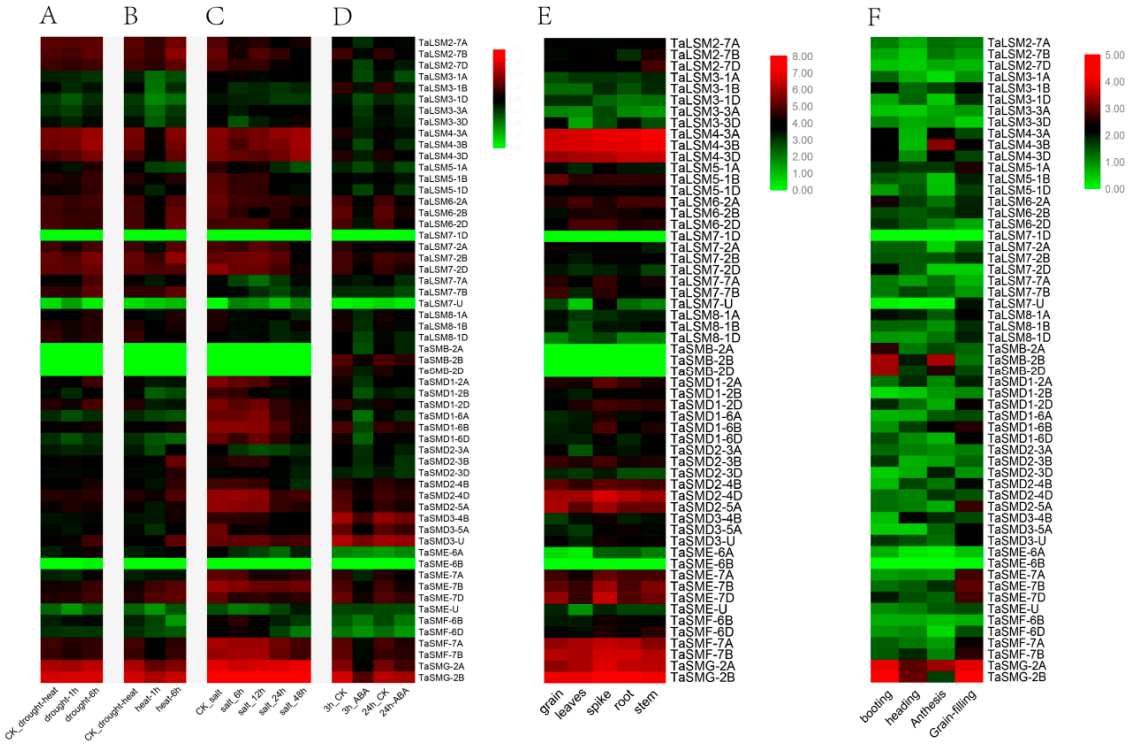

Figure 6. The expression profiles of SSM/SLSM in four stress treatments, five different tissues, and four different development stages in bread wheat. The expression profile in (**A**–**D**) for four different stress, respectively, including drought, hot, salt and ABA. The expression profile in (**E**) for five different tissues, and (**F**) four different development stages.

