

Article

# The Gene Structure and Expression Level Changes of the *GH3* Gene Family in *Brassica napus* Relative to Its Diploid Ancestors

Ruihua Wang<sup>1</sup>, Mengdi Li<sup>1</sup>, Xiaoming Wu<sup>2</sup> and Jianbo Wang<sup>1,\*</sup> 

<sup>1</sup> State Key Laboratory of Hybrid Rice, College of Life Sciences, Wuhan University, Wuhan 430072, China; wangruihua@whu.edu.cn (R.W.); mengdili@whu.edu.cn (M.L.)

<sup>2</sup> Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan 430072, China; wuxm@oilcrops.cn

\* Correspondence: jbwang@whu.edu.cn; Tel.: +86-27-68-75-22-13

Received: 3 December 2018; Accepted: 15 January 2019; Published: 17 January 2019



**Abstract:** The *GH3* gene family plays a vital role in the phytohormone-related growth and developmental processes. The effects of allopolyploidization on *GH3* gene structures and expression levels have not been reported. In this study, a total of 38, 25, and 66 *GH3* genes were identified in *Brassica rapa* ( $A_rA_r$ ), *Brassica oleracea* ( $C_oC_o$ ), and *Brassica napus* ( $A_nAC_nC_n$ ), respectively. *BnaGH3* genes were unevenly distributed on chromosomes with 39 on  $A_n$  and 27 on  $C_n$ , in which six *BnaGH3* genes may appear as new genes. The whole genome triplication allowed the *GH3* gene family to expand in diploid ancestors, and allopolyploidization made the *GH3* gene family re-expand in *B. napus*. For most *BnaGH3* genes, the exon-intron compositions were similar to diploid ancestors, while the cis-element distributions were obviously different from its ancestors. After allopolyploidization, the expression patterns of *GH3* genes from ancestor species changed greatly in *B. napus*, and the orthologous gene pairs between  $A_n/A_r$  and  $C_n/C_o$  had diverged expression patterns across four tissues. Our study provides a comprehensive analysis of the *GH3* gene family in *B. napus*, and these results could contribute to identifying genes with vital roles in phytohormone-related growth and developmental processes.

**Keywords:** *GH3* gene; *Brassica napus*; orthologous gene; cis-element; expression pattern; allopolyploidization

## 1. Introduction

Polyploidization, as a prevalent phenomenon, is an important method of speciation in the plant kingdom. In angiosperms, approximately 70 percent of species have undergone one or several rounds of polyploidization in the evolutionary process [1]. Allopolyploids are derived from the hybridization between different species, followed with chromosome doubling or the fusion of unreduced gametes between different species. Allopolyploids usually have better adaptability and superior traits than their parents, which may contribute to natural selection and crop domestication [2]. Extensive changes on genetic and gene expression levels have been reported in allopolyploids. Genetic changes contain chromosomal rearrangements, DNA sequence elimination, and amplification. Chromosomal rearrangements could lead to the variation on seed yields and flowering time in *Brassica* allotetraploids [3,4]. Comparative analysis of gene structure among polyploid wheat and model grass genomes reveals that many genes have exonic sequence acquirement and loss, and 35 percent of these gene structure rearrangements give rise to premature termination codons and frame-shift mutations in polyploid wheat [5]. In addition, allopolyploidization has a profound impact on

gene expression levels, including gene silencing, gene activation, gene non-additive expression, and gene bias expression. The gene expression levels of hexaploid wheat are compared with its parents, which reveals that the gene silencing is caused by gene regulation rather than gene loss, and one gene was specifically activated, suggesting that gene expression changes are correlative with polyploidization [6]. Non-additive gene expression has been found in diverse polyploids and is associated with allopolyploid growth vigor [7]. Genome-biased alterations in gene expression may be a way to overcome incompatibilities after different genome fusion, thus might facilitate the rapid genetic diploidization of allopolyploids [8]. In conclusion, changes on genetic and gene expression levels could contribute to adaptive enhancement during the evolution of allopolyploids.

Genes rapidly and transiently induced by auxin are regarded as early/primary auxin-responsive genes, which mainly belong to three gene families, including auxin/indole-3-acetic acid (*Aux/IAA*), Gretchen Hagen3 (*GH3*), and small auxin-up RNA (*SAUR*) gene family [9]. The GH3 proteins catalyze conjugation of amino acids with free indole-3-acetic acid (IAA), jasmonic acid (JA), and salicylic acid (SA) to modulate the concentrations of hormone bioactive forms during plant development and stress adaptation processes [10]. The GH3 proteins are classified into three groups according to protein sequence similarities and substrate specificities in *Arabidopsis thaliana* [11,12]. Group I is made up of two members: AtGH3.11, which adenylates jasmonic acid (JA) and acts as JA-amido synthetase, and AtGH3.10, which does not own these features [13]. Group II, containing eight members, adenylates indoleacetic acid (IAA) and conjugates IAA to amino acids. In addition, AtGH3.5 in group II also adenylates salicylic acid (SA) and displays SA-amino synthetase activity. Group III consists of nine members. AtGH3.12/PBS3 in group III conjugates glutamic acid (Glu) to 4-substituted benzoates, which is involved in signal transduction of SA [14], and the function of the other eight members is still unknown. The auxin-responsive elements (AuxREs) that possess core motifs (TGTCTC) exist in the *GH3* gene promoters, and the AuxRE motifs are various either in sequences or number [10]. Besides, the other hormone-related elements and stress-related elements are also found in the promoter regions of the *GH3* genes, which ensures their fast and transient responses to the other hormones and stress stimulus [15–17]. The GH3 proteins play important roles in plant development and light signal pathways [18–21]. So far, the identification and analysis of the *GH3* gene family have been performed in different plant species, such as *Arabidopsis thaliana*, rice, maize, *Medicago truncatula*, *Carica papaya*, and so on [12,17,22–24].

The divergence between *Brassica* and *Arabidopsis* is estimated to be 14.5 to 20.4 million years ago (MYA) [25]. The *Brassica* genome went through the whole genome triplication (WGT) about 15.9 MYA, followed by species divergence between *B. rapa* ( $A_rA_r$ ,  $2n = 20$ ) and *B. oleracea* ( $C_oC_o$ ,  $2n = 18$ ) about 4.6 MYA [26]. *B. napus* ( $A_nA_nC_nC_n$ ,  $2n = 38$ ) is a natural allotetraploid, formed by the hybridization between *B. rapa* and *B. oleracea* followed with chromosome doubling [27]. Some GH3 proteins are found to have crucial roles in *Brassica* plant organ development and stress responses. A few GH3 proteins in *B. napus* are involved in seedling development, controlling leaf morphology and branch angle regulation [28–30]. The *BrGH3.12* that is highly expressed in the leaf apical region regulates heading type in *B. rapa* [31]. One *GH3* gene in *Brassica juncea* is strongly induced by Cd in roots and leaves [32]. In addition, a GH3 protein in *B. rapa* takes part in resistance to drought stress [33]. To date, very little information about the *GH3* gene family in *B. napus* is known. It is unknown how the gene structures and expression levels of the *GH3* gene family change in *B. napus* compared with diploid ancestors. The availability of *Brassica* genome annotations provides an opportunity for genome-wide analysis of the *GH3* gene family, which contributes to a better understanding of the function and evolution of the *GH3* gene family in *B. napus*.

In this study, the genome-wide analysis of the *GH3* gene family in *B. napus* were carried out. This study aimed to address: (1) the identification and characteristics of *GH3* genes in the three species of *Brassica*; (2) the source analysis of the *GH3* gene family members in *B. napus*; (3) the comparative analysis of the *GH3* gene exon-intron organization, cis-element distribution, and expression patterns between *B. napus* and ancestor species. The comprehensive analysis of the *GH3* gene family would

provide a better understanding of the *GH3* gene function in polyploid growth and development, and a reference for exploring the adaptation and evolution mechanisms of polyploid.

## 2. Materials and Methods

### 2.1. Plant Materials

Plant materials contained *B. rapa* cv. Chiifu ( $A_rA_r$ ,  $2n = 20$ ), *B. oleracea* cv. Jinzaosheng ( $C_oC_o$ ,  $2n = 18$ ), and *B. napus* cv. Darmor ( $A_nA_nC_nC_n$ ,  $2n = 38$ ). The seeds of the three species of *Brassica* were acquired from the Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, China. Plant materials were planted in large flowerpots outside the greenhouse of Wuhan University, Wuhan (30.52 degrees north latitude and 114.31 degrees east longitude), China, under natural conditions. Plant materials were planted in September (autumn), and plant tissues were taken in March of the following year (spring).

### 2.2. GH3 Gene Retrieval in the Three Species of Brassica

*Brassica rapa* (v1.5), *B. oleracea* (v1.1), and *B. napus* (v5.0) genome annotation files and protein sequences were downloaded from BRAD database (<http://brassicadb.org/brad/>). A total of 19 GH3 protein sequences in *A. thaliana* were obtained from TAIR database (<https://www.arabidopsis.org/>). All protein sequences of each *Brassica* species were searched via BLASTP algorithms with E values lower than  $1e-10$ , using 19 GH3 protein sequences of *A. thaliana* as queries. Further, the Hidden Markov Model (HMM) profile of GH3 domain (Pfam: 03321) was used to search all protein sequences of each *Brassica* species using HMMER3 software. For each *Brassica* species, all GH3 protein candidates gained from these two searches were combined to make a non-redundant protein list. Finally, all non-redundant protein sequences were analyzed to confirm the presence of the entire GH3 domain by PHMMER (<http://plants.ensembl.org/hmmer/index.html>), InterProScan (<http://www.ebi.ac.uk/interpro/sequence-search>), and NCBI Conserved Domain Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). The *GH3* genes in the three species of *Brassica* were renamed on the basis of the standard gene nomenclature for the *Brassica* genus [34].

### 2.3. Chromosome Location, Gene Structure, and Protein Property Analysis

The information about the chromosomal locations of the *GH3* genes were obtained from GFF files in BRAD database. The diagrams of the gene chromosome locations were drawn using the MapInspect software. The structures of the *GH3* genes were displayed to illustrate the exon-intron composition by TBtools according to the gene annotation from GFF files. The GH3 protein properties, including molecular weight (MW), isoelectric points (pI), and grand average of hydropathy (GRAVY) were calculated by ProtParam tool (<https://web.expasy.org/protparam/>). The sequences of all GH3 proteins were carried on multiple amino acid alignment by MAFFT to find conserved regions. The protein structures of AtGH3-5 (5KOD), AtGH3-11 (4EPL), AtGH3-12 (4EWV), and AtGH3-15 (6AVH) were obtained from Protein Data Bank (<https://www.rcsb.org/>). The online tool Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) was used to predict the GH3 protein dimensional structure via homology modeling under “intensive” modelling mode [35]. The software Swiss-PdbViewer\_4.10 was used to examine the locations of  $\alpha$ -helices/ $\beta$ -strands and display the protein dimensional structure.

### 2.4. Identification of Orthologous Genes and Syntenic Genes Among the Three Species of Brassica and Arabidopsis thaliana

The most orthologous sets of *GH3* genes among *B. rapa*, *B. oleracea*, *B. napus*, and *A. thaliana* were acquired from the previous study [27], and others were identified by reciprocal BLASTP analysis. Syntenic genes were searched using online tool SynOrths (<http://brassicadb.org/brad/searchSyntenytPCK.php>) [36]. The Syntenic relationships were shown by Circos [37].

### 2.5. Analysis of Duplication Pattern and Phylogenetic Relationship of the GH3 Gene Family

Five types of gene duplications were examined using MCScanX, namely singleton, dispersed, proximal, tandem, and segmental duplication/WGD [38]. Phylogenetic tree was constructed by MEGA7 with the neighbor-joining (NJ) method and bootstrap replication of 1000.

### 2.6. Calculation of $d_N$ and $d_S$ Values

The protein sequences of each GH3 orthologous pair between *B. napus* and two diploid ancestors were aligned using MAFFT. Then, a multiple sequence alignment of proteins and the corresponding DNA sequences were inputted into PAL2NAL to convert into a codon alignment. Finally, the codon alignment was subjected to the calculation of synonymous ( $d_S$ ) and non-synonymous ( $d_N$ ) substitution rates by the codeml program in PAML [39].

### 2.7. Search of cis-Elements in the GH3 Gene Promoter Regions

For all GH3 genes, the 2000 bp of upstream genomic sequences relative to the translation start codon were extracted from the three species of *Brassica* genomes. PLACE was used to scan cis-elements on both strands of the promoter sequences [40].

### 2.8. GH3 Gene Expression Analysis

Stems, young leaves, blooming flowers, and siliques (ten days after pollination) were collected from nine independent plants (three *B. rapa*, three *B. oleracea*, and three *B. napus*) that grew for six months. There were three biological replicates for stems, leaves, flowers, and siliques in each *Brassica* species. All plant tissues were frozen in liquid nitrogen for later use. Total RNA was extracted from these plant tissues using Trizol reagent (Invitrogen) with the manufacturer's procedure. The residual DNA was digested using RNase-free DNase I (Fermentas, Canada). The thirty-six cDNA libraries were constructed, and then checked for quality using ABI StepOnePlus Real-Time PCR System Qualification and Agilent 2100 Bioanalyzer. Finally, 150 bp paired-end sequencing was performed using an Illumina HiSeq XTen. All reads were qualified: the adapter reads, the reads in which the proportion of unidentified base was greater than five percent, and the low-quality reads, in which proportion of base ( $Q \leq 15$ ) was greater than twenty percent were removed. The clean reads of the three species of *Brassica* were mapped to the *B. rapa*, *B. oleracea*, and *B. napus* reference genome sequences using Bowtie2, respectively [41]. The reads that mapped to reference genome were assembled into the transcripts using Cufflinks software. The Cuffcompare program was used to annotate the assembled transcripts according to the annotation files of the three species genome sequence. Next, RSEM was used to calculate the expression levels of GH3 genes [42]. The measurement unit was fragments per kilobase of exon per million mapped reads (FPKM). Z-value was used to normalize expression levels, and the FPKM values of all GH3 genes were converted to Z-values, then Z-values were used to construct the heatmaps by TBtools.  $Z\text{-value} = \frac{\log_2(\text{FPKM}) - \text{Mean}(\log_2(\text{FPKM}) \text{ of all samples})}{\text{standard deviation}(\log_2(\text{FPKM}) \text{ of all samples})}$ . The raw data of RNA-seq reads were deposited in the NCBI database (accession number SRR7816633-SRR7816668, unreleased).

