

ARTICLE

Divergent evolutionary patterns of the MAPK cascade genes in *Brassica rapa* and plant phylogenetics

Peng Wu, Wenli Wang, Ying Li and Xilin Hou

Mitogen-activated protein kinase (MAPK) cascade signal transduction modules play crucial roles in regulating many biological processes in plants. These cascades are composed of three classes of hierarchically organized protein kinases, MAPKKKs, MAPKKs and MAPKs. Here, we analyzed gene retention, phylogenetic, evolution and expression patterns of MAPK cascade genes in *Brassica rapa*. We further found that the MAPK branches, classes III and IV, appeared after the split of bryophytes and green algae after analyzing the MAPK cascade genes in 8 species, and their rapid expansion led to the great size of the families of MAPKs. In contrast, the ancestral class I subfamily of MAPKK gene families have been highly conserved from algae to angiosperms. For the MAPKKK family, the MEKK and Raf subfamily share a common evolutionary origin, and Raf plays a major role in the expansion of the MAPKKK gene family. The *cis*-elements and interaction network analyses showed the important function of MAPK cascade genes in development and stress responses in *B. rapa*. This study provides a solid foundation for molecular evolution analyses of MAPK cascade genes.

Horticulture Research (2017) 4, 17079; doi:10.1038/hortres.2017.79; Published online 27 December 2017

INTRODUCTION

Based on their exposure to environmental stress, plants have constructed complex signaling networks to adapt to different stress conditions.¹ Protein kinases (PKs) play important roles in both growth and reproduction processes.^{1,2} As a highly conserved sub-family of PKs, the mitogen-activated protein kinases (MAPKs) consist of the following three classes: (1) MAPKs, (2) MAPK kinases (MAPKKs), and (3) MAPKK kinases (MAPKKKs), which act as a signal transmission cascade in the cellular metabolic and transcriptional response centers.^{1,2} MAPKs are activated by MAPKKs via phosphorylation of conserved threonine and tyrosine residues in the Thr-X-Tyr (T-X-Y) motif. MAPKKs, in turn, are activated by MAPKKKs when serine and threonine residues in the S/TXXXXXS/T motif are phosphorylated.^{3–9}

In eukaryotic organisms, MAPK pathways are relatively well-characterized cascades.¹⁰ In plants, genome-wide analyses of MAPK cascades (MAPKKK—MAPKK—MAPK) have been reported in several species.^{11–18} In *Arabidopsis*, 23 MAPKs, 10 MAPKKs, and 80 putative MAPKKKs have been identified. MAPK and MAPKK genes in plants are each divided into four distinct classes (classes A, B, C, and D), whereas MAPKKK genes are classified into the following three subfamilies: (1) MEKK, (2) Raf and (3) ZIK.

Plant MAPKs are involved in both biotic stresses, such as AtMAPK4,^{18,19} MsMMK2, and MsMMK3,¹ and in abiotic stresses, such as AtMAPK4,²⁰ OsMAPK4,²¹ and AtMAPK7.²² MKK9—MPK3/MPK6 regulates ethylene signaling and may also play a role in leaf senescence.^{9,23,24} Plant MKKs (MAPKKs) contain a phosphorylation site (S/T-X5-S/T) and a putative MAPK docking site in their N-terminal. Currently, only A and C type MAPKKs, such as AtMKK1, have been demonstrated to be activated by multiple abiotic stresses.²⁵ Previous studies have examined that MEKK1, MKK2, MPK4, and/or MPK6 respond to salt, drought, and cold stress in *Arabidopsis*.²⁶ CsNMAPK

regulated NO₃⁻ stress,²⁷ ROS, and osmotic adjustment under salt stress.²⁸ Several reports revealed that MAPKKK play roles in drought resistance. A study in rice showed that DSM1, a RAF-Like MAPKKK gene has function in drought and oxidative stresses signaling.²⁹ A tobacco MAPKKK (NPK1) has been shown to enhance drought tolerance in transgenic maize.³⁰ The plant immune response is negatively controlled by the MEKK1 (AtMEKK1)—MKK1 (AtMKK1) and MKK2—MPK4—MKS1/WRKY33 cascades.^{31,32} AtMKK1—AtMPK6 are involved in ABA—induced CAT1 expression,³³ MEKK1—MKK4/5 have an important function in plant innate immunity,³⁴ and MKK6 can regulate cytokinesis.^{35,36}

