Genome-wide identification of MATH-BTB gene family and expression analysis in rye (Secale cereale L.)

Zhenbo Zhai
Hebei Normal University of Science and Technology
Yonghe Che
Hebei Normal University of Science and Technology
Yunjie Yang
Hebei Normal University of Science and Technology
Xirui Wei
Hebei Normal University of Science and Technology
Yanping Yang
Hebei Normal University of Science and Technology
Lai Wei
txameupx@163.com

Hebei Normal University of Science and Technology

Research Article

Keywords: rye, MATH-BTB gene family, bioinformatics analysis, grain width, kASP marker

Posted Date: November 12th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7922703/v1

Additional Declarations: No competing interests reported.

Abstract

Proteins encoded by the *MATH-BTB* gene family regulate plant grain development by mediating the degradation of transcription factors. This study identified the *MATH-BTB* family in rye through a homology-based search and performed a bioinformatic analysis to investigate its functions. We identified 56 *MATH-BTB* genes in the rye genome. Phylogenetic analysis classified these genes into 10 distinct subfamilies. Synteny analysis revealedno collinear genes with *Arabidopsis* or maize, one with rice, and 16 with the A subgenome of wheat. Cluster analysis of expression data revealed that this gene family is predominantly expressed in spikes and grains, suggesting its potential role in grain development. Integration with our previous findings on yield-related traits pinpointed a specific *MATH-BTB* gene, *ScWN4R01G352500.1*, which was previously associated with grain width.. This strengthens the hypothesis that this gene family plays a role in grain development. Furthermore, we developed a Kompetitive Allele-Specific PCR (KASP) marker that effectively genotypes the grain width trait. Collectively, these results provide a valuable resource for future functional studies of the *MATH-BTB* family in rye and offer a directly applicable marker for molecular breeding programs aimed at improving grain width.

Introduction

Grain yield, a primary objective in cereal breeding, is governed by a suite of traits, including panicle size, grain dimensions (length and width), and thousand-grain weight (Kumar et al. 2016; Wang et al. 2019; Guan et al. 2020). Numerous genes controlling these traits have been characterized across cereals. A prominent example is the *GS3* gene in rice (*Oryza sativa* L.), a major quantitative trait locus (QTL) for grain length and weight, whose functional homologs in wheat (*Triticum aestivum* L.) and maize (*Zea mays* L.) modulate similar traits (Fan et al. 2006; Mao et al. 2010; Li et al. 2010; Zhang et al. 2014). Likewise, loss-of-function mutations in *GW2* and *GW5* increase grain weight (Song et al. 2007; Shomura et al. 2008; Huang et al. 2013), whereas *GS5*, which encodes a serine carboxypeptidase, acts as a positive regulator of grain size (Li et al. 2011).

The *MATH-BTB* proteins constitute a subfamily of the BTB (Broad-complex, Tramtrack, and Bric-à-brac) family, characterized by an N-terminal MATH (Meprin and TRAF homology) domain and a C-terminal BTB/POZ domain that forms a characteristic fold (Artem et al. 2023). This protein family is implicated in maintaining cellular homeostasis and regulating the development of various plant organs. Initial characterization of the *MATH-BTB* family in plants was performed in maize and *Arabidopsis thaliana* (L.) Heynh (Juranič et al. 2012; Chen et al. 2014). In *Arabidopsis*, the six family members influence fatty acid metabolism and seed development by regulating the ethylene-responsive transcription factor WRI1 (WRINKLED1) (Weber et al. 2009; Chen et al. 2013). In maize, *MATH-BTB* proteins are critical for reproductive development, among other key biological processes (Juranič et al. 2014). In wheat, the gene *TaBTB-5A* has been identified as influencing yield-related traits such as plant height, grain length, and grain width (Dong, 2021). However, the functions of this protein family remain uncharacterized in rye (*Secale cereale* L.).

Rye, a member of the *Triticeae* tribe (*Poaceae* family), is often considered a secondary crop due to its close morphological resemblance to wheat and barley (*Hordeum vulgare* L.). It was domesticated in the coastal areas of the eastern Mediterranean (McElroy, 2014; Institute, 2022). Crucially, rye shares high genomic homology with wheat, which has facilitated its extensive use in wheat improvement programs, for instance, through the introgression of desirable traits (Bauer et al., 2017; Rabanus-Wallace et al., 2021).

The KASP (Kompetitive Allele-Specific PCR) system is a fluorescence-based genotyping technology that enables high-throughput, precise bi-allelic scoring of target SNPs and InDels (Dipta et al., 2024). It is widely applied in plant breeding to screen for desirable agronomic traits. For example, five KASP markers highly associated with smut resistance were identified in sugarcane (Gao et al., 2022), and 15 markers linked to powdery mildew resistance were developed in melon (Cao et al., 2021). In wheat breeding, KASP markers have been successfully used to select genotypes with enhanced drought tolerance (Eltaher et al., 2023).

This study performs a systematic identification and comprehensive bioinformatic analysis of the *MATH-BTB* gene family in rye. Furthermore, we developed KASP markers for yield-related traits, providing a practical tool for the molecular screening of wheat plants with improved agronomic characteristics.

Materials and methods

Plant materials

The three rye accessions used in this study (89R13, Z1672, and 90R13). There were obtained from and maintained by Hebei Normal University of Science and Technology. All accessions were grown under standard field conditions at the university's agricultural experiment station in Changli, Hebei Province, China, during the 2022 growing season.

Identification of the MATH-BTB gene family in rye

The protein sequences of known *MATH-BTB* genes from *Arabidopsis* obtained from the TAIR database, while those from rice and maize were retrieved from Ensembl Plants. The genome sequences of rye (Weining) and the A subgenome of wheat were downloaded from the National Genomics Data Center (NGDC) database (Li et al., 2021; Ling et al., 2013).

Identification was performed in two steps. First, the *Arabidopsis MATH-BTB* protein sequences were used as queries to perform a local BLASTP search against the rye and wheat A subgenome protein databases using TBtools software (Chen et al., 2020). Second, an online BLASTP search was conducted via the NCBI website to identify additional potential homologs. The conserved domains of all candidate proteins from Arabidopsis, wheat, and rye were verified using the NCBI's conserved domain database (CDD) batch search tool. Results were visualized with TBtools, and only proteins containing both the MATH and BTB domains were retained for subsequent analysis.

Following identification, the physicochemical properties of the deduced rye *MATH-BTB* proteins were predicted using the ExPASy ProtParam tool (Gasteiger et al., 2003).

Phylogenetic analysis

A phylogenetic tree was constructed from the aligned *MATH-BTB* protein sequences of rye, rice, wheat, maize, and *Arabidopsis* MEGA software. Multiple sequence alignment was performed with the MUSCLE algorithm, and the tree was built using the Neighbor-Joining (NJ) method. The topological robustness of the tree was assessed with 1000 bootstrap replicates. The final tree was visualized and annotated using FigTree software (Kumar et al. 2016).

Conserved motif analysis

Motif analysis of the rye *MATH-BTB* proteins was performed with the MEME Suite (max motifs = 10, default settings), followed by visualization of the motif architectures using TBtools (Bailey et al. 2009).

Collinearity analysis

Collinearity analysis of the *MATH-BTB* gene family was performed between rye and four reference species: *Arabidopsis*, maize, rice, and the A subgenome of wheat.

Expression analysis and functional validation of rye MATH-BTB genes

Expression profiles of the identified *MATH-BTB* gene family members across different rye tissues were analyzed using publicly available transcriptome data from the WheatOmics database. The expression patterns were visualized as a heatmap using TBtools to identify genes with tissue-specific or developmentally regulated expression (Ma et al., 2021).