## 3. Results

### 3.1. Identification and Chromosomal Distribution of GH3 Genes

A total of 38, 25, and 66 GH3 genes were confirmed in *B. rapa*, *B. oleracea*, and *B. napus*, respectively (Table 1). Each GH3 gene was allocated a unique name in accordance with the standardized gene nomenclature (Table S1). Compared with *A. thaliana*, *Brassica* genome experienced a unique triplication, so each *AtGH3* gene theoretically had three copies in *B. rapa* and *B. oleracea*. However, that is not the case. The *AtGH3-4* and *AtGH3-13* genes had no copy in *B. rapa* and *B. oleracea*. Most *AtGH3* genes had one or two copies in *B. rapa* and *B. oleracea*. This indicated that *B. rapa* and *B. oleracea* underwent gene loss in the process of *Brassica* WGT (whole genome triplication) after the divergence with *A. thaliana*.

In addition, *AtGH3-8*, *AtGH3-12*, and *AtGH3-19* had more than three copies in *B. rapa*, and *AtGH3-8* also had more than three copies in *B. oleracea*, suggesting that a few *GH3* genes underwent gene expansion in *B. rapa* and in *B. oleracea*. *B. napus* is derived from the hybridization between *B. rapa* and *B. oleracea* followed with chromosome doubling, and the number of *GH3* genes in *B. napus* should be the sum of two diploids in theory. However, *BnaGH3* also had expansion or loss after allopolyploidization. For example, the number of *BnaGH3-8* genes exceeded the sum of *BraGH3-8* and *BolGH3-8* genes, while the number of *BnaGH3-19* genes was one third of the total number of *BraGH3-19* and *BolGH3-19*. Although *BnaGH3* genes went through expansion or loss during the evolution process, the number of *BnaGH3* genes was close to the sum of *BraGH3* and *BolGH3* genes.

**Table 1.** The number of *GH3* genes in the three species of *Brassica*.

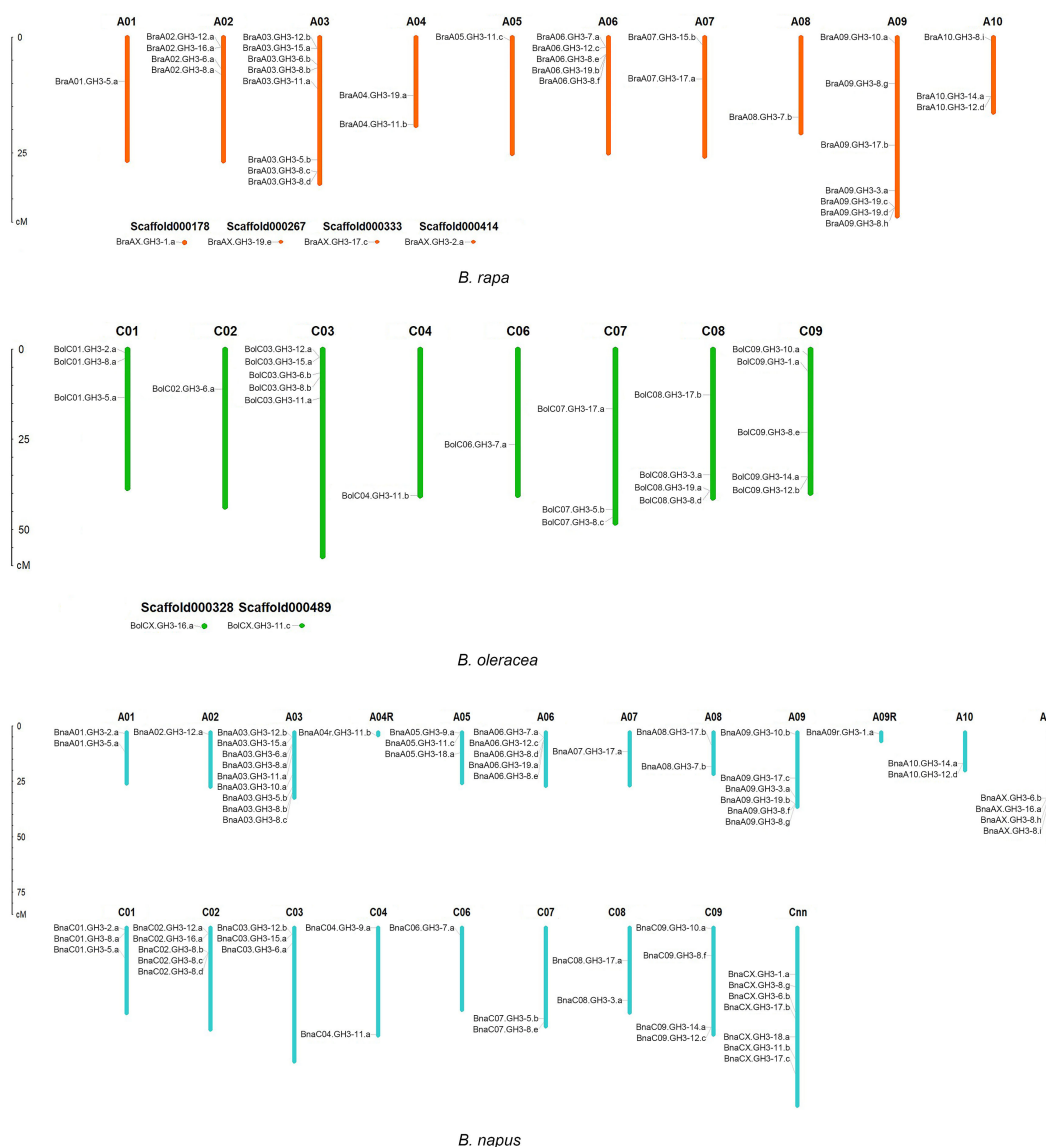
<i>GH3</i> Gene	<i>Brassicarapa</i>	<i>Brassicaoleracea</i>	<i>Brassicananapus</i>
<i>GH3-1</i>	1	1	2
<i>GH3-2</i>	1	1	2
<i>GH3-3</i>	1	1	2
<i>GH3-4</i>	0	0	0
<i>GH3-5</i>	2	2	4
<i>GH3-6</i>	2	2	4
<i>GH3-7</i>	2	1	3
<i>GH3-8</i>	9	5	16
<i>GH3-9</i>	0	0	2
<i>GH3-10</i>	1	1	3
<i>GH3-11</i>	3	3	5
<i>GH3-12</i>	4	2	7
<i>GH3-13</i>	0	0	0
<i>GH3-14</i>	1	1	2
<i>GH3-15</i>	2	1	2
<i>GH3-16</i>	1	1	2
<i>GH3-17</i>	3	2	6
<i>GH3-18</i>	0	0	2
<i>GH3-19</i>	5	1	2
Total	38	25	66

The *GH3* genes were unevenly distributed across the chromosomes in *B. napus*. The number of *BnaGH3* genes in the  $A_n$  and  $C_n$  subgenomes was 37 and 29, respectively, which was similar to that in *B. rapa* ( $A_r$ , 38) and *B. oleracea* ( $C_o$ , 25). The number distribution comparison of *GH3* genes between  $A_n$  and  $A_r$ ,  $C_n$ , and  $C_o$  was carried out. The number of *GH3* genes on each chromosome was counted, and the results were shown in Table 2. For the  $A_n$  subgenome, the top three chromosomes were  $A_n03$ ,  $A_n09$ , and  $A_n06$  in *B. napus*, which was the same as ancestor *B. rapa*, in which the top three chromosomes were  $A_r03$ ,  $A_r09$ , and  $A_r06$ . The number distribution pattern of the remaining *BnaGH3* genes in the  $A_n$  subgenome was 4 ( $A_{nn}$ -random), 3 ( $A_n05$ ), 2 ( $A_n01$ ,  $A_n08$  and  $A_n10$ ), 1 ( $A_n02$ ,  $A_n04$ -random,  $A_n07$ , and  $A_n09$ -random). The number distribution pattern of the remaining *BraGH3* genes in the  $A_r$  genome also was 4 ( $A_r02$  and scaffolds), 3 ( $A_r10$ ), 2 ( $A_r04$  and  $A_r07$ ), 1 ( $A_r01$ ,  $A_r05$ ,  $A_r08$ ). For the  $C_n$  subgenome,  $C_{nn}$ -random contained the largest number. However, for the  $C_o$  genome,  $C_o03$  and  $C_o09$  contained the largest numbers. There was no *BnaGH3* gene on  $C_n05$ , which was the same as  $C_o05$ . Excluding  $C_n02$ , the number distribution pattern of the remaining *BnaGH3* genes in the  $C_n$  subgenome was still 4 ( $C_n09$ ), 3 ( $C_n01$ ,  $C_n03$ ), 2 ( $C_n04$ ,  $C_n07$  and  $C_n08$ ), 1 ( $C_n06$ ), which was in line with the  $C_o$  genome. Chromosomal location distribution of *GH3* genes in the three species of *Brassica* was shown in Figure 1.

**Table 2.** The number of *GH3* genes on each chromosome.

The Number of <i>GH3</i> Gene	<i>B. rapa</i> Chromosome	<i>B. oleracea</i> Chromosome	<i>B. napus</i> Chromosome
1	A <sub>r</sub> 01, A <sub>r</sub> 05, A <sub>r</sub> 08	C <sub>o</sub> 02, C <sub>o</sub> 04, C <sub>o</sub> 06	A <sub>n</sub> 02, A <sub>n</sub> 04-random, A <sub>n</sub> 07, A <sub>n</sub> 09-random, C <sub>n</sub> 06
2	A <sub>r</sub> 04, A <sub>r</sub> 07	Scaffolds	A <sub>n</sub> 01, A <sub>n</sub> 08, A <sub>n</sub> 10, C <sub>n</sub> 04, C <sub>n</sub> 07, C <sub>n</sub> 08
3	A <sub>r</sub> 10	C <sub>o</sub> 01, C <sub>o</sub> 07	A <sub>n</sub> 05, C <sub>n</sub> 01, C <sub>n</sub> 03
4	A <sub>r</sub> 02, Scaffolds	C <sub>o</sub> 08	A <sub>nn</sub> -random, C <sub>nn</sub> 09
5	A <sub>r</sub> 06	C <sub>o</sub> 03, C <sub>o</sub> 09	A <sub>n</sub> 06, C <sub>n</sub> 02
6			A <sub>n</sub> 09
7	A <sub>r</sub> 09		C <sub>nn</sub> -random
8	A <sub>r</sub> 03		
9			A <sub>n</sub> 03

Scaffolds: assembled sequence that failed to be incorporated into the chromosomes; A<sub>nn</sub>-random: unmapped A chromosomes of *B. napus* genome; C<sub>nn</sub>-random: unmapped C chromosomes of *B. napus* genome; A<sub>n</sub>04-random and A<sub>n</sub>09-random: the specific positions of genes on A<sub>n</sub>04 and A<sub>n</sub>09 are still unknown.



**Figure 1.** The chromosomal location of *GH3* genes in the three species of *Brassica*. Scaffolds: assembled sequence that failed to be incorporated into the chromosomes; A04R means A04-random; A09R means A09-random; A04-random and A09-random: the specific positions of genes on A04 and A09 are still unknown; A<sub>nn</sub> means A<sub>nn</sub>-random; C<sub>nn</sub> means C<sub>nn</sub>-random; A<sub>nn</sub>-random and C<sub>nn</sub>-random: unmapped A chromosomes of *B. napus* genome and unmapped C chromosomes of *B. napus* genome.