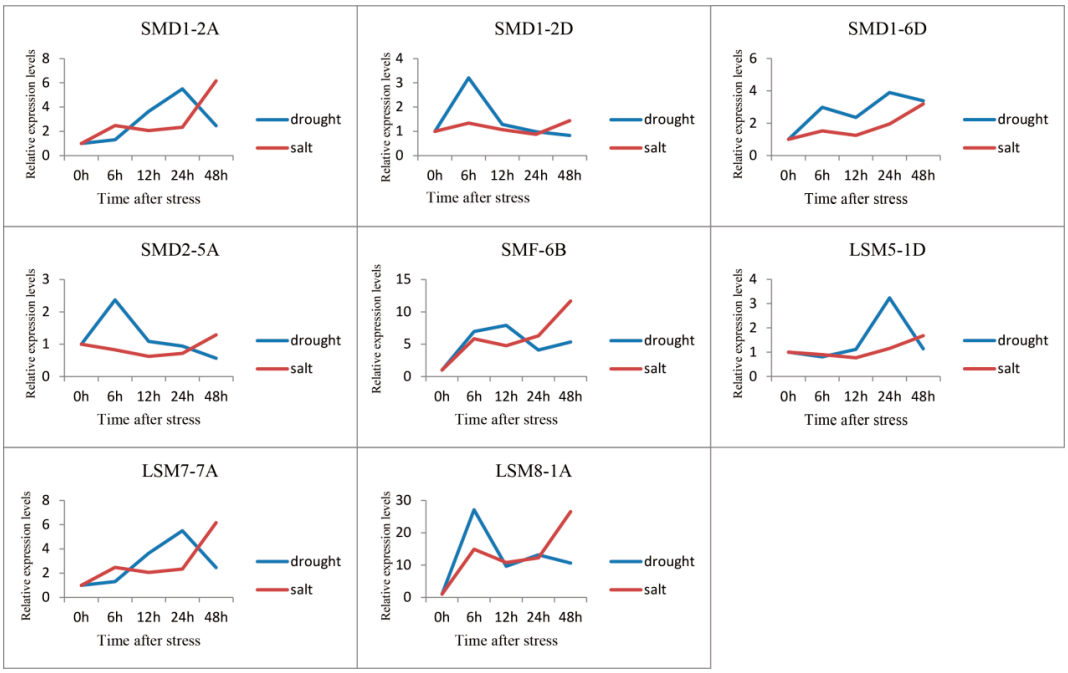

Figure 7. Real-time quantitative polymerase chain reaction (qRT-PCR) analysis of 8 TaSSM/SLSM genes. Expression levels of eight TaSSM/SLSM genes were analyzed using qPCR in the leaves. The pressure treatment was divided into 6, 12, 24, and 48 h, and the levels at 0 h are presented as the relative ratios. The experiments were repeated three times. The red line shows the gene expression under salt stress. The blue line shows the gene expression under drought stress.

3.6. Co-Expression Network Analysis of Wheat SSM/SLSM Genes

We constructed a gene co-expression network based on the wheat transcriptome in different tissues and developing periods. These SSM/SLSM genes were widely distributed in different modules without any preference. Furthermore, we found that LSM7-7A gene might be related to salt stress and drought stress. Subsequently, we screened LSM7-7A related genes in this network (Figure 8), and annotated these genes (Figure 9), showing that 47 genes related to the TaLSM7-7A genes were related to abiotic stress. Among the TaLSM7-7A related genes, TraesCS6D02G310800, highly homologous with the RCY1 gene of *Arabidopsis*, was found to be related with salt stress [55].