To leverage prior genetic evidence, we integrated these findings with results from our previous study based on SLAF-seq of a rye hybrid population, which mapped QTLs for yield-related traits and identified associated candidate genes (Che et al., 2023). The functional validation in the present study consisted of identifying the intersection between the *MATH-BTB* gene family and the candidate genes previously implicated in yield traits.

Development of molecular markers for candidate genes

Conventional PCR primers flanking the candidate gene sequences were designed using Primer Premier 5 software. Total RNA was extracted from randomly selected individuals of the rye accessions 89R41 and Z1672 and reverse-transcribed into cDNA. The target sequences were amplified by PCR, and the products were resolved by agarose gel electrophoresis. The resulting bands were examined for size

polymorphisms and/or intensity variations to preliminarily assess sequence variation in the candidate genes.

If no clear polymorphism was detected, KASP markers were developed based on the SNP loci previously linked to the candidate gene. The KASP assay, designed using the PolyMarker website, comprised two allele-specific forward primers (each labeled with a distinct fluorescent dye, FAM or HEX) and a common reverse primer. Genotyping was performed via KASP-PCR, and the fluorescence signals were quantified to identify markers associated with the target phenotypic trait.

Results

Identification and characterization of the MATH-BTB gene family in rye

A total of 56 *MATH-BTB* genes were identified in the rye genome and systematically annotated as *ScMB1* - *ScMB56*. These genes were distributed across all seven rye chromosomes.

The physicochemical properties of the encoded 56 proteins were analyzed (Table 1). The isoelectric points (pI) ranged from 4.84 (*ScMB14*) − 8.83 (*ScMB45*). The majority (50 proteins) had a pI below 7, indicating that most *MATH-BTB* proteins in rye are acidic, while only six are alkaline. Analysis of the grand average of hydropathicity (GRAVY) showed that the proteins encoded by *ScMB3* and *ScMB27* are hydrophilic, whereas the rest are hydrophobic. The molecular weights ranged from 36.05 kDa (*ScMB45*) − 59.48 kDa (*ScMB48*), corresponding to amino acid lengths of 327–546 residues. The aliphatic index varied from 75.03 (*ScMB18*) − 93.98 (*ScMB27*). Notably, 52 proteins had an index greater than 80, suggesting that the majority of the *MATH-BTB*family members are thermostable, a feature that may contribute to their functional adaptability under diverse environmental conditions.

Table 1
Physical and chemical characteristics of rye *MATH-BTB* genes and its encoded protein

Name	Gene	pI	MW / kD	No.ofaminoacid	Aliphaticindex	GRAVY
ScMB1	ScWN3R01G289700.1	5.66	39 990.62	364	87.58	-0.09
ScMB2	ScWN2R01G399400.1	5.87	40 351.34	361	85.87	-0.27
ScMB3	ScWN6R01G378100.1	6.24	40 198.27	365	90.93	0.01
ScMB4	ScWN2R01G562500.1	6.16	41 254.30	369	80.08	-0.11
ScMB5	ScWN5R01G406400.1	5.94	39 308.09	352	83.64	-0.08
ScMB6	ScWN1R01G194300.1	5.73	41 489.49	371	90.40	-0.08
ScMB7	ScWN5R01G425000.1	5.97	45 463.25	412	89.44	-0.03
ScMB8	ScWN4R01G155400.1	5.74	39 437.03	360	88.03	-0.10
ScMB9	ScWN6R01G033300.1	7.06	39 057.69	349	79.94	-0.31
ScMB10	ScWN7R01G487200.1	6.99	36 624.09	332	91.39	-0.06
ScMB11	ScWN3R01G232100.1	6.16	38 370.79	349	86.39	-0.06
ScMB12	ScWN1R01G077700.1	5.15	39 837.15	363	89.67	-0.03
ScMB13	ScWN3R01G289900.1	5.79	40 945.61	372	82.55	-0.19
ScMB14	ScWN6R01G530500.1	4.84	43 527.54	392	79.85	-0.14
ScMB15	ScWN4R01G385200.1	6.56	42 432.38	387	80.39	-0.12
ScMB16	ScWN7R01G486000.1	5.35	41 300.36	370	93.57	-0.04
ScMB17	ScWN5R01G504300.1	6.12	39 096.00	353	88.41	-0.10
ScMB18	ScWN5R01G496400.1	5.09	47	437	75.03	-0.32

Name	Gene	pI	MW / kD	No.ofaminoacid	Aliphaticindex	GRAVY
			363.71			
ScMB19	ScWN2R01G516700.1	8.7	42 640.99	384	81.30	-0.14
ScMB20	ScWN5R01G424700.1	5.66	40 010.80	365	85.21	-0.06
ScMB21	ScWN7R01G486600.1	6.99	39 676.64	360	91.03	-0.05
ScMB22	ScWN1R01G411700.1	5.73	39 961.87	356	81.35	-0.11
ScMB23	ScWN3R01G472700.1	5.05	39 606.08	360	92.14	-0.02
ScMB24	ScWN5R01G424500.1	6.38	44 702.21	404	84.98	-0.20
ScMB25	ScWN4R01G352500.1	6.4	39 779.01	355	87.94	-0.16
ScMB26	ScWN7R01G108900.1	5.89	40 954.55	370	81.97	-0.23
ScMB27	ScWN7R01G487700.1	5.53	38 200.83	352	93.98	0.07
ScMB28	ScWN7R01G388800.1	5.69	39 209.73	362	87.40	-0.04
ScMB29	ScWN6R01G535700.1	5.04	41 178.65	367	81.63	-0.24
ScMB30	ScWN1R01G116400.2	4.9	43 385.72	386	91.37	-0.04
ScMB31	ScWN5R01G453500.1	5.44	38 544.85	350	81.69	-0.04
ScMB32	ScWN7R01G487100.1	7.52	39 697.69	360	92.67	-0.04
ScMB33	ScWN6R01G535900.1	6.43	41 698.18	371	82.83	-0.22
ScMB34	ScWN3R01G289800.1	5.63	44 283.34	400	80.42	-0.10
ScMB35	ScWN7R01G484500.1	6.17	38 808.40	348	89.14	-0.20
ScMB36	ScWN5R01G051900.1	5.85	40 551.44	364	85.49	-0.19

Name	Gene	pI	MW / kD	No.ofaminoacid	Aliphaticindex	GRAVY
ScMB37	ScWN3R01G402000.1	5.61	41 322.48	365	86.30	-0.17
ScMB38	ScWN5R01G533900.1	5.77	40 764.95	363	84.63	-0.16
ScMB39	ScWN5R01G036500.1	6.41	39 319.53	353	88.39	-0.03
ScMB40	ScWN7R01G109800.1	5.89	40 954.55	370	81.97	-0.23
ScMB41	ScWN3R01G536900.1	5.24	41 786.58	377	80.40	-0.11
ScMB42	ScWN3R01G243300.1	5.29	38 473.04	345	89.33	-0.05
ScMB43	ScWN6R01G235200.1	6.19	39 669.87	351	87.49	-0.23
ScMB44	ScWN2R01G469700.1	6.57	42 605.49	376	89.73	-0.19
ScMB45	ScWN5R01G424600.1	5.78	44 265.85	400	89.92	-0.02
ScMB46	ScWN1R01G411600.3	5.83	39 448.31	356	83.88	-0.09
ScMB47	ScWN5R01G044200.1	5.95	49 786.20	446	79.01	-0.34
ScMB48	ScWN2R01G516600.1	6.58	59 477.82	546	81.98	-0.10
ScMB49	ScWN5R01G504200.1	6.4	44 029.41	393	87.66	-0.19
ScMB50	ScWN2R01G292300.1	6.08	44 140.15	396	83.18	-0.20
ScMB51	ScWN6R01G535400.1	5.04	41 178.65	367	81.63	-0.24
ScMB52	ScWN7R01G485000.1	5.78	39 862.65	360	88.08	-0.13
ScMB53	ScWN5R01G532000.1	7.58	41 156.33	364	85.22	-0.33
ScMB54	ScWN7R01G297300.1	5.67	41 097.08	366	83.66	-0.13

