



Article

# Patterns of Expansion and Expression Divergence of the Polygalacturonase Gene Family in *Brassica oleracea*

Meiling Lyu <sup>1,2</sup>, Junaid Iftikhar <sup>1</sup>, Rongfang Guo <sup>1</sup>, Binghua Wu <sup>1</sup>  and Jiashu Cao <sup>2,3,\*</sup>

<sup>1</sup> College of Horticulture, Fujian Agriculture & Forestry University, Fuzhou 350002, China; mlyu@fafu.edu.cn (M.L.); junaidiftikhar@outlook.com (J.I.); 000q813038@fafu.edu.cn (R.G.); binghua.wu@fafu.edu.cn (B.W.)

<sup>2</sup> Laboratory of Cell & Molecular Biology, Institute of Vegetable Science, Zhejiang University, Hangzhou 310058, China

<sup>3</sup> Zhejiang Provincial Key Laboratory of Horticultural Plant Integrative Biology, Zhejiang University, Hangzhou 310058, China

\* Correspondence: jshcao@zju.edu.cn

Received: 15 June 2020; Accepted: 6 August 2020; Published: 9 August 2020



**Abstract:** Plant polygalacturonases (PGs) are closely related to cell-separation events during plant growth and development by degrading pectin. Identifying and investigating their diversification of evolution and expression could shed light on research on their function. We conducted sequence, molecular evolution, and gene expression analyses of PG genes in *Brassica oleracea*. Ninety-nine *B. oleracea* PGs (BoPGs) were identified and divided into seven clades through phylogenetic analysis. The exon/intron structures and motifs were conserved within, but divergent between, clades. The second conserved domain (GDDC) may be more closely related to the identification of PGs. There were at least 79 common ancestor PGs between *Arabidopsis thaliana* and *B. oleracea*. The event of whole genome triplication and tandem duplication played important roles in the rapid expansion of the BoPG gene family, and gene loss may be an important mechanism in the generation of the diversity of BoPGs. By evaluating the expression in five tissues, we found that most of the expressed BoPGs in clades A, B, and E showed ubiquitous expression characteristics, and the expressed BoPGs in clades C, D, and F were mainly responsible for reproduction development. Most of the paralogous gene pairs (76.2%) exhibited divergent expression patterns, indicating that they may have experienced neofunctionalization or subfunctionalization. The cis-elements analysis showed that up to 96 BoPGs contained the hormone response elements in their promoters. In conclusion, our comparative analysis may provide a valuable data foundation for the further functional analysis of BoPGs during the development of *B. oleracea*.

**Keywords:** *Brassica oleracea*; polygalacturonase; gene family; evolution; expression analysis

## 1. Introduction

Pectin is the main component of the primary cell wall. Its degradation happens in a variety of periods of plant development, such as seed germination, organ abscission, fruit ripening, anther dehiscence, silique dehiscence, pollen development, pollen tube growth, and so on [1–4]. Polygalacturonases (PGs) are cell wall hydrolases. They can catalyze random hydrolysis and disassembly of the 1,4- $\alpha$ -D-galactosiduronic linkages of pectin, causing pectin degradation [5]. Therefore, PGs participate in almost all of the developmental processes of plants. In *Arabidopsis thaliana*, more than 50% of PGs with high expression levels were found in floral tissues, suggesting their important roles in flower development [6]. In the process of the maturation and shedding of oil

palm fruit, 14 endogenous PG genes were identified by the transcriptional analysis, and these genes were characterized in the shedding area when the fruit was ripe [7]. Genome-wide identification and analysis of PGs in *Solanum lycopersicum* showed that most SIPGs had specific or high expression patterns in at least one organ, and five PGs were associated with fruit development [8]. In *Brassica rapa*, the PGs BcMF2, BcMF6, BcMF16, BcMF17, and BcMF24 have been shown to play important roles in pollen development and pollen tube growth [9–13]. The PG gene *BoMF25* in *Brassica oleracea* was proven to be related to the development of the pollen wall [14]. Recently, He et al. [15] characterized the maize PG gene *ZmPGH1* and demonstrated that it functions as a suppressor of programmed cell death in plants.

PGs belong to a large gene family in plants. In *A. thaliana*, *Cucumis sativus*, *Citrullus lanatus*, *Populus*, *Oryza sativa*, *B. rapa*, and *Malus sieversii*, the PG gene family members have been widely identified and analyzed based on their genomes [6,16–22]. According to clustering maps or expression characteristics, PGs are divided into several clades (A-C or A-F). Studies have shown that different PG members in the same clades have high similarity in terms of their gene structure and molecular characteristics [18]. A whole genome duplication (WGD) event occurs when accompanied by the fractionation and rearrangement of genome fragments, providing initial resources for genome evolution and gene differentiation, and resulting in the gain and loss of PG genes. WGD and tandem duplication events also generated many duplicated PG genes [23,24]. During the evolutionary process, most PG duplicates have partially overlapping but distinct expression patterns in different tissues and organs [6]. The differences in expression regulation patterns between different members are mainly determined by three regulatory factors: the promoter sequence, cis-regulatory element, and exon-intron structure [25–27]. An in-depth analysis of the composition, expansion, and expression patterns of the PG gene family has far-reaching significance for its evolution and function.

During the evolution process, Brassicaceae species underwent two WGD events and one Brassicaceae lineage-specific whole-genome triplication (WGT) event, which involved substantial genome reshuffling and gene loss [28–30]. Therefore, Brassicaceae species comprise a good model for the study of polyploid genome evolutionary mechanisms. In the current study, we conducted genome-wide identification and analysis of PG genes in the *B. oleracea* genome. Through a comprehensive analysis of the gene sequences, gene structures, molecular evolution, a single gene, and duplicated gene expression patterns, as well as subcellular localization of PG proteins, we provided a detailed characterization of the composition, expansion, and expression divergence of the PG gene family in *B. oleracea*. The results will offer valuable data for studying the evolution and biological functions of PGs in plants.

## 2. Results

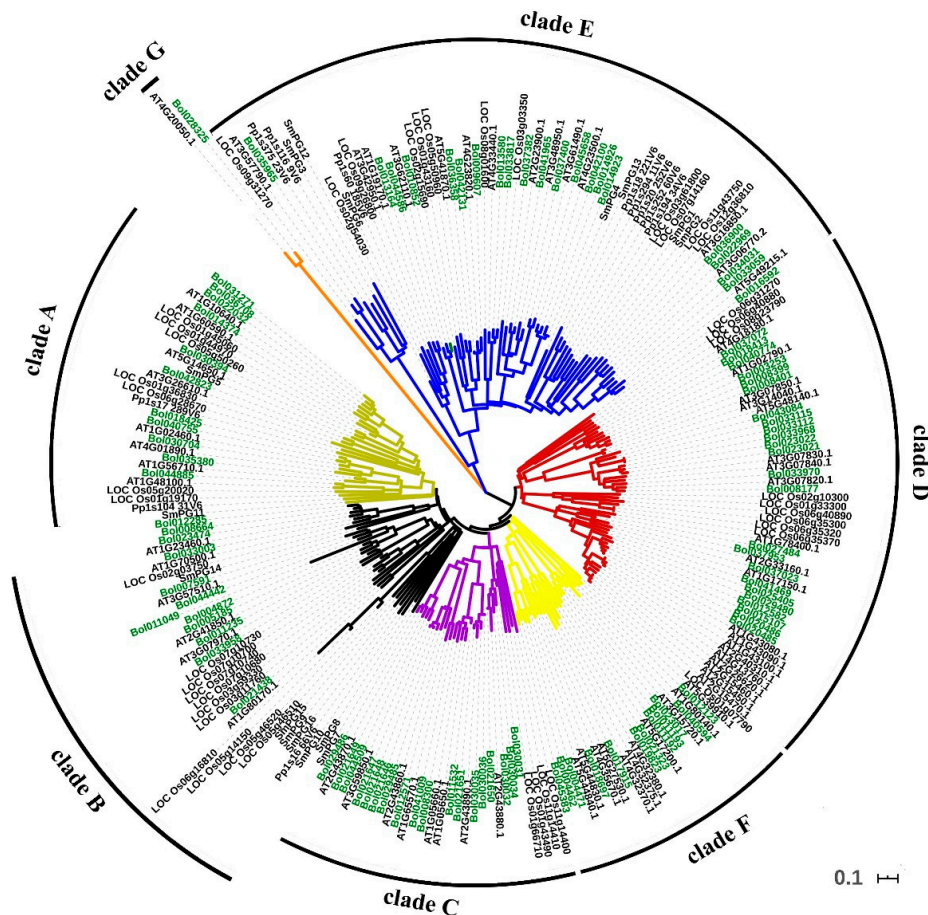
### 2.1. The Polygalacturonase Gene Family in *B. oleracea* Consists of 99 Members

Ninety-nine full-length genes encoding putative PG proteins were identified in the *B. oleracea* (Supplementary Materials Table S1). Using Pfam, the candidates were confirmed as having GH28 domains in the protein structure.

According to the sequence characteristics analysis (Supplementary Materials Table S2), the open reading frames (ORF) of *B. oleracea* PGs (BoPGs) were from 555 bp (*Bol030036*) to 1650 bp (*Bol007591*) in length. The length range of the amino acid sequence was 184 to 549 aa, and the corresponding molecular weight ranged from 19.7 to 59.1 kDa. Their isoelectric point range was 4.83 (*Bol011049*) to 8.52 (*Bol021650*). The signal peptide prediction results of the amino acid sequence showed that they all contained an N-terminal signal peptide with a length from 10 (*Bol044586*) to 58 aa (*Bol013164*), which indicated that they were all secreted proteins.

