genes

MDPI

Article

Genome-Wide Identification of *R2R3-MYB* Family Genes and Expression Analysis of the S12 Subfamily Genes Under Abiotic Stress in *Brassica juncea* L.

Ting Kuang [1,2,3,4], Tianshun Liu [1,2,3,4], Mingyu Xie [1,2,3,4], Haidong Liu [1,2,3,4], Kaixiang Li [1,2,3,4], Anqi Wang [1,2,3,4], Youjuan Quan [1,2,3,4], Lingxiong Zan [1,2,3,4], Jun Ma [1,2,3,4], Haijun Zhang [1,2,3,4] and Yanmei Yao [1,2,3,4,*]

[1] Spring Rapeseed Research Institute, Academy of Agricultural and Forestry Sciences, Qinghai University, 251 Ningda Road, Chengbei District, Xining 810016, China; kt0922@126.com (T.K.)
[2] Laboratory for Research and Utilization of Qinghai Tibet Plateau Germplasm Resources, 251 Ningda Road, Chengbei District, Xining 810016, China
[3] Key Laboratory of Spring Rapeseed Genetic Improvement of Qinghai Province, 251 Ningda Road, Chengbei District, Xining 810016, China
[4] Qinghai Spring Rape Engineering Research Center, 251 Ningda Road, Chengbei District, Xining 810016, China
* Correspondence: 2009990028@qhu.edu.cn

Abstract

Background/Objectives: *Brassica juncea* (L.) Czern. (*B. juncea*) is widely adaptable with strong tolerance to drought, poor soil, pests, and diseases. *R2R3-MYB* transcription factors govern plant development, hormone signalling, and stress responses. To date, the *BjR2R3-MYB* family has not been systematically characterised in *B. juncea*, and the functions of its S12 subfamily in glucosinolate (GLS) biosynthesis and abiotic stress resistance remain unclear. This study aimed to perform a genome-wide analysis of the *BjR2R3-MYB* family and explore the roles of S12 subfamily members in GLS accumulation and stress response. Methods: Genomic sequences, protein sequences, and GFF3 annotation files of *B. juncea* (genome assembly ASM1870372v1) were retrieved from Ensembl Plants and used to identify and characterise the *BjR2R3-MYB* family genes, including their chromosomal distribution, subcellular localisation, conserved domains, gene structure, motifs, and cis-elements. Transcriptome profiling was performed using two *B. juncea* accessions contrasting in seed GLS content (X398, high GLS; X402, low GLS). qRT-PCR was used to detect expression changes of S12 genes in X398 under PEG-induced drought, NaCl salt stress, and exogenous methyl jasmonate (MeJA) treatment. Results: In total, 151 *BjR2R3-MYB* genes were identified, unevenly distributed across 18 chromosomes and predicted to encode nuclear-localised proteins. Genes within the same subfamily shared conserved structural features. Transcriptome data showed that, compared with X402, *BjMYB68* was markedly highly expressed in ovaries, seeds, and SP at different developmental stages of high-GLS accession X398, whereas *BjMYB109* was downregulated in the SP (at both 30 DAP and 40 DAP) of X398; *BjMYB22* and *BjMYB59* showed no significant GLS-associated expression difference. Stress treatments revealed that *BjMYB68* is responsive to both drought and the MeJA signalling pathway, *BjMYB59* is responsive to both salt and MeJA signalling pathways, and *BjMYB109* is only responsive to the MeJA signalling pathway. Conclusions: These findings provide a systematic framework for understanding the *BjR2R3-MYB* gene family, identify three key S12 subfamily members (*BjMYB68*, *BjMYB109*, and *BjMYB59*) with potential roles in regulating seed GLS synthesis and abiotic stress resistance, and lay a solid foundation for the genetic improvement of *B. juncea*. Next, further functional verification for candidate genes *BjMYB68* and *BjMYB109* and *BjMYB59* need be performed through complementary assay as well as the combined analysis of GLS content and gene expressions of multiple tissues and developmental stages to elucidate their regulatory mechanisms.

Check for updates

Received: 24 October 2026
Revised: 13 November 2026
Accepted: 15 November 2026
Published: 18 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: *Brassica juncea*; BjR2R3-MYB; genome-wide identification; transcriptome analysis; qRT-PCR; abiotic stress; S12 subfamily

1. Introduction

Based on amino acid sequence conservation in *Arabidopsis thaliana* (*A. thaliana*), the *R2R3-MYB* transcription factor family is divided into distinct subfamilies with unique functions, among which the S12 subfamily (comprising *AtMYB28*, *AtMYB29*, *AtMYB34*, *AtMYB51*, *AtMYB76*, *AtMYB122*) forms a precise regulatory network for glucosinolate (GLS) biosynthesis and stress responses. Notably, S12 subfamily members exhibit pathway-specific regulation of GLS biosynthesis: *AtMYB28*, *AtMYB29*, and *AtMYB76* directly or indirectly regulate aliphatic GLS synthesis (e.g., *AtMYB28/29* activate key biosynthetic enzymes, while *AtMYB76* modulates sulphur source distribution) [1–4], whereas *AtMYB34*, *AtMYB51*, and *AtMYB122* specifically govern indole GLS biosynthesis [5]. Genetic evidence in *Arabidopsis* confirms this specificity; *myb28myb29* double mutants show reduced aliphatic GLS and pest resistance [6], while *myb34myb51myb122* triple mutants exhibit decreased indole GLS [5]. Similar regulatory patterns extend to *Brassica* crops: overexpression of *BnaC2.MYB28* in *B. napus* or *BoMYBA29* in *B. oleracea* enhanced aliphatic GLS accumulation, improving insect resistance or anti-cancer metabolite production [1,2]. In *B. juncea*, copy number variation of the *MYB28* gene is responsible for the differences in GLS accumulation in various populations [7]. In *B. oleracea*, CRISPR-Cas9 gene editing technology was employed to knock out *MYB28*, leading to down-regulation of the expression of multiple genes involved in aliphatic GLS biosynthesis [8]. In *B. juncea*, there were also S12 subfamily genes homologous to those of *A. thaliana*, such as *BjuMYB28*, which was homologous to *AtMYB28; BjuMYB51, BjuMYB34*, and *BjuMYB122* were homologous to *AtMYB51, AtMYB34*, and *AtMYB122*, respectively; and *BjuMYB59* was homologous to *AtMYB59*. Therefore, the S12 subfamily genes in *B. juncea* would also be responsible for GLS biosynthesis and stress responses.

Additionally, there are hormone-responsive elements in *MYB* promoters, such as abscisic acid and methyl jasmonate (MeJA) responsive elements, which facilitate the integration of hormonal signals with metabolic pathways. For example, treatment of pakchoi with MeJA induced *MYB28* expression, thereby elevating its aliphatic GLS levels [9]. After *A. thaliana* was infected by a pathogen, its MeJA content was improved and so as to elevate *AtMYB122* expression level [5]. Moreover, brassinosteroids could repress the expression of *MYB51* and *MYB29* in *Arabidopsis* and *Brassica* crops, thereby regulating their GLS biosynthesis pathway [10]. Therefore, if the hormone binds to hormone-responsive elements in *MYB* promoters, it will regulate the expression level of the *BjR2R3-MYB* S12 subfamily genes so as to affect the GLS biosynthesis pathway and GLS content. In addition, GLS are a class of secondary metabolites containing nitrogen and sulphur, and are widely present in plants of the *Brassicaceae* family [11]. GLS not only play a significant role in protecting plants from microbial pathogens and insect attacks [12], but their hydrolysis products are beneficial to human health by preventing chronic diseases, acting as antioxidants, anti-cancer agents, and cancer prevention, as well as regulating the body's immune system [13,14].

Brassica juncea L. (*B. juncea*), as one of the three cultivated varieties of rapeseed, has excellent characteristics such as drought resistance, cold resistance, pod cracking resistance, disease and pest resistance, and wide planting range. However, the excessively high GLS content limits the application of *B. juncea* [15,16]. Because GLS hydrolysis products will remain in the rapeseed cake meal, it will lead to goitre as well as abnormal liver and kidney function in livestock that have eaten the rapeseed cake meal [15]. Therefore, the GLS content

in seeds being less than 30 μmol/g becomes an important indicator for the registration of new rapeseed varieties [16]. In a word, genome-wide identification of *R2R3-MYB* family genes facilitates clarifying the molecular mechanism of GLS metabolism in *B. juncea*, so as to lay the foundation for breeding new *B. juncea* varieties with different GLS contents according to actual needs, such as *B. juncea* varieties with GLS contents in seeds below 30 μmol/g, whose rapeseed cake meal can be used as livestock feed, while retaining or enhancing GLS content in vegetative tissues for protecting plants from microbial pathogens and insect attacks, as well as *B. juncea* varieties with GLS contents in seeds higher than 200 μmol/g, which can be used for extracting GSL hydrolysis products that are beneficial to human health.

In addition, previous studies have shown that there were also some variations in GLS composition and accumulation between cultivated *Brassica* plant species and their wild relative species [17,18]. For example, the diversity of GLS components for cultivated *Brassica* plant species was reduced, and their GLS accumulation levels were adjusted to satisfy agronomic and edible demands through long-term domestication and directional selection. In contrast, wild relative species maintained more diversified GLS types and highly variable contents under natural selection. Moreover, these GLS metabolic divergences were tightly associated with plant development, environmental adaptation, and biochemical diversity. For instance, differential GLS abundance can coordinate tissue growth and developmental progression, while diverse GLS variants can also serve as defensive compounds against herbivores and pathogens and enhance tolerance to abiotic stresses. Therefore, GLS metabolic divergence between cultivated *Brassica* species and wild relative species not only enriched the biochemical diversity of *Brassica* plant species, but also laid the genetic foundation for dissecting the regulatory network of GLS biosynthesis in *B. juncea*, as well as the expression patterns of the *BjR2R3-MYB* S12 subfamily genes in *B. juncea* after stress treatments for this study.

As we all know, the *R2R3-MYB* transcription factor is an important regulatory factor involved in plant growth and development, secondary metabolism, hormone signal transduction, and stress responses [19–21]. However, the *BjR2R3-MYB* family in *B. juncea* has not been systematically investigated, and the regulatory roles of the S12 subfamily in GLS biosynthesis and abiotic stress resistance remain unclear [22]. Furthermore, with the completion of the genome sequencing for *B. juncea*, *B. rapa*, and *B. nigra*, it becomes possible to research the *R2R3-MYB* S12 subfamily genes in *B. juncea* through bioinformatics [23–25]. Consequently, we analysed the phylogenetic relationships, cis-regulatory elements, conserved motifs, and homology of the *BjR2R3-MYB* family genes in *B. juncea*, as well as the expression patterns of the *BjR2R3-MYB* S12 subfamily genes in *B. juncea* after treatments with three different abiotic stresses. Additionally, RNA-Seq analysis was conducted using two *B. juncea* accessions with high (X398) and low (X402) GLS content, respectively, to investigate the differential expression of the *BjR2R3-MYB* S12 subfamily genes between X398 and X402. These findings provide a systematic framework for understanding the *BjR2R3-MYB* gene family and lay a theoretical basis for investigating the potential biological functions of the *BjR2R3-MYB* S12 subfamily genes in *B. juncea* for GLS content and abiotic stress resistance.