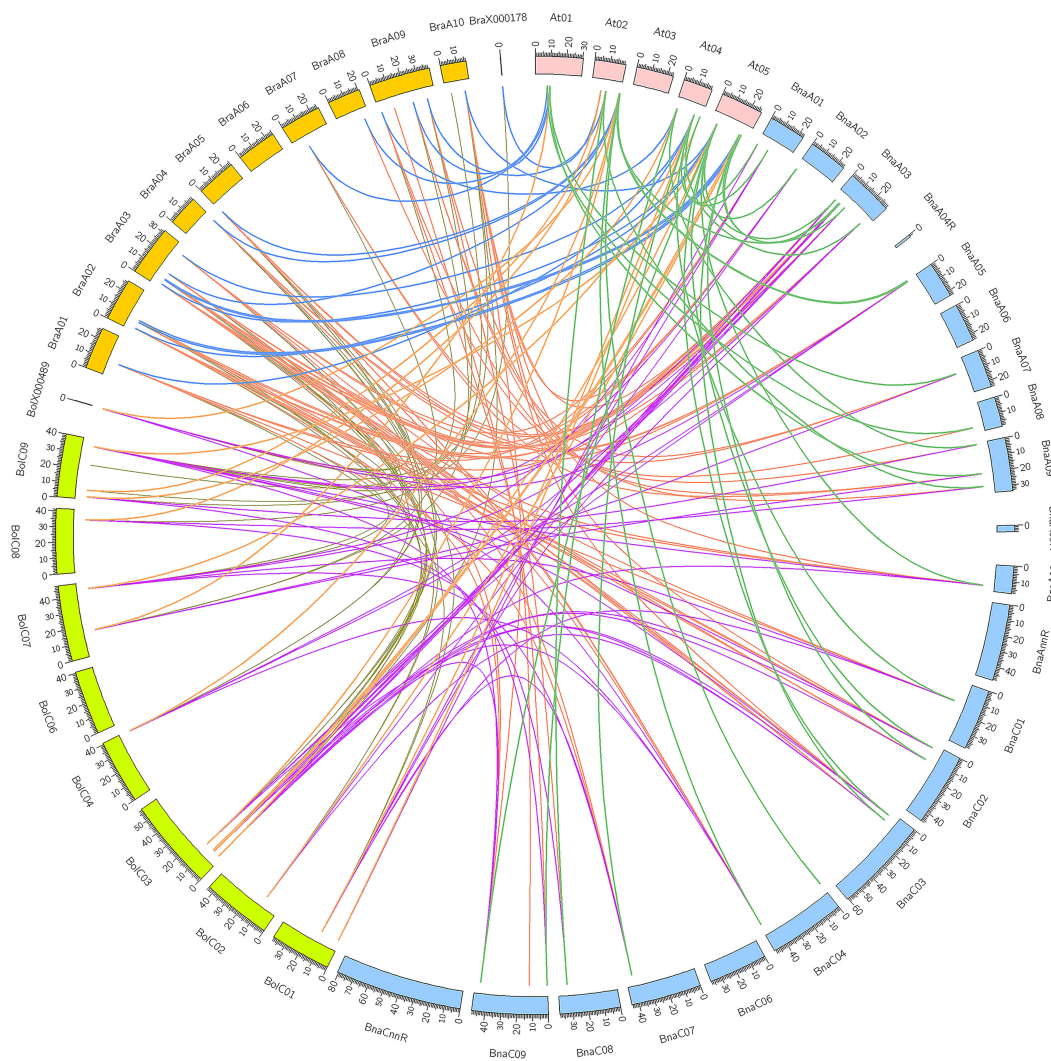
### 3.2. Orthologous Relationship and Synteny Analysis of GH3 Genes

Thirty-three orthologous *GH3* gene pairs were identified between  $A_n$  and  $A_r$ , and twenty orthologous *GH3* gene pairs were found between  $C_n$  and  $C_o$  (Table S2). Chalhoub et al. reported that 27,360 of 32,699 orthologous gene pairs (83.7%) between ancestor *B. rapa* and *B. oleracea* were retained as homologous gene pairs in *B. napus* [27]. Our results revealed that most of the orthologous *GH3* gene pairs (82.6%) between *B. rapa* and *B. oleracea* were retained as homologous gene pairs in *B. napus* (Table S2). These results suggested that most of the *GH3* genes remained intact during the formation and evolution of allotetraploid *B. napus*.

Thirteen *BnaGH3* genes had no corresponding *BraGH3* or *BolGH3* orthologous genes (four *BnaGH3* genes on the A subgenome had no corresponding *BraGH3* orthologous gene, and nine *BnaGH3* genes on the C subgenome had no corresponding *BolGH3* orthologous gene in Table S2), and we analyzed the origins of these thirteen genes. Seven *BnaGH3* genes (Table S3) had corresponding *B. rapa* or *B. oleracea* orthologous genes, but the proteins that these ancestral orthologous genes corresponded to did not have a complete GH3 domain. There may have been two possibilities. In the first possibility, the sequences of the *GH3* genes from the ancestors changed after allopolyploidization, which resulted in the formation of a complete GH3 domain. In the second possibility, before the occurrence of allopolyploidization, the proteins that the seven ancestral genes corresponded to had complete GH3 domains. After allopolyploidization, the proteins that the seven genes from the ancestors corresponded to still had complete GH3 domains in *B. napus*, and the GH3 domains had been kept complete. However, for the ancestors, the proteins that the seven genes corresponded to lost the complete GH3 domains in the recent evolutionary process. The remaining six *BnaGH3* genes (*BnaC02.GH3-8.b*, *BnaC02.GH3-8.c*, *BnaA09.GH3-8.g*, *BnaCX.GH3-8.g*, *BnaAX.GH3-8.i*, and *BnaCX.GH3-17.c*) had no corresponding *B. rapa* or *B. oleracea* orthologous genes, indicating that these six genes may appear as new genes during the formation of *B. napus*. In the next section, we carried out the analysis of the gene duplication, and *BnaA09.GH3-8.g*, *BnaC02.GH3-8.b*, and *BnaC02.GH3-8.c* were newly generated tandem repeat genes after allopolyploidization.

Syntenic genes are orthologous genes that locate in syntenic fragments between different species that derive from a shared ancestor. They usually have similar functions, so syntenic gene analysis is used to understand gene function of newly sequenced genomes. Besides, syntenic gene analysis can also reveal species genomic evolution. Three subgenomes are produced because of genome triplication in the diploid *B. rapa* and *B. oleracea*, and then a lot of genes are lost. According to the extent of gene retention, three subgenomes are specified as LF (maximum gene retention), MF1, and MF2 (least gene retention). The results of syntenic gene analysis showed that *AtGH3-11* had three syntenic orthologs in both *B. rapa* and *B. oleracea*, which confirmed the occurrence of *Brassica* genome triplication. *AtGH3-12* also had three syntenic orthologs in *B. rapa*. The other *AtGH3* genes had 0, 1, or 2 syntenic orthologs in *B. rapa* or *B. oleracea*, which suggested that *GH3* genes underwent loss after *Brassica* genome triplication. After allopolyploidization, each *AtGH3* gene theoretically had six syntenic orthologs in *B. napus*, but only *AtGH3-12* had six syntenic orthologs in *B. napus*. The other *AtGH3* genes had zero to four syntenic orthologs in *B. napus*. There were three ways to lose syntenic *BnaGH3* orthologs. First, after the *Brassica* genome triplication, the loss of the *GH3* genes in the diploid *B. rapa* and *B. oleracea* resulted in the loss of syntenic *BnaGH3* orthologs. For example, *AtGH3-3* had one syntenic ortholog in both *B. rapa* and *B. oleracea*, so it had two syntenic *BnaGH3* orthologs. Second, the loss of the *GH3* genes after the allopolyploidization. For example, *AtGH3-11* had three syntenic orthologs in both *B. rapa* and *B. oleracea*, while it had three syntenic orthologs in *B. napus*. Finally, the *Brassica* genome triplication process and the allopolyploidization process together caused the loss of syntenic *BnaGH3* orthologs. For example, *AtGH3-1* had one syntenic ortholog in both *B. rapa* and *B. oleracea*, while it had no syntenic ortholog in *B. napus*. In addition, allopolyploidization could make *AtGH3* genes gain syntenic *BnaGH3* orthologs. *AtGH3-9* had no syntenic ortholog in both *B. rapa* and *B. oleracea*, while it had two syntenic orthologs in *B. napus*. *AtGH3-12* had three and two syntenic orthologs in *B. rapa* and *B. oleracea*, while

it had six syntenic orthologs in *B. napus*. The syntenic relationship pairs between *GH3* genes from the three species of *Brassica* and *A. thaliana* were shown in Figure 2.



**Figure 2.** The synteny analysis of *GH3* genes in the three species of *Brassica* and *Arabidopsis thaliana* chromosomes.

To comprehend whether natural selection acted on the evolution of the *GH3* gene family in *B. napus*, selection pressure analysis was performed on the orthologous *GH3* gene pairs between  $A_n$  and  $A_r$ ,  $C_n$ , and  $C_o$ . The non-synonymous rate ( $d_N$ ) and synonymous rate ( $d_S$ ) value were calculated. The  $d_N/d_S$  ratio  $> 1$  represents positive selection, the  $d_N/d_S$  ratio  $= 1$  represents neutral selection, and the  $d_N/d_S$  ratio  $< 1$  represents purifying selection [43]. The  $d_N/d_S$  ratios of the orthologous gene pairs were shown in Table S4. Seven *BnaGH3* genes had  $d_N/d_S$  ratios that were  $\gg 1$ , suggesting that these genes experienced stronger positive selection pressure during the evolution process. In addition, the  $d_N/d_S$  ratios of *BnaC09.GH3-10.a* and *BnaC06.GH3-7.a* were 1.3000 and 1.6412, respectively, which indicated that these two genes were subject to positive selection pressure. The positive selection could promote the function changes of these nine *BnaGH3* genes to survive, and mutations were advantageous. The rest of *BnaGH3* genes had  $d_N/d_S$  ratios that  $< 1$ , which indicated that they underwent purifying selection during the evolution process and may preferentially conserve function, with mutations being disadvantageous. The average  $d_N/d_S$  ratio of *BnaGH3* genes on the  $A_n$  subgenome was higher than that on the  $C_n$  subgenome, showing that *BnaGH3* genes on the  $A_n$  subgenome underwent higher selection pressure during the evolution process of *B. napus*.



### 3.3. The Duplication Pattern Analysis of GH3 Genes

The duplication patterns of *GH3* genes were analyzed in *B. napus* and two ancestor *Brassica* species by MCScanX. The four duplicated types were found in each *Brassica* species, namely dispersed, proximal, tandem, and segmental duplication/WGD (Table S5). We found that 55% of *BraGH3* genes, 72% of *BolGH3* genes, and 62% of *BnaGH3* genes were generated by WGD, while 26% of *BraGH3* genes, 12% of *BolGH3* genes, and 24% of *BnaGH3* genes were generated by tandem duplication, suggesting that the duplication pattern of *GH3* genes was high WGD and low tandem duplication. The previous studies showed that WGD contributed the most to the expansion of the gene families in *B. napus* [44,45]. Our findings were consistent with the previous studies, and the WGD type had the largest number of *GH3* genes, which indicated that the main cause of the expansion of *GH3* genes was WGD. The ancestral *Brassica* genome underwent a triplication event, which could greatly promote the expansion of the *GH3* gene family in *B. rapa* and *B. oleracea*. Next, *B. napus* was formed by the hybridization between *B. rapa* and *B. oleracea* followed with chromosome doubling. Therefore, the expansion of the *GH3* gene family in *B. napus* was affected by *Brassica* WGT and allopolyploidization. Four, one, and five clusters of *GH3* tandem repeat genes were found in *B. rapa*, *B. oleracea*, and *B. napus*, respectively (Table 3). Three clusters of tandem repeat genes from *B. rapa* were completely retained in *B. napus*. Although BnaA09g48090D was the orthologous gene of BraA09.GH3-19.d, it was not a member of the *GH3* gene family, because BnaA09g48090D protein had a GH3 domain with the missing N-terminal, so it was not in Table 3. The orthologous gene of BraA06.GH3-8.f was BnaA06.GH3-8.e. Due to the change in the position of BnaA06.GH3-8.e after allopolyploidization, it was not in cluster 2, so this cluster only retained two genes (BnaA06.GH3-19.a and BnaA06.GH3-8.d). The loss of *GH3* tandem repeat genes showed the changes in gene sequence and location after allopolyploidization. In *B. napus*, tandem duplication resulted in the generation of a new gene, namely BnaA09.GH3-8.g. The cluster of tandem repeat genes in *B. oleracea* was not retained in *B. napus*. A new cluster of tandem repeat genes appeared in *B. napus*, namely cluster 5, in which BnaC02.GH3-8.c and BnaC02.GH3-8.d were newly generated genes.

**Table 3.** The clusters of *GH3* tandem repeat genes in the three species of *Brassica*.

Cluster	<i>B. rapa</i>	<i>B. oleracea</i>	<i>B. napus</i>
Cluster 1	BraA06.GH3-12c <sup>a</sup>	BolC08.GH3-8.d	BnaA06.GH3-12.c <sup>a</sup>
	BraA06.GH3-7.a <sup>b</sup>	BolC08.GH3-19.a	BnaA06.GH3-7.a <sup>b</sup>
Cluster 2	BraA06.GH3-8.f		BnaA06.GH3-19.a <sup>c</sup>
	BraA06.GH3-19.b <sup>c</sup>		BnaA06.GH3-8.d <sup>d</sup>
	BraA06.GH3-8.e <sup>d</sup>		
Cluster 3	BraA09.GH3-8.h <sup>e</sup>		BnaA09.GH3-8.f <sup>e</sup>
	BraA09.GH3-19.d		BnaA09.GH3-19.b <sup>f</sup>
	BraA09.GH3-19.c <sup>f</sup>		BnaA09.GH3-8.g
Cluster 4	BraA03.GH3-8.d <sup>g</sup>		BnaA03.GH3-8.c <sup>g</sup>
	BraA03.GH3-8.c <sup>h</sup>		BnaA03.GH3-8.b <sup>h</sup>
Cluster 5			BnaC02.GH3-8.b
			BnaC02.GH3-8.c
			BnaC02.GH3-8.d

Genes with the same superscript letter belong to the same set of orthologous gene. The letters "a" to "h" represent orthologous gene pairs.