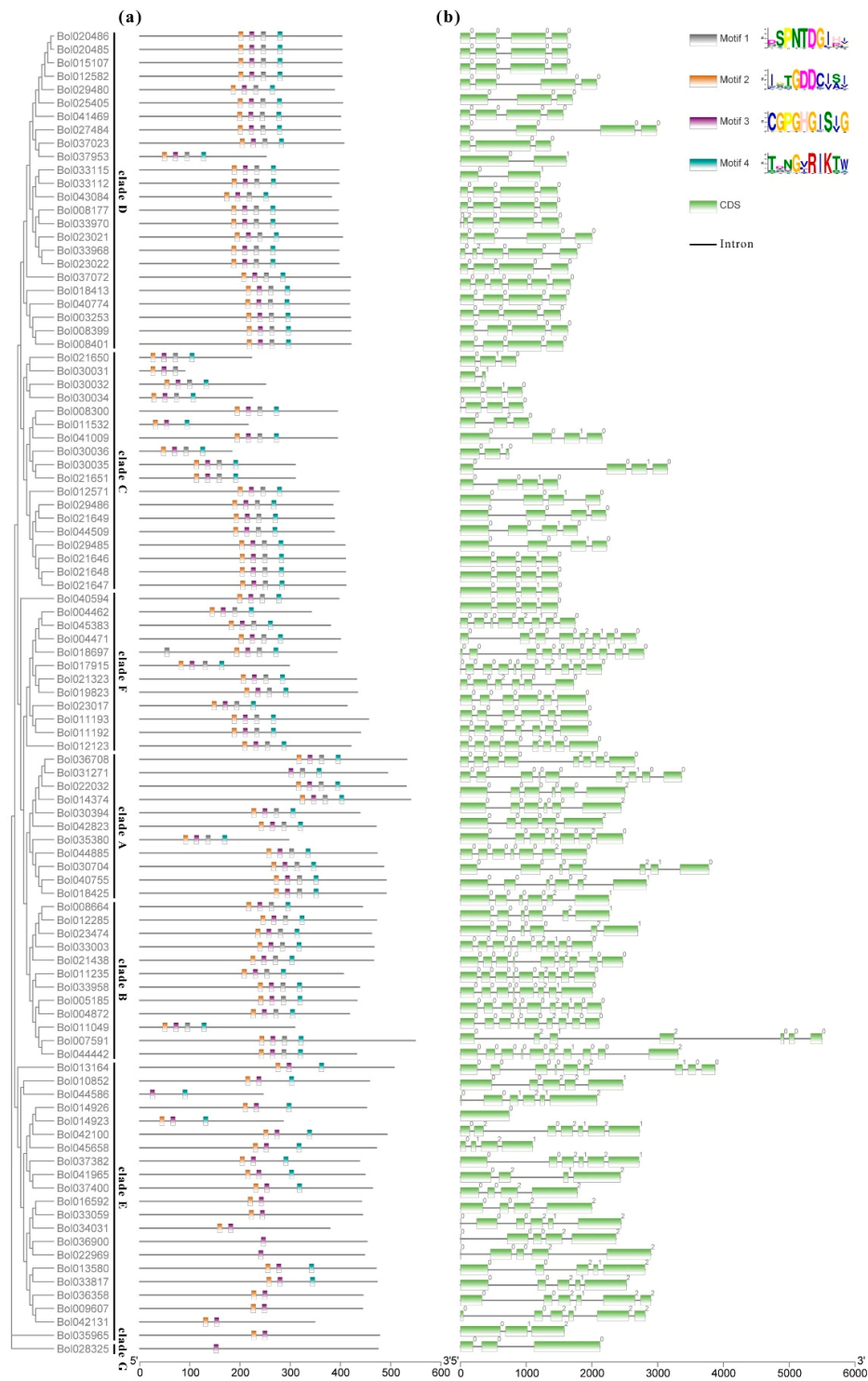
To investigate the relationships between PGs, we performed a phylogenetic analysis of five land plant species, namely *A. thaliana*, *B. oleracea*, *O. sativa*, *Selaginella moellendorffii* (spikemoss, lycophyte), and *Physcomitrella patens* (moss, bryophyte). All 238 PGs fell into seven distinct clades (A–G) (Figure 1). We counted the PGs of different clades in these five species (Supplementary Materials Table S3).

The results showed that the PGs of dicotyledonous plants *A. thaliana* and *B. oleracea* were distributed in A~G clades, and the clade G only contained PG genes of dicotyledonous plants. The PGs of monocotyledonous rice were distributed in A~F clades. According to our result, the number of members of the PG gene family increased rapidly from monocotyledons to dicotyledons. Compared with the number in the *O. sativa* genome, the PG members increased by 24 in *A. thaliana* and by up to 55 in *B. oleracea*. Meanwhile, a marked difference was also found in the PG gene family size among bryophyte, lycophyte, and angiosperms. The *P. patens* and *S. moellendorffii* PGs only existed in clades A, B, and E.



**Figure 1.** Phylogenetic tree of the 238 polygalacturonases (PGs) from five land plant species. The tree was constructed by using the neighbor-joining (NJ) method. The bar indicates the relative divergence of the sequences examined.

One gene containing at least one of the four conserved domains (I, SPNTDGI; II, GDGC; III, CGPGHGIS; and IV, RIK) can be identified as PG. The conserved domain analysis revealed that almost all of the BoPGs in clades A, B, C, D, and F contained all of the four conserved domains of BoPG (Supplementary Materials Table S2 and Figure 2a). The II domain was the most conserved domain, and it could be found in all of the BoPGs. It is worth mentioning that the homologous protein of AtQRT3 (Bol028325), which was the only member of clade G, only had the II domain. In contrast, the III domain had the largest number of changes. All of the 21 members in clade E lacked the III domain. Among them, 14 (Bol037400, Bol041965, Bol370382, Bol045658, Bol042100, Bol014923, Bol014926, Bol013580, Bol033817, Bol036358, Bol009607, Bol042131, Bol013164, and Bol010852) only contained the I, II, and IV conserved domains; four (Bol016592, Bol033059, Bol034031, and Bol035965) only contained the I and II conserved domains; and two (Bol036900 and Bol022969) only contained the II conserved domain.



**Figure 2.** Analysis of the conserved motifs and gene structures of *Brassica oleracea* PGs (BoPGs). (a) Four motifs of PG genes were investigated by the online tool MEME. Motif I to motif IV (I, SPNTDGI; II, GDDC; III, CGPGHGIS; and IV, RIK) are indicated by different color boxes, respectively. (b) Exon and intron organization of PG genes. Exons and introns are represented by green boxes and black lines, respectively. Intron phases are also analyzed from the exon information. Phase 0 is designated as introns between codons, phase 1 is designated as introns between the first and the second nucleotide in a codon, and phase 2 is designated as introns between the second and third nucleotides in a codon.





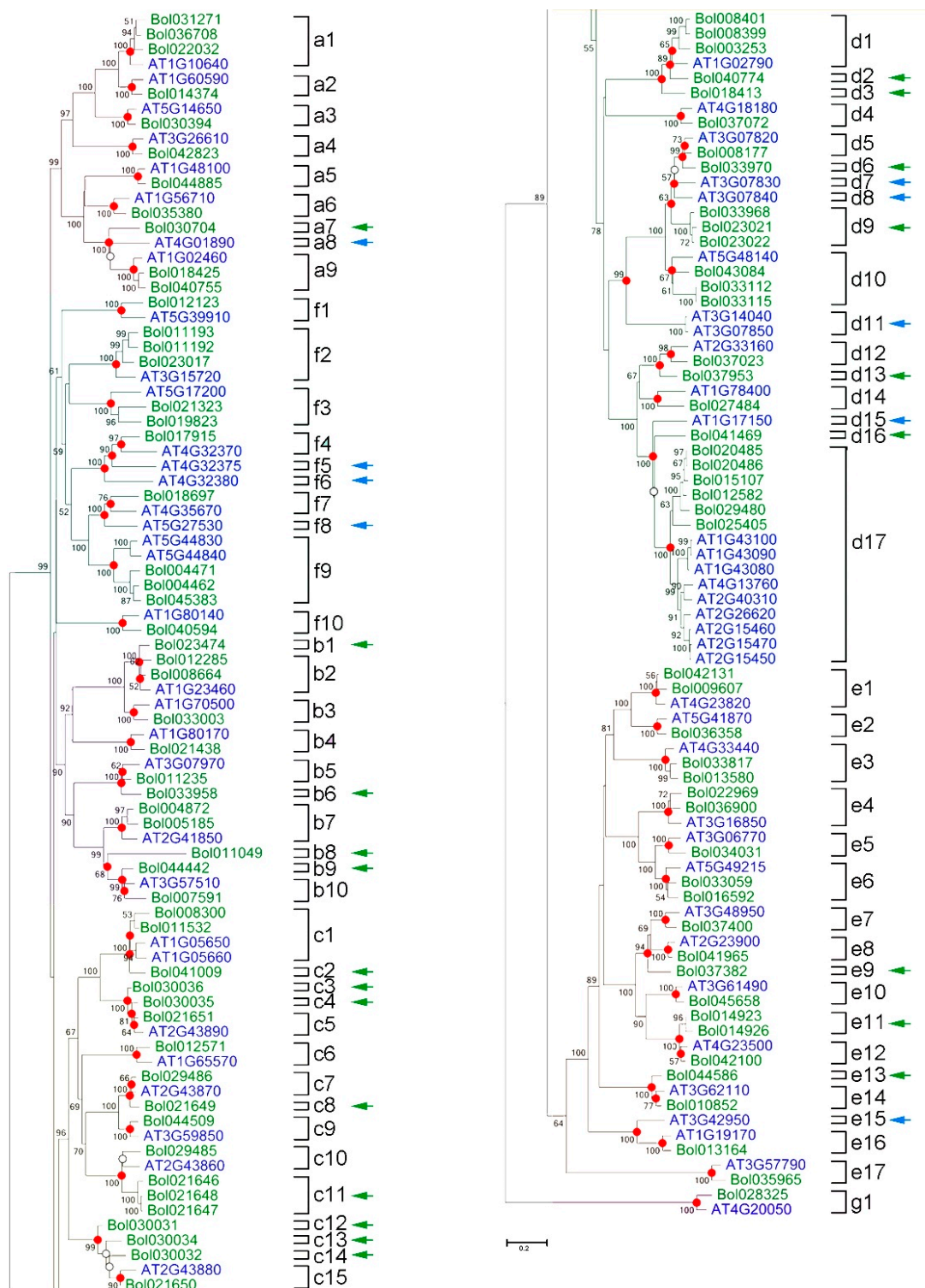
## 2.2. Expansion Mechanisms Accounting for the PG Family in *B. oleracea*