2. Materials and Methods

2.1. Plant Materials

Two *B. juncea* inbred accessions, X398 was collected from Drolma District, Lhasa city, Xizang province in China, and X402 was from Xining city of Qinghai province in China, were both used for tissue-specific expression analysis of *BjR2R3-MYB S12* subfamily genes. X398 was also used for stress treatments and qRT-PCR analysis. These two *B. juncea* inbred

accessions have been self-pollinated for more than seven generations and are fixed for GLS content. All materials were provided by the Qinghai Academy of Agricultural and Forestry Sciences at Qinghai University, China.

2.2. Determination of the Total GLS Content

The total GLS content of X398 and X402 was determined using high-performance liquid chromatography (HPLC) with a chromatographic instrument (Agilent Technologies, Santa Clara, CA, USA) and test standard (ISO9167-1) [26]. A reversed-phase chromatographic column TC-C18 (4.60 × 250 mm, 5 um particle size; Agilent Technologies, CA, USA) was used.

2.3. Genome-Wide Identification and Systematic Characterisation of the BjR2R3-MYB Gene Family in Brassica juncea

Genomic sequences, protein sequences, and GFF3-format gene-model annotations of *B. juncea* (genome assembly ASM1870372v1, containing 18 pseudomolecules from A and B subgenomes) were downloaded from Ensembl Plants (https://plants.ensembl.org/, accessed on 16 June 2025). A total of 126 *A. thaliana R2R3-MYB* protein sequences were retrieved from PlantTFDB (https://planttfdb.gao-lab.org/, accessed on 16 June 2025), and Swiss-Prot reference data were obtained from UniProt release 2025_05 (https://www.uniprot.org/, accessed on 16 June 2025).

Candidate *BjR2R3-MYB* members were identified using TBtools-II (v2.142) [27], via a multi-step filtering pipeline with fixed parameters for reproducibility. First, BLASTP (BLAST+ v2.142) searches were performed using *A. thaliana R2R3-MYB* proteins as queries against the full proteome of *B. juncea* ASM1870372v1 with an E-value $\leq 1 \times 10^{-10}$, and all other parameters were kept at default [28]. BLASTP hits were further filtered against the Swiss-Prot subset of UniProt release 2025_05, and sequences without curated *MYB* family annotations were discarded. Second, hidden Markov model searches were conducted in HMMER (v3.3.2) using the Pfam MYB-DNA-binding profile PF00249 [29]; sequences with a bit score ≥ 50 and an E-value $\leq 1 \times 10^{-10}$ were retained [30]. Third, domain architectures of all candidate proteins were individually verified using the Normal mode of the SMART server (http://smart.embl-heidelberg.de/, accessed on 17 June 2025) and NCBI Batch CD-Search (v3.20) against the Conserved Domain Database (CDD; https://www.ncbi.nlm.nih.gov/cdd, accessed on 18 June 2025) [29]. Only proteins containing two complete, non-overlapping R2 and R3 *MYB* repeats were kept; sequences with a single repeat or truncated *MYB* domains were excluded. After filtering, a non-redundant set of 151 *BjR2R3-MYB* genes was obtained and sequentially named *BjMYB1-BjMYB151* (Supplementary Table S1).

Protein physicochemical properties, including amino-acid length, molecular weight, theoretical isoelectric point (pI), instability index, and aliphatic index, were calculated with the Protein Properties tool embedded in TBtools-II (v2.142) [27]. Proteins were classified as acidic (pI < 7.0), neutral (pI = 7.0), or alkaline (pI > 7.0) using pI = 7.0 as the classification threshold. Subcellular localisation was predicted by WoLF PSORT (https://wolfpsort.hgc.jp/, accessed on 18 June 2025) [31] with the plant-specific database and default settings, and the compartment with the highest prediction score was recorded for each protein.

For phylogenetic analysis, 151 *BjR2R3-MYB* and 126 *AtR2R3-MYB* protein sequences were aligned by ClustalW 2.1. A neighbour-joining (NJ) tree was constructed in MEGA-X (v10.2.6) with the p-distance model and 1000 bootstrap replicates to evaluate branch reliability [32,33]. The phylogenetic tree was exported and visualised in iTOL v7 (https://itol.embl.de/index.shtml, accessed on 18 June 2025). Chromosomal locations of *BjR2R3-MYB* genes were plotted in TBtools-II (v2.142) [27] based on positional information

parsed from the ASM1870372v1 GFF3 file. Exon-intron organisation (3′-UTR, CDS, intron) was extracted from the same GFF3 annotation. Conserved protein motifs were detected by MEME Suite 5.5.5 with a maximum of 10 motifs under default parameters [34] and domain-architecture diagrams were generated from Batch CD-Search outputs within TBtools-II (v2.142).

2.4. Collinearity Analysis and Cis-Acting Element Prediction

Genomic sequences and GFF3 annotations of *B. rapa* (Chiifu_V4.1) and *B. nigra* (Ni100_V2) were downloaded from BRAD (http://brassicadb.cn, accessed on 16 June 2025) [35]. The *A. thaliana* genome (TAIR10) and its GFF3 annotation were obtained from Ensembl Plants, and all datasets were used exactly as released, without re-annotation. Interspecies collinearity between *BjR2R3-MYB* genes and those from *A. thaliana*, *B. rapa*, and *B. nigra*, as well as intraspecies collinearity within *B. juncea*, were analysed using TBtools-II (v2.142) [27]. Cis-acting elements in the 2000 bp promoter regions (upstream of ATG) of *BjR2R3-MYB* genes were predicted via PlantCARE (https://ngdc.cncb.ac.cn/databasecommons/database/id/5261, accessed on 17 June 2025) with default parameters, and each element was assigned to a hormone-, stress-, development-, or metabolism-related category according to the PlantCARE description and counted per gene [36] and visualised with TBtools-II (v2.142) [27].

2.5. Tissue-Specific Expression Analysis of BjR2R3-MYB S12 Subfamily Genes

For identifying potential roles in regulating seed GLS synthesis of *BjR2R3-MYB* S12 subfamily genes in *B. juncea*, two *B. juncea* inbred accessions with high (X398, GLS = 154.88 μmol/g) and low (X402, GLS = 25.54 μmol/g) GLS content were both cultivated in the experimental field of Qinghai Academy of Agricultural and Forestry Sciences in Xining city of Qinghai province in China (altitude 2225 m, 101°49′ N, 36°34′ E, temperature ranged from 2 to 25 °C, and the duration of sunshine was between 7 and 8.5 h per day) from April to July 2025. Each accession was grown in 5-rowed plots consisting of 60 plants. The length and width of the rows were 2 and 0.3 m, respectively. For transcriptome analysis, during the initial flowering period, ovaries at 0 and 10 days after pollination (DAP) were sampled, and during the blooming period, seed and SP (silique pericarp) at 20, 30, and 40 DAP were sampled, respectively. Three biological replicates were set for each tissue at each developmental stage. All samples were immediately frozen in liquid nitrogen, then stored at −80 °C before lyophilisation. All clean reads were mapped to the *B. juncea* reference genome (ASM1870372v1) using HISAT2 (v2.2.1, https://daehwankimlab.github.io/hisat2/, accessed on 18 June 2025), and gene expression levels were quantified as fragments per kilobase of transcript per million mapped reads (FPKM) using StringTie (v1.3.6, https://ccb.jhu.edu/software/stringtie/, accessed on 18 June 2025) through the standard RNA-seq analysis pipeline. The expression patterns of four *BjR2R3-MYB* S12 subfamily genes (*BjMYB68*, *BjMYB109*, *BjMYB22*, *BjMYB59*) that may be involved in GLS biosynthesis were analysed using FPKM values, and a heatmap was developed using TBtools-II (v2.142) software [27].

2.6. Different Stress Treatments and qRT-PCR Analysis

Seeds of a high GLS *B. juncea* inbred accession (X398) were sown in three pots with sterilized mixed soil (soil:vermiculite = 3:1). Then three pots were placed in growth chamber with the condition of 25 °C and 16 h light/8 h dark photoperiod for 30 days, and three pots of plantlets of 30 days (at the six-leaf stage) were treated with 200 mmol/L NaCl, 10% PEG 6000, or 200 μmol/L methyl jasmonate (MeJA), respectively [37]. Untreated plantlets were used as a control (CK). Three biological replicates were set for each treatment. For each treatment, 3 leaves from 3 different plantlets were sampled at 3, 6, 9, 12, and 24 h after

treatment, respectively, snap-frozen in liquid nitrogen, and stored at −80 °C [38]. Total RNA from leaf tissue was extracted using RNAprep Pure Plant Kit (DP432) (Tiangen Biotech, Beijing, China) and reverse-transcribed to cDNA using a FastKing cDNA First Strand Synthesis Kit (Genome Removal) (KR116) (Tiangen Biotech, Beijing, China). The reference gene *BjuB036217*, a Tonoplastic intrinsic protein, was used as an internal control [39]. The primers were designed using Primer-3 software (https://primer3.ut.ee/, accessed on 20 June 2025) (Supplementary Table S4). RT-qPCR was performed as previously described; three technical triplicates were set per sample [40]. The expression fold change of the differentially expressed transcripts was estimated using the $2^{-\Delta\Delta CT}$ method [41,42].

2.7. Statistical Analysis

For the stress treatment experiment, it was comprised of two crossed factors, treatment (four levels, including control) and sampling time (six levels), with three independent biological replicates for each treatment. For each gene, relative expression was compared among different sampling times within a treatment by one-way ANOVA (factor: sampling time, six levels) followed by Duncan's multiple range test in SPSS 26.0 (IBM, USA), and differences were declared significant at * $p < 0.05$ and ** $p < 0.01$. Values are means ± standard error (SE) of the three biological replicates, each of which was the mean of three technical replicates. Normality of the residuals and homogeneity of variances were checked with the Shapiro–Wilk and Levene tests before ANOVA; no data transformation was applied because the assumptions for parametric analysis were satisfied.

The transcriptome analysis experiment also comprised three biological replicates for every accession × tissue × stage combination; FPKM values of the four S12 subfamily genes were compared between X398 and X402 using independent-samples *t*-tests with false discovery rate (FDR) correction for multiple comparisons, and a gene was scored as not detected (ND) when its FPKM was zero in all three replicates.

3. Results

3.1. Identification and Physicochemical Characterisation of BjR2R3-MYB Genes

The conservative domain homology analysis was performed using the amino acid sequences of *R2R3-MYBs* from *A. thaliana* as query sequences via the HMM [29] and the conserved domains from the CDD [29]. Ultimately, 151 *BjR2R3-MYB* family genes were identified in *B. juncea*, renamed from *BjMYB1* to *BjMYB151*. For the *BjR2R3-MYB* family genes, the numbers of amino acids ranged from 146aa (*BjMYB124*) to 567aa (*BjMYB149*); the molecular weights ranged from 17032.25Da (*BjMYB124*) to 62742.41Da (*BjMYB149*); the theoretical isoelectric points ranged from 10.48 (*BjMYB124*) to 4.70 (*BjMYB11*) (Supplementary Table S1). For the *BjR2R3-MYB* family proteins, 60.26% were acidic, 39.07% were alkaline and only one was neutral (*BjMYB131*) while the theoretical isoelectric point (pI = 7.0) was set as the cutoff, and the instability indexes and aliphatic indexes ranged from 32.8 (*BjMYB130*) to 77.43 (*BjMYB148*) and 43.67 (*BjMYB45*) to 81.84 (*BjMYB102*), respectively (Supplementary Table S1). Furthermore, we found that the predicted proteins were all located in the nucleus.