Studies on the evolution would be another interesting topic because their evolutionary history can shed light on functions. In land plants, the number of MAPKs and MAPKKs has expanded.^{37,38} Before the divergence of the rice and Brachypodium, MAPKKK genes experienced the segment duplication.³⁸ The conservation of MAPKKKs have also implied by the strong purifying selection during the evolution.³⁹ During the long evolutionary history, most of higher plant lineages have undergone polyploidization.⁴⁰ In parallel with the whole-genome duplication (WGD) in angiosperm genomes, the three large families (MAPK, MAPKK and MAPKKK) experienced expansion and functional divergence. If the genes were involved in networks or function in a dose-sensitive manner, which should be retained that gene dosage hypothesis predicts.⁴¹ *Brassica rapa* has undergone two duplication events (WGD α and β) and one whole-genome triplication event (WGT γ) within the Brassicaceae lineage.⁴² Specifically, the WGT caused extensive fractionation in the *B. rapa* genome, which has the chance to study the molecular evolution of the MAPK cascade. Comparative analyses of many other plant species can yield the origin and evolution of these genes. Therefore, in addition to the detailed study of the MAPK cascade in *Brassica rapa*, we comparatively analyzed MAPK, MAPKK, and MAPKKK genes

from eight representative plant species, including an algae (*Volvox carter*), the ancestor of land plants,⁴³ a bryophyte (*Physcomitrella patens*), the earliest sequenced land plant,⁴⁴ a lycophyte (*Selaginella moellendorffii*), an early vascular plant,⁴⁵ and five angiosperm plants: *Amborella trichopoda*, one of the earliest angiosperm, *Vitis vinifera*, the basal eucosid, and *Populus trichocarpa*, *Carica papaya* and *A. thaliana*.

In this study, we comprehensively analyzed the MAPK cascade genes in the *B. rapa* genome, as well as representative genomes in the tree of plant life, to understand the cascade retention and expansion patterns following the WGT event. In addition, divergent tissue-specific expression patterns and response to stresses of these genes will help us to understand the function of the MAPK cascade genes.

MATERIALS AND METHODS

Retrieval of sequences

To identify MAPK cascade genes in *B. rapa*, *Arabidopsis*, and *O.sativa*, the sequences were retrieved from BRAD (<http://brassicadb.org/brad/>),⁴² TAIR (<http://www.arabidopsis.org/>) and RGAP (<http://rice.plantbiology.msu.edu/>),⁴⁶ respectively. The sequences of others species were downloaded from Phytozome v9.1 (<http://www.phytozome.net/>).⁴⁷ MAPK cascade genes of *Arabidopsis* were performed against the proteins from other species by

BLASTP ($e\text{-value} < 1 \times 10^{-10}$, identity $> 70\%$). The serine/threonine-protein kinase-like domain (PF00069), was downloaded from Pfam (<http://pfam.sanger.ac.uk/>), and used to search the *B. rapa* protein database by HMMER v3.0 (<http://hmmer.janelia.org/>). Potential sequences were detected by SMART (<http://smart.embl-heidelberg.de/>).

Identification of gene synteny and duplicated MAPK cascade genes

According to previous reports, gene synteny and duplicated types were detected by the MScanX (Multiple Collinearity Scan toolkit).⁴⁸ The positions of syntenic MAPK cascade genes in the *B. rapa* on the ancestral genomic blocks were searched from the BRAD website (<http://brassicadb.org/brad/searchSynteny.php>).⁴² Circos software was used to draw syntenic diagram.⁴⁹

Phylogenetic analysis and characterization of the MAPK cascade genes

MEGA v6.0 was used to construct the maximum-likelihood (ML) trees with a bootstrap value of 1000 replications. MEGA v6.0 was used to calculate genetic distances. The MEME v4.10.1 tool (<http://meme.sdsc.edu/meme/>) was used to identify conserved motifs in the MAPK cascade proteins. The MAPK cascade gene structures were drawn by the online tool Gene Structure Display (GSDS, <http://gsds.cbi.pku.edu.cn/>). The STRING tool was used to construct interaction network of MAPK cascade proteins in *B. rapa* (<http://string-db.org/>).⁵⁰