According to the results, we found that the distribution of PG genes on *B. oleracea* chromosomes appears to be nonrandom. According to the phylogenetic relationships of BoPGs (Figure 2), the clusters with high densities were observed on several chromosomes, especially chromosomes C03, C04, and C08. Moreover, seven sets of tandem replication gene pairs were located on five chromosomes (Figure 3). Among them, there were two sets on chromosome C01 (*Bol014926* and *Bol014923*; and *Bol023022* and *Bol023021*), one set on chromosome C02 (*Bol033115* and *Bol033112*), two sets on chromosome C03 (*Bol020485* and *Bol020486*; and *Bol008399* and *Bol008401*), one set on chromosome C04 (*Bol021647* and *Bol021648*), and one set on chromosome C05 (*Bol011192* and *Bol011193*). Among these replication gene pairs, there were four sets belonging to clade D (*Bol023022* and *Bol023021*; *Bol033115* and *Bol033112*; *Bol020485* and *Bol020486*; and *Bol008399* and *Bol008401*). The remaining three belonged to clade E (*Bol014926* and *Bol014923*), clade C (*Bol021647* and *Bol021648*), and clade F (*Bol011192* and *Bol011193*).

After splitting with *A. thaliana*, Brassicaceae species underwent an addition of whole genome triplication (WGT) [28]. Therefore, in theory, the genome size of *B. oleracea* should be three times the size of the *A. thaliana* genome. However, the WGT of Brassicaceae species in the evolution process involved substantial genomic rearrangement and gene loss, resulting in a large variation in the number of chromosomes and the size of the genome. The triplicated blocks resulting from WGT were partitioned into three subgenomes: the least fractionated subgenome (LF), medium fractionated subgenome (MF1), and most fractionated subgenome (MF2) [28–30]. In our study, more PGs were detected in *B. oleracea* than in *A. thaliana*, but far less than three times the number of *A. thaliana* PGs. By analyzing the syntenic PG genes between the two species and the distribution of BoPGs in three subgenomes, we found that 55 out of 68 PGs (80.9%) of *A. thaliana* were syntenic with BoPGs, and 76 out of 99 BoPGs (76.8%) can map to the three subgenomes (Supplementary Materials Table S4). The numbers of BoPGs on the three subgenomes (LF, MF1, and MF2) were 33, 23, and 20, respectively. In order to analyze the variation of BoPGs during evolution, we calculated the proportions of them retained in each of the three subgenomes relative to *A. thaliana*. We found that the LF subgenome retained 16.2% of the PGs found in *A. thaliana*, whereas the MF1 and MF2 subgenomes retained substantially lower proportions of retained genes (11.3% and 9.8%, respectively). These results suggested that many BoPGs were lost during the evolution of *B. oleracea*.

Meanwhile, we analyzed the extent of lineage-specific expansion in *B. oleracea* and *A. thaliana*. According to the phylogenetic relationships, we identified the nodes that lead to *A. thaliana*-specific and *B. oleracea*-specific branches, which represented the most recent common ancestral genes before the split. During evolution, some PGs suffered from gene loss or gene gain. We found that 22 branches contained only BoPGs (Figure 4, green arrows), and nine branches contained only *Arabidopsis* PG genes (Figure 4, blue arrows), indicating that gene losses could have occurred in these branches. An analysis of the number of nodes on the phylogenetic tree, with 73 branches of common ancestral gene nodes containing higher confidence values (>50%) (Figure 4, red dots), indicated that there were at least 73 putative common ancestral genes before the isolation of *B. oleracea* and *A. thaliana*. In addition, there were six branches of common ancestor gene nodes with lower confidence values (<50%) (Figure 4, white dots). If these low confidence nodes are also correct, then *B. oleracea* and *A. thaliana* contain at least 79 putative common ancestral genes.

Based on phylogenetic analysis, the common ancestral genes of *B. oleracea* and *A. thaliana* obtained or lost during evolution were calculated, respectively (Supplementary Materials Figure S1). There were at least nine most recently common ancestral clade A PG genes in *B. oleracea* and *A. thaliana*. After the split, *B. oleracea* gained three genes and lost no genes. In contrast, *A. thaliana* gained no genes and lost one gene. Similarly, the numbers of genes gained and lost from the most recently common ancestral clade B to clade G PG genes in the *B. oleracea* lineage and *A. thaliana* lineage were calculated. The events of gene gain and loss have resulted in the rapid expansion of PG genes in these two plants. After the split, the number of genes gained in the *B. oleracea* lineage was greater than that in the *A. thaliana* lineage, except for clade G PGs, for which neither gain nor loss was observed in either species.



**Figure 4.** Phylogenetic tree of *B. oleracea* and *Arabidopsis thaliana* PGs. Numbers on branches indicate the bootstrap percentage values calculated from 1000 replicates, and only values higher than 50% are shown. The nodes that represent the most recent common ancestral genes before the *B. oleracea* and *A. thaliana* split are indicated by red dots (bootstrap support >50%) and white dots (bootstrap support <50%). Clades that contain only *B. oleracea* and *A. thaliana* PG genes are indicated by green and blue arrows, respectively.



### 2.3. Expression Divergence Implies the Involvement of BoPGs in Different Tissues

We investigated the expression patterns of BoPGs by qRT-PCR under normal growth conditions. Firstly, we examined the expression of all 99 BoPGs in five tissues, including roots, stems, leaves, inflorescences, and tender siliques. We found that there was a substantially higher variation in the expression patterns between BoPGs. The result showed that all 99 BoPGs can be divided into eight groups (Group 1 to 8) according to their tissue specificity and relative gene expression levels (Figure 5). Group 1 contained eight genes, which had a higher expression level in inflorescences and tender siliques. Group 2 contained 21 genes whose expression could mainly be detected in inflorescences. Five members in Group 3 had a higher expression level in tender siliques. The expression of BoPG members in Group 4 and Group 5 displayed higher variation in roots, stems, inflorescences, and tender siliques, while almost all of them were not expressed in leaves. BoPG members in Group 6 did not express or had a very low expression level in every tissue. Group 7 contained one member, which was mainly expressed in one tissue—stems. The remaining four members that belonged to Group 8 were only expressed in inflorescences and stems. Taken together, approximately 30% of BoPGs did not express in any of the five tissue types tested, and almost all of the BoPGs did not express in leaves. Approximately 40% of expressed BoPGs had a high level of expression in inflorescences, 5% in root tissues, 8% in stems, and 14% in tender siliques. These findings indicate that most of the expressed BoPGs have a higher expression level in reproductive tissues (inflorescences and tender siliques).

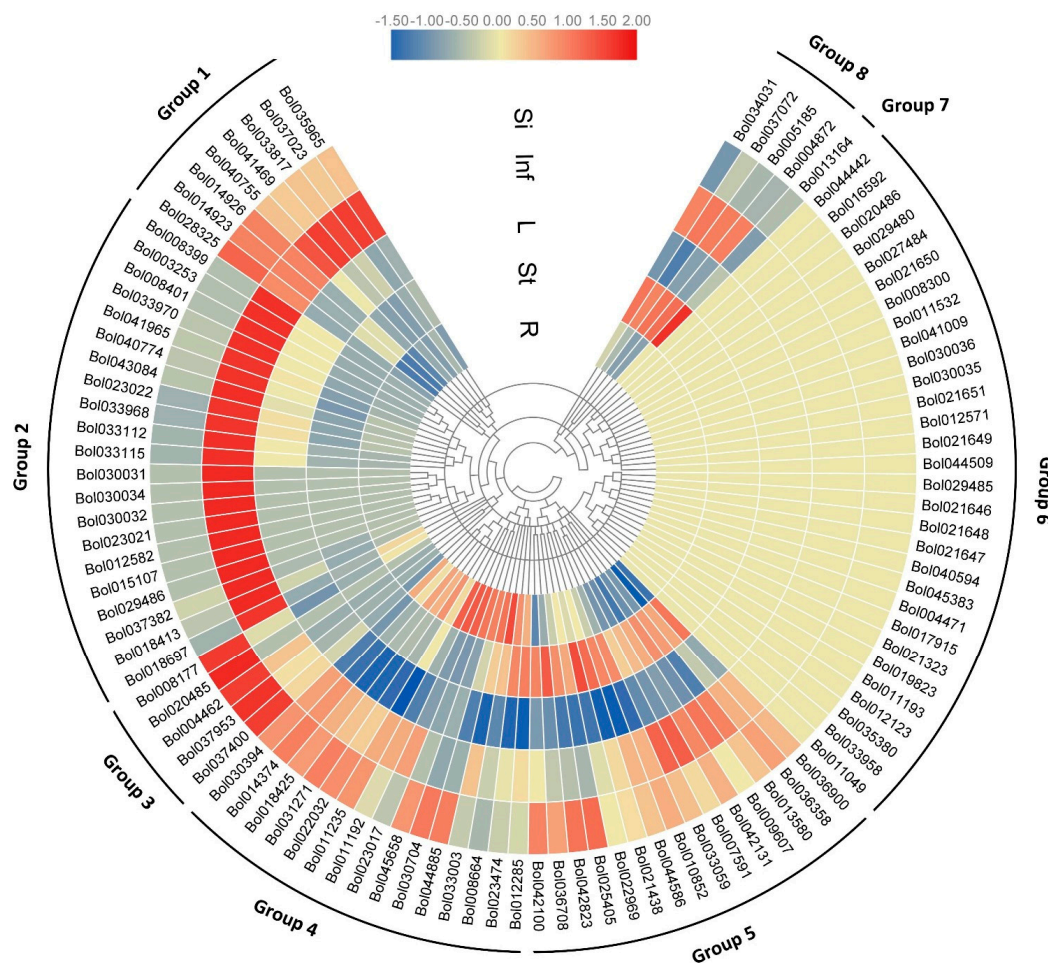
The expression patterns of BoPGs were also analyzed according to the seven distinct clades (clades A–G) (Supplementary Materials Figure S2). According to our result, the expressed BoPGs in clades C, D, and F showed a higher expression level in reproductive tissues (inflorescences and tender siliques), while the expressed BoPGs in clades A, B, and E showed ubiquitous expression patterns. In particular, the 21 members in clade E exhibited all the expression characteristics of the eight groups. The only one member of clade G was specifically expressed in stems.