Name	Gene	pI	MW / kD	No.ofaminoacid	Aliphaticindex	GRAVY
ScMB55	ScWN6R01G233900.1	6.11	41 251.63	366	87.65	-0.23
ScMB56	ScWN3R01G019000.1	7.16	41 249.43	374	86.60	-0.04

Phylogenetic analysis of rye MATH-BTB proteins

A phylogenetic tree was constructed using *MATH-BTB* protein sequences from rye (56), wheat A subgenome (72), rice (74), maize (40), and *Arabidopsis* (6) to elucidate their evolutionary relationships (Fig. 1). The proteins clustered into ten distinct subfamilies (*G1–G10*). Proteins from the monocot species (rye, wheat, rice, and maize) were distributed across all subfamilies, whereas those from the dicot *Arabidopsis* confined to *G5*. This distribution suggests that the diversification of the MATH-BTB family occurred after the monocot-dicot divergence. The number of rye genes in each subfamily (*G1–G10*) was 4, 3, 9, 1, 2, 4, 8, 5, 6, and 14.

Conserved motif analysis

Conserved motif analysis of the 56 rye *MATH-BTB* proteins revealed a high degree of sequence conservation (Fig. 2). All members shared motifs 1, 2, 3, 4, 5, 7, and 10. Only minor variations were observed: the proteins encoded by *ScMB30* and *ScMB18* lacked motif 6, and the protein from *ScMB10* lacked motif 8. The overall minimal variation in domain composition across subfamilies indicates strong structural conservation within this protein family in rye.

Analysis of genomic synteny

Synteny analysis of the rye *MATH-BTB* genes with four reference species revealed a distinct pattern: no collinearity with *Arabidopsis* and maize, one syntenic pair with rice, and 16 pairs with the wheat A subgenome (Fig. 3). The strong collinearity observed specifically with wheat underscores their close phylogenetic relationship and highlights the value of rye genetic resources for wheat improvement.

Tissue-specific expression analysis of rye MATH-BTB genes

Expression patterns of the MATH-BTB genes were analyzed using RNA-seq data from the WheatOmics database, which included root, stem, leaf, and spike tissues at the heading stage, as well as seeds at 10, 20, 30, and 40 days after pollination. A heatmap generated with TBtools revealed distinct expression profiles (Fig. 4). Fourteen genes (*ScMB2, ScMB6, ScMB9, ScMB15, ScMB19, ScMB24, ScMB26, ScMB35, ScMB37, ScMB40, ScMB41, ScMB44, ScMB45, ScMB47*) showed no expression across all examined tissues. In contrast, the remaining genes were expressed in various tissues, with particularly high

expression levels in spikes and developing seeds, collectively suggesting a potential role for this gene family in grain development.

Functional validation of the rye MATH-BTB gene family

Building on our previous SLAF-seq study that identified QTLs for yield-related traits, we integrated the genetic data with the 56 *MATH-BTB* genes identified here. This integration pinpointed a strong candidate gene, *ScWN4R01G352500.1* (*ScMB25*), which is located within a QTL interval for grain width. The QTL was mapped to chromosome 4 (77.871 cM), with a LOD score of 3.63 and a PVE of 15%. A total of 27 SNPs were linked to this locus. This result provides genetic evidence that the *MATH-BTB* gene family, particularly *ScMB25*, may regulate grain width in rye.

Development of molecular markers

To validate the presence of the candidate gene *ScWN4R01G352500.1*, we designed a gene-specific primer pair. cDNA from the wide-grain wild rye 89R41 and the narrow-grain cultivated rye Z1672 was used as a template for PCR amplification. Electrophoresis confirmed the presence of the target gene in both accessions, but no polymorphism was detected.

Consequently, we developed seven KASP markers (KM38–KM44) targeting SNP loci linked to the grain width QTL. Initial screening in a hybrid population (89R41 × Z1672) showed that KM38 and KM42 could genotype the trait. These two markers were subsequently validated in a larger population from the wild rye 90R13 (n = 311). A t-test comparing the genotyping results with the grain width phenotypic data revealed a significant association between the KM42 marker and grain width ($P = 0.018$, $P < 0.05$), indicating its utility for selecting rye plants with desirable grain width. The SNP corresponding to this marker is located at position 749,349,504 bp on chromosome 4. This marker can therefore be used for selecting rye plants with desirable grain width.

Phenotypic analysis revealed that allele 1 of KM42 positively influences grain width compared to allele 2, with a contribution rate of 102.2%. Detailed results are presented in Tables 2–3 and Fig. 5. The environmental stability of KM42 is under investigation.

Table 2 KASP primer information table

Primer	Sequences
Pre-primer sequence 1 (5□□3□)	GAAGGTGACCAAGTTCATGCTtgatagaggagtcgtcgacC
Pre-primer sequence 2 (5□□3□)	GAAGGTCGGAGTCAACGGATTtgatagaggagtcgtcgacT
Co-primers (5□□3□)	cttgatggtgacctcggagg

Note: Pre-primer sequence 1 connects the FAM-type fluorescent motif sequence GAAGGTGACCAAGTTCATGCT at the 5' end, and pre-primer sequence 2 connects the HEX-type fluorescent motif sequence GAAGGTCGGGAGTCAACGGATT at the 5' end, and C/T stands for SNP site.

Table 3
The results of KASP molecular marker test

Typing	Amplification rate	Amplified samples	Phenotypic data (cm)	Contribution rate	*p*
Homozygous allele 1	86.50%	122	0.232 ± 0.019	102.2%	0.018
Homozygous allele 2		147	0.227 ± 0.018		

Discussion

We identified 56 *MATH-BTB* genes in the rye genome, a number comparable to other grasses but larger than in *Arabidopsis*, likely reflecting a *Poaceae* specific expansion. We systematically analyzed the gene structures, phylogenetic relationships, conserved protein motifs, synteny, and expression patterns of these 56 genes. Our findings establish a crucial foundation for future functional studies, such as the cloning and validation of individual rye *MATH-BTB* genes, and provide a valuable resource for elucidating their roles in rye growth and development.

Our integrated analysis yields several key insights into the *MATH-BTB* family in rye. First, the prevalence of acidic, hydrophobic, and thermostable proteins suggests an adaptation to diverse environmental stresses. Second, the phylogenetic isolation of the *Arabidopsis* genes in a single subfamily, contrasted with the broad distribution in cereals, strongly supports a lineage-specific expansion after the monocot-dicot split. The conserved domain architecture within subfamilies implies conserved biological functions. Most significantly, the extensive collinearity between rye and wheat provides compelling genomic evidence of their close evolutionary relationship. These findings establish the *MATH-BTB* family as a subject for future functional studies in rye and firmly position rye as a valuable reservoir of genetic diversity for the enhancement of wheat.

The tight linkage between gene expression and function was evident in our analysis. The predominant expression of rye *MATH-BTB* genes in spikes and seeds strongly suggests a primary role in reproductive development. An intriguing pattern emerged from integrating phylogenetic and expression data: two members of the *G5* subfamily displayed significantly higher and more ubiquitous expression. The fact that the entire *Arabidopsis MATH-BTB* family also clusters within *G5* provides a key evolutionary context. This allows us to infer that these two rye *G5* members are orthologs of the ancestral genes present before the monocot-dicot divergence, and their strong expression may reflect a fundamental, conserved function. In contrast, the more restricted expression of other subfamily members likely indicates functional specialization following lineage-specific expansions.