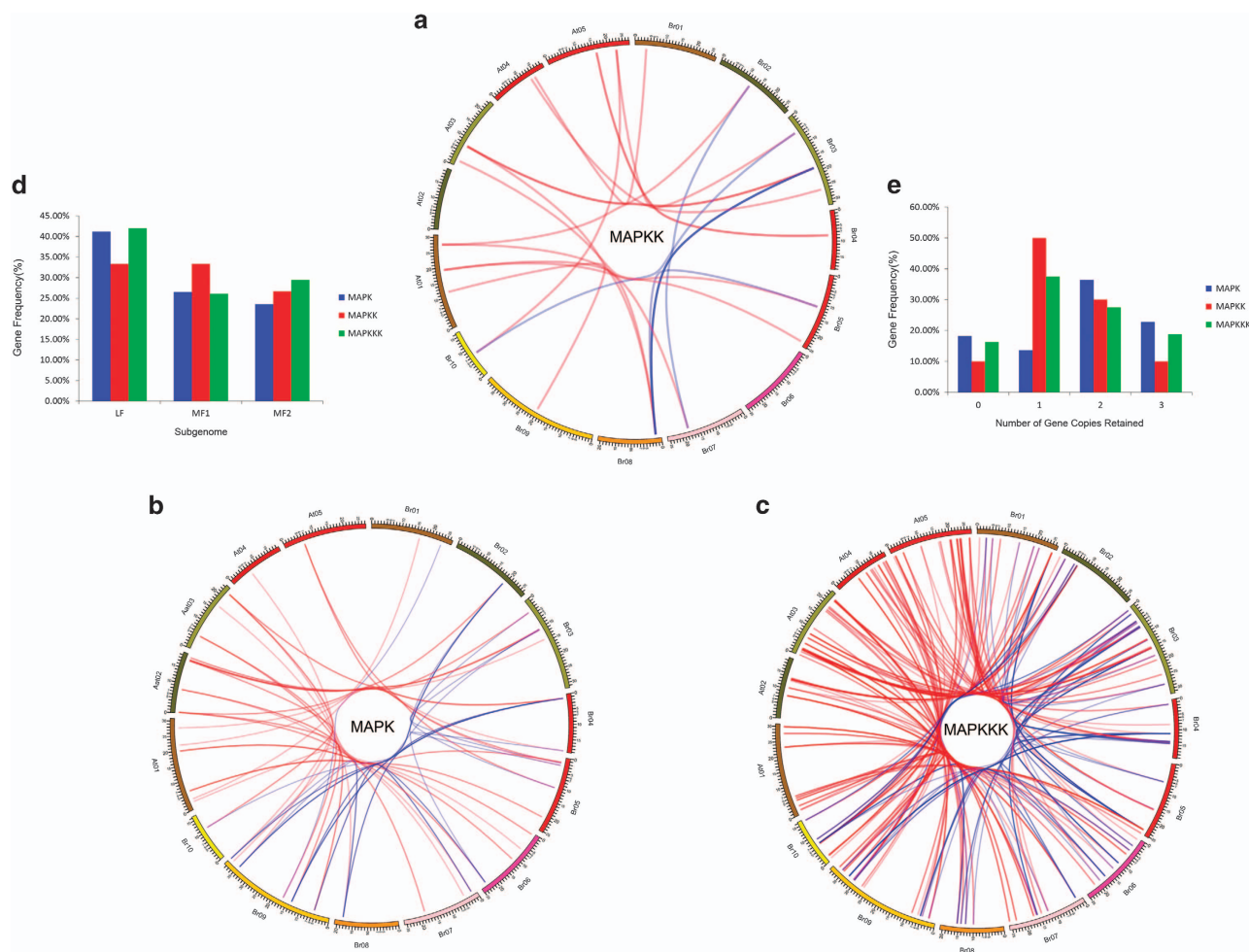


Figure 1. The syntenic MAPK cascade genes between *Brassica rapa* and *Arabidopsis thaliana* and their different retentions. (a–c) MAPK, MAPKK, MAPKKK syntenic genes lines are shown among the 10 *B. rapa* chromosomes (Br01–Br10) and the five *A. thaliana* chromosomes (At01–At05). The red lines represent the syntenic genes pairs between *B. rapa* and *Arabidopsis*, and the blue lines represent the syntenic genes in *B. rapa*. (d) Retention of homoeologs of MAPK cascade genes in the three subgenomes of LF (least fractionized subgenome), MF1 (moderately fractionized subgenome), and MF2 (most fractionized subgenome) in *B. rapa*. (e) Numbers of MAPK cascade genes after genome triplication and fractionation in *B. rapa*.