### 3.4. Phylogenetic Relationship of *GH3* Genes in *Arabidopsis* and *Brassica*

To explore the evolutionary relationships of the *GH3* genes from the three species of *Brassica* and *A. thaliana*, a phylogenetic tree was generated with 19 *A. thaliana*, 38 *B. rapa*, 25 *B. oleracea*, and 66 *B. napus* members, which would contribute to understanding the potential roles of the *GH3* genes in *Brassica* species. The phylogenetic tree was divided into three groups (I, II, and III), and this pattern of



### 3.5. Gene Structure Analysis and Predicted Protein Characteristics

The 66 *BnaGH3* genes ranged from 1578 to 9498 bp in ORF (open reading frame) length, with an average of 2992 bp. These genes ranged from 1314 to 2379 in CDS (coding domain sequence) length, with an average of 1744 bp (Table S6). Thus, the change in the intron length was broader than that in flanking exon length. To examine the differences in *GH3* ORF and CDS length between *B. napus* and the diploid ancestors, the ORF and CDS length of *GH3* genes in the diploid ancestors were analyzed (Table S6). The 38 *BraGH3* genes ranged from 1888 to 4242 bp in ORF length, with an average of 1754 bp. The *BraGH3* genes ranged from 1530 to 1839 in CDS length, with an average of 1574 bp. The ORF length of the 25 *BolGH3* genes ranged from 1965 to 5355 bp, with an average of 2651 bp. The CDS length of the *BolGH3* genes ranged from 1413 to 1974 bp, with an average of 1756 bp. These results revealed that the average ORF length of the *BnaGH3* genes had larger differences from that of the *GH3* genes in the diploid ancestors, while the average CDS length was similar to the diploid ancestors, suggesting that intron length variation of the *BnaGH3* genes was more extensive than that of the *GH3* genes in the diploid ancestors.

The comparison of the structural pattern of exon-intron composition between *B. napus* and the diploid ancestors was performed. The exon number of *BraGH3* genes was from two to six, with most genes (63.2%) having four exons (Figure 4a). The exon number of *BolGH3* genes was from two to between six and eight. The number of *BolGH3* genes with three exons and genes with four exons were both 9 (37%) (Figure 4b). In *B. napus*, the number of *GH3* gene exons was from two to seven, and the genes (42.4%) with four exons was the most prevalent (Figure 4c). A new exon-intron organization appeared in *B. napus*, and two *BnaGH3* genes had seven exons. The variation of exon number was similar, and the difference was the proportion of genes with the same exon number between *B. napus* and the diploid ancestors. Finally, the gene structure between orthologous gene pairs (33  $A_n$ - $A_r$  pairs and 20  $C_n$ - $C_o$  pairs) was compared (Figure 5). For 33  $A_n$ - $A_r$  pairs, there were three patterns, namely, *BnaGH3* genes had the same exon number as *BraGH3* genes, *BnaGH3* genes had more exons than *BraGH3* genes, and *BnaGH3* genes had less exons than *BraGH3* genes (Table S7). Twenty  $A_n$ - $A_r$  orthologous pairs (60.6%) had the same exon number, in which 15 pairs had the same CDS length, but they had different ORF lengths, suggesting that differences in intron length led to the differences in orthologous gene structure. The remaining 13  $A_n$ - $A_r$  orthologous pairs had different exon numbers, in which 11 *BnaGH3* genes had more exons than *BraGH3* genes. Although a few *BnaGH3* genes had more exons than *BraGH3* genes, their ORF lengths were almost the same as that of *BraGH3* genes (Table S7). This may be because a part of the exon of *BraGH3* genes became an intron after allopolyploidization, such as BnaA01.GH3-5.a/BraA01.GH3-5.a. In addition, there is another possibility, and a part of the intron of *BraGH3* genes became an exon after allopolyploidization, such as BnaA03.GH3-11.a/BraA03.GH3-11.a. Finally, we found that BnaA06.GH3-12.c had less exons than BraA06.GH3-12.c, and the CDS length of BnaA06.GH3-12.c was also shorter than that of BraA06.GH3-12.c, but the ORF length of BnaA06.GH3-12.c was 1.7 times longer than that of the BraA06.GH3-12.c gene. This is because that the increase in the intron length lengthened the length of the *BnaGH3* gene. For 20  $C_n$ - $C_o$  pairs, there were two patterns, namely, *BnaGH3* genes had the same exon number as *BolGH3* genes, and *BnaGH3* genes had more exons than *BolGH3* genes. Fifteen pairs of 20 (75%) had the same exon number, and eleven of these fifteen  $C_n$ - $C_o$  pairs also had the same CDS length (Table S7). Two pairs not only had the same CDS length, but also owned the same ORF length. In the remaining 5  $C_n$ - $C_o$  pairs, 5 *BnaGH3* genes had more exons than *BolGH3* genes, but not all *BnaGH3* genes had longer CDS than *BolGH3* genes.

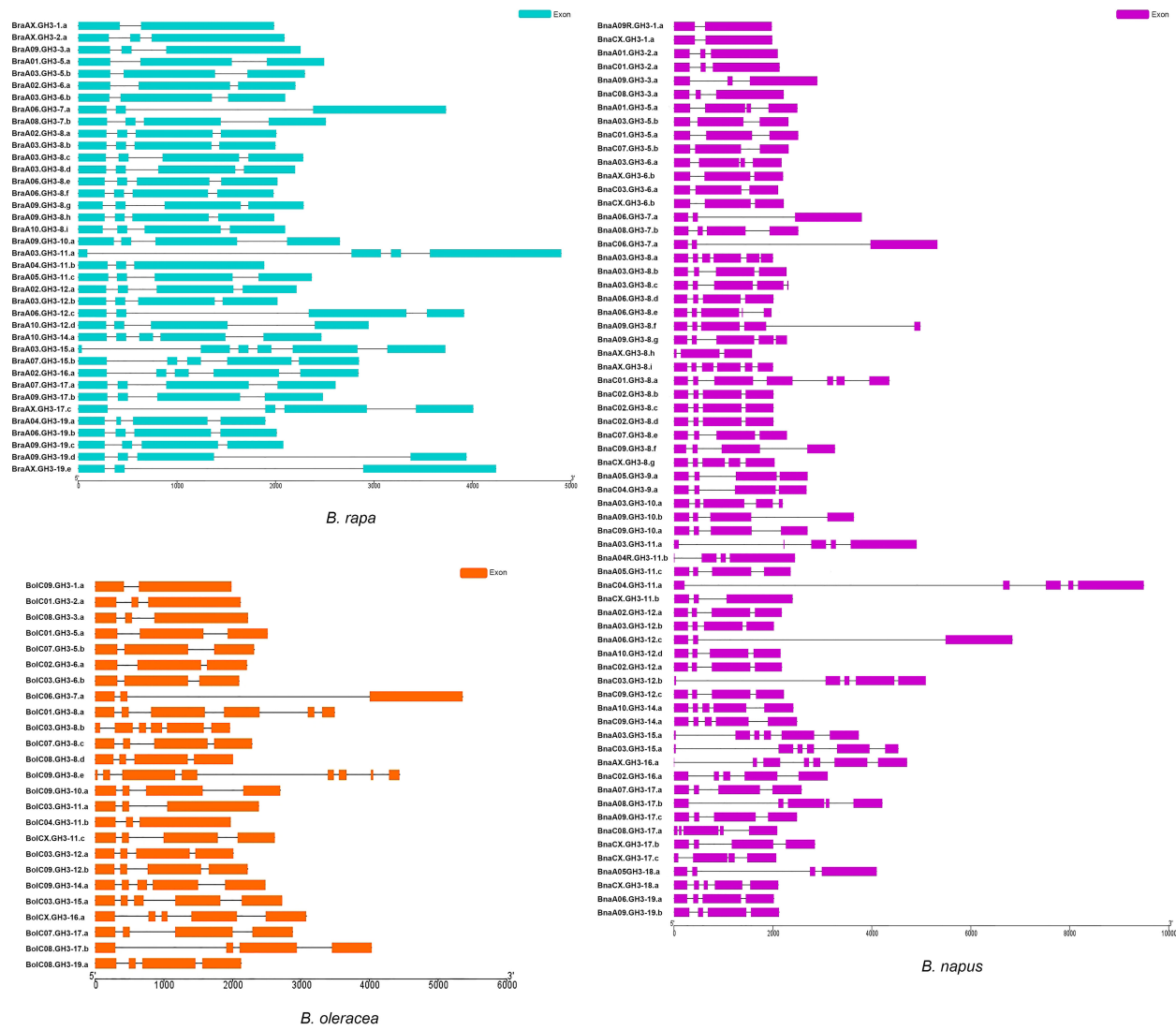
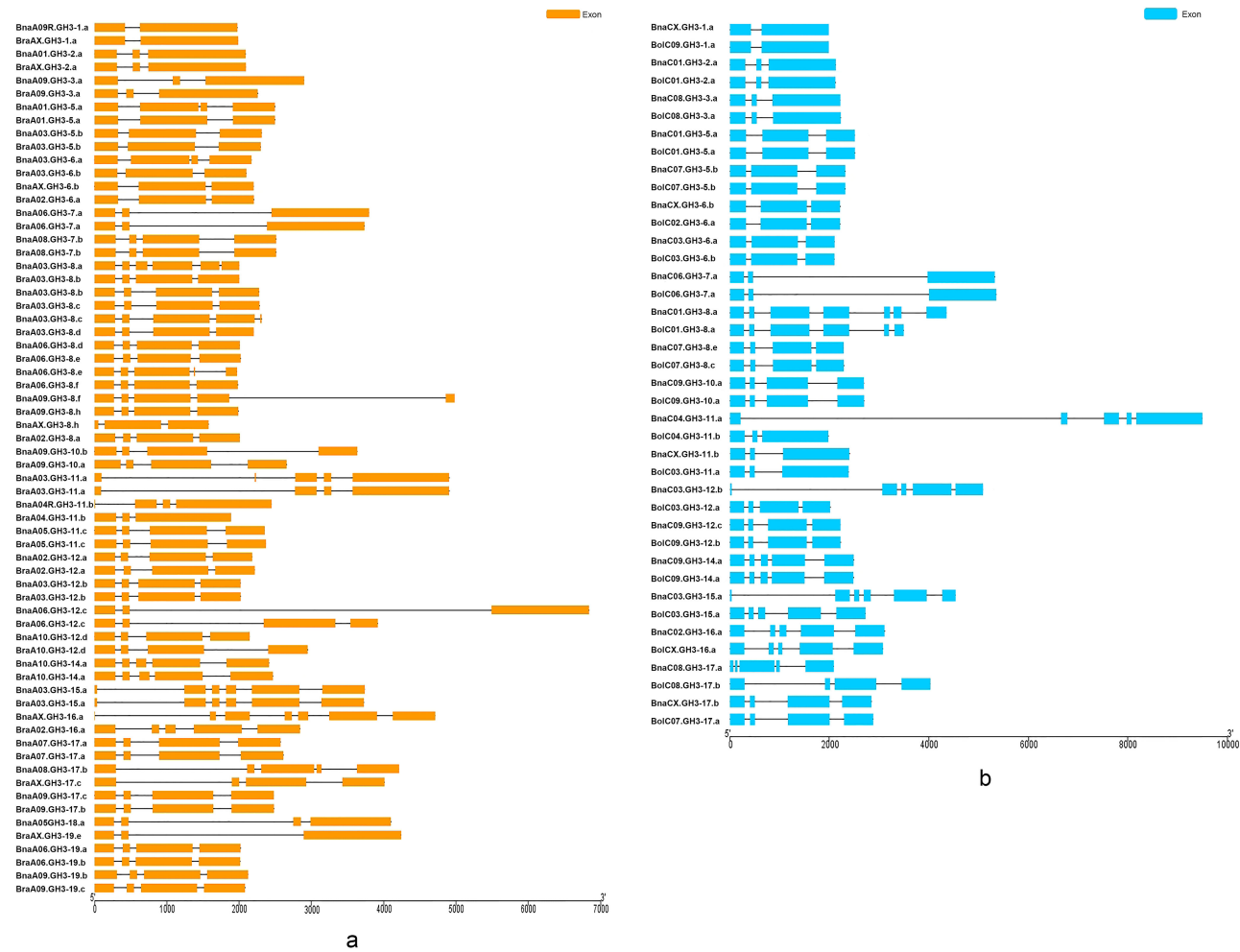


Figure 4. The exon–intron compositions of GH3 genes in the three species of Brassica.



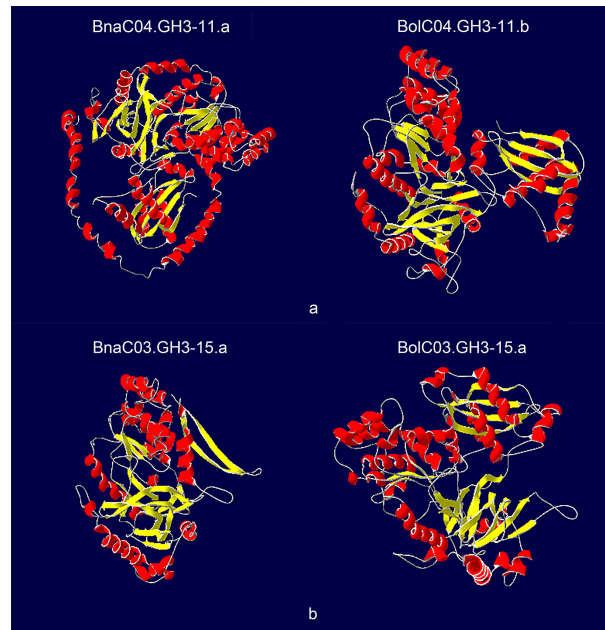
**Figure 5.** The exon–intron compositions of 33 *GH3* A<sub>n</sub>-A<sub>r</sub> orthologous pairs and 20 *GH3* C<sub>n</sub>-C<sub>o</sub> orthologous pairs. (a) Thirty-three *GH3* A<sub>n</sub>-A<sub>r</sub> orthologous pairs; (b) twenty *GH3* C<sub>n</sub>-C<sub>o</sub> orthologous pairs.

BnaC03.GH3-15.a owned shorter CDS, however, the increase in intron length allowed its ORF to be longer than BolC03.GH3-15.a. In summary, most *GH3* orthologous gene pairs between *B. napus* and the diploid ancestors had an almost identical gene structure pattern with regard to exon number and CDS length. These results demonstrated that the *GH3* gene family may be relatively conservative with regard to exon-intron composition, possibly because of their important roles in plant growth and development.