3.2. Chromosomal Localisation of the BjR2R3-MYB Family Genes

In the present study, systematic analysis of the chromosomal distribution of *BjR2R3-MYB* family genes was conducted utilising the genomic annotation data of *B. juncea* and TBtools-II software (v2.142) [27]. The results revealed that A and B subgenomes harboured 75 and 76 *BjR2R3-MYB* family genes, respectively (Figure 1). So, no skewed distribution was found for *BjR2R3-MYB* family genes between the A and B subgenomes. Nevertheless, the *BjR2R3-MYB* family genes exhibited an asymmetric distribution across the 18 chromosomes; some chromosomes contained a larger number of *BjR2R3-MYB* genes, for example, 16, 14, 12 and 13 *BjR2R3-MYB* genes on ChrA03, ChrA09, ChrB02 and ChrB08, respectively. Only 3, 3, 4, and 4 *BjR2R3-MYB* genes were on ChrA04, ChrA05, ChrA08 and ChrA10, respectively, which contained fewer *BjR2R3-MYB* genes. In addition, for four *BjR2R3-MYB* S12 subfamily genes, *BjMYB22* was located on chromosome ChrA03; both *BjMYB68* and *BjMYB59* were located on chromosome ChrA09; and *BjMYB109* was located on chromosome ChrB04.

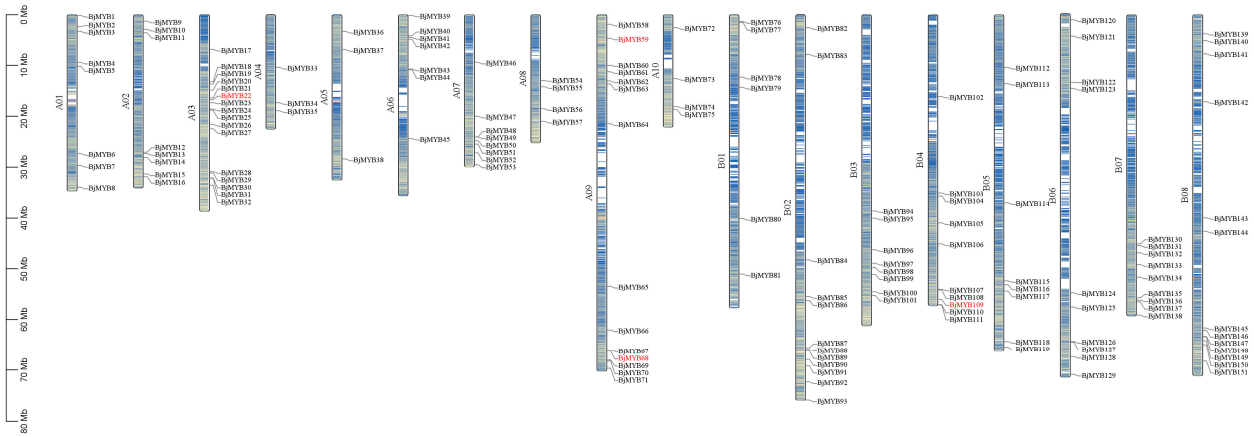


Figure 1. Chromosomal localisation of *BjR2R3-MYB* family genes. The scale bar on the left represents the lengths of the chromosomes (Mb). The candidate S12-subfamily *BjR2R3-MYB* genes (*BjMYB22*, *BjMYB59*, *BjMYB68* and *BjMYB109*) are highlighted in red.

3.3. Phylogenetic Analysis of the BjR2R3-MYB Gene Family

We constructed a phylogenetic tree with amino acid sequences of 151 *BjR2R3-MYB* proteins (Supplementary File S10) in *B. juncea* and 126 *AtR2R3-MYB* proteins in *A. thaliana* (Figure 2), which were divided into 26 subgroups according to sequence similarity and topological structure; the 26 subgroups were named B1-B26. No *BjR2R3-MYB* proteins were identified in subgroups B3 (S5) and B10 (S7), which exclusively comprised *AtR2R3-MYB* proteins. In contrast, the remaining 24 subgroups contained both *BjR2R3-MYB* and *AtR2R3-MYB* proteins. Interestingly, the number of *BjR2R3-MYB* proteins was about 1.5–2.5 times that of *AtR2R3-MYB* proteins in B1 (S15), B2 (S18), B6 (S22), B9 (S17), B19 (S1) and B26 (S4) subgroups, therefore, *BjR2R3-MYB* proteins could undergo specific amplification compared to *Arabidopsis*. Furthermore, subgroups B9 (S17), B19 (S1), B24 (S10) and B25 (S12) all contain two distinct clades. This divergence is attributed to differences in C-terminal regulatory regions, conserved motif compositions, gene structures, and cis-acting elements, implying functional specialisation. However, these clades were retained within the same subgroups due to high homology in the core R2/R3 domains and consistent topological clustering with *Arabidopsis R2R3-MYB* subfamilies [18,41].

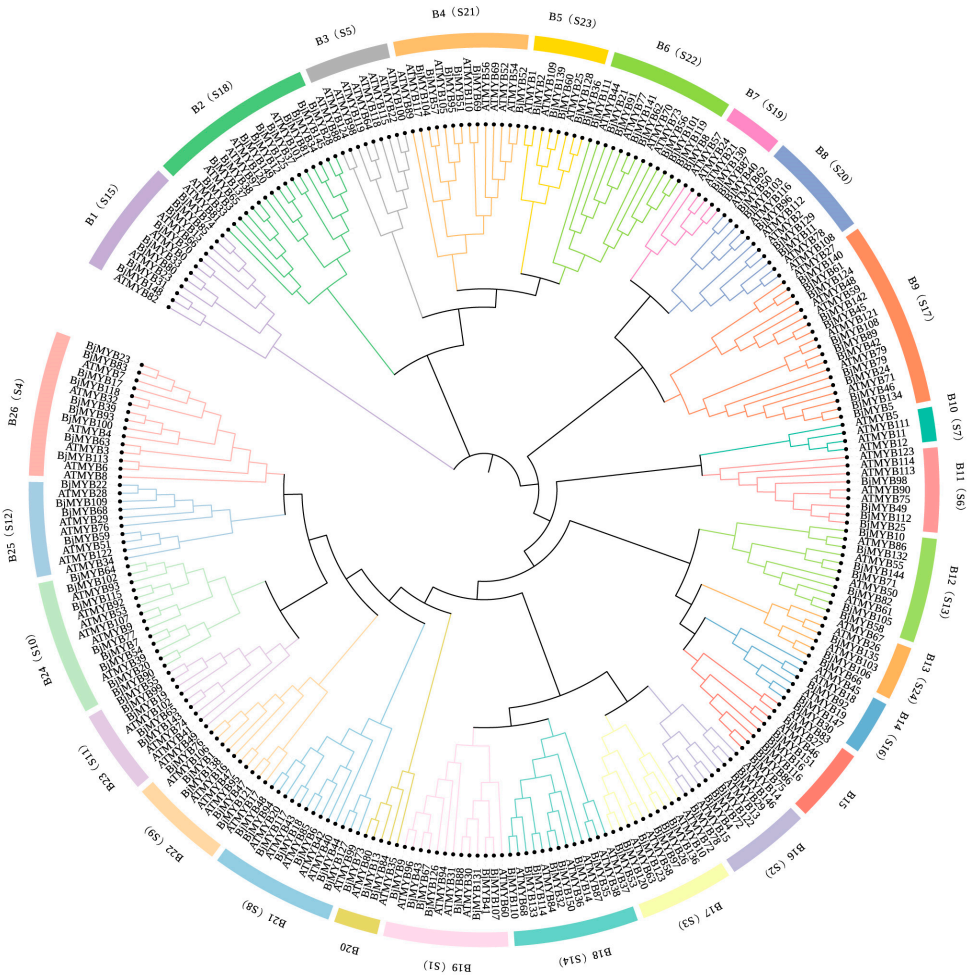


Figure 2. Phylogenetic analysis of *BjR2R3-MYB* proteins in *B. juncea* and *A. thaliana*. Different colours represent different subfamilies, and the corresponding subfamily names were marked on the outer circle.

3.4. Motif Composition, Conserved Domains, and Gene Structure Analysis of the BjR2R3-MYB Gene Family

We predicted 10 conserved motifs in 151 *BjR2R3-MYB* family proteins using the online MEME program and visualised the length and conserved sequences of each motif using TBtools-II software (v2.142) [27] (Figure 3). Our research revealed that all members of the *BjR2R3-MYB* family contain motifs 1 and 2, with the types and numbers of these motifs exhibiting a high degree of conservation within each subfamily. The conservation of the core DNA-binding pattern reflects the evolutionary stability of the basic function of *BjR2R3-MYB* proteins [41]. Furthermore, each subfamily has a specific motif pattern. For instance, motifs 8 and 9 are only present in the B19 (S1) subfamily, whereas motif 7 is exclusively identified in the B6 (S22) subfamily. The emergence of subfamily-specific motifs is a sign of functional specialisation during the expansion of the gene family [41]. For example, the motifs 8 and 9 specially existing in the B19 (S1) subfamily may help form specific regulatory protein complexes, thereby finely regulating downstream gene expression in the specific biological processes of *B. juncea*. Motif 7, specifically existing in the B6 (S22) subfamily genes, was found to be associated with subfamily-specific regulatory functions.