According to the results presented above, many BoPGs had higher expression levels in inflorescences tissue. To further elucidate their expression patterns in reproductive growth, we selected seven BoPGs which were specifically and largely expressed in inflorescences to examine their expression level in the flower buds (Bud 1 to Bud 5) (Supplementary Materials Figure S3). The result indicated that five members (*Bol370382*, *Bol041965*, *Bol033968*, *Bol023022*, and *Bol018697*) had the highest expression level in Bud 5; one member (*Bol035965*) had higher expression levels in Bud 1, Bud 2, and Bud 3; one member (*Bol033970*) showed a high expression level in Bud 1 and Bud 2; and one member (*Bol040774*) was only expressed in Bud 3. Subsequently, the five members, which were specific and dominantly expressed in Bud 5, were selected to further analyze their expression patterns in sepal, petal, filament, anther, and pistil of Bud 5. Among them, two members (*Bol041965* and *Bol018697*) were specific and dominantly expressed in the anther (Supplementary Materials Figure S4).

The expansion of the PG family raises an intriguing question on the mechanisms of duplicate retention and their functions in plants [31]. There may be similar expression patterns between more closely related genes. To test this, we investigated the expression patterns of 21 paralogous gene pairs which were at the terminal nodes of the phylogenetic tree, and the similarity was higher than 80% for identification analysis. By comparing the temporal and spatial expression characteristics, it was found that the expression differences of them were classified into four categories (I, II, III, and IV) (Figure 2 and Table 1). In the I category, the gene pairs displayed the same expression characteristic, which did not express (*Bol033112* and *Bol033115*; *Bol008399* and *Bol008401*; and *Bol010852* and *Bol044586*) or expressed in the same organ and with the same expression level (*Bol021648* and *Bol021647*; and *Bol021323* and *Bol019823*). In the II category, the gene pairs were expressed in the same organ but with a divergent expression level. Seven pairs belonged to this category. In the III category, the gene pairs were selectively expressed in different tissue or exhibited partial overlap. One pair displayed this expression pattern. In the IV category, one gene was not expressed, and the other was selectively expressed in certain tissue. Eight gene pairs belonged to this category. Taken together,



among the 21 paralogous gene pairs, 16 pairs (76.2%) showed divergent expression patterns. These findings suggest that most of the retained gene pairs exhibited expression divergence.



**Figure 5.** Hierarchical clustering and heatmap showing the expression levels of BoPGs. The scale bars represent the relative expression level. The blue, yellow, and red shading indicates a relatively low, medium, and high expression, respectively. Their expression characteristics are divided into eight groups. R: roots; St: stems; L: leaves; Inf: inflorescences; and Si: tender siliques.

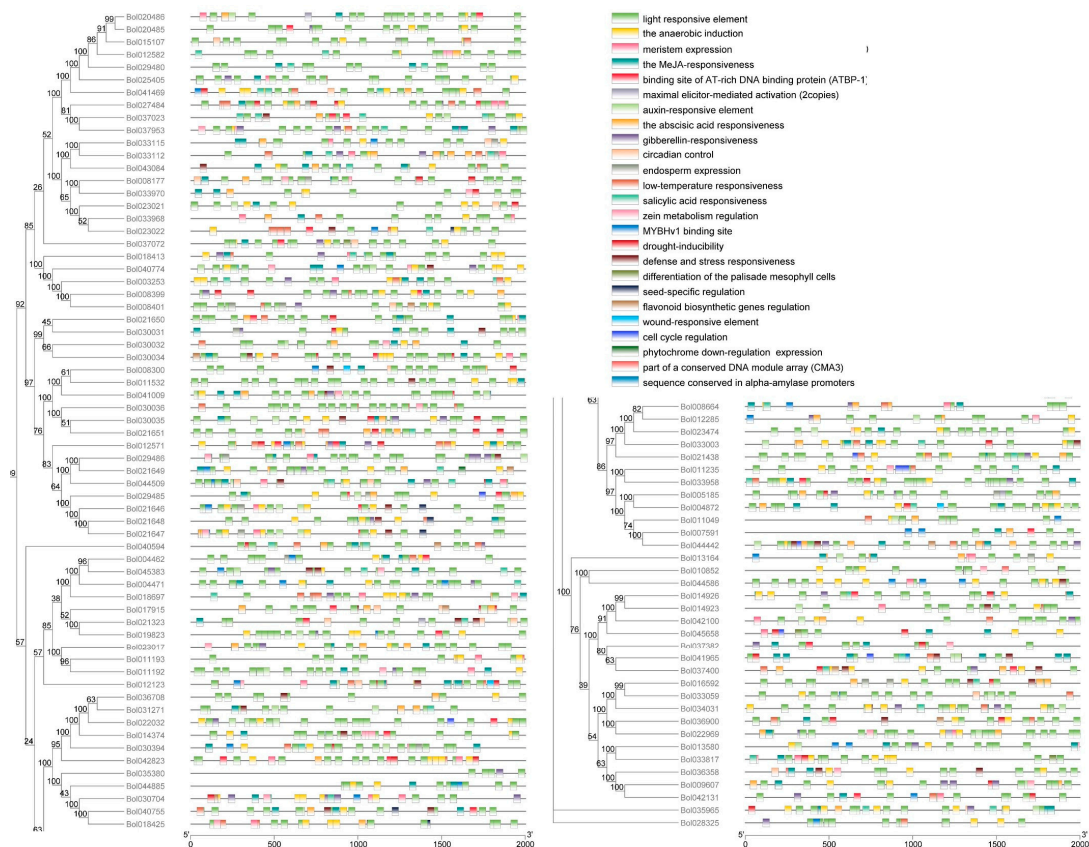
To explore the relationship between the nonsynonymous substitution rates ( $K_a$ ), synonymous substitution rates ( $K_s$ ), and expression pattern,  $K_a$ ,  $K_s$ , and  $K_a/K_s$  of the 21 paralogous gene pairs were calculated (Table 1). The result showed that the  $K_a/K_s$  values of 20 pairs were less than 1, indicating that these genes suffered negative selection during evolution. There was one pair (*Bol033115* and *Bol033112*) with  $K_a$  and  $K_s$  values equal to 0, indicating that the base has not been substituted during evolution. However, according to our result, there was not a significant correlation between the  $K_a/K_s$  and the gene expression divergence.

**Table 1.** The synonymous substitution rate ( $K_a$ ) and nonsynonymous substitution rate ( $K_s$ ) between paralogous PG gene pairs of *B. oleracea*. Gene expression patterns were categorized into four categories (I, II, III, and IV).

Paralogous Gene Pairs		$K_a$	$K_s$	$K_a/K_s$	Categories
<i>Bol020486</i>	<i>Bol020485</i>	0.0021	0.0264	0.0808	IV
<i>Bol037023</i>	<i>Bol037953</i>	0.1202	0.5223	0.2301	II
<i>Bol033115</i>	<i>Bol033112</i>	0	0	—	I
<i>Bol008177</i>	<i>Bol033970</i>	0.0547	0.267	0.2047	II
<i>Bol008399</i>	<i>Bol008401</i>	0.0221	0.0661	0.3347	I
<i>Bol029486</i>	<i>Bol021649</i>	0.0383	0.2782	0.1377	IV
<i>Bol021648</i>	<i>Bol021647</i>	0.0063	0.0072	0.8854	I
<i>Bol004462</i>	<i>Bol045383</i>	0.0316	0.1065	0.297	IV
<i>Bol021323</i>	<i>Bol019823</i>	0.1487	0.3771	0.3942	I
<i>Bol011193</i>	<i>Bol011192</i>	0.0504	0.1386	0.3636	IV
<i>Bol040755</i>	<i>Bol018425</i>	0.0359	0.2634	0.1361	II
<i>Bol008664</i>	<i>Bol012285</i>	0.0383	0.367	0.1044	II
<i>Bol011235</i>	<i>Bol033958</i>	0.0934	0.3071	0.3041	IV
<i>Bol005185</i>	<i>Bol004872</i>	0.0471	0.2669	0.1766	I
<i>Bol007591</i>	<i>Bol044442</i>	0.1061	0.323	0.3283	IV
<i>Bol010852</i>	<i>Bol044586</i>	0.0528	0.3528	0.1498	I
<i>Bol014926</i>	<i>Bol014923</i>	0.003	0.0152	0.201	IV
<i>Bol016592</i>	<i>Bol033059</i>	0.0526	0.3394	0.155	IV
<i>Bol036900</i>	<i>Bol022969</i>	0.0622	0.2988	0.2082	II
<i>Bol013580</i>	<i>Bol033817</i>	0.0427	0.3785	0.1129	III
<i>Bol009607</i>	<i>Bol042131</i>	0.0294	0.3163	0.0928	II

#### 2.4. Cis-Elements Analysis of the BoPG Promoters

Promoters can control the gene expression at the right time, place, and level [32]. To further investigate the potential functions of BoPGs, the cis-elements in their promoter sequences were identified and analyzed. Twenty-seven cis-elements related to biotic and abiotic stresses, as well as hormone responses, were detected (Figure 6 and Supplementary Materials Tables S5 and S6). Up to 96 promoters contained the hormone-response cis-elements (the MeJA-responsiveness (CGTCA-motif, and CGTCA-motif), auxin-responsive element (TGA-box and AuxRR-core), gibberellin-responsiveness (GARE-motif, P-box, and TATC-box), salicylic acid responsiveness (TCA-element), or abscisic acid responsiveness (ABRE)). In terms of the responsiveness to various abiotic stresses, 45 promoters containing the cis-acting element related to low-temperature responsiveness (LTR) and 36 promoters, including the drought-inducibility element (MBS) were identified. All of the promoters of BoPGs contained a light-responsive element. The MYBHv1 binding sites (CCAAT-box) were detected in 33 promoters. The cis-element essential for anaerobic induction (anaerobic responsive element, ARE) was detected in up to 86 promoters, whereas only three promoters covered the wound-responsive element (wound (WUN)-motif). In addition, 37 promoters containing zein metabolism regulation elements (O2-site) were identified. Flavonoid biosynthetic gene regulation elements were detected in 10 promoters. There were seed-specific regulation elements (RY-element) and root-specific elements (motif I) in six and one promoters, respectively. In general, 27 kinds of cis-elements were detected in the promoters of 99 BoPGs.