SLAF-seq technology enables high-throughput development and genotyping of molecular markers (SNPs and InDels) and is widely applied in QTL mapping and crop breeding (Sun et al. 2013; Guo et al. 2022; Anilkumar et al. 2022; Zhang et al. 2016). By integrating previously identified yield-related QTLs, linked

SNPs, and candidate genes obtained through SLAF-seq with the 56 *MATH-BTB* genes identified in this study, we identified a candidate gene, *ScWN4R01G352500.1* (*ScMB25*), which was associated with grain width in the SLAF-seq results. PCR analysis of this gene revealed no sequence variation among different rye varieties; however, under consistent cDNA concentrations, the band intensity was stronger in seeds with wider grains compared to those with narrower grains (Fig. 6). Quantification using ImageJ showed average grayscale values of 123.0, 155.7, 110.3, and 158 for lanes 2, 3, 4, and 5, respectively, indicating differential expression of this gene in seeds of varying sizes. Higher expression levels appear to promote increased grain width, a hypothesis that warrants further validation through transgenic experiments. These findings suggest that the function of the *MATH-BTB* protein encoded by this gene in rye is similar to that observed in *Arabidopsis* and wheat, supporting the hypothesis that *MATH-BTB* proteins influence grain width in rye. Additionally, the KASP marker KM42 developed in this study provides a practical tool for selecting rye varieties with desirable grain width traits.

Note

Lane 1 shows the DNA Marker (50–500 bp), with bands from top to bottom corresponding to 500 bp, 400 bp, 300 bp, 250 bp, 200 bp, 150 bp, 100 bp, and 50 bp. Lanes 2–5 display the target gene band (410 bp) in the upper region, lanes 2 and 4 are electrophoresis bands of low-grain width rye plants, the average gray levels were 123 and 110.3, lanes 3 and 5 are electrophoresis bands of high-grain width rye plants, the average gray levels were 155.7 and 158.

Figure 6 Electrophoresis map of *ScMB25* gene

Conclusions

This study provides the first genome-wide identification and characterization of the *MATH-BTB* gene family in rye, comprising 56 members. Integrated bioinformatic and phylogenetic analyses, combined with prior genetic mapping data, pinpointed *ScWN4R01G352500.1* as a key candidate gene regulating grain width. Furthermore, we developed and validated a functional KASP marker (KM42) for this gene, offering a practical molecular tool for the rapid selection of improved grain width in rye breeding programs.

Declarations

Author contributions

LW, ZZ, YC: contributed to the management and manuscript review. ZZ, YY, XW: performed the experiment; ZZ, XW, YY: designed experiments as well as provided the methodology of data collection and analysis. All authors have read and agreed to the published version of the manuscript.

Funding

This research supported by the “S&T Program of Hebei Department of Science and Technology” (21326340D).

Data availability

The protein sequences of Arabidopsis *MATH-BTB* gene family members were obtained from the TAIR database (http://www.arabidopsis.org), the protein sequences of the *MATH-BTB* gene family members of rice and maize were obtained from the Ensembl Plants database (http://plants.ensembl.org/index.html). the data of wheat A genome and rye (Weining) genome were obtained from the NGDC database (https://ngdc. cncb.ac.cn/), the transcript expression profiles of different tissues of rye were obtained from the WheatOmics database (http://wheatomics.sdau.edu.cn/). All data generated or analysed during this study are included in this published article.

References

1. Anilkumar C, Sah RP, Azharudheen TM, Behera S, Singh N, Prakash NR, Sunitha NC, Devanna BN, Marndi BC, Patra BC, Nair SK (2022) Understanding complex genetic architecture of rice grain weight through QTL-meta analysis and candidate gene identification. Scientific Reports 12: 1332. https://doi.org/10.1038/s41598-022-17402-w
2. Artem B, Konstantin B, Pavel Georgiev (2023) BTB domains: A structural view of evolution, multimerization, and protein–protein interactions. BioEssays 45(2): 2200179. https://doi.org/10.1002/bies.202200179
3. Bauer E, Schmutzer T, Barilar I, Mascher M, Gundlach H, Martis MM, Twardziok SO, Hackauf B, Gordillo A, Wilde P, Schmidt M, Korzun V, Mayer KF, Schmid K, Schön CC, Scholz U (2017) Towards a whole-genome sequence for rye (*Secale cereale* L.). The Plant Journal 89(5): 853–869. https://doi.org/10.1111/tpj.13436
4. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE:Tools for motif discovery and searching. Nucleic Acids Research 37: 202–208. https://doi.org/10.1093/nar/gkp335
5. Cao Y Y, Diao Q N, Chen Y Y, Jin HJ, Zhang YP, Zhang HM (2021) Development of KASP markers and identification of a QTL underlying powdery mildew resistance in melon (*Cucumis melo* L.) by bulked segregant analysis and RNA-seq. Frontiers in Plant Science 11: 593207. https://doi.org/10.3389/fpls.2020.593207
6. Che YH, Yang Y J, Wei L, Guo J, Yang XM, Li XQ, Liu WH, Li LH (2023) Construction of a high-density genetic map and mapping of a spike length locus for rye. PLoS One 18(10): e0293604. https://doi.org/10.1371/journal.pone.0293604
7. Chen C J, Chen H, Zhang Y, Thomas HR, Frank MH, He YH, Xia R (2020) TBtools:An integrative toolkit developed for interactive analyses of big biological data. Molecular Plant 13(8): 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

8. Chen L, Bernhardt A, Lee J, Hellmann H (2014) Identification of Arabidopsis MYB56 as a novel substrate for CRL3(BPM) E3 ligases. Molecular Plant 8(2): 242–250. https://doi.org/10.1016/j.molp.2014.10.004
9. Chen LY, Lee JH, Weber H, Tohge T, Witt S, Roje S, Fernie AR, Hellmann H (2013) Arabidopsis BPM proteins function as substrate adaptors to a CULLIN3-based E3 ligase to affect fatty acid metabolism in plants. The Plant Cell 25(6): 2253–2264. https://doi.org/10.1105/tpc.112.107292
10. Dipta B, Sood S, Mangal V, Bhardwaj V, Thakur AK, Kumar V, Singh B (2024) KASP: A high-throughput genotyping system and its applications in major crop plants for biotic and abiotic stress tolerance. Molecular Biology Reports 51(1): 508. https://doi.org/10.1007/s11033-024-09455-z
11. Dong CY (2021) Isolation and superior allelic variation analysis of wheat *TaBTB-5A* gene. Taian: Shandong Agricultural University.
12. Eltaher S, Hashem M, Ahmed AAM, Baenziger PS, Börner A, Sallam A (2023) Effectiveness of *TaDreb-B1* and *1-FEH w3* KASP markers in spring and winter wheat populations for marker-assisted selection to improve drought tolerance. *International Journal of Molecular Sciences* 24(10): 8986. https://doi.org/10.3390/ijms24108986
13. Fan CC, Xing YZ, Mao HL Lu TT, Han B, Xu CG, Li XH, Zhang QF (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theoretical and Applied Genetics 112(6): 1164–1171. https://doi.org/10.1007/s00122-006-0218-1
14. Gao YJ, Zhou S, Huang YX, Zhang BQ, Xu YH, Zhang GM, Lakshmanan P, Yang RZ, Zhou H, Huang DL, Liu JX, Tan HW, He WZ, Yang CF, Duan WX (2022) Quantitative trait loci mapping and development of KASP marker smut screening assay using high-density genetic map and bulked segregant RNA sequencing in sugarcane (*Saccharum* spp.). Frontiers in Plant Science 12: 796189. https://doi.org/10.3389/fpls.2021.796189
15. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy:The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research 31(13): 3784–3788. https://doi.org/10.1093/nar/gkg563
16. Guo X, Wu CN, Wang DH, Wang GY, Jin KT, Zhao YJ, Tian JC, Deng ZY (2022) Conditional QTL mapping for seed germination and seedling traits under salt stress and candidate gene prediction in wheat. Scientific Reports 12: 21010. https://doi.org/10.1038/s41598-022-25703-3
17. Guan PF, Shen XY, Mu Q, Wang YF, Wang XB, Chen YM, Zhao Y, Chen XY, Zhao AJ, Mao WW, Guo YW, Xin MM, Hu ZR, Yao YY, Ni ZF, Sun QX, Peng HR (2020) Dissection and validation of a QTL cluster linked to Rht-B1 locus controlling grain weight in common wheat (*Triticum aestivum* L.) using near-isogenic lines. Theoretical and Applied Genetics 133(9): 2639–2653. https://doi.org/10.1007/s00122-020-03622-z
18. Huang R Y, Jiang L R, Zheng J S, Wang TS, Wang HC, Huang YM, Hong ZL (2013) Genetic bases of rice grain shape: So many genes, so little known. Trends in Plant Science 18(4): 218–226. https://doi.org/10.1016/j.tplants.2012.11.001