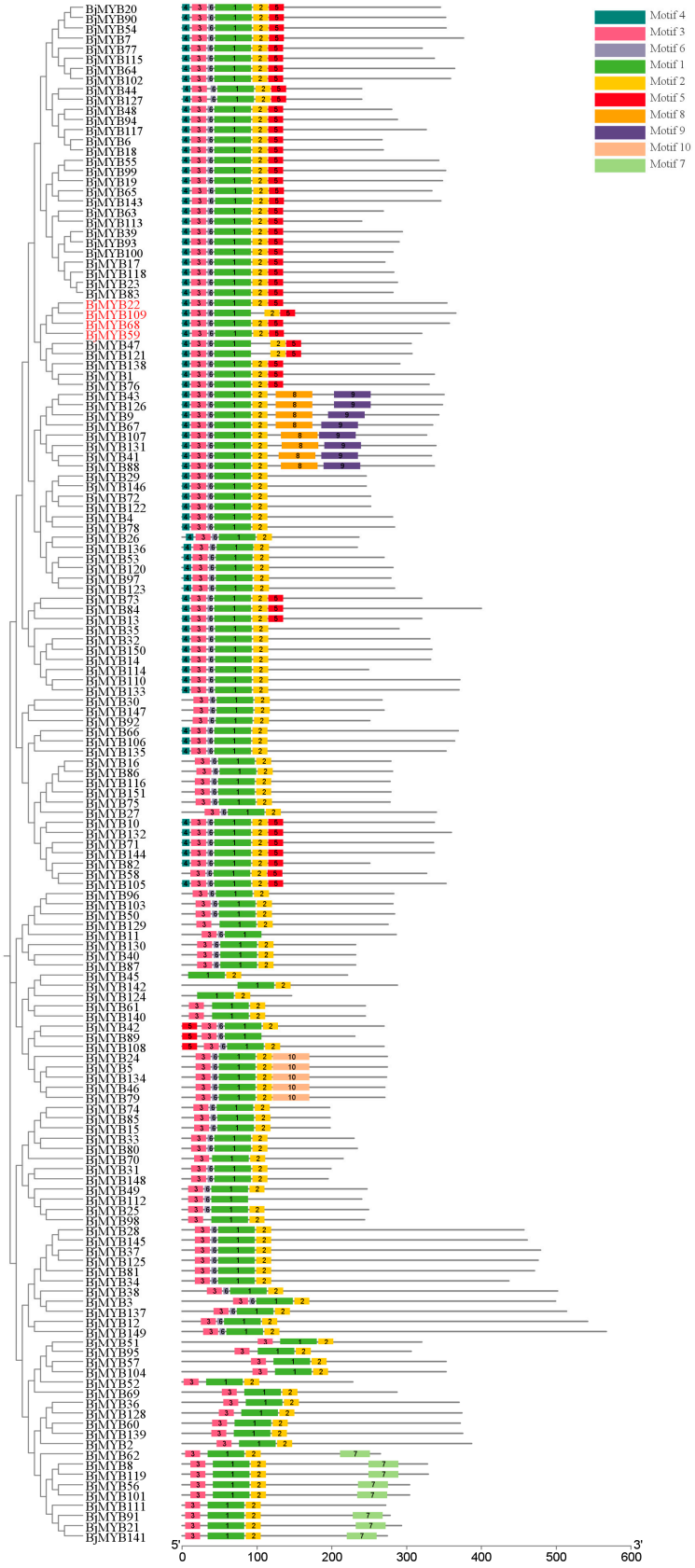


Figure 3. Phylogenetic tree of 151 *BjR2R3-MYB* proteins and conserved motifs for *BjR2R3-MYB* proteins in *B. juncea*. Different motifs are shown with different coloured boxes and numbers (1–10). The grey lines represent the non-conserved sequences. The lengths of motifs can be estimated using the scale at the bottom. The candidate S12-subfamily *BjR2R3-MYB* genes (*BjMYB22*, *BjMYB59*, *BjMYB68* and *BjMYB109*) are highlighted in red.

To further explore the conserved structural domains of the *BjR2R3-MYB* proteins, multiple comparisons of the 151 *BjR2R3-MYB* protein sequences were conducted using TBtools-II software (v2.142) [27]. The results show that the vast majority of genes (99%, that is, 150 genes) contain 1 to 4 coding regions, among which 101 *BjR2R3-MYB* genes have 3 coding regions (Figure 4). In contrast, the *BjMYB84* gene has five coding regions, which is the highest number of coding regions among all 151 *BjR2R3-MYB* genes. While, nine *BjR2R3-MYB* genes, including *BjMYB62*, *BjMYB8*, *BjMYB119*, *BjMYB56*, *BjMYB101*, *BjMYB111*, *BjMYB91*, *BjMYB21*, *BjMYB141*, all lack introns and have only one exon, moreover, they all belong to the B6 (S22) subfamily. Interestingly, the B6 (S22) subfamily genes not only have no introns but also contain the unique motif 7; motif 7 may confer subfamily-specific regulatory functions, while intron deletion further optimises the expression dynamics, thereby prompting the B6 (S22) subfamily genes to possess a coordinated evolutionary adaptation. In addition, for *BjR2R3-MYB* genes, most sequence lengths were less than 6 kb, with only one gene, *BjMYB84*, exceeding 15 kb because of its longer introns (Figure 4). Consequently, despite variations in the lengths of coding regions and introns, a significant degree of structural similarity, including their distribution and quantity, is evident among *BjR2R3-MYB* genes within the same subfamily.

3.5. Analysis of Cis-Acting Elements in BjR2R3-MYB Gene Promoters

In this study, a total of 32 cis-acting elements were identified in the promoter region of the *BjR2R3-MYB* gene (Supplementary Table S5). These elements are associated with various biological processes, including hormone signalling, environmental stress response, growth and development regulation, metabolic control, and light signal transduction (Figure 5). Specifically, five hormone-responsive elements associated with auxin, gibberellin, abscisic acid, salicylic acid, and MeJA signalling pathways were detected. Additionally, six elements were implicated in environmental stress responses, encompassing defence mechanisms, low-temperature adaptation, anaerobic induction, drought tolerance, and wound response. Regulatory elements associated with growth and development were also identified, including those involved in cell cycle regulation, seed-specific regulation, meristematic tissue expression, root specificity, and endosperm expression, among others. Furthermore, regulatory elements involved in flavonoid biosynthesis and maize alcohol-soluble protein metabolism were also identified in the *BjR2R3-MYB* genes. The coexistence of these elements reflects the complexity and versatility of *BjR2R3-MYB* genes. In addition, 32 types of cis-acting elements were identified in *BjR2R3-MYB* gene promoters (Figure 5); furthermore, there is a clear correlation between cis-element type and subfamily classification. For example, the B6 (S22) subfamily genes contained a high proportion of stress-responsive elements and seed-specific regulatory elements, which is consistent with the fact that they are a type of genes related to rapid stress adaptation and developmental regulation [19].

3.6. Collinearity Analysis of the BjR2R3-MYB Family Genes

In order to systematically identify the homology retention and differentiation of the *BjR2R3-MYB* family genes in *B. juncea* (AABB) during the tetraploid formation process. This study conducted a genome-wide collinearity analysis by comparing *B. juncea* with its diploid progenitors, *B. rapa* (AA) and *B. nigra* (BB), and the model species *A. thaliana* (Figure 6, Supplementary File S6). The results showed that a total of 402, 409 and 218 homologous gene pairs were identified between *B. juncea* and *B. rapa*, *B. nigra*, as well as *A. thaliana*, respectively. Therefore, the *B. juncea BjR2R3-MYB* retained the gene collinearity characteristics with its diploid progenitors (both *B. rapa* and *B. nigra*), reflecting genetic conservation at the subgenomic level.



Figure 4. Phylogenetic tree of 151 *BjR2R3-MYB* proteins; conserved domains of the *BjR2R3-MYB* genes in *B. juncea*; gene structure of *BjR2R3-MYBs* in *B. juncea*. Positions of the 3′UTR, CDS, and intron are indicated by blue squares, pink squares and grey lines. The candidate S12-subfamily *BjR2R3-MYB* genes (*BjMYB22*, *BjMYB59*, *BjMYB68* and *BjMYB109*) are highlighted in red.



Figure 5. Predicted cis-Acting Elements in *BjR2R3-MYB* promoters. Different colours represent the different types of cis-Acting Elements.

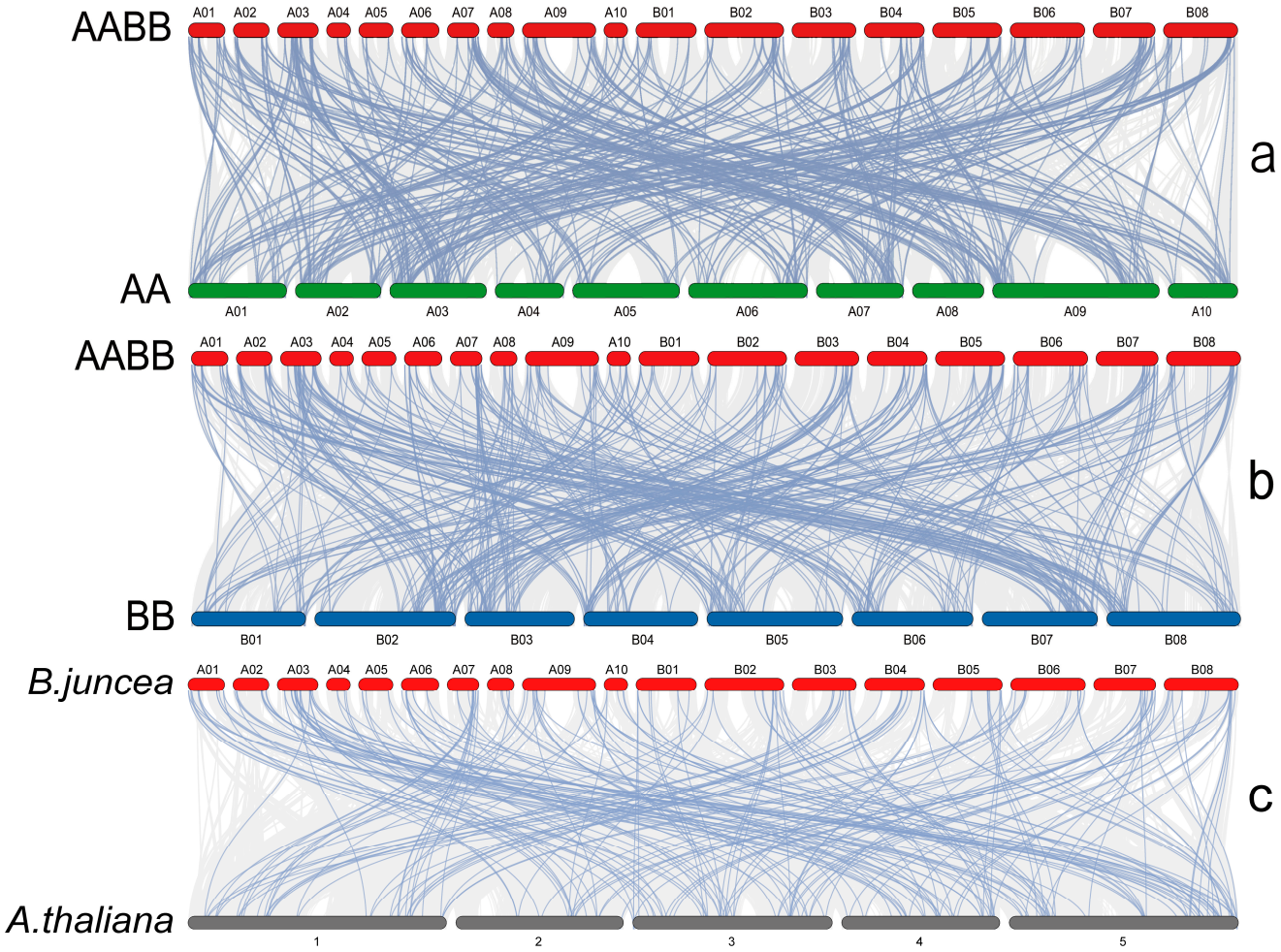


Figure 6. Collinearity analysis of *BjR2R3-MYB* genes between *B. juncea* (AABB) and *B. rapa* (AA), *B. nigra* (BB), and *A. thaliana*. (**a**) Collinearity between *B. juncea* and *B. rapa*; (**b**) Collinearity between *B. juncea* and *B. nigra*; (**c**) Collinearity between *B. juncea* and *A. thaliana*. Letters and numbers denote chromosome IDs of each species. Grey background lines represent collinear gene pairs between the *B. juncea* genome and the other reference genomes, while blue lines highlight collinear *BjR2R3-MYB* gene pairs between *B. juncea* and these species.