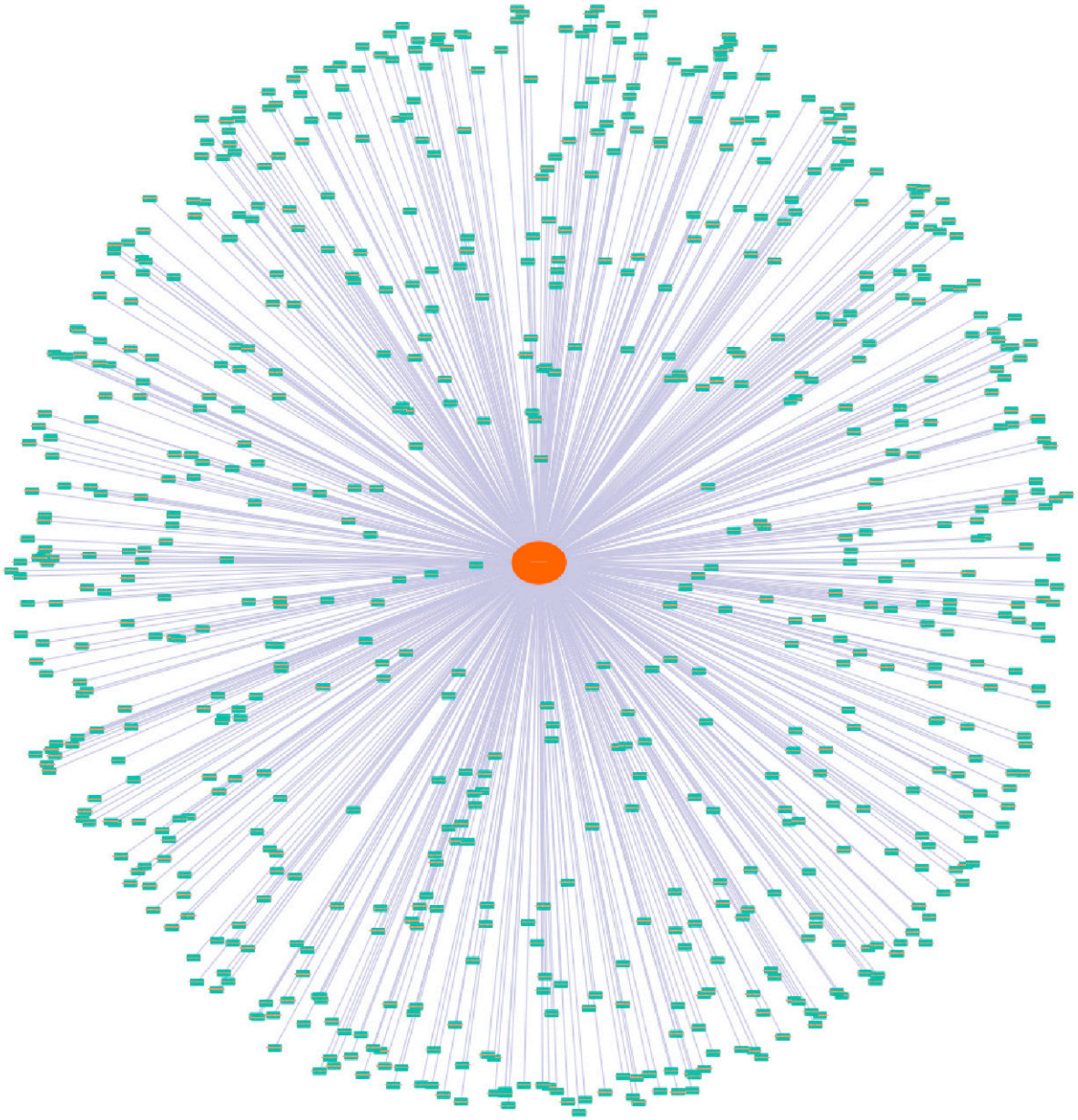

Figure 8. The LSM7-7A related genes in the co-expression network. The orange dots represent the LSM7-7A gene, and the green rectangles represent the genes associated with LSM7-7A.