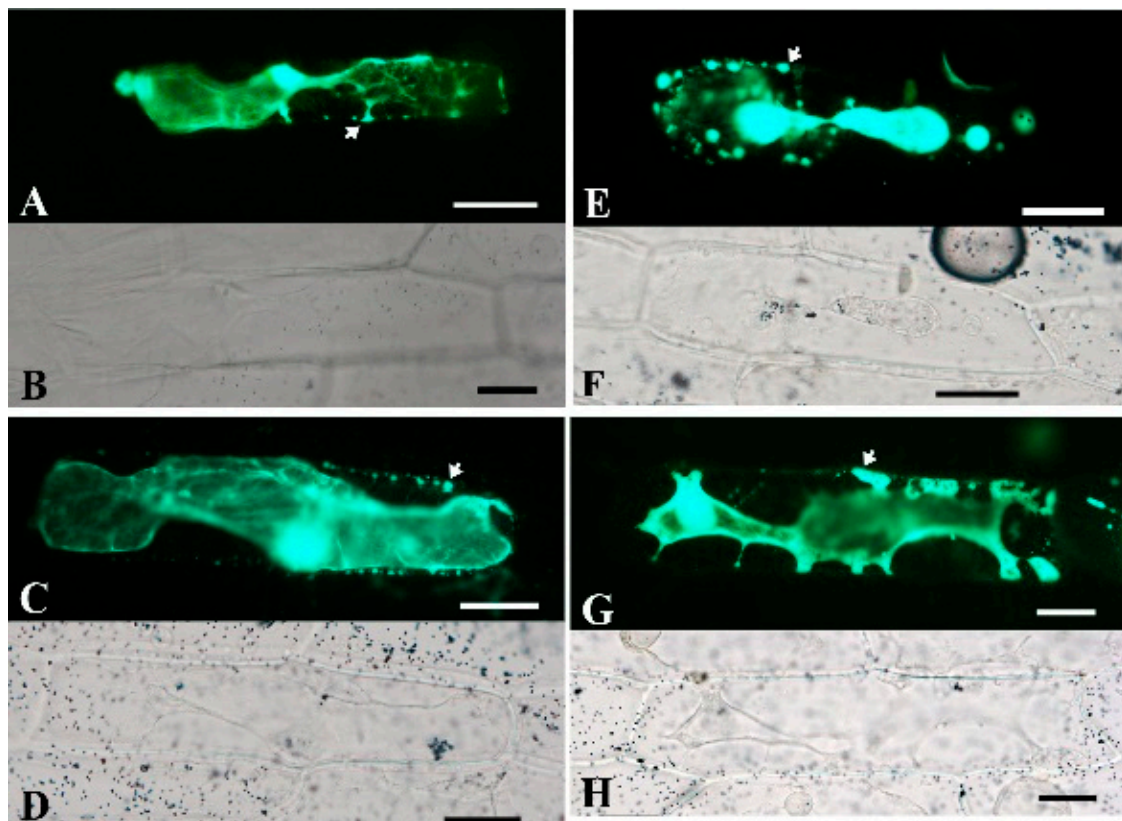


**Figure 6.** The cis-acting elements predication in the 2 kb sequences upstream of the “ATG” of the BoPGs. The types of elements are marked by different background colors.

### 2.5. Subcellular Localization of BoPG Proteins

The PGs localize to the plant cell wall and are involved in the degradation of pectin [33]. The presence of the amino acid N-terminal precursor sequence keeps the PG protein in an inhibitory state until it is transported to the final site of action, or localizes the PG protein to a specific site on the cell wall [34]. In this study, TargetP 1.1 Server was used to predict the subcellular location of eukaryotic proteins (Supplementary Materials Table S2). All of the 99 BoPG proteins were predicted to contain a signal peptide, indicating that these proteins might be secreted proteins. To examine the accuracy of the prediction of subcellular locations of BoPGs, we selected the four PG proteins (Bol041965, Bol035965, Bol040774, and Bol041469), which had a higher expression level in inflorescences to construct subcellular localization fusion expression vectors. After transient expression of the fusions in onion epidermis cells, we visualized the green fluorescent protein (GFP) signal by confocal microscopy. Observations indicated that the fluorescent signals of the four proteins were ubiquitously distributed in the onion epidermal cells. To determine if these proteins could be localized on the cell wall, we performed a plasmolysis experiment by treatment with  $0.3 \text{ g}\cdot\text{mL}^{-1}$  sucrose. After plasmolysis, the fluorescent signals could be observed in the cell wall (Figure 7). The results above indicated that the four PG proteins were secreted proteins and could localize on the cell wall.





**Figure 7.** Subcellular localization of four BoPG-green fluorescent protein (GFP) fusion proteins in onion epidermal cells. (A, E, C, and G) show fluorescence images of plasmolyzed transgenic cells. The fluorescence signal can be observed in the cell wall (white arrows). (B, F, D, and H) are bright-field images of the corresponding onion epidermal cells. Scale bars = 50  $\mu\text{m}$ .

### 3. Discussion

PGs play a key role in degrading pectin of the cell wall in plant development. The gene family of PG has recently been identified in many higher plants. Therefore, it is understandable that PG is a very important enzyme in the study of cell wall hydrolysis. Here, we identified 99 full-length PG genes from *B. oleracea*. Phylogenetic trees were developed to study the evolution of PGs between *B. oleracea* and *A. thaliana*. Furthermore, based on qRT-PCR data, BoPGs displayed specific temporal and spatial expression patterns, which provided clues for investigating their specific functions.

The phylogenetic analysis constructed by amino acid sequences from five species showed that all of the BoPGs were clustered into seven clades. This is similar to the results in other species, such as *B. rapa* [21] and *S. lycopersicum* [8]. By analyzing the PG numbers of the five species in different clades, we can speculate that clades A, B, and E contained PGs of land plants, and clade C and clade D only included PG genes of flowering plants. This result was consistent with the report of Park et al. [18]. According to our result, clade G only contained the PGs of dicotyledons. This may indicate that PG genes in clade G appeared later during evolution, while clades A, B, E, and F are more primitive. Meanwhile, in our result, the rapid expansion events of PG genes were found from monocotyledons to dicotyledons, which might have resulted from lineage-specific expansion after the split of monocotyledonous plants from their dicotyledonous counterparts about 140–150 Mya [35]. Furthermore, the PG expansion occurred after the split of lycophytes to euphyllophytes.

Divergences in coding regions, especially those that can change the function of the gene, can be caused by amino acid altering substitutions and/or alterations in the exon–intron structure [27]. In our study, the exon numbers of the 21 members in clade E exhibited the largest variation, varying from 1 to 6. When we analyzed the expression characteristics of BoPGs in different clades, we found that the



expressed BoPGs of clade E showed ubiquitous expression characteristics. Therefore, the changes of the exon–intron structure might relate to the expression of BoPGs. It is possible that PG functional divergence can be, in part, attributed to expression divergence [36]. Therefore, it can be speculated that the BoPGs in clade E might participate in multiple stages of the growth and development of *B. oleracea*. This further proved the previous theory that the clade E members of PG family are possibly ancient proteins and are fundamental and indispensable in almost all plant organs of different species [18]. It is worth mentioning that clade F members have been shown to maybe be related to flowering and probably fruit development in previous studies [18,21]. In our study, the expressed BoPGs in clades D, C, and F were mainly responsible for reproduction development (inflorescences or tender siliques). Moreover, the BoPG *Bol018697* that belonged to clade F was dominantly expressed in Bud 5 in our result. This BoPG has been demonstrated to be closely related to pollen development by in situ hybridization and promoter expression analysis [14]. The possible functions of other BoPGs, especially the genes with higher expression levels, need further study.

According to previous research reports, domains I and II may be components of the catalytic site, domain III is thought to participate in the catalytic reaction, and domain IV (RIK) constitutes a possible candidate for interactions with the ionic groups of the carboxylic acid groups in the substrate [16,37–39]. In our results, almost all of the BoPGs in clades A, B, C, D, and F contained all of the four conserved domains, while all of the clade E members lacked the III domain. This is similar to the results obtained for other species, with domain III showing lower conservation and being missing in clade E PGs [8,20,21]. The reason why the expression characteristics of clade E PGs displayed large variation may be associated with the lack of the III domain. In addition, some BoPG members only have the two (I and II) or one (II) domain. This indicates that these genes might have catalytic activity, but lost their interaction ability, with the substrate containing the ionic groups of the carboxylic acid groups. Since the protein domain is closely related to its function, the II domain may be more closely related to the identification of PG.