Furthermore, an intraspecific collinearity analysis was conducted to elucidate the replication mechanism of the *BjR2R3-MYB* gene family in *B. juncea* (Figure 7). A total of 252 intraspecific collinear gene pairs were identified in *B. juncea* (Supplementary File S11). The largest number of collinear gene pairs was found on chromosomes A3 and A1, with 47 and 36 pairs, respectively, whereas chromosomes B4 and B6 exhibited the smallest number, with only five pairs each (Supplementary File S11). The nonsynonymous (Ka) and synonymous (Ks) substitution rates, as well as their ratio (Ka/Ks), were estimated for the 252 collinear gene pairs. As a result, the Ka/Ks values ranged from 0.054 to 0.822, with a median of 0.255, and notably, all 252 gene pairs (100%) all exhibited a Ka/Ks value of less than 1 (Supplementary File S11). These results indicate that the collinear *BjR2R3-MYB* gene pairs have predominantly undergone strong purifying selection during evolution, thereby retaining functionally conserved protein sequences, which is consistent with the stable functional differentiation of the *BjR2R3-MYB* gene family. Notably, A3 and A1 may serve as core regions for the expansion of the *BjR2R3-MYB* gene family, whereas B4 and B6 may represent evolutionarily conserved regions or regions with low replication activity. Interchromosomal collinearity analysis revealed that chromosome A1 was primarily collinearly associated with chromosomes A3 and B1, whereas chromosome A3 was pre-

dominantly collinearly associated with chromosomes B2 and B8. These patterns collectively suggest that the *B. juncea* genome has undergone whole-genome duplication events or allopolyploidisation processes.

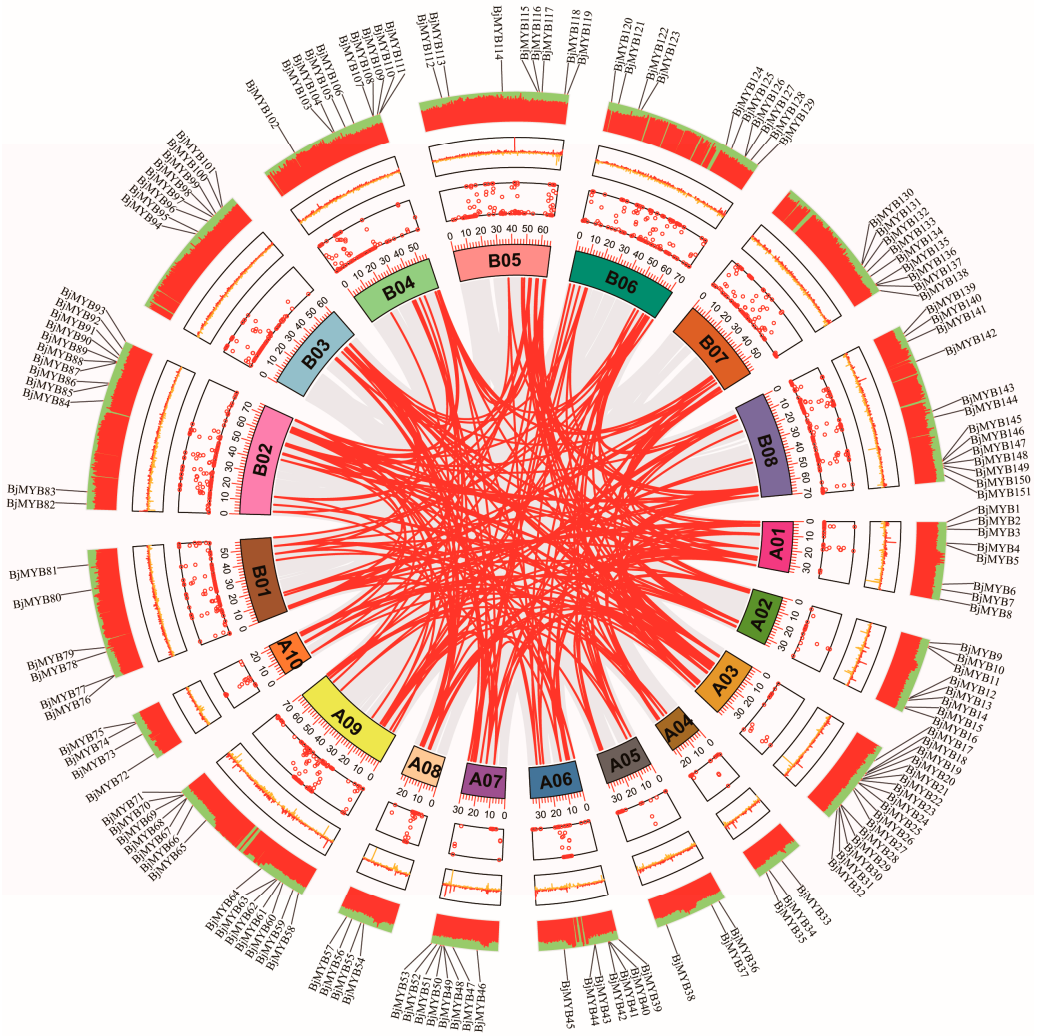


Figure 7. Intraspecific collinearity analysis of *BjR2R3-MYB* genes in *B. juncea*. Red lines indicate collinear *BjR2R3-MYB* gene pairs. The coloured blocks correspond to different chromosomes of *B. juncea*. From the inside out, the scatter points show the N ratio (Supplementary File S7), the lines show GC skew (Supplementary File S8), and the bars show the GC ratio (Supplementary File S9).

3.7. Expression Pattern Analysis of the BjR2R3-MYB S12 Subfamily Genes

Previous studies have shown that *R2R3-MYB* S12 subfamily genes may significantly influence GLS content in *A. thaliana* or Brassica crops. Furthermore, four *BjR2R3-MYB* S12 subfamily gene members, including B25(S12), *BjMYB68*, *BjMYB109*, *BjMYB22*, and *BjMYB59*, were identified in *B. juncea* using phylogenetic analysis of the *R2R3-MYB* gene family (Figure 2). According to gene annotation, *BjMYB68*, *BjMYB109*, and *BjMYB22* are all homologous to *AtMYB28* (*AT5G61420*) of *A. thaliana* and are involved in the biosynthesis of aliphatic glucosinolates. *BjMYB59* is homologous to *AtMYB51* (*AT1G18570*) of *A. thaliana* and participates in the biosynthesis of indole glucosinolates. Therefore, the four *BjR2R3-MYB* S12 subfamily genes identified in *B. juncea* may also be involved in the biosynthesis of glucosinolates.

Two *B. juncea* accessions with high-GLS (X398) and low-GLS (X402) were used for transcriptome analysis (Figure 8; Table S2), as a result, of the four S12 subfamily genes,

BjMYB68 was only detected to be markedly highly expressed in ovaries, seeds and SP at different developmental stages of high-GLS accession X398 (Figure 8, Table 1), wherein, with relatively high expression levels in the ovary (e.g., 26.36 ± 2.16 at 10 DAP) and the SP (e.g., 53.50 ± 5.28 at 20 DAP, 40.95 ± 12.08 at 30 DAP) in X398, whereas it was barely detected in low-GLS accession X402, indicating that the transcriptional activation of gene *BjMYB68* may be an important basis for GLS accumulation in seeds of high-GLS accession X398. In contrast, *BjMYB109* was down-regulated in SP (at both 30 DAP and 40 DAP) of X398 in comparison with X402; the expression level of *BjMYB109* in X402 was approximately 4.3- and 8.1-fold higher than in X398, respectively (Table 1). Beyond that, neither *BjMYB22* nor *BjMYB59* exhibited significant expression differences between X398 and X402. In consequence, two of the four *BjR2R3-MYB* S12 subfamily genes in *B. juncea* that may be involved in glucosinolate biosynthesis, *BjMYB68* and *BjMYB109*, exhibited significant expression differences between X398 and X402, indicating that the total GLS content in seeds of *B. juncea* was correlated with the expression of genes *BjMYB68* and *BjMYB109*.

Table 1. Relative expression levels (mean ± SE) of *BjMYB68* and *BjMYB109* in different tissues at various developmental stages of high-GLS (X398) and low-GLS (X402) accessions of *B. juncea*.

Tissue /Developmental Stage	*BjMYB68* (X398)	*BjMYB68* (X402)	*BjMYB109* (X398)	*BjMYB109* (X402)
ZF00	3.31 ± 0.07	ND	3.87 ± 0.29	3.42 ± 0.36
ZF10	26.36 ± 2.16	ND	15.31 ± 0.86	10.61 ± 1.70
ZZ20	0.17 ± 0.04	ND	0.01 ± 0.01	0.37 ± 0.24
ZZ30	0.28 ± 0.07	ND	0.04 ± 0.03	ND
ZZ40	0.18 ± 0.09	ND	ND	ND
JGP20	53.50 ± 5.28	ND	18.31 ± 0.47	35.04 ± 8.31
JGP30	40.95 ± 12.08	ND	13.23 ± 1.62	57.37 ± 9.91
JGP40	5.48 ± 1.41	ND	3.72 ± 0.78	30.01 ± 3.24

Note: ZF00 indicates ovary at 0 DAP; ZF10 indicates ovary at 10 DAP; ZZ20, ZZ30, and ZZ40 indicate seeds at 20, 30, and 40 DAP, respectively; JGP20, JGP30, and JGP40 indicate SP at 20, 30, and 40 DAP, respectively. ND, not detected (FPKM = 0 in all three biological replicates). Values are means ± SE of three biological replicates. Data are derived from the transcriptome analysis (Table S2).

3.8. Analysis of the BjR2R3-MYB S12 Subfamily Genes Under Abiotic Stress

To investigate the biological functions of the *BjR2R3-MYB* S12 subfamily genes in *B. juncea*, the *B. juncea* accession (X398) was utilised as the experimental material in this study. The plants were subjected to three treatments: simulated drought stress (10% PEG-6000), salt stress (200 mmol/L NaCl), and exogenous application of MeJA (200 µmol/L). The expression patterns of the four S12 subfamily genes were analysed using qRT-PCR (Figure 9, Table S3).

As a result, *BjMYB68* was significantly upregulated at 6 h after treatment with 10% PEG-6000 compared with 0 h. In contrast, no statistically significant differences were observed for the expression levels of *BjMYB22*, *BjMYB59*, and *BjMYB109*. *BjMYB59* exhibited significant upregulation at 12 h after treatment with 200 mmol/L NaCl compared with 0 h, whereas the expression levels of the other three genes, *BjMYB22*, *BjMYB68*, and *BjMYB109*, showed no notable fluctuations. Following treatment with 200 µmol/L MeJA, *BjMYB59* showed continuous and significant upregulation from 9 h to 24 h after treatment and reached its peak at 24 h, while both *BjMYB68* and *BjMYB109* exhibited significant upregulation at 3 h after treatment compared to 0 h; however, no significant changes in expression for *BjMYB22* were detected. In conclusion, the S12 subfamily genes exhibited distinct gene-specific and spatiotemporal expression differentiation characteristics under various treatment conditions, which is related to regulatory elements in the promoter regions of the *BjR2R3-MYB* S12 subfamily genes (Table S5), because different regulatory elements are responsible for different biological functions.