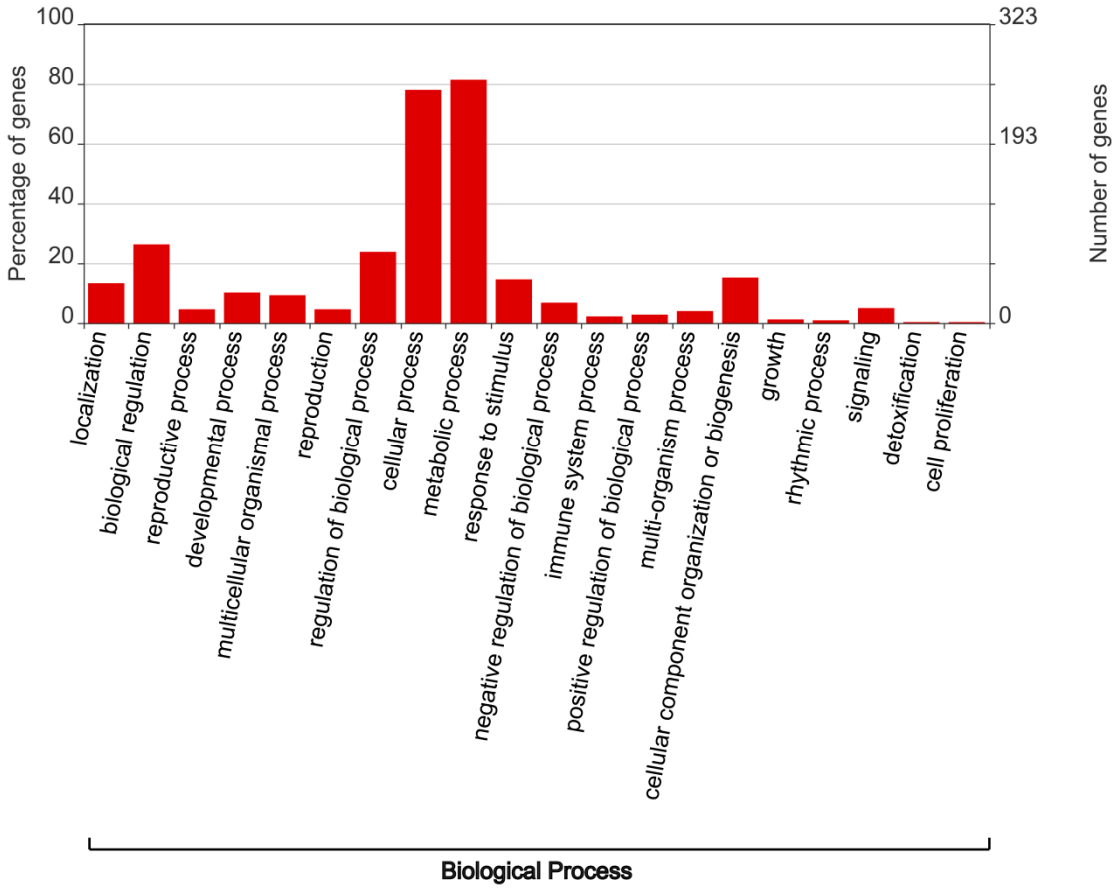

Figure 9. Go enrichment of genes related to LSM7 in the co-expression network. The figure illustrates the biological process enrichment of LSM7-related genes.