The various physicochemical parameters of GH3 proteins, including protein length, molecular weight, isoelectric point (pI), grand average of hydropathicity (GRAVY), and instability index, were calculated using ExPASy (Table S8). The BnaGH3 protein length ranged from 437 aa to 792 aa, with an average of 580 aa. The molecular weight varied from 49.00 KD to 88.45KD, with an average of 65.51 KD. Except BnaA10.GH3-14.a and BnaC09.GH3-14.a, which had  $pI > 7$ , the BnaGH3 proteins had relatively low pI ( $pI < 7$ ). The GRAVY value of BnaA09.GH3-8.f and BnaA06.GH3-8.e was greater than zero, which showed that these two proteins are hydrophobic. The rest of BnaGH3 proteins were hydrophilic, with a GRAVY value  $< 0$ . Forty-four (66.7%) BnaGH3 proteins possessed an instability index  $> 40$ , indicating that these proteins may be unstable. The physicochemical parameters of GH3 proteins in diploid ancestors were also calculated (Table S8). The BraGH3 protein length ranged from 509 aa to 612 aa, with an average of 584 aa, and the BolGH3 protein length ranged from 470 aa to 657 aa, with an average of 584 aa. The average molecular weight of BraGH3 proteins and BolGH3 proteins were 65.84 KD and 65.99 KD, respectively. Only BraA10.GH3-14.a and BolC09.GH3-14.a had  $pI > 7$ , and their *B. napus* orthologous genes were BnaA10.GH3-14.a and BnaC09.GH3-14.a, respectively. Only BraA06.GH3-8.f had GRAVY  $> 0$ , and its *B. napus* orthologous gene was BnaA06.GH3-8.e. Twenty-five BraGH3 proteins (65.8%) and eighteen BolGH3 proteins (72%) owned an instability index exceeding 40. In a word, there was little difference between BnaGH3 proteins and GH3 proteins of two ancestors with respect to the average protein length and molecular weight. Two BnaGH3 proteins that had  $pI > 7$  were derived from two diploid ancestors. One of two hydrophobic BnaGH3 proteins originated from *B. rapa*. The proportion of unstable proteins was the highest in *B. oleracea*, and the ratio in *B. napus* was similar to that in *B. rapa*.

Multiple amino acid alignment of 148 GH3 proteins (19 *A. thaliana*, 38 *B. rapa*, 25 *B. oleracea*, and 66 *B. napus* members) showed many relatively conserved sites and sequences, such as <sup>26</sup>E and <sup>82</sup>PYIDRI<sup>87</sup> (Figure S1). There are seven highly conserved sites (<sup>255</sup>G, <sup>336</sup>Y, <sup>340</sup>E, <sup>346</sup>N, <sup>362</sup>P, <sup>380</sup>L, <sup>386</sup>G) and one highly conserved sequence (<sup>366</sup>YFEF<sup>369</sup>), which could be required for protein function. Chang et al. reported that the conserved motifs were essential to the adenylation activity for acyl-adenylate/thioester-forming enzyme superfamily [47]. The sequence conservation could correspond to functional conservation of GH3 proteins.

The availability of crystal structures of four *Arabidopsis* GH3 proteins (AtGH3-5, AtGH3-11, AtGH3-12, and AtGH3-15) helped us to compare the differences in secondary structure elements between the corresponding orthologous proteins. A total of 17 pairs of orthologous proteins were used for comparative analysis. Only one pair (BnaC09.GH3-12.c/BolC09.GH3-12.b) had the same number and location of  $\alpha$ -helices and  $\beta$ -strands, suggesting that the secondary structure elements of most GH3 proteins changed after allopolyploidization (Table S9). Two pairs (BnaC04.GH3-11.a/BolC04.GH3-11.b and BnaC03.GH3-15.a/BolC03.GH3-15.a) had larger differences in the protein secondary structure (Figure 6). BnaC04.GH3-11.a had six more  $\alpha$ -helices and one more  $\beta$ -strand than BolC04.GH3-11.b, while BnaC03.GH3-15.a had four less  $\alpha$ -helices and two less  $\beta$ -strands than BolC03.GH3-15.a. We found that BnaC04.GH3-11.a/BolC04.GH3-11.b and BnaC03.GH3-15.a/BolC03.GH3-15.a had different exon-intron composition, and the CDS length of these two pairs differed by several hundred bp, which may lead to larger differences in protein secondary structure. The protein 3-dimensional structures of the other pairs were shown in Figure S2.



**Figure 6.** The 3-dimensional structures of BnaC04.GH3-11.a/BolC04.GH3-11.b and BnaC03.GH3-15.a/BolC03.GH3-15.a. The  $\alpha$ -helices,  $\beta$ -strands, and random coils were colored by red, yellow, and grey, respectively.

### 3.6. Differences in Cis-Acting Elements of Promoter Regions

Cis-acting elements that are bound by transcription factors are associated with plant growth, development, and stress response [48]. Some phytohormone-related and stress-related cis-acting elements have been found in the promoters of *GH3* genes, including W box, G-box, CG box, AuxRE, and SARE [17,23]. To understand the potential functions of three *Brassica* *GH3* genes, phytohormone-related and stress-related cis-acting elements of the *GH3* promoter regions were screened out using PLACE. Seventy-eight, eighty, and eighty-five types of phytohormone-related and stress-related cis-acting elements were discovered in the promoters of *BraGH3*, *BolGH3*, and *BnaGH3* genes, respectively (Table S10). *BraGH3* genes had one specific hypersensitive response element (S000466) and one specific auxin response element (S000368). *BolGH3* genes had one specific wounding response element (S000495) and two specific ABA response elements (S000015 and S000145), while *BnaGH3* genes had two specific G-box elements (S000345 and S000224) and one specific ABA response element (S000500). These results showed that the differences in cis-acting element types of *GH3* genes was very small among the three species of *Brassica*.

All *BraGH3* and *BolGH3* genes contained six common cis-acting elements, while all *BnaGH3* genes only had two common cis-acting elements, suggesting that changes in the promoter sequences of *BnaGH3* genes led to the loss of common cis-acting elements after allopolyploidization. There were 12 types of auxin response elements in ancestor species, and only one type (S000368) lost in *BnaGH3* genes, which indicated the types of auxin response elements had no significant differences between ancestor species and *B. napus*. The types of cis-acting elements were compared between orthologous gene pairs. For  $A_n$ - $A_r$  orthologous gene pairs, nine of thirty-three pairs had the same type of cis-acting elements (Table S11). For five of these nine pairs, we found that the number of each type was also the same. After allopolyploidization, the promoter regions of these five *BnaGH3* genes retained the ancestral types of phytohormone-related and stress-related cis-acting elements and the number of each type, showing highly conservative behavior. The remaining orthologous gene pairs owned the different types of cis-acting element. Compared to corresponding *B. rapa* orthologous genes, some *BnaGH3* genes had specific types of cis-acting elements, such as BnaA03.GH3-5.b. Only two *BnaGH3* had no specific type. For  $C_n$ - $C_o$  orthologous gene pairs, only one pair had the same type of cis-acting element, and the number of each type was also the same. The rest of the orthologous gene pairs owned

the different types of cis-acting element (Table S11). Seventeen *BnaGH3* genes had specific types of cis-acting elements relative to corresponding *B. oleracea* orthologous genes, such as BnaC04.GH3-11.a. Similarly, there are only two *BnaGH3* genes with no specific type. To sum up, most *BnaGH3* genes had specific cis-acting element types relative to orthologous diploid genes, and the promoter sequence changes of *BnaGH3* genes on the  $C_n$  subgenome was slightly larger than that of *BnaGH3* genes on the  $A_n$  subgenome. Although obvious differences in cis-acting element types between orthologous gene pairs existed, all types of cis-acting elements in *B. napus* had no obvious difference compared with the ancestor species, suggesting that the *GH3* gene family remained phytohormone-related and stress-related cis-acting elements from diploid ancestors after allopolyploidization.

### 3.7. Expression Patterns of GH3 Genes in the Three Species of Brassica

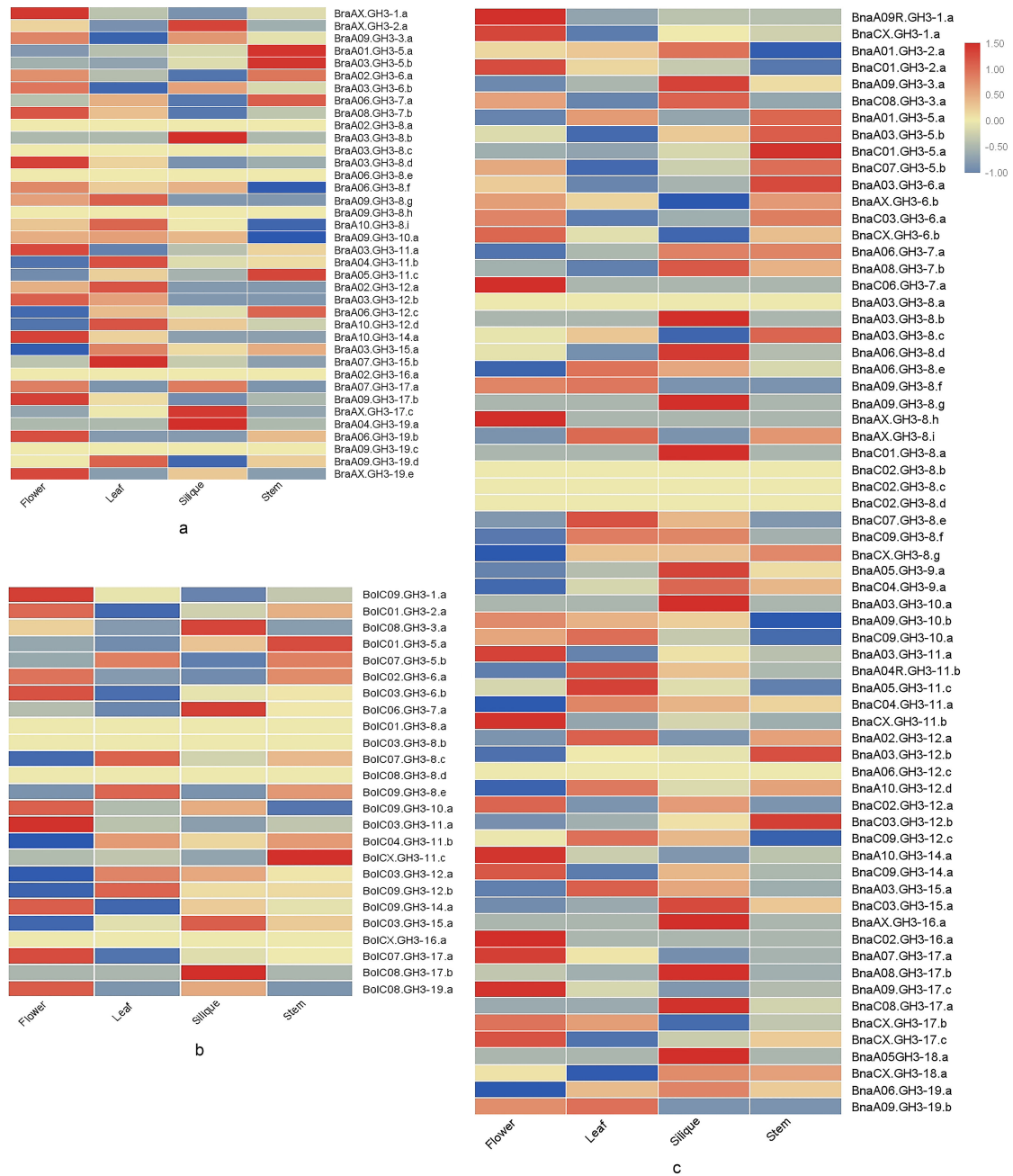
To understand the expression changes of *GH3* genes, we analyzed *GH3* gene expression patterns across four different tissues (stem, leaf, flower, and silique) in the three species of *Brassica*, and compared the expression pattern of *GH3* genes between duplicated genes based on RNA-seq data (Table S12). The *GH3* gene expression patterns in four different tissues for each *Brassica* species were shown in Figure 7. Six *BraGH3* genes, four *BolGH3* genes, and five *BnaGH3* genes had no expression in four tissues, of which eleven were *GH3-8* genes (4 *BraGH3-8* genes, 3 *BolGH3-8* genes, and 4 *BnaGH3-8* genes). The number of *GH3-8* genes was the highest in all three species. To avoid functional redundancy, some *GH3-8* genes may be non-functional or inducible. Compared with ancestor species, *B. napus* had more tissue-specific expressed genes, suggesting that allopolyploidization increased the spatial specificity of gene expression.

To understand how the expression of ancestral genes changed after allopolyploidization, the expression patterns of the orthologous gene pairs were compared between  $A_n$  and  $A_r$ ,  $C_n$ , and  $C_o$ . The results demonstrated that the vast majority of the orthologous gene pairs had disparate expression patterns across four tissues, which reflected gene expression changes after allopolyploidization (Figure 8). Some genes lacked expression in all four tissues in ancestors, however, their orthologous genes were expressed at low levels in *B. napus*, such as BraA06.GH3-8.e/BnaA06.GH3-8.d. In addition, a few genes in *B. rapa* had lower expression levels, while their orthologous genes in *B. napus* were not expressed in all four tissues, such as BraA03.GH3-8.b/BnaA03.GH3-8.a. Finally, some ancestral genes showed high expression levels in tissues, whereas their orthologous genes in *B. napus* showed low expression levels. For example, BolC07.GH3-5.b had higher expression levels in leaf, while BnaC07.GH3-5.b had the lowest expression level in leaf. On the contrary, some ancestral genes showed low expression levels in tissues, whereas their orthologous genes in *B. napus* showed high expression levels. For example, BraA06.GH3-7.a was expressed at the lowest level in silique, while BnaA06.GH3-7.a was expressed at the highest level in silique.