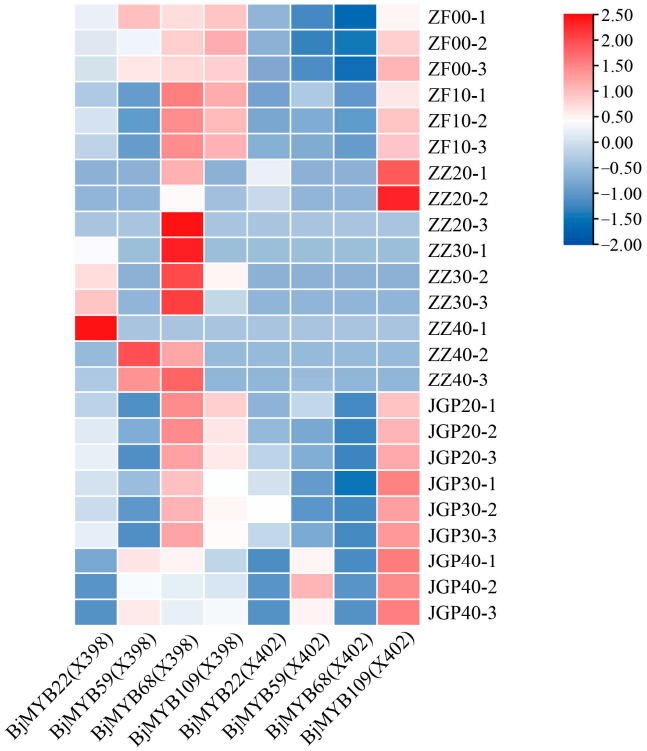


Figure 8. The expression of *BjR2R3-MYB* S12 subfamily genes. FPKM values for *BjR2R3-MYB* genes were transformed by log10. Note: ZF00 indicates ovary at 0 DAP; ZF10 indicates ovary at 10 DAP; ZZ20, ZZ30 and ZZ40 indicate seeds at 20, 30 and 40 DAP, respectively; JGP20, JGP30 and JGP40 indicate SP at 20, 30, and 40 DAP, respectively.

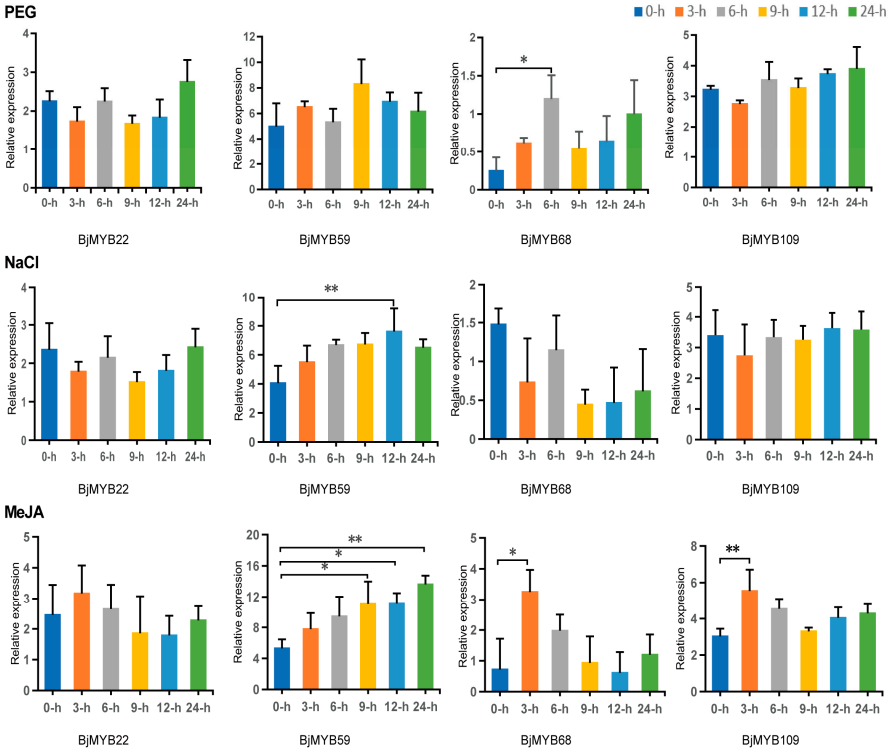


Figure 9. Relative expression levels of *BjR2R3-MYB* S12 subfamily genes under Abiotic stress at different treatment times (3 h, 6 h, 9 h, 12 h, 24 h) compared to the control (0 h). The mean (±SE) expression values were calculated from three independent biological replicates and three technical replicates (*, $p < 0.05$; **, $p < 0.01$).

4. Discussion

The *R2R3-MYB* family is one of the largest transcription factor families in plants. To date, whole-genome identification of this family has been successfully conducted across multiple species, including 126 in *A. thaliana* [41], 157 in *maize* [25], 244 in *soybean* [43], 174 in *tobacco* [44], 181 in *Spatholobus suberectus* [34], and 144 in *Santalum album* [45]. In this study, a total of 151 *BjR2R3-MYB* genes were identified in *B. juncea*. The identification process was based on sequence similarity analysis and conserved domain screening, as well as comparisons with other species. Notably, the Swiss-Prot database was utilised during the sequence similarity analysis to filter out low-quality candidate sequences, thereby effectively reducing the false positive rate. In terms of threshold control, a stringent criterion of E-value $\leq 1 \times 10^{-1}$ was applied to exclude non-specific matches and ensure that only sequences highly homologous to those from *A. thaliana* were captured. During the domain verification stage, double validation was performed using the PF00249 HMM and data from the SMART/CDD database. It was required that candidate genes possessed complete R2 and R3 MYB domains; sequences containing only a single domain or incomplete domains were excluded. This strategy is particularly effective for addressing issues related to domain loss or fragmentation caused by gene duplication in polyploid species. A total of 75 and 76 *BjR2R3-MYB* genes were identified within the A and B subgenomes of *B. juncea*, respectively (Figure 1), with no significant bias distribution observed. This finding indicates that the screening threshold did not introduce any subgenomic preferences and ensured comprehensive coverage of family members. Similar methods have also been documented in tobacco [45], thereby corroborating the reliability of the results obtained for the *BjR2R3-MYB* gene family in *B. juncea*.

The evolution of gene families is closely related to the structural characteristics of genes. Most *BjR2R3-MYB* genes identified in *B. juncea* have a typical three-exon, two-intron structure, with introns inserted within the conserved R2 and R3 repeat regions; this intron insertion pattern is a rare insertion event derived from the ancestral non-intron MYB genes and represents a conserved evolutionary feature [23–25]. In the motif analysis, it is highly conserved within subfamilies but shows differences between subfamilies (Figure 3). Motifs 1 and 2 are ubiquitous in all subfamilies and correspond to the core R2 and R3 DNA-binding domains, which are essential for MYB proteins to bind to target gene promoters and exert transcriptional regulatory functions [38,46]. Motifs 8 and 9 are only present in the B19 (S1) subfamily. The conservation of the core DNA-binding pattern reflects the evolutionary stability of the basic function of *BjR2R3-MYB* proteins, while the emergence of subfamily-specific motifs is a sign of functional specialisation during the expansion of the gene family [41]. These unique C-terminal pattern sequences usually mediate protein-protein interactions, transcriptional activation/inhibition, or responses to specific developmental and environmental signals [47]; for example, specific pattern sequences in the flavonoid-related MYB subfamily of *Arabidopsis* mediate key interactions with bHLH partner factors to activate the flavonoid biosynthesis pathway [48,49]. Therefore, the motifs 8 and 9 in the B19 (S1) subfamily may help form specific regulatory protein complexes, thereby finely regulating downstream gene expression in the specific biological processes of *B. juncea*. The B6 (S22) subfamily not only has the unique motif 7 but also its subfamily genes have no introns, which is consistent with a coordinated evolutionary adaptation for this subfamily: motif 7 may confer specific regulatory functions to B6 (S22) subfamily members, and intron deletion further optimises the expression dynamics of these genes to meet their functional requirements. This pattern confirms the role of the coordinated evolution of gene structure and motif composition in functional differentiation of gene families [50,51], emphasising that the B6 (S22) subfamily may have evolved a specialised function in *B. juncea*, mediating

rapid and precise transcriptional regulation in non-biotic stress responses and secondary metabolism [19].

Cis-acting elements in promoter regions are pivotal for mediating gene expression specificity and responsiveness, and their type, quantity, and distribution often correlate with subfamily classification and functional divergence [50]. In this study, 32 types of cis-acting elements were identified in *BjR2R3-MYB* gene promoters, covering hormone signalling, environmental stress responses, growth and development, and metabolic regulation (Figure 5). Further analysis revealed a clear correlation between cis-element profiles and subfamily classification. For example, the B6 (S22) subfamily contained a high proportion of stress-responsive elements (e.g., drought-inducibility, low-temperature responsiveness) and seed-specific regulatory elements, which is consistent with its predicted role in rapid stress adaptation and developmental regulation [19].

Direct genomic evidence for the allotetraploid origin of *B. juncea* was provided by interspecies collinearity analysis. For example, a similar number of collinear gene pairs were identified between *B. juncea* and its diploid ancestors, with *B. rapa* 402 pairs and *B. nigra* 409 pairs, respectively. This is consistent with *B. juncea* being formed by interspecific hybridisation between *B. rapa* (A genome) and *B. nigra* (B genome) followed by genome doubling, and further supports the complete transmission and retention patterns of genetic material during heterologous polyploidy events [52]. In contrast, 218 collinear gene pairs between *B. juncea* and *A. thaliana* were significantly less those between *B. juncea* and its two diploid ancestors, which demonstrated that evolutionary divergence between distant species had occurred in the early stage, retaining a certain degree of homology. Intra-species collinearity analysis revealed that fragment duplication serves as the primary driving force for the expansion of the *BjR2R3-MYB* gene family in *B. juncea*; this is consistent with similar expansion mechanisms observed in Brassica crops such as *Brassica oleracea* [53,54]. Fragment duplications resulting from polyploidy events provide essential raw materials for innovating gene functions. Some duplicated genes may undergo neofunctionalisation or subfunctionalisation for adapting to environmental selection pressures. For instance, in apple plants, duplicated *BjR2R3-MYB* genes arising from polyploidy have undergone functional differentiation related to fruit quality regulation [55].

Transcriptome analysis revealed that two of the four *BjR2R3-MYB* S12 subfamily genes in *B. juncea* that may be involved in GLS biosynthesis, *BjMYB68* and *BjMYB109*, exhibited significant expression differences between high-GLS (X398) and low-GLS (X402) accessions, indicating that the total GLS content in seeds was correlated with the expression of genes *BjMYB68* and *BjMYB109*. However, we only detected the total GLS content of mature seeds, but not that of other tissues, such as roots, leaves, ovary, SP and developing seeds, and dynamic sampling for different tissues at different developmental stages separately was not performed either. Therefore, the current data may not distinguish whether the expression difference of genes *BjMYB68* and *BjMYB109* occur in the process of GLS synthesis in the seed or in the process of GLS synthesis and transport in other tissues. It also could not be determined whether the expression differences of genes *BjMYB68* and *BjMYB109* are due to developmental timing differences or specific regulation of the GLS pathway. Subsequently, through the combined analysis of GLS content and gene expressions of multiple tissues and developmental stages, the regulatory mechanism of genes *BjMYB68* and *BjMYB109* can be further elucidated.