4. Discussion

In this study, we systematically identified 17 SSM/SLSM genes in diploid *T. urartu*, 19 in diploid *A. tauschii*, 41 in tetraploid emmer wheat (*T. dicoccoides*), and 57 in hexaploid wheat (*T. aestivum*) at the whole-genome scale. There was no multiplier increase of SSM/SLSM genes' number during wheat polyploidy, indicating that they may undergo gene loss [56] and recombination between homologous chromosomes [57] during the polyploidy process.

The wheat SSM/SLSM genes were mainly distributed on all chromosomes except for 4A, 5D, and 5B. There were no SSM/SLSM genes on the chromosome 4A of *T. urartu*. None of these genes were identified on the chromosome 5 of *Ae. tauschii* and the 5B of *T. dicoccoides*. There was no any TdSMD3-4A gene expression evidence by the subsequent RNA-seq data analysis, which may indicate a pseudo gene. From the NJ tree, we found that the pair-wise relationships between the LSM and SM genes were as follows: LSM3-SMD2, LSM4-SMD3, LSM6-SMF, LSM7-SMG, and LSM8-SMB. This was similar to Veretnik's results [9,16]. Based on the physical and chemical properties of these genes, the SSM/SLSM protein sequence lengths of bread wheat were shorter than those of its relatives, and their theoretical pI were smaller than the other relatives.

Chromosome translocation plays an important role in wheat breeding [58]. The 4AL-5AL-7BS chromosome translocation model exists naturally in wheat and was obtained after two translocations. However, the specific function of 4AL-5AL-7BS chromosome regions remains unclear. The TaSMD2-5A gene on chromosome 5A was closely related to the TaSMD2-4D gene on chromosome 4D and the TaSMD2-4B gene on chromosome 4B, compared with other SMD2 genes in wheat and its relatives. The sequence similarity of the SMD2 gene in wheat was studied, where the sequence similarity between the TaSMD2-5A

and TaSMD2-4B gene was 99.1%, 98.5% between TaSMD2-5A and TaSMD2-4D, which was higher than others. This result suggested that the SMD2 gene may be translocated between chromosomes 5A and 4A, which was consistent with Zhou's study [59]. We found that TaSMD2-5A was translocated to chr5A during the first translocation event [60]. The discovery of the TaSMD2-5A gene location contributes to the understanding of natural translocation chromosome models in wheat breeding.

The ratio of non-synonymous to synonymous substitution rates (dN/dS) can represent the evolutionary relationship under selection pressure, which is commonly used to identify protein sites that experience purifying selection (dN/dS < 1), evolve neutrally (dN/dS ≈ 1), or experience (dN/dS > 1) [61–64]. The trend diagram suggests that the LSM/SM gene with dN/dS value greater than one occurred with positive selection in the polyploidy of bread wheat. From the result of dN/dS, we found that most of the SM genes were purely selected during polyploidization, except in the first stage of polyploidization SMD1 were positively selected and in the second stage of polyploidization SMB were positively selected, while spliceosomal LSM genes mostly undergo positive selection except for LSM3 and LSM7 genes.

There is evidence that knockout SMD3 delayed flowering time and completion of the life cycle [65]. There is evidence that SMD1 relates to the formation of giant cells, is required for successful nematode infection, and facilitates posttranscriptional gene silencing (PTGS), SMD1 mutants are embryo-lethal [66,67]. Therefore, SMD1 and SMD3 may be related to spike development in bread wheat. The functions of other genes in plants are still unclear and require further verification in the later stage.

According to the qPCR validation of the selected genes, we found that the TaSMF-6B, TaLSM7-7A, TaSMD1-2A, and TaLSM8-1A in salt and drought stress were highly expressed.

Cis-regulatory elements are composed of DNA (typically, non-coding DNA) containing binding sites for TFs and/or other regulatory molecules that are needed to activate and sustain transcription. Zhang indicated that ABREs are also determinant cis-elements for stress-related transcription regulations. TaSMF-6B, TaLSM7-7A, TaSMD1-2A, and TaLSM8-1A all contain ABREs. Meanwhile TaSMD1-2A and TaLSM7-7A have the drought-inducibility cis-element MBS. Furthermore, we found that the TaLSM7-7A gene was associated with the salt stress gene (ATRCY1) by gene co-expression network analysis. The TaLSM7-7A gene was speculated to be the salt tolerance and drought-resistance gene.