After splitting with *A. thaliana*, Brassicaceae underwent the addition of WGT, which was thought to have occurred between 13 and 17 million years ago [28,29]. The WGT of Brassicaceae species in evolution involved genomic rearrangement and gene loss [28,40]. The number of BoPGs was 99, which was far less than three times the number of *A. thaliana* PGs. Moreover, by analyzing the proportions of BoPGs retained in each of the three subgenomes relative to *A. thaliana*, we found that most of the BoPGs were lost during evolution. Therefore, we speculated that gene loss may be an important mechanism in the generation of diversity of BoPGs. This was consistent with the result obtained from *Populus*, showing that gene loss might be a general evolutionary pattern of large gene family evolution [19,41]. Furthermore, the seven tandem-duplicated pairs were only present in C01–C05, indicating that the expansion modes of BoPGs in different chromosomes after splitting with *A. thaliana* were not the same during the evolutionary process. After calculating the obtained or lost BoPGs in different clades when split with *A. thaliana*, we found that their numbers in clade A to clade E were all changed. In particular, the tandem repeat event mainly occurred in clade C and clade D. This illustrated that there are also differences in the expansion mode of BoPGs in different clades.

By comparing the common ancestor genes of different clades and the number of BoPGs obtained or lost during evolution, we found that, although a large number of PG genes were lost, many duplicates have been retained in the *B. oleracea* genome. The retention and loss of duplicate genes are related to the functional needs of species. In general, the variation in expression is believed to be the initial step in functional divergence among duplicate genes and thereby increases the probability of the existence of duplicate genes in the genome [42]. Moreover, the phenomena of neofunctionalization and subfunctionalization may also have been adapted by the retained duplicate genes. The neofunctionalization model postulates that the gain of new function(s) is a major factor in the retention of both copies of duplicate genes in a genome. The subfunctionalization model is also known as the duplication-degeneration-complementation (DDC) model, and it assumes that the two duplicate genes undergo complementary degeneration of their cis-regulatory motifs, so that

both copies are required to produce the full complement of the cis-regulatory motifs of the ancestral gene [36,42,43]. By analyzing the temporal and spatial expression characteristics of 21 paralogous BoPG gene pairs, 16 pairs (76.2%) showed divergent expression patterns. According to our statistics, some paralogs were selectively expressed in different tissues or partial overlap, indicating that they may have undergone neofunctionalization in the process of evolution, or that both of them have undergone subfunctionalization and have complementary functions for the growth and development of plants. In addition, in some paralogs, one gene was not expressed, and the other was selectively expressed in certain tissue. This might be due to one of the paralog genes having become a pseudogene or the other one having obtained a new function during the process of evolution. Moreover, some paralogs displayed the same expression characteristic (did not express or expressed in the same organ and with the same expression level), indicating that the two genes may both be pseudogenes or have the same function.

Transcriptional regulation is a crucial contributor to evolutionary change in the genotype-phenotype relationship [32]. At the transcription level, the combination of cis-acting elements and transcription factors in the promoter region is involved in the regulation of gene expression. For example, two duplicated PG genes, *BcMF26a* and *BcMF26b*, in *B. rapa*, with amino acid sequences, had 92.19% similarity and showed a divergent expression and function, which may be responsible for their promoters having a very large divergence (63.20% similarity) [44]. In our study, the cis-elements in BoPGs' promoter sequences were identified and analyzed. We found that up to 96 promoters contained the hormone response cis-elements, including ABA, IAA, MeJA, and GA, and different stresses. Some papers have characterized the relationship between PG and hormones. Hou et al. [45] and Lin et al. [46] showed that the inhibitor of ethylene perception 1-methylcyclopropene (1-MCP) can distinctly suppress the expression levels of PaPG1 and decrease the PG activities of apricots and plums during cold storage. During the germination of tomato seeds, there is a delay in ABA degradation, and the activity of the four cell-wall-degrading enzymes (including exo-polygalacturonase) evaluated is inhibited [47]. Moreover, previous research results showed that ethylene, JA, and ABA were required for the normal expression of the PG gene *QRT2* in floral organ abscission zones. While ethylene may directly regulate *QRT2* expression in the floral organ abscission zones, JA and ABA may act indirectly via roles in promoting flower senescence [48]. Therefore, the hormones may participate in plant development by regulating the PGs' activities. However, relevant research reports are still lacking, especially on reproductive development. According to our result, hormones may also be important for the growth and development of *B. oleracea*. However, the regulation of PGs' function by hormone signaling needs to be further clarified.

## 4. Materials and Methods

### 4.1. Identification and Sequence Analysis

Gene identifiers of 68 *A. thaliana* PG genes were obtained from Liang et al. [21]. The corresponding protein sequences were obtained from The Arabidopsis Information Resource (<http://www.arabidopsis.org/>). The target PG family members were searched for using the *B. oleracea* genome database (<http://www.ocri-genomics.org/bolbase/index.html>) in the TBLASTP program, with default algorithm parameters. Then, together with the flank regions of 2 kb upstream and downstream, the candidates were re-annotated by using FGENESH (<http://linux1.softberry.com/berry.phtml?group=programs&subgroup=gfind&topic=fgenesh>). All *B. oleracea* PG candidates collected were primarily analyzed by using the protein families' database (Pfam) to confirm the presence of GH28 domains in their protein structures. Candidate genes with at least one PG conserved domain (domain I, II, III, and IV) were collected. The well-known PG AtQRT3 [49], which does not contain one conserved domain of PG, was also analyzed.

The whole sequences of chromosomes and scaffolds were downloaded from the *B. oleracea* database. Together with the results of BLASTP, we obtained information of genomic sequences, cDNA sequences,

intron distribution patterns, phases, and intron/exon boundaries of the deduced BoPGs. Together with the sequence information provided above, we located all of the BoPGs on the chromosomes and scaffolds. An exon/intron map was constructed by using the software TBtools. The conserved domains were analyzed, using the MEME website (<http://meme.sdsc.edu/meme/meme.html>). Their protein molecular weight and isoelectric point were predicted by using PROTEIN CALCULATOR v3.4 (<http://protcalc.sourceforge.net/>). Their signal peptide sequence prediction was performed by using SignalP-5.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>).

To analyze the cis-acting elements, we first obtained the *B. oleracea* genome sequence from the Bolbase (<http://www.ocri-genomics.org/bolbase/index.html>). By using the TBtools software, we extracted and obtained 2 kb sequences upstream of the BoPGs' start codon. Then, the cis-elements in the promoter sequences were predicted by PlantCare (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). The predicted cis-acting elements information was visualized by using TBtools software.

#### 4.2. Phylogenetic and Molecular Evolution

Multiple sequence alignment for full-length PG protein sequences of five species was generated by using ClustalX. The full-length PG protein sequences of *P. patens* (moss and bryophyte), *S. moellendorffii* (spikemoss and lycophyte), and *O. sativa* were obtained from a previous study [6,19]. Phylogenetic relationships were developed by the MEGA7 program, using a neighbor-joining (NJ) procedure with the following parameters: Poisson correction, pairwise deletion, and bootstrap analysis with 1000 replicates. Sequence gaps were treated as missing characters. Then, the online software iTOL (<https://itol.embl.de/>) was used to decorate the phylogenetic tree. According to the method of Park et al. [18], the different clades were divided.

*Brassicaceae* species have three subgenomes which share the same diploid ancestor of the model species *A. thaliana*. Among these three subgenomes, there are biased gene fractionation (gene loss), named the least fractionized subgenome as LF, the moderate gene fractionized as MF1, and the most gene fractionized as MF2. According to the report of Cheng et al. [50], we searched for syntenic genes between *A. thaliana* and *B. oleracea*, using the "Syntenic gene search" in the Brassica Database (<http://brassicadb.org/brad/searchSyteny.php>). The distribution of syntenic BoPGs of *A. thaliana* in the three subgenomes was recorded.

The phylogenetic relationship between *B. oleracea* and *A. thaliana* was analyzed by the MEGA7 program by using a neighbor-joining (NJ) procedure with the following parameters: Poisson correction, pairwise deletion, and bootstrap analysis with 1000 replicates. Using the method of Kim et al. [31], the extent of lineage-specific expansion of the PG genes between the two species was investigated. The branches were defined by identifying nodes representing speciation events (dots). For those branches, a bootstrap value higher than 50% indicated the criteria for the least number of common ancestral PGs between *B. oleracea* and *A. thaliana*. The branches containing only sequences for one of the two plants indicated that gene loss had occurred during evolution.

In this study, when the length of highly similar sequences between two closely linked genes was more than 80% of the longer one, whilst their similarity was greater than 80%, they were characterized as tandem replication genes. The nonsynonymous substitution rates ( $Ka$ ) and synonymous substitution rates ( $Ks$ ) between paralog pairs within BoPGs were calculated by using the  $KaKs$  Calculator program of TBtools software. Additionally, by using TBtools, we conducted synteny analysis and generated a diagram of the genomic distribution in *B. oleracea*.

#### 4.3. Gene Expression Analysis

*B. oleracea* cv. Sanxiong plants were cultivated on the experimental farm of the Zhejiang Academy of Agricultural Sciences. Flower buds (Bud 1 to Bud 5) at five pollen developmental stages (stage 1, pollen mother cell stage; stage 2, tetrad stage; stage 3, uninucleate microspore stage; stage 4, binucleate microspore stage; and stage 5, mature pollen stage) were selected by cytological detection, as described

by Huang et al. [51]. At the flowering stage, the samples, including flower buds (Bud 1 to Bud 5), five different organs (roots, stems, leaves, inflorescences, and germinal siliques), and five floral parts (sepals, petals, filaments, anthers, and pistils) of Bud 5 were collected and stored at  $-80^{\circ}\text{C}$ .

The total RNA of the above samples was purified by using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). First-strand cDNA was synthesized by using the PrimeScript RT Reagent Kit (TaKaRa, Dalian, China), in accordance with the manufacturer's instructions.