BjMYB68 and *BjMYB59*, two members of the S12 subfamily with distinct GLS metabolic roles, exhibit specific abiotic stress response patterns rooted in functional specialisation, cis-acting element specificity, and the evolutionary coordination between GLS metabolism and stress adaptation. *BjMYB68*, homologous to *AtMYB28*, is responsible for aliphatic GLS biosynthesis [1,7] is specifically upregulated under 10% PEG-6000 (drought) at 6 h; this

specificity arises from its promoter harbouring drought-inducible elements [50], which was consistent with findings that homologs of *AtMYB28* in *Brassica napus* mediated drought adaptation [1]. In contrast, *BjMYB59*, homologous to *AtMYB51* responsible for indole GLS biosynthesis [5], is induced and upregulated by 200 mmol/L NaCl (12 h) and 200 µmol/L MeJA (9–24 h) but not drought. Therefore, *BjMYB59* would specifically integrate salt stress signalling and MeJA-mediated defence with indole GLS metabolism, arising from its promoter harbouring salt-responsive elements and MeJA-responsive elements (Supplementary Table S5). Furthermore, indole GLS and their derivatives play critical roles in plant salt tolerance and pathogen defence [5,12]. The response to drought stress, however, primarily relies on aliphatic GLS for osmotic adjustment [7]. Collectively, these results highlight the functional differentiation of S12 subfamily genes in specific GLS metabolic branches in response to distinct abiotic stress, thereby optimising adaptive strategies of *B. juncea* through precise spatiotemporal and signal-dependent regulation [7,10].

As we all know, on the one hand, GLS hydrolysis products remaining in rapeseed cake meal have adverse effects on the health of livestock (e.g., goitre, liver, and kidney dysfunction). Yet on the other hand, GLS not only play a significant role in protecting plants from microbial pathogens and insect attacks [12], but its hydrolysis products are beneficial to human health by preventing chronic diseases, antioxidant, anti-cancer and cancer-preventing effects, as well as regulating the body's immune system [13,14]. Therefore, for the genetic improvement of *B. juncea*, we should regulate GLS content for different purposes, rather than reduce it indiscriminately; furthermore, precise regulation of GLS composition and accumulation is indispensable for the comprehensive utilisation of *B. juncea*. In this study, four key candidate S12 subfamily genes were identified in *B. juncea*, wherein *BjMYB68*, *BjMYB109*, and *BjMYB22* were all homologous to *AtMYB28* (*AT5G61420*) of *A. thaliana*, which dominates the biosynthesis of aliphatic GLS [1–4], and *BjMYB59* was homologous to *AtMYB51* (*AT1G18570*) of *A. thaliana*, which specifically regulates indole GLS metabolism [5]. In addition, transcriptome analysis indicated that *BjMYB68* and *BjMYB109* exhibited significant expression differences between high-GLS accession X398 and low-GLS accession X402, suggesting that the total GLS content in seeds of *B. juncea* was correlated with the expression of genes *BjMYB68* and *BjMYB109;* this is consistent with the fact that aliphatic GLS are dominant in *B. juncea* total GLS. Beyond that, the expression pattern analysis of the four S12 subfamily genes under abiotic stress indicated that *BjMYB68* is responsive to both drought and the MeJA signalling pathway, *BjMYB59* is responsive to both salt and MeJA signalling pathways, *BjMYB109* is only responsive to the MeJA signalling pathway, while no significant response was detected to any of the all three abiotic stresses for BjMYB22. Therefore, enhancing GLS content in vegetative tissues (leaves, roots) moderately through regulating drought and MeJA response genes *BjMYB68*, sal, and MeJA response gene *BjMYB59*, as well as MeJA response gene *BjMYB109*, can strengthen GLS-mediated defence against pathogens and pests as well as tolerance to abiotic stresses, so as to reduce pesticide reliance and improve field adaptability of *B. juncea*. In a word, further functional verification of these three core genes, such as *BjMYB68*, *BjMYB109*, and *BjMYB59*, will help clarify their precise roles in spatiotemporal GLS metabolism and provide promising genetic targets for the molecular breeding of low-toxicity, high-quality, stress-tolerant, and wide-adaptable *B. juncea* varieties. Additionally, intron structure also affects the expression of gene *BjPur* encoding an *R2R3-MYB* transcription factor in *B. juncea* [56], the regulatory mechanism of introns for candidate genes *BjMYB68*, *BjMYB109*, and *BjMYB59* can be further explored.

5. Conclusions

In this study, a total of 151 *BjR2R3-MYB* genes were identified in *B. juncea* and classified into 26 subfamilies, reflecting the evolutionary expansion and functional diversification of this family in the allotetraploid genome. 151 *BjR2R3-MYB* genes are unevenly distributed across 18 chromosomes but show a balanced distribution between the A (75 genes) and B (76 genes) subgenomes, which is consistent with the genetic conservation of the diploid progenitors of *B. juncea* (*B. rapa* and *B. nigra*). 402, 409, and 218 homologous gene pairs between *B. juncea* and *B. rapa*, *B. nigra*, and *A. thaliana*, respectively, were identified by collinearity analysis, confirming the allotetraploid origin of *B. juncea* and the conservation of core *MYB* functions. The promoter regions of *BjR2R3-MYB* genes contain 32 types of cis-acting elements, spanning hormone signalling, environmental stress responses, growth and development, and metabolism, highlighting their multifunctional regulatory potential. Transcriptome analysis revealed that two of the four *BjR2R3-MYB* S12 subfamily genes in *B. juncea* that may be involved in glucosinolate biosynthesis, *BjMYB68* and *BjMYB109*, exhibited significant expression differences between high-GLS (X398) and low-GLS (X402) accessions, indicating that the total GLS content in seeds was correlated with the expression of genes *BjMYB68* and *BjMYB109*. Furthermore, the results of qRT-PCR analysis demonstrated that the four S12 subfamily genes exhibited distinct responses to different abiotic stresses: *BjMYB68* is responsive to both drought and the MeJA signalling pathway, *BjMYB59* is responsive to both salt and MeJA signalling pathways, *BjMYB109* is only responsive to the MeJA signalling pathway, while no significant response to any of the all three abiotic stresses was detected in *BjMYB22*. Collectively, this study systematically characterised the *BjR2R3-MYB* gene family and identified genes related to GLS synthesis and abiotic stress within the S12 subfamily, providing valuable insights for unravelling the molecular mechanisms underlying GLS synthesis and abiotic stress resistance in *B. juncea*, as well as laying a theoretical foundation for subsequent genetic improvement of *B. juncea*.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes17111455/s1, Table S1: The physicochemical properties of 151 *BjR2R3-MYBs*; Table S2: Transcriptome data of the ovary, seeds and pericarp of four genes in the S12 subfamily; Table S3: qRT-PCR data of four genes in the S12 subfamily under abiotic stress; Table S4: Specifc primers of four genes in the S12 subfamily for qRT-PCR; Table S5: Data of cis-acting elements of the 151 *BjR2R3-MYBs* gene promoter; File S6: Homologous gene pairs of *BjR2R3-MYB* genes among *B. juncea*, *B. rapa*, *B. nigra* and *A. thaliana*; File S7: The ratio of unknown bases(N) in the genome of *B. juncea*; File S8: The skew of GC in the genome of *B. juncea*; File S9: The ratio of GC in the genome of *B. juncea*; File S10: The protein sequence of 151 *BjR2R3-MYBs*; File S11: Intraspecific collinear gene pairs and ka/ks ratio of *B. juncea*.

Author Contributions: Formal analysis, Y.Y., T.K., T.L., and M.X.; investigation, H.L., K.L., A.W., Y.Q., L.Z., J.M., and H.Z.; resources, H.L., K.L., A.W., Y.Q., L.Z., J.M., and H.Z.; writing—original draft preparation, Y.Y., T.K., T.L., and M.X.; writing—review and editing, Y.Y. and T.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the 2023 Western Light Talent Cultivation Program, the National Natural Science Foundation of China (Grant No. 32360516), and the National Rapeseed Industry Technology System (China Agriculture Research System, Grant No. CARS-12).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