19. Institute JK (2022) 100 Years of Chromosome Research in Rye, *Secale* L. Plants 11(13): 1753. https://doi.org/10.3390/plants11131753
20. Juranič M, Dresselhaus T (2014) Phylogenetic analysis of the expansion of the MATH-BTB gene family in the grasses. Plant Signal Behav 9(4): e28242. https://doi.org/10.4161/psb.28242
21. Juranič M, Srilunchang KO, Krohn NG, Leljak-Levanic D, Sprunck S, Dresselhaus T (2012) Germline-specific MATH-BTB substrate adaptor MAB1 regulates spindle length and nuclei identity in maize. The Plant Cell 24(12): 4974–4991. https://doi.org/10.1105/tpc.112.107169
22. Kumar A, Mantovani EE, Seetan R, Soltani A, Echeverry-Solarte M, Jain S, Simsek S, Doehlert D, Alamri MS, Elias EM, Kianian SF, Mergoum M (2016) Dissection of genetic factors underlying wheat kernel shape and size in an Elite × Nonadapted cross using a high density SNP linkage map. The Plant Genome 9(1): 27898771. – 3939. https://doi.org/10.3835/plantgenome2015.09.0081
23. Kumar S, Stecher G, Tamura K (2016) MEGA7:Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33(7): 1870–1874. https://doi.org/10.1093/molbev/msw054
24. Li GW, Wang LJ, Yang JP, *et al* (2021) A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. Nature Genetics 53(4): 574–584. https://doi.org/10.1038/s41588-021-00808-z
25. Ling HQ, Zhao SC, Liu DC, *et al* (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu* [J]. Nature 496: 87–90. https://doi.org/10.1038/nature11997
26. Li YB, Fan CC, Xing YZ, Jiang YH, Luo LJ, Sun L, Shao D, Xu CJ, Li XH, Xiao JH, He YQ, Zhang QF (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. Nature Genetics 43(12): 1266–1269. https://doi.org/10.1038/ng.977
27. Li Q, Yang XH, Bai GH, Warburton ML, Mahuku G, Gore M, Dai JR, Li JS, Yan JB (2010) Cloning and characterization of a putative *GS3* ortholog involved in maize kernel development. Theoretical and Applied Genetics 120(4): 753–763. https://doi.org/10.1007/s00122-009-1196-x
28. Ma S, Wang M, Wu J, Guo W, Chen Y, Li G, Wang Y, Shi W, Xia G, Fu D, Kang Z, Ni F (2021) WheatOmics:A platform combining multiple omics data to accelerate functional genomics studies in wheat. Molecular Plant 14(12): 1965–1968. https://doi.org/10.1016/j.molp.2021.10.006
29. Marchlwr-Bauera B, Derbyshire M K, Gonzales N R, *et al* (2015) CDD:NCBI's conserved domain database. Nucleic Acids Research 43: 222–226. https://doi.org/10.1093/nar/gku1221
30. Mao HL, Sun SY, Yao JL, Wang CR, Yu SB, Xu CG, Li XH, Zhang QF (2010) Linking differential domain functions of the GS3 protein to natural variation in grain size in rice. Proceedings of the National Academy of Sciences of the United States of America 107(45): 19579–19584. https://doi.org/10.1073/pnas.1014419107
31. McElroy JS (2014) Vavilovian Mimicry: nikolai vavilov and his little-known impact on weed science. Weed Science 62(2): 207–216. https://doi.org/10.1614/WS-D-13-00122.1
32. Rabanus-Wallace M T, Hackauf B, Mascher M, *et al* (2021) Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. Nature Genetics 53(4): 564–

573. https://doi.org/10.1038/s41588-021-00807-0

33. Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nature Genetics 40(8): 1023–1028. https://doi.org/10.1038/ng.169
34. Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nature Genetics 39(5): 623–630. https://doi.org/10.1038/ng2014
35. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X, Zheng H (2013) SLAF-seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. PLoS One 8: 58700. https://doi.org/10.1371/journal.pone.0058700
36. XQ, Dong LH, Hu JM, Xiao GL, Ma X, Kong XY, Jia JZ, Wang HW, Kong LG (2019) Dissecting genetic loci affecting grain morphological traits to improve grain weight via nested association mapping. Theoretical and Applied Genetics 132(11): 3115–3128. https://doi.org/10.1007/s00122-019-03410-4
37. Weber H, Hellmann H (2009) Arabidopsis thaliana BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. The FEBS Journal 276(22): 6624–6635. https://doi.org/10.1111/j.1742-4658.2009.07373.x
38. Zhang Z, Shang H, Shi Y, Huang L, Li J, Ge Q, Gong J, Liu A, Chen T, Wang D, Wang Y, Palanga KK, Muhammad J, Li W, Lu Q, Deng X, Tan Y, Song W, Cai J, Li P, Rashid Ho, Gong W, Yuan Y (2016) Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (*Gossypium hirsutum*.). BMC Plant Biology 16: 79. https://doi.org/10.1186/s12870-016-0741-4
39. Zhang YJ, Liu JD, Xia XC, He ZH (2014) *TaGS-D1*, an ortholog of rice *OsGS3*, is associated with grain weight and grain length in common wheat. Molecular Breeding 34(3): 1097. https://doi.org/10.1007/s11032-014-0102-7