5. Conclusions

In summary, we identified systematically the members of the SSM/SLSM gene family in bread wheat and its three progenitors, and analyzed and compared their characters. A good collinearity of SSM/SLSM genes was found among bread wheat and its progenitors, and evolutionary pressures on genes were further investigated based on dN/dS analysis of the orthologous pairs. Based on the RNA-seq data, the SSM/SLSM genes exhibited distinct tissue-specific expression patterns, and LSM7, LSM8, SMF, and SMD1 were induced by diverse abiotic stresses. This investigation provided comprehensive SSM/SLSM genes in wheat and its progenitors for further functional analysis, and contributes to a better understanding of the evolution mechanism of the SSM/SLSM genes during wheat polyploidy process.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/agronomy11071429/s1, Supplemental Table S1: Details of the cis-acting element in the promoter, Supplemental Table S2: The accession numbers and sample information of the RNA-seq data, Supplemental Table S3: Primers were designed for qPCR experiments, Supplemental Figure S1: Melt curve of eight genes at qPCR, Supplemental Figure S2: The diagram showing the whole steps of the materials and methods part.

Author Contributions: Conceptualization, W.T.; Formal analysis, R.G. and N.S.; Resources, W.T. and W.S.; validation, W.P.; Data curation, Q.B., Z.L. and X.N.; Writing-original draft preparation, R.G. and N.S.; Writing-review and editing, W.T. and X.N.; Supervision, W.T. and W.S.; Funding acquisition, W.T. and W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was mainly funded by the Modern Crop and Seed Project of Shaanxi Province Department of Agriculture in 2017 (Grant number: 20171010000004), Chinese Universities Scientific Fund (2014ZZ010, 2452015121) and the Young Faculty Initiation Fund from Shaanxi Province. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: This study did not involve humans or animals.

Informed Consent Statement: This study did not involve humans.

Data Availability Statement: All the related sequence data in this study were downloaded from public database, detailed in Material and methods.

Acknowledgments: We appreciated the High-Performance Computing center of Northwest A&F University for providing computational resources.

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

References

1. Wahl, M.C.; Will, C.L.; Luhrmann, R. The spliceosome: Design principles of a dynamic RNP machine. *Cell* **2009**, *136*, 701–718. [CrossRef]
2. Nilsen, T.W. The spliceosome: The most complex macromolecular machine in the cell? *Bioessays* **2003**, *25*, 1147–1149. [CrossRef]
3. Reeves, W.H.; Narain, S.; Satoh, M. Henry Kunkel, Stephanie Smith, clinical immunology, and split genes. *Lupus* **2003**, *12*, 213–217. [CrossRef]
4. Urlaub, H.; Raker, V.A.; Kostka, S.; Luhrmann, R. Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.* **2001**, *20*, 187–196. [CrossRef]
5. Will, C.L.; Luhrmann, R. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **2001**, *13*, 290–301. [CrossRef]
6. Hackl, W.; Fischer, U.; Luhrmann, R. A 69-Kd Protein That Associates Reversibly with the Sm Core Domain of Several Spliceosomal Snrnp Species. *J. Cell Biol.* **1994**, *124*, 261–272. [CrossRef] [PubMed]
7. He, W.; Parker, R. Functions of Lsm proteins in mRNA degradation and splicing. *Curr. Opin. Cell Biol.* **2000**, *12*, 346–350. [CrossRef]
8. Ru, Y.; Wang, B.B.; Brendel, V. Spliceosomal Proteins in Plants. *Curr. Top. Microbiol.* **2008**, *326*, 1–15.
9. Veretnik, S.; Wills, C.; Youkharibache, P.; Valas, R.E.; Bourne, P.E. Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome. *PLoS Comput. Biol.* **2009**, *5*, e1000315. [CrossRef]
10. Golisz, A.; Sikorski, P.J.; Kruszka, K.; Kufel, J. Arabidopsis thaliana LSM proteins function in mRNA splicing and degradation. *Nucleic Acids Res.* **2013**, *41*, 6232–6249. [CrossRef] [PubMed]
11. Seraphin, B. Sm and Sm-Like Proteins Belong to a Large Family-Identification of Proteins of the U6 as Well as the U1, U2, U4 and U5 Snrnps. *EMBO J.* **1995**, *14*, 2089–2098. [CrossRef]
12. Cooper, M.; Johnston, L.H.; Beggs, J.D. Identification and characterization of Uss1p (Sdb23p): A novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J.* **1995**, *14*, 2066–2075. [CrossRef]
13. Hermann, H.; Fabrizio, P.; Raker, V.A.; Foulaki, K.; Hornig, H.; Brahms, H.; Luhrmann, R. snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J.* **1995**, *14*, 2076–2088. [CrossRef]
14. Kambach, C.; Walke, S.; Young, R.; Avis, J.M.; de la Fortelle, E.; Raker, V.A.; Luhrmann, R.; Li, J.; Nagai, K. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **1999**, *96*, 375–387. [CrossRef]
15. Cao, J.; Shi, F.; Liu, X.; Jia, J.; Zeng, J.; Huang, G. Genome-wide identification and evolutionary analysis of Arabidopsis sm genes family. *J. Biomol. Struct. Dyn.* **2011**, *28*, 535–544. [CrossRef] [PubMed]
16. Li, X.; Chen, Y.; Zhang, S.T.; Su, L.Y.; Xu, X.P.; Chen, X.H.; Lai, Z.X.; Lin, Y.L. Genome-wide identification and expression analyses of Sm genes reveal their involvement in early somatic embryogenesis in Dimocarpus longan Lour. *PLoS ONE* **2020**, *15*, e0230795. [CrossRef] [PubMed]
17. Chen, Y.; Cao, J. Comparative genomic analysis of the Sm gene family in rice and maize. *Gene* **2014**, *539*, 238–249. [CrossRef]
18. Wang, B.B.; Brendel, V. The ASRG database: Identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biol.* **2004**, *5*, R102. [CrossRef] [PubMed]
19. Chen, L. Identification and Survey of Pre-mRNA Splicing-Related Proteins in 10 Plant Species. Master's Thesis, Iowa State University, Ames, IA, USA, 2011.
20. Sharpe, N.G.; Williams, D.G.; Norton, P.; Latchman, D.S. Expression of the SmB' splicing protein in rodent cells capable of following an alternative RNA splicing pathway. *FEBS Lett.* **1989**, *243*, 132–136. [CrossRef]
21. Anne, J. Arginine methylation of SmB is required for Drosophila germ cell development. *Development* **2010**, *137*, 2819–2828. [CrossRef]