GLS	Glucosinolate
MeJA	methyl jasmonate
SP	silique pericarp

References

1. Zhou, X.; Zhang, H.; Xie, Z.; Liu, Y.; Wang, P.; Dai, L.; Zhang, X.; Wang, Z.; Wang, Z.; Wan, L.; et al. Natural variation and artificial selection at the *BnaC2.MYB28* locus modulate *Brassica napus* seed Glucosinolate. *Plant Physiol.* **2023**, *191*, 352–368. [CrossRef]
2. Zuluaga, D.L.; Graham, N.S.; Klinder, A.; van Ommen Kloeke, A.E.E.; Marcotrigiano, A.R.; Wagstaff, C.; Verkerk, R.; Sonnante, G.; Aarts, M.G.M. Overexpression of the *MYB29* transcription factor affects aliphatic Glucosinolate synthesis in *Brassica oleracea*. *Plant Mol. Biol.* **2019**, *101*, 65–79. [CrossRef]
3. Gigolashvili, T.; Engqvist, M.; Yatusevich, R.; Müller, C.; Flügge, U.I. *HAG2/MYB76* and *HAG3/MYB29* exert a specific and coordinated control on the regulation of aliphatic Glucosinolate biosynthesis in *A. thaliana*. *New Phytol.* **2008**, *177*, 627–642. [CrossRef]
4. Li, Y.; Sawada, Y.; Hirai, A.; Sato, M.; Kuwahara, A.; Yan, X.; Hirai, M.Y. Novel insights into the function of *Arabidopsis R2R3-MYB* transcription factors regulating aliphatic Glucosinolate biosynthesis. *Plant Cell Physiol.* **2013**, *54*, 1335–1344. [CrossRef]
5. Frerigmann, H.; Gigolashvili, T. *MYB34*, *MYB51*, and *MYB122* distinctly regulate indolic Glucosinolate biosynthesis in *A. thaliana*. *Mol. Plant* **2014**, *7*, 814–828. [CrossRef]
6. Coleto, I.; Bejarano, I.; Marín-Peña, A.J.; Medina, J.; Rioja, C.; Burow, M.; Marino, D. *A. thaliana* transcription factors *MYB28* and *MYB29* shape ammonium stress responses by regulating Fe homeostasis. *New Phytol.* **2021**, *229*, 1021–1035. [CrossRef]
7. Yang, J.; Wang, J.; Li, Z.; Li, X.; He, Z.; Zhang, L.; Sha, T.; Lyu, X.; Chen, S.; Gu, Y.; et al. Genomic signatures of vegetable and oilseed allopolyploid *Brassica juncea* and genetic loci controlling the accumulation of Glucosinolate. *Plant Biotechnol. J.* **2021**, *19*, 2619–2628. [CrossRef]
8. Neequaye, M.; Stavnstrup, S.; Harwood, W.; Lawrenson, T.; Hundleby, P.; Irwin, J.; Troncoso-Rey, P.; Saha, S.; Traka, M.H.; Mithen, R.; et al. CRISPR-Cas9-Mediated Gene Editing of *MYB28* Genes Impair Glucoraphanin Accumulation of *Brassica oleracea* in the Field. *CRISPR J.* **2021**, *4*, 416–426. [CrossRef]
9. Ren, L.; Zhao, Y.; Liu, H.; Wang, H. Effect of Exogenous Methyl Jasmonate on Pakchoi *Myb28* Expression. In Proceedings of the Conference on Horticulture Science and Technology, Beijing, China, 24–25 November 2012.
10. Wang, M.; Cai, C.; Li, Y.; Tao, H.; Meng, F.; Sun, B.; Miao, H.; Wang, Q. Brassinosteroids fine-tune secondary and primary sulfur metabolism through *BZR1*-mediated transcriptional regulation. *J. Integr. Plant Biol.* **2023**, *65*, 1153–1169. [CrossRef]
11. Nguyen, V.P.T.; Stewart, J.; Lopez, M.; Ioannou, I.; Allais, F. Glucosinolates: Natural Occurrence, Biosynthesis, Accessibility, Isolation, Structures, and Biological Activities. *Molecules* **2020**, *25*, 4537. [CrossRef]
12. Clay, N.K.; Adio, A.M.; Denoux, C.; Jander, G.; Ausubel, F.M. Glucosinolate metabolites required for an *Arabidopsis* innate immune response. *Science* **2009**, *323*, 95–101. [CrossRef]
13. Calabrese, E.J.; Kozumbo, W.J. The phytoprotective agent sulforaphane prevents inflammatory degenerative diseases and age-related pathologies via *Nrf2*-mediated hormesis. *Pharmacol. Res.* **2021**, *163*, 105283. [CrossRef]
14. Kaiser, A.E.; Baniasadi, M.; Giansiracusa, D.; Giansiracusa, M.; Garcia, M.; Fryda, Z.; Wong, T.L.; Bishayee, A. Sulforaphane: A *Broccoli* Bioactive Phytocompound with Cancer Preventive Potential. *Cancers* **2021**, *13*, 4796. [CrossRef]
15. Downey, R.K. Brassica oilseed breeding-achievements and opportunities. *Plant Breed. Abstr.* **1990**, *60*, 1165–1170.
16. Woods, D.L.; Capcara, J.J.; Downey, R.K. The potential of *mustard* (*Brassica juncea* (L.) *Coss*) as an edible oil crop on the Canadian Prairies. *Can. J. Plant Sci.* **1991**, *71*, 195–198. [CrossRef]
17. Garcia, G.; Treccarichi, S.; Arena, D.; Ammar, H.B.; Maggioni, L.; Branca, F. Capturing the *Brassica oleracea* L. wild relatives diversity for improving nutraceutical traits of cole crops. *Genet. Resour. Crop Evol.* **2025**, *72*, 8855–8871. [CrossRef]
18. Amegan Missinou, A.; Ferreira de Carvalho, J.; Marnet, N.; Delhaye, T.; Hamzaoui, O.; Abdel Sayed, D.; Guitton, Y.; Lebreton, L.; Langrume, C.; Laperche, A.; et al. Identification and Quantification of Glucosinolates and Phenolics in a Large Panel of Brassica napus Highlight Valuable Genetic Resources for Chemical Ecology and Breeding. *J. Agric. Food Chem.* **2022**, *27*, 5245–5261. [CrossRef]
19. Chen, Y.; Yang, X.; He, K.; Liu, M.; Li, J.; Gao, Z.; Lin, Z.; Zhang, Y.; Wang, X.; Qiu, X.; et al. The *MYB* transcription factor superfamily of *Arabidopsis*: Expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol. Biol.* **2006**, *60*, 107–124. [CrossRef]
20. Martin, C.; Paz-Ares, J. *MYB* transcription factors in plants. *Trends Genet.* **1997**, *13*, 67–73. [CrossRef]

21. Jin, H.; Martin, C. Multifunctionality and diversity within the plant *MYB*-gene family. *Plant Mol. Biol.* **1999**, *41*, 577–585. [CrossRef]
22. Augustine, R.; Mukhopadhyay, A.; Bisht, N.C. Targeted silencing of *BjMYB28* transcription factor gene directs development of low glucosinolate lines in oilseed *Brassica juncea*. *Plant Biotechnol. J.* **2013**, *11*, 855–866. [CrossRef]
23. Jiang, C.; Gu, J.; Chopra, S.; Gu, X.; Peterson, T. Ordered origin of the typical two- and three-repeat *Myb* genes. *Gene* **2004**, *326*, 13–22. [CrossRef]
24. Fedorova, L.; Fedorov, A. Introns in gene evolution. *Genetica* **2003**, *118*, 123–131. [CrossRef]
25. Du, H.; Feng, B.R.; Yang, S.S.; Huang, Y.B.; Tang, Y.X. The *R2R3-MYB* transcription factor gene family in *maize*. *PLoS ONE* **2012**, *7*, e37463. [CrossRef]
26. *ISO 9167-1:1992*; Rapeseed-Determination of Glucosinolates Content-Part 1: Method Using High-Performance Liquid Chromatography. International Organization for Standardization: Geneva, Switzerland, 1992.
27. Chen, C.; Wu, Y.; Li, J.; Wang, X.; Zeng, Z.; Xu, J.; Liu, Y.; Feng, J.; Chen, H.; He, Y. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* **2023**, *16*, 1733–1742. [CrossRef]
28. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [CrossRef]
29. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [CrossRef]
30. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef]
31. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I.; et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222–D226. [CrossRef]
32. Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, W585–W587. [CrossRef]
33. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
34. Yang, J.; Zhang, B.; Gu, G.; Yuan, J.; Shen, S.; Jin, L.; Lin, Z.; Lin, J.; Xie, X. Genome-wide identification and expression analysis of the *R2R3-MYB* gene family in *tobacco* (*Nicotiana tabacum* L.). *BMC Genom.* **2022**, *23*, 432. [CrossRef]
35. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef]
36. Chen, H.; Wang, T.; He, X.; Cai, X.; Lin, R.; Liang, J.; Wu, J.; King, G.; Wang, X. BRAD V3.0: An upgraded Brassicaceae database. *Nucleic Acids Res.* **2022**, *50*, D1432–D1441. [CrossRef]
37. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; VandePeer, Y.; Rouzé, 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]
38. Wang, L.; Zhao, Z.; Li, H.; Pei, D.; Huang, Z.; Wang, H.; Xiao, L. Genome-Wide Identification of *NDPK* Family Genes and Expression Analysis under Abiotic Stress in *Brassica napus*. *Int. J. Mol. Sci.* **2024**, *25*, 6795. [CrossRef]
39. Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. *MYB* transcription factors in *Arabidopsis*. *Trends Plant Sci.* **2010**, *15*, 573–581. [CrossRef]
40. Dixit, S.; Jangid, V.K.; Grover, A. Evaluation of suitable reference genes in *Brassica juncea* and its wild relative Camelina sativa for qRT-PCR analysis under various stress conditions. *PLoS ONE* **2019**, *14*, e0222530. [CrossRef]
41. Balhara, R.; Verma, D.; Kaur, R.; Singh, K. *MYB* transcription factors, their regulation and interactions with non-coding RNAs during drought stress in *Brassica juncea*. *BMC Plant Biol.* **2024**, *24*, 999. [CrossRef]
42. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef]
43. Fan, J.; Jiang, C.; Xing, J.; Dong, J. Structure and function of the 22nd subfamily in *Arabidopsis R2R3-MYB* family. *Yi Chuan* **2014**, *36*, 985–994. Available online: https://www.chinagene.cn/CN/10.3724/SP.J.1005.2014.0985 (accessed on 16 June 2025). (In Chinese) [CrossRef]
44. Du, H.; Yang, S.S.; Liang, Z.; Feng, B.R.; Liu, L.; Huang, Y.B.; Tang, Y.X. Genome-wide analysis of the *MYB* transcription factor superfamily in *soybean*. *BMC Plant Biol.* **2012**, *12*, 106. [CrossRef]
45. Qin, S.; Wei, F.; Liang, Y.; Tang, D.; Lin, Q.; Miao, J.; Wei, K. Genome-wide analysis of the *R2R3-MYB* gene family in *Spatholobus suberectus* and identification of its function in flavonoid biosynthesis. *Front. Plant Sci.* **2023**, *14*, 1219019. [CrossRef]
46. Tang, M.; Liu, L.; Hu, X.; Zheng, H.; Wang, Z.; Liu, Y.; Zhu, Q.; Cui, L.; Xie, S. Genome-wide characterization of *R2R3-MYB* gene family in *Santalum album* and their expression analysis under cold stress. *Front. Plant Sci.* **2023**, *14*, 1142562. [CrossRef]
47. Stracke, R.; Werber, M.; Weisshaar, B. The *R2R3-MYB* gene family in *A. thaliana*. *Curr. Opin. Plant Biol.* **2001**, *4*, 447–456. [CrossRef]

48. Grotewold, E. The genetics and biochemistry of floral pigments. *Annu Rev. Plant Biol.* **2006**, *57*, 761–780. [CrossRef]
49. Zimmermann, I.M.; Heim, M.A.; Weisshaar, B.; Uhrig, J.F. Comprehensive identification of *A. thaliana MYB* transcription factors interacting with R/B-like *BHLH* proteins. *Plant J.* **2004**, *40*, 22–34. [CrossRef]
50. Baudry, A.; Heim, M.A.; Dubreucq, B.; Caboche, M.; Weisshaar, B.; Lepiniec, L. *TT2*, *TT8*, and *TTG1* synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *A. thaliana*. *Plant J.* **2004**, *39*, 366–380. [CrossRef]
51. Jiang, C.; Gu, X.; Peterson, T. Identification of conserved gene structures and carboxy-terminal motifs in the *Myb* gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. *Genome Biol.* **2004**, *5*, R46. [CrossRef]
52. Islam, S.; Sajib, S.D.; Jui, Z.S.; Arabia, S.; Ghosh, A. Genome-wide identification of glutathione S-transferase gene family in *pepper*, its classification, and expression profiling under different anatomical and environmental conditions. *Sci. Rep.* **2019**, *9*, 9101. [CrossRef]
53. Kang, L.; Qian, L.; Zheng, M.; Chen, L.; Chen, H.; Yang, L.; You, L.; Yang, B.; Yan, M.; Gu, Y.; et al. Genomic insights into the origin, domestication and diversification of *Brassica juncea*. *Nat. Genet.* **2021**, *53*, 1392–1402. [CrossRef]
54. Luo, D.; Mei, D.; Wei, W.; Liu, J. Identification and Phylogenetic Analysis of the *R2R3-MYB* Subfamily in *Brassica napus*. *Plants* **2023**, *12*, 886. [CrossRef]
55. Zhang, Y.; Qian, K.; Yu, Q.; Chen, X.; Liang, J.; Liu, Z.; Dong, Z.; Liu, Y.; Sun, Y.; Guo, Z.; et al. Subgenomic divergence and functional innovation following whole-genome duplication in Maleae species of *Rosaceae*. *Plant J.* **2025**, *123*, e70499. [CrossRef]
56. Heng, S.; Cheng, Q.; Zhang, T.; Liu, X.; Huang, H.; Yao, P.; Liu, Z.; Wan, Z.; Fu, T. Fine-mapping of the *BjPur* gene for purple leaf color in *Brassica juncea*. *Theor. Appl. Genet.* **2020**, *133*, 2989–3000. [CrossRef]

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