The qRT-PCR analysis was used to examine the transcript levels of BoPGs. Based on the multiple-sequence alignment of BoPG gene sequences, specific primers were designed by using Primer premier 5.0 (Supplementary Materials Table S7). The primers of the PG genes which have a high sequence similarity were designed in the non-conserved regions of the gene coding sequence or the 5'/3'-untranslated region for expression analysis. Two highly similar genes that do not have large-scale different sequences suitable for primer design, respectively, were used for the same primer pairs for the qRT-PCR (the five sets with a yellow back label in Supplementary Materials Table S7). The constitutively expressed gene *GAPDH* [52] was selected as an internal control. The qRT-PCR was performed with the CFX96 Real Time System (Bio-Rad, California, Hercules, USA) machine. PCR conditions were 30 s at  $95^{\circ}\text{C}$ , followed by 40 cycles of 5 s at  $95^{\circ}\text{C}$ , 30 s at  $65^{\circ}\text{C}$ , and 1 min at  $72^{\circ}\text{C}$ . Each sample was run in triplicate with 15  $\mu\text{L}$  of the reaction volume by using the SYBR Premix Ex Taq Kit (TaKaRa, Dalian, China). For qRT-PCR analysis, three technical replications were applied. Expression analysis of low- and non-expressed genes was conducted at least twice with independent experiments. The  $2^{-\Delta\Delta\text{Ct}}$  method was used to calculate the relative expression levels. The software TBtools was employed to generate the heatmap.

#### 4.4. Subcellular Localization

We selected four BoPGs which had higher expression levels in inflorescences, to investigate their subcellular localizations by generating C-terminal green fluorescent protein (GFP) fusions. Based on their CDS provided by the *Brassica* database, we designed four specific primers (Supplementary Materials Table S8) for nucleotide fragment amplification. The PCR amplification reaction was carried out by using the high-fidelity enzyme KOD-plus-Neo, in accordance with the manufacturer's instruction. The resulting fragment was cloned into the fusion expression vector pFGC-GFP under the control of the constitutive CaMV35S promoter, to create the fusion vector. Colonies containing the appropriate insert were identified by sequencing. The fusion vector was then transiently transformed into onion epidermal cells by particle bombardment. The fluorescence signal was analyzed after 16 h of incubation with the Nikon Eclipse 90i Fluorescent Microscope (ECLIPSE 90i; Nikon). The onion epidermal cells were plasmolyzed by infiltrating  $0.3\text{ g}\cdot\text{mL}^{-1}$  sucrose.

**Supplementary Materials:** Supplementary Materials can be found at <http://www.mdpi.com/1422-0067/21/16/5706/s1>. Figure S1. The copy number changes of *B. oleracea* and *A. thaliana* PG genes in A to G clades. Numbers in ellipses and rectangles represent the numbers of PG genes in extant and ancestral species, respectively. Numbers on branches with plus and minus symbols represent the numbers of gene gains and losses, respectively. Figure S2. Hierarchical clustering and heat map showing the expression levels of BoPGs in seven clades. The scale bars represent relative expression level. The blue, yellow, and red shading indicate relatively low, medium, and high expression respectively. R: roots, St: stems, L: leaves, Inf: inflorescences, Si: siliques. Figure S3. The expression analysis of BoPGs which had specific or high expression levels in inflorescences. B1 (Bud1) - B5 (Bud5) indicate the flower buds at different pollen development stages. B1, pollen mother cell stage; B2, tetrad stage; B3, uninucleate microspore stage; B4, binucleate microspore stage; and B5, mature pollen stage. Figure S4. The expression analysis of *Bol041965* and *Bol018697* in the five organs of Bud5. Se, Sepals; Pe, petals; Fil, Filaments; An, Anthers; Pi, pistils. Table S1. Full-length amino acid sequence of PGs identified from *B. oleracea* genome. Table S2. The molecular details analysis of BoPGs. Table S3. The number of PG members of clade A ~ clade G in five terrestrial plant species. Table S4. The syntenic PG genes between *B. oleracea* and *A. thaliana* and the distributed of BoPGs in three subgenomes (LF, MF1 and MF2). The orange background indicates the recurring PG genes. Table S5 & Table S6. The cis-element analysis of BoPGs' promoters. Table S7. Primers used in qRT-PCR analysis of PG genes in *B. oleracea*. The sets with yellow background indicate the genes used the same primer pairs for qRT-PCR. Table S8. The four pairs of primer sequences for PCR amplification in subcellular localization experiment. The underlined base represents the cleavage site.



**Author Contributions:** Conceptualization, J.C.; data curation, M.L., J.I., and R.G.; formal analysis, J.I.; investigation, M.L. and J.I.; methodology, R.G.; supervision, J.C.; visualization, M.L. and R.G.; writing—original draft, M.L.; writing—review and editing, B.W. and J.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Natural Science Foundation of China (No. 31272176) and the Fujian Natural Science Foundation (No. 2017J01431). The APC was funded by the College of Horticulture at Fujian Agriculture & Forestry University.

**Acknowledgments:** We thank Xinmin Zhong, Zhejiang Academy of Agricultural Sciences, for providing the plant material *Brassica oleracea* cv. Sanxiong for this experiment.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Demura, T.; Tashiro, G.; Horiguchi, G.; Kishimoto, N.; Kubo, M.; Matsuoka, N.; Minami, A.; Nagata-Hiwatashi, M.; Nakamura, K.; Okamura, Y.; et al. Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15794–15799. [[CrossRef](#)] [[PubMed](#)]
2. Hadfield, K.A.; Bennett, A.B. Polygalacturonases: Many genes in search of a function. *Plant Physiol.* **1998**, *117*, 337–343. [[CrossRef](#)] [[PubMed](#)]
3. Sander, L.; Child, R.; Ulvskov, P.; Albrechtsen, M.; Borkhardt, B. Analysis of a dehiscence zone endo-polygalacturonase in oilseed rape (*Brassica napus*) and *Arabidopsis thaliana*: Evidence for roles in cell separation in dehiscence and abscission zones, and in stylar tissues during pollen tube growth. *Plant Mol. Biol.* **2001**, *46*, 469–479. [[CrossRef](#)] [[PubMed](#)]
4. Sitrit, Y.; Hadfield, K.A.; Bennett, A.B.; Bradford, K.J.; Downie, A.B. Expression of a polygalacturonase associated with tomato seed germination. *Plant Physiol.* **1999**, *121*, 419–428. [[CrossRef](#)] [[PubMed](#)]
5. Fischer, R.L.; Bennett, A.B. Role of cell-wall hydrolases in fruit ripening. *Annu. Rev. Plant Physiol. Mol. Biol.* **1991**, *42*, 675–703. [[CrossRef](#)]
6. Kim, J.; Patterson, S.E. Expression divergence and functional redundancy of polygalacturonases in floral organ abscission. *Plant Signal. Behav.* **2006**, *1*, 281–283. [[CrossRef](#)]
7. Roongsattham, P.; Morcillo, F.; Jantasuriyarat, C.; Pizot, M.; Moussu, S.; Jayaweera, D.; Collin, M.; Gonzalez-Carranza, Z.H.; Amblard, P.; Tregear, J.W.; et al. Temporal and spatial expression of polygalacturonase gene family members reveals divergent regulation during fleshy fruit ripening and abscission in the monocot species oil palm. *BMC Plant Biol.* **2012**, *12*. [[CrossRef](#)]
8. Ke, X.; Wang, H.; Li, Y.; Zhu, B.; Zang, Y.; He, Y.; Cao, J.; Zhu, Z.; Yu, Y. Genome-wide identification and analysis of polygalacturonase genes in *Solanum lycopersicum*. *Int. J. Mol. Sci.* **2018**, *19*, 2290. [[CrossRef](#)]
9. Huang, L.; Cao, J.; Zhang, A.; Ye, Y.; Zhang, Y.; Liu, T. The polygalacturonase gene BcMF2 from *Brassica campestris* is associated with intine development. *J. Exp. Bot.* **2009**, *60*, 301–313. [[CrossRef](#)]
10. Huang, L.; Ye, Y.; Zhang, Y.; Zhang, A.; Liu, T.; Cao, J. BcMF9 a novel polygalacturonase gene, is required for both *Brassica campestris* intine and exine formation. *Ann. Bot.* **2009**, *104*, 1339–1351. [[CrossRef](#)]
11. Zhang, A.; Qiu, L.; Huang, L.; Yu, X.; Lu, G.; Cao, J. Isolation and characterization of an anther-specific polygalacturonase gene, BcMF16, in *Brassica campestris* ssp. *chinensis*. *Plant Mol. Biol. Rep.* **2012**, *30*, 330–338. [[CrossRef](#)]
12. Zhang, A.; Chen, Q.; Huang, L.; Qiu, L.; Cao, J. Cloning and expression of an anther-abundant polygalacturonase gene BcMF17 from *Brassica campestris* ssp. *chinensis*. *Plant Mol. Biol. Rep.* **2011**, *29*, 943–951. [[CrossRef](#)]
13. Yu, Y.; Lv, M.; Liang, Y.; Xiong, X.; Cao, J. Molecular cloning and characterization of a novel polygalacturonase gene, BcMF24, involved in pollen development of *Brassica campestris* ssp. *chinensis*. *Plant Mol. Biol. Rep.* **2014**, *32*, 476–486. [[CrossRef](#)]
14. Lyu, M.; Liang, Y.; Yu, Y.; Ma, Z.; Song, L.; Yue, X.; Cao, J. Identification and expression analysis of BoMF25, a novel polygalacturonase gene involved in pollen development of *Brassica oleracea*. *Plant Reprod.* **2015**, *28*, 121–132. [[CrossRef](#)]