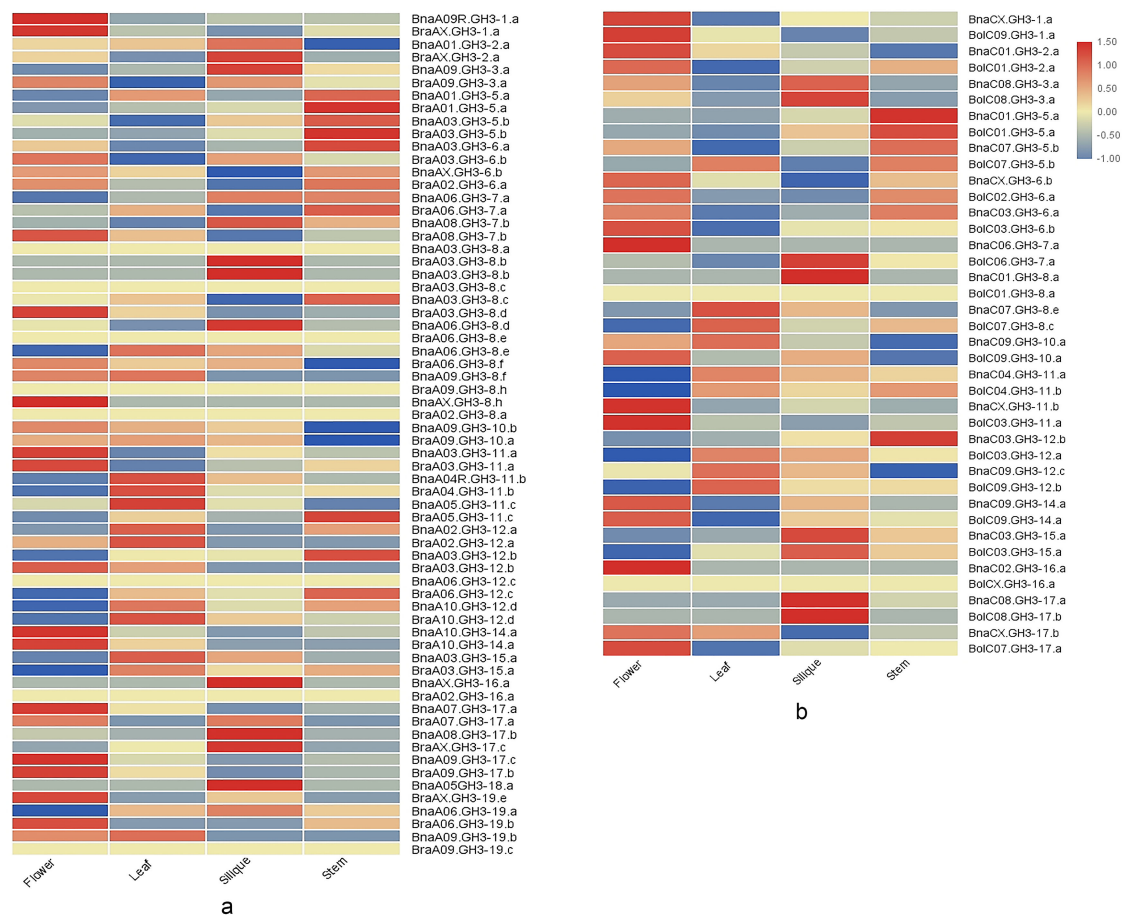
Zhou et al. reported that some duplicated genes of *bZIP* transcription factor family in *B. napus* had opposite expression patterns in seventeen tissues [49]. To comprehend whether the *GH3* gene family in *B. napus* also had this phenomenon, we compared the expression patterns of the duplicated genes. The duplicated genes had a diverged expression pattern. Tandem duplication gene BnaA06.GH3-7.a and BnaA06.GH3-12.c showed opposite expression patterns. BnaA06.GH3-12.c lacked expression in all tissues, however, BnaA06.GH3-7.a had high expression in all tissues. The expression patterns of the most homologous gene pairs between  $A_n$  and  $C_n$  were diverse. For example, BnaA03.GH3-12.b had lower expression in all tissues (FPKM < 1), and the expression levels in the leaf and the silique were comparable. Its duplication gene BnaC03.GH3-12.b had higher expression in all tissues, and the expression level in the silique is 1.8 times higher than that of the leaf. In addition, the tissue-specific expression of the duplicated genes was also different. BnaC02.GH3-16.a was the flower-specific expression, while BnaAX.GH3-16.a was the silique-specific expression. Finally, a few duplicated gene pairs shared similar expression patterns in four tissues, such as BnaA03.GH3-11.a and BnaCX.GH3-11.b.



In summary, the expression levels of the ancestral species *GH3* genes changed obviously after allopolyploidization. Duplicated gene pairs in *B. napus* showed diverse expression patterns, which could be due to the functional divergence during the process of evolution.



**Figure 7.** The expression profiles of *GH3* genes in different tissues for the three species of *Brassica*. (a) *B. rapa*; (b) *B. oleracea*; (c) *B. napus*.



**Figure 8.** The comparison of expression patterns between orthologous pairs. (a) Thirty-three *GH3* A<sub>n</sub>-A<sub>r</sub> orthologous pairs; (b) twenty *GH3* C<sub>n</sub>-C<sub>o</sub> orthologous pairs.

#### 4. Discussion

In recent years, increasing studies have showed that *GH3* genes played an important role in plant growth and development, but the genome-wide comprehensive analysis of the *GH3* gene family in *B. napus*, and how the exon-intron organization, cis-element distribution, and gene expression level of the *BnaGH3* genes change after allopolyploidization still remain largely unknown. In this study, we confirmed the origin of *BnaGH3* genes and analyzed changes in *BnaGH3* gene exon-intron composition, cis-element distribution, and expression pattern relative to diploid ancestors. These results may provide a reference for exploring the molecular mechanisms of the adaptive enhancement in *Brassica* allopolyploid.

##### 4.1. The Brassica Genome Triplication and Allopolyploidization Bring about the Expansion and Loss of *GH3* Genes

The whole genome duplication (*Brassica* WGT and allopolyploidization) contributes the most to the expansion of gene families in *B. napus*, such as the *Aux/IAA* gene family, the *BES1* gene family, and the *AP2/ERF* gene family [44,45,50]. Our results showed that the expansion of the *GH3* gene family was also mainly affected by the whole genome duplication. Previous research reported that whole genome duplication had a significant effect on the amplification of the *GH3* gene family in legumes and apples [46,51].

The *Brassica* genus underwent an extra whole genome triplication relative to *Arabidopsis*, therefore, one *A. thaliana* gene should possess three corresponding *B. rapa* or *B. oleracea* orthologous genes and six *B. napus* orthologous genes [52]. With 19 *GH3* genes in *A. thaliana*, there should be theoretically about 60 and 120 *GH3* genes in *B. rapa/B. oleracea* and *B. napus*. However, only 38, 25,

and 66 *GH3* genes were found in *B. rapa*, *B. oleracea*, and *B. napus*, respectively. The number of *GH3* genes in the  $A_n$  and  $C_n$  was 39 and 27 in *B. napus*, which was almost the same as that in  $A_r$  (*B. rapa*) and  $C_o$  (*B. oleracea*) genomes. These results demonstrated that the loss of most *GH3* genes happened after *Brassica* genome triplication. For example, *AtGH3-1* had only one corresponding *B. rapa* and *B. oleracea* orthologous gene, so only two *GH3-1* genes remained in *B. napus*. *AtGH3-4* had no corresponding *B. rapa* and *B. oleracea* orthologous gene, which could lead to the absence of *GH3-4* in *B. napus*. After *Brassica* genome triplication, 35% genes were lost in the *Brassica* lineage, most likely through a deletion mechanism [53]. The *Brassica* WGT resulted in more *GH3* genes in the *Brassica* diploid progenitors than in *A. thaliana*, which was followed by gene loss. Figure S3 showed the time of WGT and allopolyploidization events and the number of *GH3* genes that were missing or new in the different times.

We found that some *GH3* genes in diploid ancestors were lost after allopolyploidization. Five *BraGH3* genes and five *BolGH3* genes had no corresponding *B. napus* orthologous genes, which suggested that these genes could be lost during the formation of allotetraploid. There were three possibilities for the absence of ancestor *GH3* genes. The first possibility, *B. napus* went through rapid inter-chromosomal rearrangements and chromosomal fragment losses following allopolyploidization, which resulted in these *GH3* genes' losses from the *B. napus* genome [54]. The second possibility, these *GH3* genes became pseudogenes via accumulating deleterious mutants, and pseudogenes could be deleted or diverge too extensively to be confirmed during the evolution process of *B. napus* [55]. The final possibility, the absence or incompleteness of the *GH3* functional domain caused these ancestor *GH3* proteins to no longer be *GH3* proteins after allopolyploidization, which could be due to the changes of ancestor *GH3* gene structures or sequences. The gene loss has been regarded as an advantageous event during the genome evolution [56]. The loss of a gene encoding an anthocyanin pathway enzyme was involved in the transition from blue to red flowers, which led to phenotypic differences among *Andean Iochroma* species [57]. The deletion of a penicillin-binding protein was associated with the resistance to the cephalosporin drug ceftazidime [58]. The loss of these ancestor *GH3* genes may play beneficial and adaptive roles in the evolution process of *B. napus*.

Several molecular mechanisms are involved in the production of new genes, such as gene duplication, exon shuffling, lateral gene transfer, and so on [59]. In our results, thirteen new *BnaGH3* genes appeared after allopolyploidization. Changes in the gene structures or sequences of seven ancestor *GH3* genes (their corresponding proteins had incomplete functional domains) resulted in the reacquisition of the complete *GH3* domains during the allopolyploidization. For example, Bra004432 protein had an incomplete *GH3* domain, so it was not considered as a *GH3* protein, while its orthologous protein BnaA05.GH3-9.a had a complete *GH3* domain after the formation of *B. napus*. The tandem duplication was one of the expansion mechanisms of the *GH3* gene family in *Medicago truncatula* [17]. We also found that three new *BnaGH3* genes originated from the tandem duplication. The production of the remaining three new *BnaGH3* genes may be due to the following molecular mechanisms. The first molecular mechanism was the retroposition. MRNAs are reverse-transcribed into double strand cDNAs and then incorporated back into new genomic positions to form retrogenes. Many retrogenes have been identified in rice, and most of them go through negative selection, suggesting that these retrogenes are likely functional [60]. The second molecular mechanism was the lateral gene transfer. The exchanges of genes between different species, organelles, and nuclei can create new genes. Mitochondrial genes that encode respiratory and ribosomal proteins are subject to lateral transfer between flowering plants [61]. The final molecular mechanism was the exon shuffling. Exons from different genes are reassembled to form a novel exon–intron structure. A defensin gene was inserted into an exon that was from a gene belonging to another subfamily, and the tissue specificity expression of this gene was still preserved [62]. Since new genes play an important role in plant growth, development, and defense [63–65], the appearance of new *BnaGH3* genes may contribute to the adaptability improvement of *B. napus*.

#### 4.2. Variant Expression Patterns of BnaGH3 Genes May Contribute to the Adaptability Improvement in *Brassica napus*

Gene expression changes in allopolyploids could be an adaptive mechanism, which makes allopolyploids evolve rapidly and steadily in nature [66]. The expression changes in genes mediated by circadian clock genes resulted in strong growth vigor and increased biomass in *Arabidopsis* allotetraploids [67]. Compared with *Brassica* diploids, *Brassica* allopolyploids showed different gene expression patterns, which could contribute to survival and evolution [68,69]. In addition, the members of the gene families also exhibited distinct expression patterns across different tissues in *Brassica* allotetraploid, suggesting that some genes may be sub-functionalized or neo-functionalized during the process of the evolution [44,70,71].

After allopolyploidization, the expression levels of some *GH3* genes showed significant changes. Besides, the expression patterns of most orthologous pairs are different. Compared with diploid ancestors, the expression levels of a few *GH3* genes were greatly enhanced in *B. napus*. For example, BraA06.GH3-7.a had the lower expression levels (FPKM < 1.2) in four tissues, while the FPKM of its orthologous BnaA06.GH3-7.a ranged from 24.21 to 83.66 in four tissues. It is reported that the *GH3-7* gene could play important roles in drought and salt stress [46,51]. The high expression of BnaA06.GH3-7.a may increase the stress tolerance of *B. napus*. In addition, Zeng et al. reported that *GH3-7* had important functions in peach fruit ripening [72]. The expression level of BnaA06.GH3-7.a was the highest in the silique, showing that it could play a vital role in controlling the silique development in *B. napus*. The *GH3-12* gene participated in salicylic acid metabolism and activated defense responses in *Arabidopsis* leaves [73]. BolC03.GH3-12.a had FPKM < 1 in leaf, while its orthologous BnaC03.GH3-12.b had a relatively high level (FPKM = 4.9) in leaf, suggesting that the increased expression level of BnaC03.GH3-12.b may improve the defense ability in *B. napus* leaf. Relative to diploid ancestors, the expression levels of a few *GH3* genes in *B. napus* were greatly decreased. For example, the FPKM of BolC07.GH3-5.b was 63.77, while the FPKM of BnaC07.GH3-5.b was 1.49 in the stem. The FPKM of BolC03.GH3-15.a ranged from 54.42 to 87.51 across four tissues, while the FPKM of BnaC03.GH3-15.a varied from 6.05 to 9.96. The large drop in the expression levels may be due to the increase in the number of genes in *B. napus*, and measures were taken to avoid gene function redundancy.