Figures

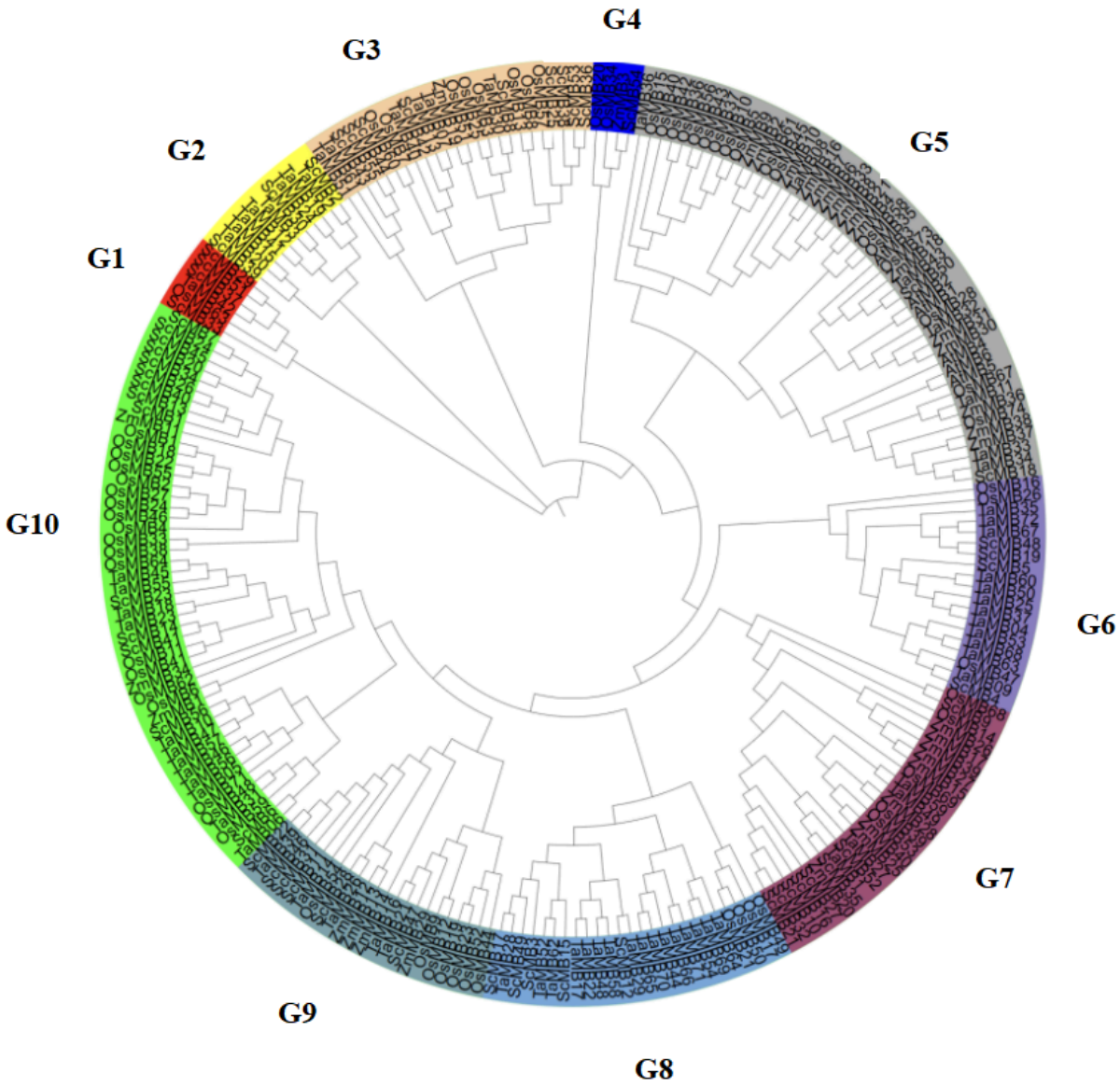

Figure 1

Phylogenetic analysis of the *MATH-BTB*gene family in rye

Note: Sc: *S. cereale* ; At: *Arabidopsis thaliana*; Ta: *Triticum aestivum*; Os: *Oryza sativa* ; Zm: *Zea mays*; *G1~G10*represent 10 subclasses.

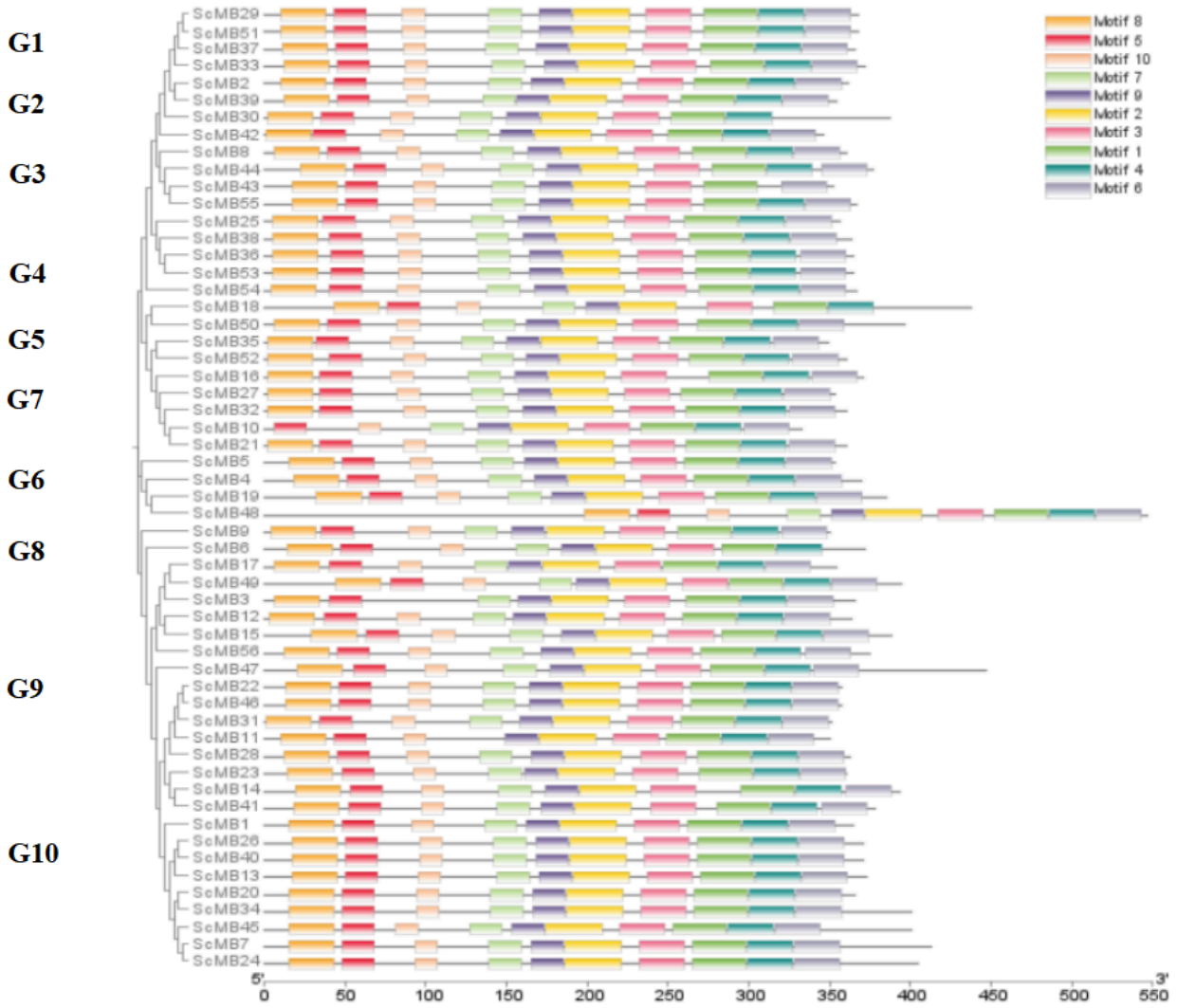

Figure 2

Conserved motif analysis of rye *MATH-BTB*protein

Note: Motif1-10 represents a conserved domain.