15. He, Y.; Karre, S.; Johal, G.S.; Christensen, S.A.; Balint-Kurti, P. Amaize polygalacturonase functions as a suppressor of programmed cell death in plants. *BMC Plant Biol.* **2019**, *19*, 310. [[CrossRef](#)] [[PubMed](#)]
16. Torki, M.; Mandaron, P.; Mache, R.; Falconet, D. Characterization of a ubiquitous expressed gene family encoding polygalacturonase in *Arabidopsis thaliana*. *Gene* **2000**, *242*, 427–436. [[CrossRef](#)]
17. Park, K.C.; Kwon, S.J.; Kim, P.H.; Bureau, T.; Kim, N.S. Gene structure dynamics and divergence of the polygalacturonase gene family of plants and fungus. *Genome* **2008**, *51*, 30–40. [[CrossRef](#)]
18. Park, K.C.; Kwon, S.J.; Kim, N.S. Intron loss mediated structural dynamics and functional differentiation of the polygalacturonase gene family in land plants. *Genes Genom.* **2010**, *32*, 570–577. [[CrossRef](#)]
19. Yang, Z.L.; Liu, H.J.; Wang, X.R.; Zeng, Q.Y. Molecular evolution and expression divergence of the *Populus* polygalacturonase supergene family shed light on the evolution of increasingly complex organs in plants. *New Phytol.* **2013**, *197*, 1353–1365. [[CrossRef](#)]
20. Yu, Y.; Liang, Y.; Lv, M.; Wu, J.; Lu, G.; Cao, J. Genome-wide identification and characterization of polygalacturonase genes in *Cucumis sativus* and *Citrullus lanatus*. *Plant Physiol. Biochem.* **2014**, *74*, 263–275. [[CrossRef](#)]
21. Liang, Y.; Yu, Y.; Shen, X.; Dong, H.; Lyu, M.; Xu, L.; Ma, Z.; Liu, T.; Cao, J. Dissecting the complex molecular evolution and expression of polygalacturonase gene family in *Brassica rapa* ssp *chinensis*. *Plant Mol. Biol.* **2015**, *89*, 629–646. [[CrossRef](#)] [[PubMed](#)]
22. Chen, H.; Shao, H.; Fan, S.; Ma, J.; Zhang, D.; Han, M. Identification and phylogenetic analysis of the polygalacturonase gene family in Apple. *Hortic. Plant J.* **2016**, *2*, 241–252. [[CrossRef](#)]
23. Thompson, P.; Patel, G. Compensatory regulation of two closely related hemoglobin loci in *Chironomus tentans*. *Genetics* **1972**, *70*, 275–290. [[PubMed](#)]
24. Kassahn, K.S.; Dang, V.T.; Wilkins, S.J.; Perkins, A.C.; Ragan, M.A. Evolution of gene function and regulatory control after whole-genome duplication: Comparative analyses in vertebrates. *Genome Res.* **2009**, *19*, 1404–1418. [[CrossRef](#)] [[PubMed](#)]
25. Blanc, G.; Wolfe, K.H. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **2004**, *16*, 1679–1691. [[CrossRef](#)] [[PubMed](#)]
26. Duarte, J.M.; Cui, L.Y.; Wall, P.K.; Zhang, Q.; Zhang, X.H.; Leebens-Mack, J.; Ma, H.; Altman, N.; de Pamphilis, C.W. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Mol. Biol. Evol.* **2006**, *23*, 469–478. [[CrossRef](#)]
27. Xu, G.; Guo, C.; Shan, H.; Kong, H. Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1187–1192. [[CrossRef](#)]
28. Lysak, M.A.; Koch, M.A.; Pecinka, A.; Schubert, I. Chromosome triplication found across the tribe Brassiceae. *Genome Res.* **2005**, *15*, 516–525. [[CrossRef](#)]
29. 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. *PMC Plant Cell* **2006**, *18*, 1348–1359. [[CrossRef](#)]
30. Wang, X.; Wang, H.; Wang, J.; Sun, R.; Wu, J.; Liu, S.; Bai, Y.; Mun, J.H.; Bancroft, I.; Cheng, F.; et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **2011**, *43*, 1035–1039. [[CrossRef](#)]
31. Kim, J.; Shiu, S.H.; Thoma, S.; Li, W.H.; Patterson, S.E. Patterns of expansion and expression divergence in the plant polygalacturonase gene family. *Genome Biol.* **2006**, *7*, R87. [[CrossRef](#)] [[PubMed](#)]
32. Wray, G.A.; Hahn, M.W.; Abouheif, E.; Balhoff, J.P.; Pizer, M.; Rockman, M.V.; Romano, L.A. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **2003**, *20*, 1377–1419. [[CrossRef](#)] [[PubMed](#)]
33. Qi, M.F.; Xu, T.; Chen, W.Z.; Li, T.L. Ultrastructural localization of polygalacturonase in ethylene-stimulated abscission of tomato pedicel explants. *Sci. World J.* **2014**, *2014*. [[CrossRef](#)] [[PubMed](#)]
34. Dellapenna, D.; Bennett, A.B. In vitro synthesis and processing of tomato fruit polygalacturonase. *Plant Physiol.* **1998**, *86*, 1057–1063. [[CrossRef](#)]
35. Chaw, S.M.; Chang, C.C.; Chen, H.L.; Li, W.H. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol. Evol.* **2004**, *58*, 424–441.
36. Li, W.H.; Yang, J.; Gu, X. Expression divergence between duplicate genes. *Trends Genet.* **2005**, *21*, 602–607. [[CrossRef](#)]
37. Rexová-Benková, L. Evidence for the role of carboxyl groups in activity of endopolygalacturonase of *Aspergillus niger*. Chemical modification by carbodiimide reagent. *Collect. Czech. Chem. Commun.* **1990**, *55*, 1389–1395. [[CrossRef](#)]

38. Rao, M.N.; Kembhavi, A.A.; Pant, A. Implication of tryptophan and histidine in the active site of endo-polygalacturonase from *Aspergillus ustus*: Elucidation of the reaction mechanism. *Biochim. Biophys. Acta* **1996**, *1296*, 167–173.
39. Bussink, H.J.; Buxton, F.P.; Visser, J. Expression and sequence comparison of the *Aspergillus niger* and *Aspergillus tubigenensis* genes encoding polygalacturonase II. *Curr. Genet.* **1991**, *19*, 467–474. [[CrossRef](#)]
40. Liu, S.; Liu, Y.; Yang, X.; Tong, C.; Edwards, D.; Parkin, I.A.P.; Zhao, M.; Ma, J.; Yu, J.; Huang, S. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* **2014**, *5*, 3930. [[CrossRef](#)]
41. Zou, C.; Lehti-Shiu, M.D.; Thibaud-Nissen, F.; Prakash, T.; Buell, C.R.; Shiu, S.H. Evolutionary and expression signatures of pseudogenes in Arabidopsis and Rice. *Plant Physiol.* **2009**, *151*, 3–15. [[CrossRef](#)] [[PubMed](#)]
42. Lynch, M.; Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **2000**, *290*, 1151–1155. [[CrossRef](#)] [[PubMed](#)]
43. Force, A.; Lynch, M.; Pickett, F.B.; Amores, A.; Yan, Y.L.; Postlethwait, J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **1999**, *151*, 1531–1545. [[PubMed](#)]
44. Lyu, M.; Yu, Y.; Jiang, J.; Song, L.; Liang, Y.; Ma, Z.; Xiong, X.; Cao, J. BcMF26a and BcMF26b are duplicated polygalacturonase genes with divergent expression patterns and functions in pollen development and pollen tube formation in *Brassica campestris*. *PLoS ONE* **2015**, *10*, e0131173. [[CrossRef](#)]
45. Hou, Y.; Wu, F.; Zhao, Y.; Shi, L.; Zhu, X. Cloning and expression analysis of polygalacturonase and pectin methylesterase genes during softening in apricot (*Prunus armeniaca* L.) fruit. *Sci. Hortic.* **2019**, *256*, 108607. [[CrossRef](#)]
46. Lin, Y.; Lin, Y.; Lin, H.; Lin, M.; Li, H.; Yuan, F.; Chen, Y.; Xiao, J. Effects of paper containing 1-MCP postharvest treatment on the disassembly of cell wall polysaccharides and softening in Younai plum fruit during storage. *Food Chem.* **2018**, *264*, 1–8. [[CrossRef](#)]
47. Lara-Nunez, A.; Ventura-Gallegos, J.; Luisa, A.A.; Cruz-Ortega, R. Phytotoxicity of sicyos deppei during tomato germination and its effects on the role of ABA and cell wall enzymes. *Bot. Sci.* **2015**, *93*, 771–781. [[CrossRef](#)]
48. Ogawa, M.; Kay, P.; Wilson, S.; Swain, S.M. Arabidopsis dehiscence zone Polygalacturonase1 (ADPG1), ADPG2, and QUARTET2 are polygalacturonases required for cell separation during reproductive development in Arabidopsis. *Plant Cell* **2009**, *21*, 216–233. [[CrossRef](#)]
49. Rhee, S.Y.; Osborne, E.; Poindexter, P.D.; Somerville, C.R. Microspore separation in the quartet 3 mutants of Arabidopsis is impaired by a defect in a developmentally regulated polygalacturonase required for pollen mother cell wall degradation. *Plant Physiol.* **2003**, *133*, 1170–1180. [[CrossRef](#)]
50. Cheng, F.; Wu, J.; Fang, L.; Wang, X. Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. *Front. Plant Sci.* **2012**, *3*. [[CrossRef](#)]
51. Huang, L.; Zhao, X.; Liu, T.; Dong, H.; Cao, J. Developmental characteristics of floral organs and pollen of Chinese cabbage (*Brassica campestris* L. ssp. *chinensis*). *Plant. Syst. Evol.* **2010**, *286*, 103–115. [[CrossRef](#)]
52. Suzuki, T.; Higgins, P.J.; Crawford, D.R. Control selection for RNA quantitation. *Biotechniques* **2000**, *29*, 332–337. [[CrossRef](#)] [[PubMed](#)]