The molecular mechanisms of gene expression changes in polyploidies have been explored. The first molecular mechanism is altered regulatory networks. Gene expression is controlled by regulatory networks composed of various interrelated regulatory factors, such as transcription factors, and altered regulatory networks could affect the phenotype and contribute to heterosis in polyploids [74]. The genetic changes are associated with alterations in DNA sequences, which also have effects on gene expression [74]. After the formation of *Brassica* allopolyploids, the gene structures had undergone rapid and extensive changes caused by DNA rearrangements, homoeologous recombination, gene conversion-like events, etc. [75]. Insertion of transposons into genes or locations near the genes could alter the gene expression pattern or gene structure [76]. Genome rearrangements significantly changed the gene expression level in *B. napus*, which could cause phenotypic variability [77]. The final molecular mechanism is epigenetic changes, such as DNA methylation, miRNA regulation, and so on. Alterations in DNA methylation patterns brought about transcriptional changes and phenotypic instability in *Arabidopsis* allotetraploids [78]. Ha et al. reported that changes in miRNA expression could give rise to variation in gene expression and adaptation [79]. In our results, genetic structure changes may be one of the reasons for changes in *BnaGH3* gene expression. The exon-intron organization of BnaA05GH3-18.a was different from BraAX.GH3-19.e after allopolyploidization, which may lead to  $d_N/d_S \gg 1$ . BnaA05GH3-18.a was subjected to strong positive selection, so it had a distinct expression pattern. BnaA03.GH3-8.a and BnaAX.GH3-8.h may be also like this. In addition, gene expression could be regulated by cis-elements in response to various environmental stress or developmental programs [80]. We found that the majority of orthologous gene

pairs between *B. napus* and ancestor species had differences in the types and number of cis-elements, which might lead to variation in *BnaGH3* gene expression.

A direct result of allopolyploidization is gene duplication. There are several cases for duplicated genes in allopolyploids: (1) duplicated genes retain original function; (2) one of duplicated genes is neo-functionalized; (3) one of the duplicated genes is silenced; (4) duplicated genes are sub-functionalized [81,82]. The study on homologous genes derived from three diploid species in allohexaploid wheat indicated that homologous genes not only had similar expression patterns, but also had distinct expression patterns in different tissues [83]. The functional divergence of duplicated genes is a main trait during the long-term evolution process for polyploids [84]. In our results, a few homologous gene pairs had similar expression patterns. *BnaA03.GH3-11.a* and *BnaCX.GH3-11.b* had the highest expression level in the flower, followed by the silique, stem, and leaf, indicating that they could play an important role in the flower development. Many homologous gene pairs had different expression patterns. For example, *BnaA06.GH3-7.a* had higher expression level across four tissues, while *BnaC06.GH3-7.a* was barely expressed across the four tissues, showing the emergence of functional divergence. This differential expression between homologous genes in polyploids may have a profound effect on plant evolution. On the one hand, the functional divergence of homologous genes could protect redundant genes from being eliminated in natural selection during long-term evolution. On the other hand, the distinct expression patterns of homologous genes in different tissues may also lead to the production of new phenotypes, promoting plant diversity.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/1/58/s1>. Figure S1: The sequence alignment of 148 proteins (19 *A. thaliana*, 38 *B. rapa*, 25 *B. oleracea*, and 66 *B. napus* members). Figure S2: The 3-dimensional structures of 15 orthologous protein pairs. Figure S3: The time of WGT and allopolyploidization events and the number of *GH3* genes that were missing or new in the different time. Table S1: The information of *GH3* genes identified in the three species of *Brassica*. Table S2: The orthologous sets of *GH3* genes among *B. rapa*, *B. oleracea*, *B. napus*. Table S3: Seven ancestor orthologous genes whose proteins with incomplete *GH3* domains. Table S4: The synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) nucleotide substitution rates for orthologous pairs. Table S5: The four duplicated types of *GH3* genes in the three species of *Brassica*. Table S6: The CDS and ORF length of *GH3* genes in the three species of *Brassica*. Table S7: Exon number, CDS length, and ORF length of 33  $A_n-A_r$  pairs and 20  $C_n-C_o$  pairs. Table S8: The physicochemical parameters of *GH3* proteins in the three species of *Brassica*. Table S9: The locations of  $\alpha$ -helices and  $\beta$ -strands for 17 orthologous protein pairs. Table S10: The cis-elements of *GH3* gene promoter regions in the three species of *Brassica*. Table S11: The number of cis-acting element types of 33  $A_n-A_r$  pairs and 20  $C_n-C_o$  pairs. Table S12: The FPKM values of *GH3* genes in different tissues for the three species of *Brassica*.

**Author Contributions:** J.W. and R.W. designed the experiment. R.W. carried out the field sampling and data analysis. R.W. wrote the manuscript draft and J.W. revised the manuscript. M.L. provided the RNA-seq data. X.W. provided the experimental materials. All authors read and approved the final manuscript. Methodology, R.W.; project administration, J.W.; resources, M.L. and X.W.; writing—original draft, R.W.; writing—review and editing, J.W.

**Funding:** This research was funded by the National Natural Science Foundation of China (31570539, 31370258).

**Acknowledgments:** We are thankful for funding from the National Natural Science Foundation of China (31570539, 31370258).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## References

1. Masterson, J. Stomatal size in fossil plants—Evidence for polyploidy in majority of angiosperms. *Science* **1994**, *264*, 421–424. [CrossRef] [PubMed]
2. Ng, D.W.K.; Lu, J.; Chen, Z.J. Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr. Opin. Plant Biol.* **2012**, *15*, 154–161. [CrossRef] [PubMed]
3. Osborn, T.C.; Butrulle, D.V.; Sharpe, A.G.; Pickering, K.J.; Parkin, I.A.P.; Parker, J.S.; Lydiate, D.J. Detection and effects of a homeologous reciprocal transposition in *Brassica napus*. *Genetics* **2003**, *165*, 1569–1577. [PubMed]

4. Pires, J.C.; Zhao, J.W.; Schranz, M.E.; Leon, E.J.; Quijada, P.A.; Lukens, L.N.; Osborn, T.C. Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). *Biol. J. Linn. Soc.* **2004**, *82*, 675–688. [[CrossRef](#)]
5. Akhunov, E.D.; Sehgal, S.; Liang, H.Q.; Wang, S.C.; Akhunova, A.R.; Kaur, G.; Li, W.L.; Forrest, K.L.; See, D.; Simkova, H.; et al. Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. *Plant Physiol.* **2013**, *161*, 252–265. [[CrossRef](#)] [[PubMed](#)]
6. He, P.; Friebe, B.R.; Gill, B.S.; Zhou, J.M. Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant Mol. Biol.* **2003**, *52*, 401–414. [[CrossRef](#)] [[PubMed](#)]
7. Yoo, M.J.; Liu, X.X.; Pires, J.C.; Soltis, P.S.; Soltis, D.E. Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* **2014**, *48*, 485–517. [[CrossRef](#)]
8. Xu, Y.H.; Zhong, L.; Wu, X.M.; Fang, X.P.; Wang, J.B. Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta* **2009**, *229*, 471–483. [[CrossRef](#)]
9. Woodward, A.W.; Bartel, B. Auxin: Regulation, action, and interaction. *Ann. Bot.-Lond.* **2005**, *95*, 707–735. [[CrossRef](#)]
10. Wang, H.; Tian, C.E.; Duan, J.; Wu, K.Q. Research progresses on *GH3s*, one family of primary auxin-responsive genes. *Plant Growth Regul.* **2008**, *56*, 225–232. [[CrossRef](#)]
11. Staswick, P.E.; Tiryaki, I.; Rowe, M.L. Jasmonate response locus JAR1 and several related Arabidopsis genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. *Plant Cell* **2002**, *14*, 1405–1415. [[CrossRef](#)] [[PubMed](#)]
12. Staswick, P.E.; Serban, B.; Rowe, M.; Tiryaki, I.; Maldonado, M.T.; Maldonado, M.C.; Suza, W. Characterization of an Arabidopsis enzyme family that conjugates amino acids to indole-3-acetic acid. *Plant Cell* **2005**, *17*, 616–627. [[CrossRef](#)] [[PubMed](#)]
13. Staswick, P.E.; Tiryaki, I. The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in *Arabidopsis*. *Plant Cell* **2004**, *16*, 2117–2127. [[CrossRef](#)] [[PubMed](#)]
14. Okrent, R.A.; Brooks, M.D.; Wildermuth, M.C. *Arabidopsis* GH3.12 (PBS3) conjugates amino acids to 4-substituted benzoates and is inhibited by salicylate. *J. Biol. Chem.* **2009**, *284*, 9742–9754. [[CrossRef](#)] [[PubMed](#)]
15. Liu, K.D.; Kang, B.C.; Jiang, H.; Moore, S.L.; Li, H.X.; Watkins, C.B.; Setter, T.L.; Jahn, M.M. A *GH3*-like gene, *CcGH3*, isolated from *Capsicum chinense* L. fruit is regulated by auxin and ethylene. *Plant Mol. Biol.* **2005**, *58*, 447–464. [[CrossRef](#)] [[PubMed](#)]
16. Wang, S.K.; Bai, Y.H.; Shen, C.J.; Wu, Y.R.; Zhang, S.N.; Jiang, D.A.; Guilfoyle, T.J.; Chen, M.; Qi, Y.H. Auxin-related gene families in abiotic stress response in *Sorghum bicolor*. *Funct. Integr. Genom.* **2010**, *10*, 533–546. [[CrossRef](#)] [[PubMed](#)]
17. Yang, Y.J.; Yue, R.Q.; Sun, T.; Zhang, L.; Chen, W.; Zeng, H.Q.; Wang, H.Z.; Shen, C.J. Genome-wide identification, expression analysis of *GH3* family genes in *Medicago truncatula* under stress-related hormones and *Sinorhizobium meliloti* infection. *Appl. Microbiol. Biot.* **2015**, *99*, 841–854. [[CrossRef](#)]
18. Hsieh, H.L.; Okamoto, H.; Wang, M.L.; Ang, L.H.; Matsui, M.; Goodman, H.; Deng, X.W. *FIN219*, an auxin-regulated gene, defines a link between phytochrome A and the downstream regulator COP1 in light control of *Arabidopsis* development. *Gene Dev.* **2000**, *14*, 1958–1970. [[PubMed](#)]
19. Khan, S.; Stone, J.M. *Arabidopsis thaliana* GH3.9 influences primary root growth. *Planta* **2007**, *226*, 21–34. [[CrossRef](#)]
20. Park, J.E.; Seo, P.J.; Lee, A.K.; Jung, J.H.; Kim, Y.S.; Park, C.M. An *Arabidopsis* *GH3* gene, encoding an auxin-conjugating enzyme, mediates phytochrome B-regulated light signals in hypocotyl growth. *Plant Cell Physiol.* **2007**, *48*, 1236–1241. [[CrossRef](#)]
21. Zhang, S.N.; Wang, S.K.; Xu, Y.X.; Yu, C.L.; Shen, C.J.; Qian, Q.; Geisler, M.; Jiang, D.A.; Qi, Y.H. The auxin response factor, OsARF19, controls rice leaf angles through positively regulating *OsGH3-5* and *OsBR11*. *Plant Cell Environ.* **2015**, *38*, 638–654. [[CrossRef](#)] [[PubMed](#)]
22. Jain, M.; Kaur, N.; Tyagi, A.K.; Khurana, J.P. The auxin-responsive *GH3* gene family in rice (*Oryza sativa*). *Funct. Integr. Genom.* **2006**, *6*, 36–46. [[CrossRef](#)]
23. Feng, S.G.; Yue, R.Q.; Tao, S.; Yang, Y.J.; Zhang, L.; Xu, M.F.; Wang, H.Z.; Shen, C.J. Genome-wide identification, expression analysis of auxin-responsive *GH3* family genes in maize (*Zea mays* L.) under abiotic stresses. *J. Integr. Plant Biol.* **2015**, *57*, 783–795. [[CrossRef](#)] [[PubMed](#)]