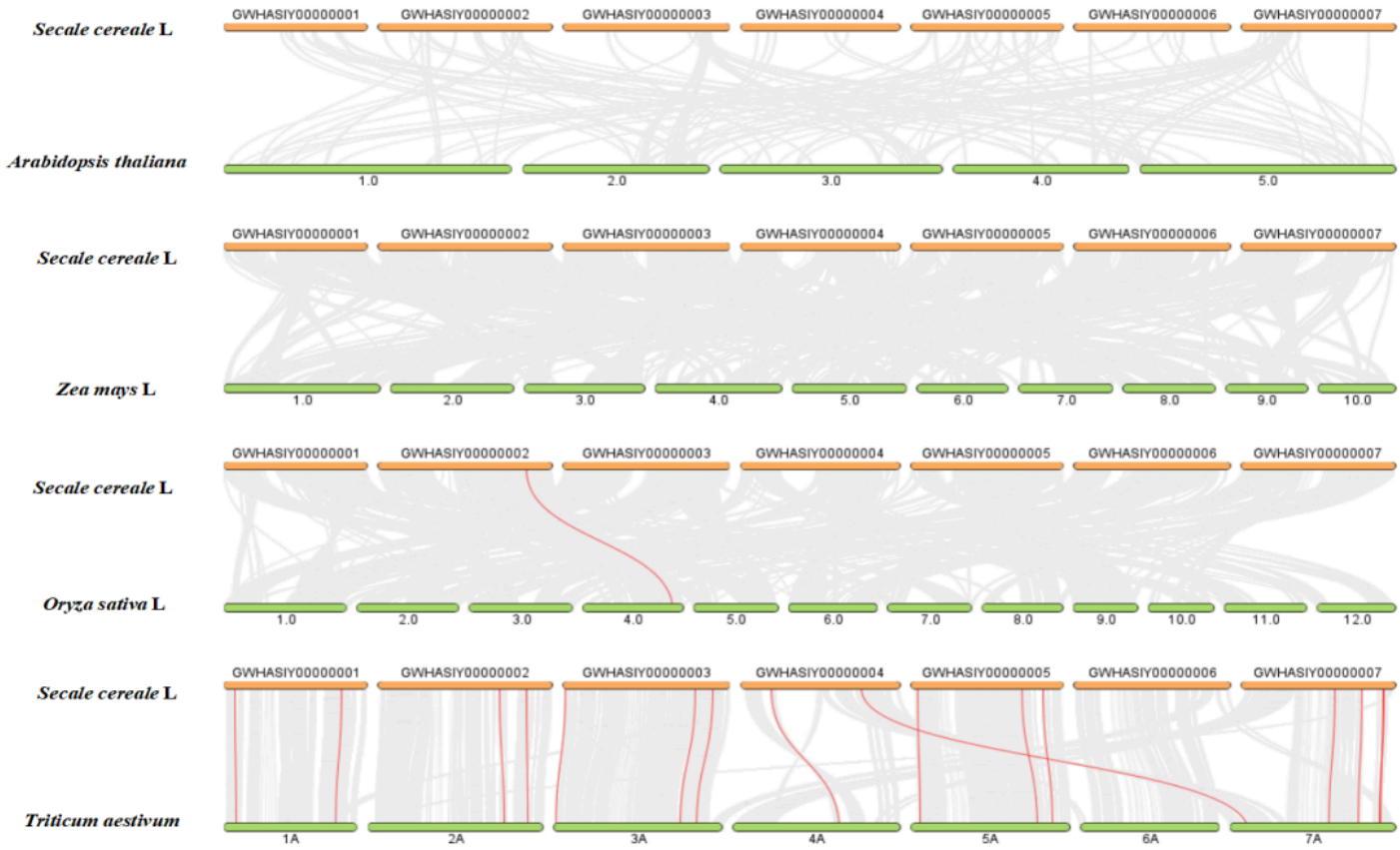

Figure 3

Syntenic relationships of *MATH-BTB* Genes between rye and four other plant species

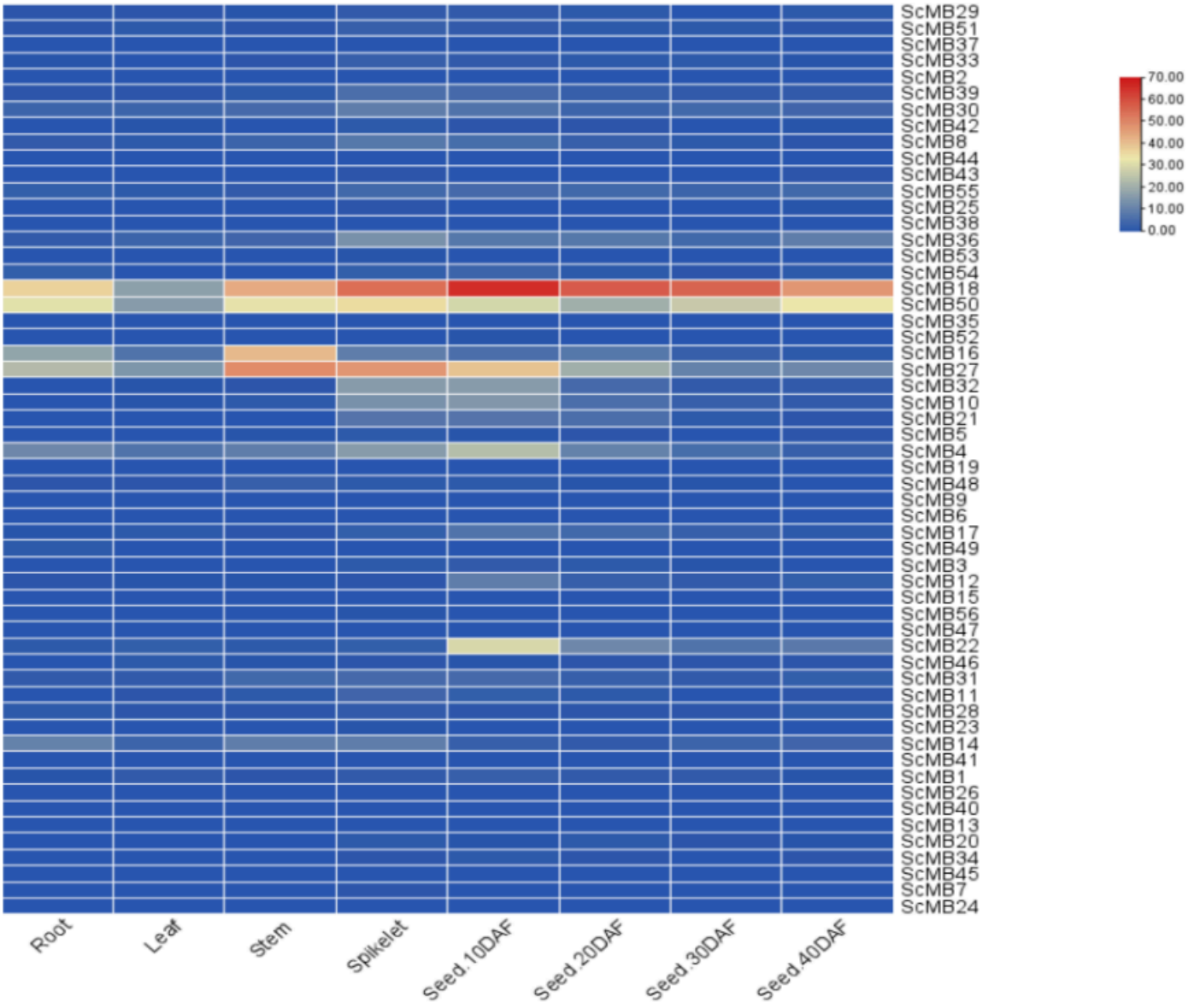

Figure 4

Expression of 56 *MATH-BTB*genes in different tissues of rye

Note: 1. Root: Root; 2. Leaf: Leaf blade; 3. Stem: Stem; 4. Spikelet: Spike; 5. Seed.10DAF: Seed at 10d after flowering; 6. Seed.20DAF: Seed at 20d after flowering; 7. Seed.30DAF: Seed at 30d after flowering; 8. Seed.40DAF: Seed at 40d after flowering; 9. Red for high expression, blue for low expression.