22. Scruggs, B.S.; Michel, C.I.; Ory, D.S.; Schaffer, J.E. SmD3 Regulates Intronic Noncoding RNA Biogenesis. *Mol. Cell. Biol.* **2012**, *32*, 4092–4103. [CrossRef]
23. Li, Z.; Putzer, B.M. Spliceosomal protein E regulates neoplastic cell growth by modulating expression of Cyclin E/CDK2 and G2/M checkpoint proteins. *J. Cell. Mol. Med.* **2008**, *12*, 2427–2438. [CrossRef]
24. Mabonga, L.; Kappo, A.P. The oncogenic potential of small nuclear ribonucleoprotein polypeptide G: A comprehensive and perspective view. *Am. J. Transl. Res.* **2019**, *11*, 6702–6716.
25. Perez-Santangelo, S.; Mancini, E.; Francey, L.J.; Schlaen, R.G.; Chernomoretz, A.; Hogenesch, J.B.; Yanovsky, M.J. Role for LSM genes in the regulation of circadian rhythms. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15166–15171. [CrossRef] [PubMed]
26. Huertas, R.; Catala, R.; Jimenez-Gomez, J.M.; Castellano, M.M.; Crevillen, P.; Pineiro, M.; Jarillo, J.A.; Salinas, J. Arabidopsis SME1 Regulates Plant Development and Response to Abiotic Stress by Determining Spliceosome Activity Specificity. *Plant Cell* **2019**, *31*, 537–554. [CrossRef]
27. Cui, P.; Zhang, S.; Ding, F.; Ali, S.; Xiong, L. Dynamic regulation of genome-wide pre-mRNA splicing and stress tolerance by the Sm-like protein LSm5 in Arabidopsis. *Genome Biol.* **2014**, *15*, R1. [CrossRef]
28. Zhang, Z.L.; Zhang, S.P.; Zhang, Y.; Wang, X.; Li, D.; Li, Q.L.; Yue, M.H.; Li, Q.; Zhang, Y.E.; Xu, Y.Y.; et al. Arabidopsis Floral Initiator SKB1 Confers High Salt Tolerance by Regulating Transcription and Pre-mRNA Splicing through Altering Histone H4R3 and Small Nuclear Ribonucleoprotein LSM4 Methylation. *Plant Cell* **2011**, *23*, 396–411. [CrossRef]
29. Perea-Resa, C.; Carrasco-Lopez, C.; Catala, R.; Tureckova, V.; Novak, O.; Zhang, W.P.; Sieburth, L.; Jimenez-Gomez, J.M.; Salinas, J. The LSM1-7 Complex Differentially Regulates Arabidopsis Tolerance to Abiotic Stress Conditions by Promoting Selective mRNA Decapping. *Plant Cell* **2016**, *28*, 505–520. [CrossRef]
30. Carrasco-Lopez, C.; Hernandez-Verdeja, T.; Perea-Resa, C.; Abia, D.; Catala, R.; Salinas, J. Environment-dependent regulation of spliceosome activity by the LSM2-8 complex in Arabidopsis. *Nucleic Acids Res.* **2017**, *45*, 7416–7431. [CrossRef] [PubMed]
31. Matsuoka, Y. Evolution of Polyploid Triticum Wheats under Cultivation: The Role of Domestication, Natural Hybridization and Allopolyploid Speciation in their Diversification. *Plant Cell Physiol.* **2011**, *52*, 750–764. [CrossRef] [PubMed]
32. Feldman, M.; Levy, A.A. Genome Evolution Due to Allopolyploidization in Wheat. *Genetics* **2012**, *192*, 763–774. [CrossRef]
33. Pfeifer, M.; Kugler, K.G.; Sandve, S.R.; Zhan, B.; Rudi, H.; Hvidsten, T.R.; International Wheat Genome Sequencing Consortium; Mayer, K.F.; Olsen, O.A. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **2014**, *345*, 1250091. [CrossRef] [PubMed]
34. He, F.; Pasam, R.; Shi, F.; Kant, S.; Keeble-Gagnere, G.; Kay, P.; Forrest, K.; Fritz, A.; Hucl, P.; Wiebe, K.; et al. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat. Genet.* **2019**, *51*, 896–904. [CrossRef]
35. Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. HMMER web server: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W200–W204. [CrossRef]
36. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef]
37. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [CrossRef]
38. Zhang, Z.; Xiao, J.F.; Wu, J.Y.; Zhang, H.Y.; Liu, G.M.; Wang, X.M.; Dai, L. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **2012**, *419*, 779–781. [CrossRef]
39. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]
40. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef]
41. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **2001**, *25*, 402–408. [CrossRef]
42. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]
43. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]
44. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef] [PubMed]
45. Ye, J.; Zhang, Y.; Cui, H.; Liu, J.; Wu, Y.; Cheng, Y.; Xu, H.; Huang, X.; Li, S.; Zhou, A.; et al. WEGO 2.0: A web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* **2018**, *46*, W71–W75. [CrossRef]
46. Lescot, M.; Dehais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouze, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]
47. Chen, S.; Luo, Y.; Wang, G.; Feng, C.; Li, H. Genome-wide identification of expansin genes in Brachypodium distachyon and functional characterization of BdEXPA27. *Plant Sci.* **2020**, *296*, 110490. [CrossRef]
48. Ding, F.; Cui, P.; Wang, Z.; Zhang, S.; Ali, S.; Xiong, L. Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis. *BMC Genom.* **2014**, *15*, 431. [CrossRef]