24. Liu, K.D.; Wang, J.X.; Li, H.L.; Zhong, J.D.; Feng, S.X.; Pan, Y.L.; Yuan, C.C. Identification, expression and IAA-Amide synthetase activity analysis of Gretchen Hagen 3 in papaya fruit (*Carica papaya* L.) during postharvest process. *Front. Plant Sci.* **2016**, *7*, 1555. [[CrossRef](#)] [[PubMed](#)]
25. Yang, Y.W.; Lai, K.N.; Tai, P.Y.; Li, W.H. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *J. Mol. Evol.* **1999**, *48*, 597–604. [[CrossRef](#)] [[PubMed](#)]
26. Liu, S.Y.; Liu, Y.M.; Yang, X.H.; Tong, C.B.; Edwards, D.; Parkin, I.A.P.; Zhao, M.X.; Ma, J.X.; Yu, J.Y.; Huang, S.M.; et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* **2014**, *5*, 3930. [[CrossRef](#)] [[PubMed](#)]
27. Chalhoub, B.; Denoeud, F.; Liu, S.Y.; Parkin, I.A.P.; Tang, H.B.; Wang, X.Y.; Chiquet, J.; Belcram, H.; Tong, C.B.; Samans, B.; et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **2014**, *345*, 950–953. [[CrossRef](#)] [[PubMed](#)]
28. Nguyen, T.C.T.; Obermeier, C.; Friedt, W.; Abrams, S.R.; Snowdon, R.J. Disruption of germination and seedling development in *Brassica napus* by mutations causing severe seed hormonal imbalance. *Front. Plant Sci.* **2016**, *7*, 322. [[CrossRef](#)]
29. Jian, H.J.; Yang, B.; Zhang, A.X.; Zhang, L.; Xu, X.F.; Li, J.N.; Liu, L.Z. Screening of candidate leaf morphology genes by integration of QTL mapping and RNA sequencing technologies in oilseed rape (*Brassica napus* L.). *PLoS ONE* **2017**, *12*, e0169641. [[CrossRef](#)]
30. Cheng, H.T.; Hao, M.Y.; Wang, W.X.; Mei, D.S.; Wells, R.; Liu, J.; Wang, H.; Sang, S.F.; Tang, M.; Zhou, R.J.; et al. Integrative RNA-and miRNA-profile analysis reveals a likely role of BR and auxin signaling in branch angle regulation of *B. napus*. *Int. J. Mol. Sci.* **2017**, *18*, 887. [[CrossRef](#)]
31. Gu, A.X.; Meng, C.; Chen, Y.Q.; Wei, L.; Dong, H.; Lu, Y.; Wang, Y.H.; Chen, X.P.; Zhao, J.J.; Shen, S.X. Coupling Seq-BSA and RNA-Seq analyses reveal the molecular pathway and genes associated with heading type in Chinese cabbage. *Front. Genet.* **2017**, *8*, 176. [[CrossRef](#)] [[PubMed](#)]
32. Minglin, L.; Zhang, Y.X.; Chai, T.Y. Identification of genes up-regulated in response to Cd exposure, in *Brassica juncea* L. *Gene* **2005**, *363*, 151–158. [[CrossRef](#)]
33. Kwon, S.W.; Kim, M.; Kim, H.; Lee, J. Shotgun quantitative proteomic analysis of proteins responding to drought stress in *Brassica rapa* L. (inbred line “Chiifu”). *Int. J. Genom.* **2016**, *2016*, 4235808.
34. Ostergaard, L.; King, G.J. Standardized gene nomenclature for the *Brassica* genus. *Plant Methods* **2008**, *4*, 10. [[CrossRef](#)] [[PubMed](#)]
35. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)]
36. Cheng, F.; Wu, J.; Fang, L.; Wang, X.W. Syntenic gene analysis between *Brassica rapa* and other Brassicaceae specie. *Front. Plant Sci.* **2012**, *3*, 198. [[CrossRef](#)] [[PubMed](#)]
37. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [[CrossRef](#)]
38. Wang, Y.P.; Tang, H.B.; DeBarry, J.D.; Tan, X.; Li, J.P.; Wang, X.Y.; Lee, T.H.; Jin, H.Z.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)]
39. Goldman, N.; Yang, Z.H. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **1994**, *11*, 725–736.
40. Higo, K.; Ugawa, Y.; Iwamoto, M.; Korenaga, T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* **1999**, *27*, 297–300. [[CrossRef](#)]
41. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)]
42. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]
43. Nekrutenko, A.; Makova, K.D.; Li, W.H. The K-A/K-S ratio test for assessing the protein-coding potential of genomic regions: An empirical and simulation study. *Genome Res.* **2002**, *12*, 198–202. [[CrossRef](#)]
44. Li, H.T.; Wang, B.; Zhang, Q.H.; Wang, J.; King, G.J.; Liu, K.D. Genome-wide analysis of the auxin/indoleacetic acid (*Aux/IAA*) gene family in allotetraploid rapeseed (*Brassica napus* L.). *BMC Plant Biol.* **2017**, *17*, 204. [[CrossRef](#)]

45. Song, X.M.; Ma, X.; Li, C.J.; Hu, J.J.; Yang, Q.H.; Wang, T.; Wang, L.; Wang, J.P.; Guo, D.; Ge, W.N.; et al. Comprehensive analyses of the *BES1* gene family in *Brassica napus* and examination of their evolutionary pattern in representative species. *BMC Genom.* **2018**, *19*, 346. [[CrossRef](#)] [[PubMed](#)]
46. Singh, V.K.; Jain, M.; Garg, R. Genome-wide analysis and expression profiling suggest diverse roles of *GH3* genes during development and abiotic stress responses in legumes. *Front. Plant Sci.* **2014**, *5*, 789. [[CrossRef](#)] [[PubMed](#)]
47. Chang, K.H.; Xiang, H.; Dunaway-Mariano, D. Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: A site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate: Coenzyme A ligase. *Biochemistry* **1997**, *36*, 15650–15659. [[CrossRef](#)]
48. Ibraheem, O.; Botha, C.E.J.; Bradley, G. In silico analysis of *cis*-acting regulatory elements in 5' regulatory regions of sucrose transporter gene families in rice (*Oryza sativa* Japonica) and *Arabidopsis thaliana*. *Comput. Biol. Chem.* **2010**, *34*, 268–283. [[CrossRef](#)]
49. Zhou, Y.; Xu, D.X.; Jia, L.D.; Huang, X.H.; Ma, G.Q.; Wang, S.X.; Zhu, M.C.; Zhang, A.X.; Guan, M.W.; Lu, K.; et al. Genome-wide identification and structural analysis of *bZIP* transcription factor genes in *Brassica napus*. *Genes* **2017**, *8*, 288. [[CrossRef](#)]
50. Song, X.M.; Wang, J.P.; Ma, X.; Li, Y.X.; Lei, T.Y.; Wang, L.; Ge, W.N.; Guo, D.; Wang, Z.Y.; Li, C.J.; et al. Origination, expansion, evolutionary trajectory, and expression bias of *AP2/ERF* superfamily in *Brassica napus*. *Front. Plant Sci.* **2016**, *7*, 1186. [[CrossRef](#)]
51. Yuan, H.Z.; Zhao, K.; Lei, H.J.; Shen, X.J.; Liu, Y.; Liao, X.; Li, T.H. Genome-wide analysis of the *GH3* family in apple (*Malus × domestica*). *BMC Genom.* **2013**, *14*, 297. [[CrossRef](#)] [[PubMed](#)]
52. Cheng, F.; Wu, J.; Wang, X.W. Genome triplication drove the diversification of *Brassica* plants. *Hortic. Res.-Engl.* **2014**, *1*, 14024. [[CrossRef](#)] [[PubMed](#)]
53. Town, C.D.; Cheung, F.; Maiti, R.; Crabtree, J.; Haas, B.J.; Wortman, J.R.; Hine, E.E.; Althoff, R.; Arbogast, T.S.; Tallon, L.J.; et al. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *Plant Cell* **2006**, *18*, 1348–1359. [[CrossRef](#)] [[PubMed](#)]
54. Chester, M.; Gallagher, J.P.; Symonds, V.V.; da Silva, A.V.C.; Mavrodiev, E.V.; Leitch, A.R.; Soltis, P.S.; Soltis, D.E. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1176–1181. [[CrossRef](#)] [[PubMed](#)]
55. Zhang, J.Z. Evolution by gene duplication: An update. *Trends Ecol. Evol.* **2003**, *18*, 292–298. [[CrossRef](#)]
56. Wang, X.X.; Grus, W.E.; Zhang, J.Z. Gene losses during human origins. *PLoS Biol.* **2006**, *4*, 366–377. [[CrossRef](#)] [[PubMed](#)]
57. Smith, S.D.; Rausher, M.D. Gene loss and parallel evolution contribute to species difference in flower color. *Mol. Biol. Evol.* **2011**, *28*, 2799–2810. [[CrossRef](#)]
58. Torok, M.E.; Chantratita, N.; Peacock, S.J. Bacterial gene loss as a mechanism for gain of antimicrobial resistance. *Curr. Opin. Microbiol.* **2012**, *15*, 583–587. [[CrossRef](#)]
59. Long, M.; Betran, E.; Thornton, K.; Wang, W. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **2003**, *4*, 865–875. [[CrossRef](#)]
60. Wang, W.; Zheng, H.K.; Fan, C.Z.; Li, J.; Shi, J.J.; Cai, Z.Q.; Zhang, G.J.; Liu, D.Y.; Zhang, J.G.; Vang, S.; et al. High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* **2006**, *18*, 1791–1802. [[CrossRef](#)] [[PubMed](#)]
61. Bergthorsson, U.; Adams, K.L.; Thomason, B.; Palmer, J.D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **2003**, *424*, 197–201. [[CrossRef](#)] [[PubMed](#)]
62. Houlne, G.; Meyer, B.; Schantz, R. Alteration of the expression of a plant defensin gene by exon shuffling in bell pepper (*Capsicum annuum* L.). *Mol. Gen. Genet.* **1998**, *259*, 504–510. [[CrossRef](#)] [[PubMed](#)]
63. Andersson, C.R.; Jensen, E.O.; Llewellyn, D.J.; Dennis, E.S.; Peacock, W.J. A new hemoglobin gene from soybean: A role for hemoglobin in all plants. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5682–5687. [[CrossRef](#)]
64. Jones, J.D.G. Putting knowledge of plant disease resistance genes to work. *Curr. Opin. Plant Biol.* **2001**, *4*, 281–287. [[CrossRef](#)]
65. Wu, D.D.; Wang, X.; Li, Y.; Zeng, L.; Irwin, D.M.; Zhang, Y.P. “Out of pollen” hypothesis for origin of new genes in flowering plants: Study from *Arabidopsis thaliana*. *Genome Biol. Evol.* **2014**, *6*, 2822–2829. [[CrossRef](#)] [[PubMed](#)]
66. Pikaard, C.S. Genomic change and gene silencing in polyploids. *Trends Genet.* **2001**, *17*, 675–677. [[CrossRef](#)]



67. Ni, Z.F.; Kim, E.D.; Ha, M.S.; Lackey, E.; Liu, J.X.; Zhang, Y.R.; Sun, Q.X.; Chen, Z.J. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* **2009**, *457*, 327–331. [[CrossRef](#)]
68. Zhao, Q.; Zou, J.; Meng, J.L.; Mei, S.Y.; Wang, J.B. Tracing the transcriptomic changes in synthetic trigonomic allohexaploids of *Brassica* using an RNA-Seq approach. *PLoS ONE* **2013**, *8*, e68883. [[CrossRef](#)]
69. Jiang, J.J.; Wang, Y.; Zhu, B.; Fang, T.T.; Fang, Y.J.; Wang, Y.P. Digital gene expression analysis of gene expression differences within *Brassica* diploids and allopolyploids. *BMC Plant Biol.* **2015**, *15*, 22. [[CrossRef](#)]
70. Cheng, H.T.; Hao, M.Y.; Wang, W.X.; Mei, D.S.; Tong, C.B.; Wang, H.; Liu, J.; Fu, L.; Hu, Q. Genomic identification, characterization and differential expression analysis of *SBP-box* gene family in *Brassica napus*. *BMC Plant Biol.* **2016**, *16*, 196. [[CrossRef](#)]
71. Wang, J.X.; Wang, X.M.; Geng, S.Y.; Singh, S.K.; Wang, Y.H.; Pattanaik, S.; Yuan, L. Genome-wide identification of hexokinase gene family in *Brassica napus*: Structure, phylogenetic analysis, expression, and functional characterization. *Planta* **2018**, *248*, 171–182. [[CrossRef](#)] [[PubMed](#)]
72. Zeng, W.F.; Pan, L.; Niu, L.; Lu, Z.H.; Cui, G.C.; Wang, Z.Q. Bioinformatics analysis and expression of the Nectarine Indole-3-acetic acid-amido synthase (*GH3*) gene family during fruit development. *Acta Hortic. Sin.* **2015**, *42*, 33–842.
73. Nobuta, K.; Okrent, R.A.; Stoutemyer, M.; Rodibaugh, N.; Kempema, L.; Wildermuth, M.C.; Innes, R.W. The *GH3* acyl adenylase family member *PBS3* regulates salicylic acid-dependent defense responses in *Arabidopsis*. *Plant Physiol.* **2007**, *144*, 1144–1156. [[CrossRef](#)] [[PubMed](#)]
74. Osborn, T.C.; Pires, J.C.; Birchler, J.A.; Auger, D.L.; Chen, Z.J.; Lee, H.S.; Comai, L.; Madlung, A.; Doerge, R.W.; Colot, V.; et al. Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* **2003**, *19*, 141–147. [[CrossRef](#)]
75. Song, K.M.; Lu, P.; Tang, K.L.; Osborn, T.C. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7719–7723. [[CrossRef](#)] [[PubMed](#)]
76. Lockton, S.; Gaut, B.S. The contribution of transposable elements to expressed coding sequence in *Arabidopsis thaliana*. *J. Mol. Evol.* **2009**, *68*, 80–89. [[CrossRef](#)]
77. Gaeta, R.T.; Pires, J.C.; Iniguez-Luy, F.; Leon, E.; Osborn, T.C. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* **2007**, *19*, 3403–3417. [[CrossRef](#)]
78. Madlung, A.; Masuelli, R.W.; Watson, B.; Reynolds, S.H.; Davison, J.; Comai, L. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiol.* **2002**, *129*, 733–746. [[CrossRef](#)]
79. Ha, M.; Lu, J.; Tian, L.; Ramachandran, V.; Kasschau, K.D.; Chapman, E.J.; Carrington, J.C.; Chen, X.M.; Wang, X.J.; Chen, Z.J. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17835–17840. [[CrossRef](#)]
80. Soh, H.; Choi, Y.; Lee, T.K.; Yeo, U.D.; Han, K.; Auh, C.; Lee, S. Identification of unique cis-element pattern on simulated microgravity treated *Arabidopsis* by in silico and gene expression. *Adv. Space Res.* **2012**, *50*, 397–407. [[CrossRef](#)]
81. Adams, K.L.; Cronn, R.; Percifield, R.; Wendel, J.F. Genes duplicated by polyploid show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4649–4654. [[CrossRef](#)] [[PubMed](#)]
82. Adams, K.L.; Percifield, R.; Wendel, J.F. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* **2004**, *168*, 2217–2226. [[CrossRef](#)] [[PubMed](#)]
83. Mochida, K.; Yamazaki, Y.; Ogihara, Y. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol. Genet. Genom.* **2004**, *270*, 371–377. [[CrossRef](#)] [[PubMed](#)]
84. Blanc, G.; Wolfe, K.H. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **2004**, *16*, 1679–1691. [[CrossRef](#)] [[PubMed](#)]