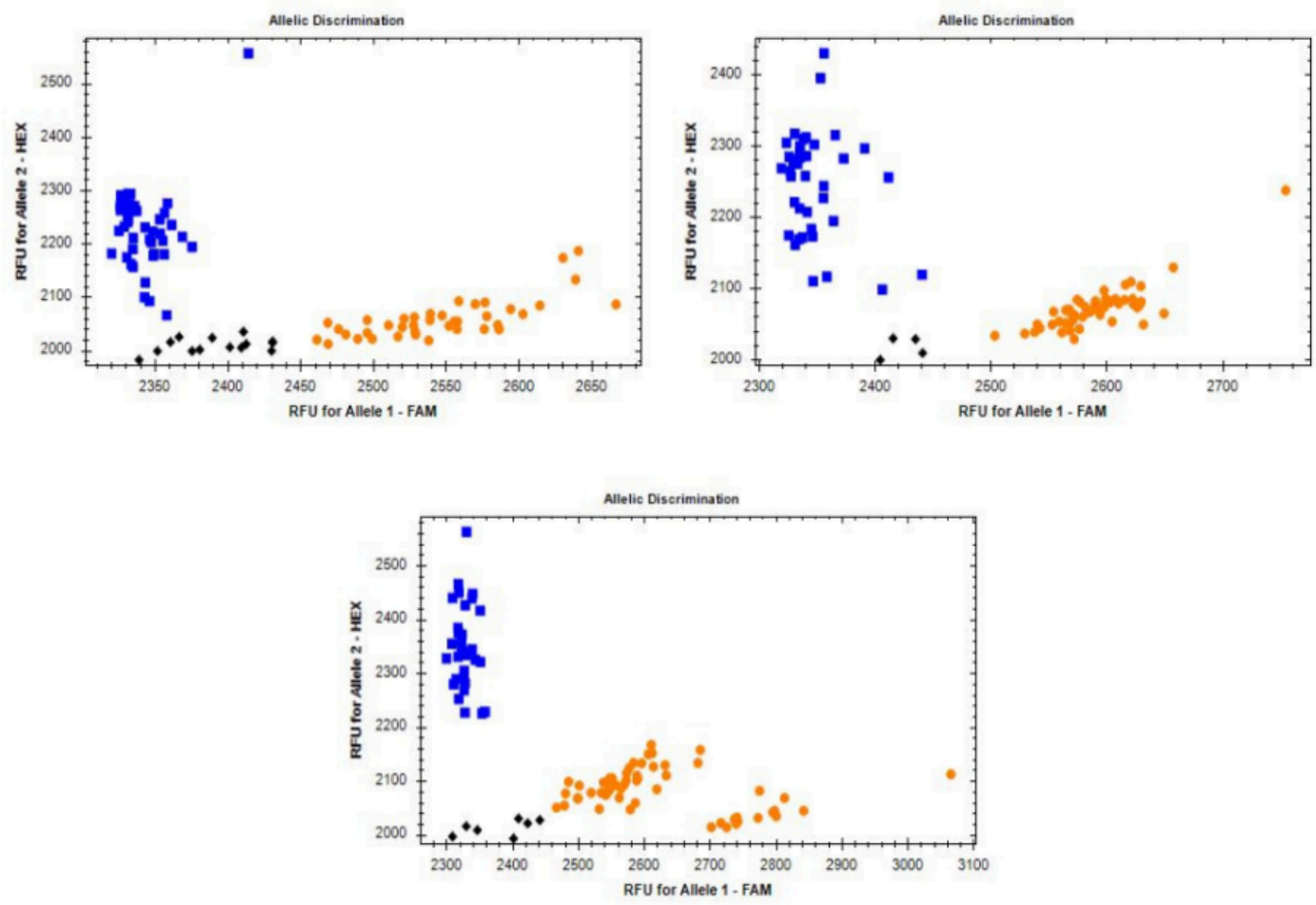

Figure 5

KASP labeling 3 sampling typing results

Note: 1. X-axis, allele 1, reported by FAM-type fluorescence; 2. Y-axis, allele 2, reported by HEX-type fluorescence; 3. Blue dots represent homozygous allele group 1; 4. Orange dots represent homozygous allele group 2; and 5. Black dots represent no response.

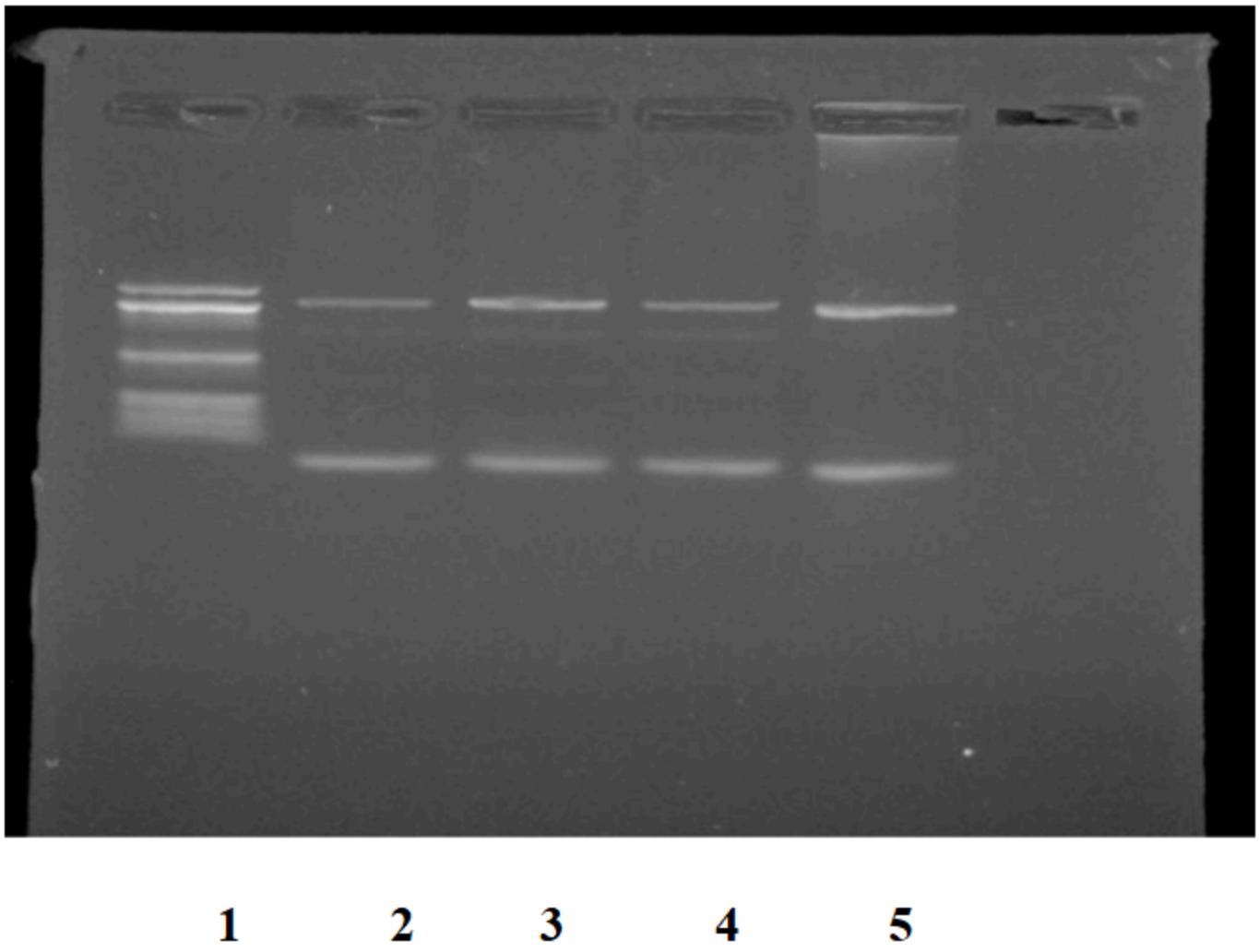

Figure 6

Electrophoresis map of *ScMB25* gene

Note: Lane 1 shows the DNA Marker (50–500 bp), with bands from top to bottom corresponding to 500 bp, 400 bp, 300 bp, 250 bp, 200 bp, 150 bp, 100 bp, and 50 bp. Lanes 2-5 display the target gene band (410 bp) in the upper region, lanes 2 and 4 are electrophoresis bands of low-grain width rye plants, the average gray levels were 123 and 110.3, lanes 3 and 5 are electrophoresis bands of high-grain width rye plants, the average gray levels were 155.7 and 158.