49. Feng, J.; Li, J.; Gao, Z.; Lu, Y.; Yu, J.; Zheng, Q.; Yan, S.; Zhang, W.; He, H.; Ma, L.; et al. SKIP Confers Osmotic Tolerance during Salt Stress by Controlling Alternative Gene Splicing in Arabidopsis. *Mol. Plant* **2015**, *8*, 1038–1052. [CrossRef]
50. Zhan, X.Q.; Qian, B.L.; Cao, F.Q.; Wu, W.W.; Yang, L.; Guan, Q.M.; Gu, X.B.; Wang, P.C.; Okusolubo, T.A.; Dunn, S.L.; et al. An Arabidopsis PWI and RRM motif-containing protein is critical for pre-mRNA splicing and ABA responses. *Nat. Commun.* **2015**, *6*, 8139. [CrossRef]
51. Thatcher, S.R.; Danilevskaya, O.N.; Meng, X.; Beatty, M.; Zastrow-Hayes, G.; Harris, C.; Van Allen, B.; Habben, J.; Li, B.L. Genome-Wide Analysis of Alternative Splicing during Development and Drought Stress in Maize. *Plant Physiol.* **2016**, *170*, 586–599. [CrossRef]
52. Jiang, J.F.; Liu, X.N.; Liu, C.H.; Liu, G.T.; Li, S.H.; Wang, L.J. Integrating Omics and Alternative Splicing Reveals Insights into Grape Response to High Temperature. *Plant Physiol.* **2017**, *173*, 1502–1518. [CrossRef]
53. Keller, M.; Hu, Y.J.; Mesihovic, A.; Fragkostefanakis, S.; Schleiff, E.; Simm, S. Alternative splicing in tomato pollen in response to heat stress(aEuro). *DNA Res.* **2017**, *24*, 205–217. [CrossRef]
54. Zhu, F.Y.; Chen, M.X.; Ye, N.H.; Shi, L.; Ma, K.L.; Yang, J.F.; Cao, Y.Y.; Zhang, Y.J.; Yoshida, T.; Fernie, A.R.; et al. Proteogenomic analysis reveals alternative splicing and translation as part of the abscisic acid response in Arabidopsis seedlings. *Plant J.* **2017**, *91*, 518–533. [CrossRef]
55. Forment, J.; Naranjo, M.A.; Roldan, M.; Serrano, R.; Vicente, O. Expression of Arabidopsis SR-like splicing proteins confers salt tolerance to yeast and transgenic plants. *Plant J.* **2002**, *30*, 511–519. [CrossRef]
56. Shaked, H.; Kashkush, K.; Ozkan, H.; Feldman, M.; Levy, A.A. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* **2001**, *13*, 1749–1759. [CrossRef]
57. Szadkowski, E.; Eber, F.; Huteau, V.; Lode, M.; Huneau, C.; Belcram, H.; Coriton, O.; Manzanares-Dauleux, M.J.; Delourme, R.; King, G.J.; et al. The first meiosis of resynthesized Brassica napus, a genome blender. *New Phytol.* **2010**, *186*, 102–112. [CrossRef]
58. Wang, J.; Liu, Y.; Su, H.; Guo, X.; Han, F. Centromere structure and function analysis in wheat-rye translocation lines. *Plant J.* **2017**, *91*, 199–207. [CrossRef]
59. Zhou, C.; Dong, Z.; Zhang, T.; Wu, J.; Yu, S.; Zeng, Q.; Han, D.; Tong, W. Genome-Scale Analysis of Homologous Genes among Subgenomes of Bread Wheat (*Triticum aestivum* L.). *Int. J. Mol. Sci.* **2020**, *21*, 3015. [CrossRef]
60. Chen, Y.; Song, W.; Xie, X.; Wang, Z.; Guan, P.; Peng, H.; Jiao, Y.; Ni, Z.; Sun, Q.; Guo, W. A Collinearity-Incorporating Homology Inference Strategy for Connecting Emerging Assemblies in the Triticeae Tribe as a Pilot Practice in the Plant Pangenomic Era. *Mol. Plant* **2020**, *13*, 1694–1708. [CrossRef] [PubMed]
61. Spielman, S.J.; Wilke, C.O. The Relationship between dN/dS and Scaled Selection Coefficients. *Mol. Biol. Evol.* **2015**, *32*, 1097–1108. [CrossRef] [PubMed]
62. Mugal, C.F.; Wolf, J.B.; Kaj, I. Why time matters: Codon evolution and the temporal dynamics of dN/dS. *Mol. Biol. Evol.* **2014**, *31*, 212–231. [CrossRef] [PubMed]
63. Ratnakumar, A.; Mousset, S.; Glemin, S.; Berglund, J.; Galtier, N.; Duret, L.; Webster, M.T. Detecting positive selection within genomes: The problem of biased gene conversion. *Philos. Trans. R. Soc. B* **2010**, *365*, 2571–2580. [CrossRef]
64. Yang, Z.; Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **2000**, *17*, 32–43. [CrossRef] [PubMed]
65. Swaraz, A.M.; Park, Y.D.; Hur, Y. Knock-out mutations of Arabidopsis SmD3-b induce pleotropic phenotypes through altered transcript splicing. *Plant Sci.* **2011**, *180*, 661–671. [CrossRef] [PubMed]
66. Elvira-Matelot, E.; Bardou, F.; Ariel, F.; Jauvion, V.; Bouteiller, N.; Le Masson, I.; Cao, J.; Crespi, M.D.; Vaucheret, H. The Nuclear Ribonucleoprotein SmD1 Interplays with Splicing, RNA Quality Control, and Posttranscriptional Gene Silencing in Arabidopsis. *Plant Cell* **2016**, *28*, 426–438. [CrossRef] [PubMed]
67. Mejias, J.; Bazin, J.; Truong, N.M.; Chen, Y.; Marteu, N.; Bouteiller, N.; Sawa, S.; Crespi, M.D.; Vaucheret, H.; Abad, P.; et al. The root-knot nematode effector MiEFF18 interacts with the plant core spliceosomal protein SmD1 required for giant cell formation. *New Phytol.* **2020**. [CrossRef] [PubMed]