The MAPK cascade genes Ka/Ks analysed

The MAPK cascade genes from *B. rapa* were aligned using ClustalX 2.0 with the default parameters.⁵¹ The KaKs_calculator was used to calculate the Ks and Ka values.⁵² The Ks values for all the MAPK cascade homologous genes in *B. rapa* and *A. thaliana* were shown by boxplot using the R program (boxplot package).⁵³ The following formula: $T = Ks/2r$ was used to calculate divergence time ($r = 1.5 \times 10^{-8}$, synonymous substitutions per site per year).⁵⁴ Nucleotide distance was used to analyze the relationships among Classes I, II, III and IV.

Expression pattern analysis for MAPK cascade genes

Illumina RNA-seq data⁵⁵ were used to characterize the expression of the MAPK cascade genes in *B. rapa*. The expression level of MAPK cascade genes were analysed by the fragments per kilobase of exon model per million mapped reads (FPKM). The AtGenExpress Visualization Tool (AVT; <http://jsp.weigelworld.org/expviz/expviz.jsp>) was used to analyse the expression level⁵⁶ in *A. thaliana*. The R program (VennDiagram package) was used to draw Venn diagrams.

Promoter sequence analysis

Search for *cis*-regulatory elements in promoter of MAPK cascade genes sequences was performed using the PLACE Web Signal ScanPLACE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

RESULTS

Different retention of MAPK cascade genes following the WGT event in *Brassica rapa*

BLAST software was used to identify the number of MAPK, MAPKK, and MAPKKK genes in *A. thaliana* and *B. rapa*. Then, the MAPK cascade genes containing the serine/threonine-protein kinase-like domain was searched by an HMM (PF00069) in *B. rapa*. Finally, all the homogeneous candidate MAPK cascade sequences were subjected to Pfam, SMART and NCBI database analyses. After

multiple screening and validation steps, we finally identified 34 BrMAPK, 14 BrMAPKK, and 112 BrMAPKKK genes (Supplementary Table S1).

To investigate different syntenic genes in the MAPK cascade during the *B. rapa* WGT event, we completed a syntenic analysis with MCScanX. In total, 31 MAPKs, 14 MAPKKs, and 100 MAPKKKs were located in the syntenic regions (Figure 1). About 91% (31/34) of the MAPK genes, 93% (13/14) of the MAPKK genes, and 89% (100/112) of the MAPKKK were retained in the syntenic regions (Figure 1). We also counted the number of genes and analyzed the distribution of the three subgenomes by comparing the retention of MAPKs, MAPKKs and MAPKKKs (Figures 1d and e). The proportion of retained MAPK cascade genes was higher in the least fractionated (LF) subgenome than in the medium fractionated (MF1) and most fractionated (MF2) subgenome. This finding is consistent with a previous report in which the degree of retained genes in these three sub-genomes (LF, MF1, and MF2) decreased gradually.⁴² More MAPKKKs were retained in the MF2 sub-genome than in the MF1 sub-genome. MAPK (59%), MAPKK (40%), and MAPKKK (46%) genes retained more than two homologous genes, and the MAPK genes had higher retention than the two other gene families (MAPKK and MAPKKK). However, some MAPK (18%), MAPKK (10%) and MAPKKK (16%) genes were lost (Figure 1e). Therefore, MAPK genes were preferentially retained compared to the other two gene families. The MAPKK genes were similarly retained to MAPKKK genes following the WGT in *B. rapa* (Figure 1, Supplementary Table S2).

Characteristic structure analysis of BrMAPK cascade proteins

To study the characteristic of MAPK cascade genes in *B. rapa*, we used the Maximum Likelihood method to construct phylogenetic trees (Figure 2). All the BrMAPKs were classified into four classes, I, II, III and IV (Figure 2). We searched using the MEME tool to identify

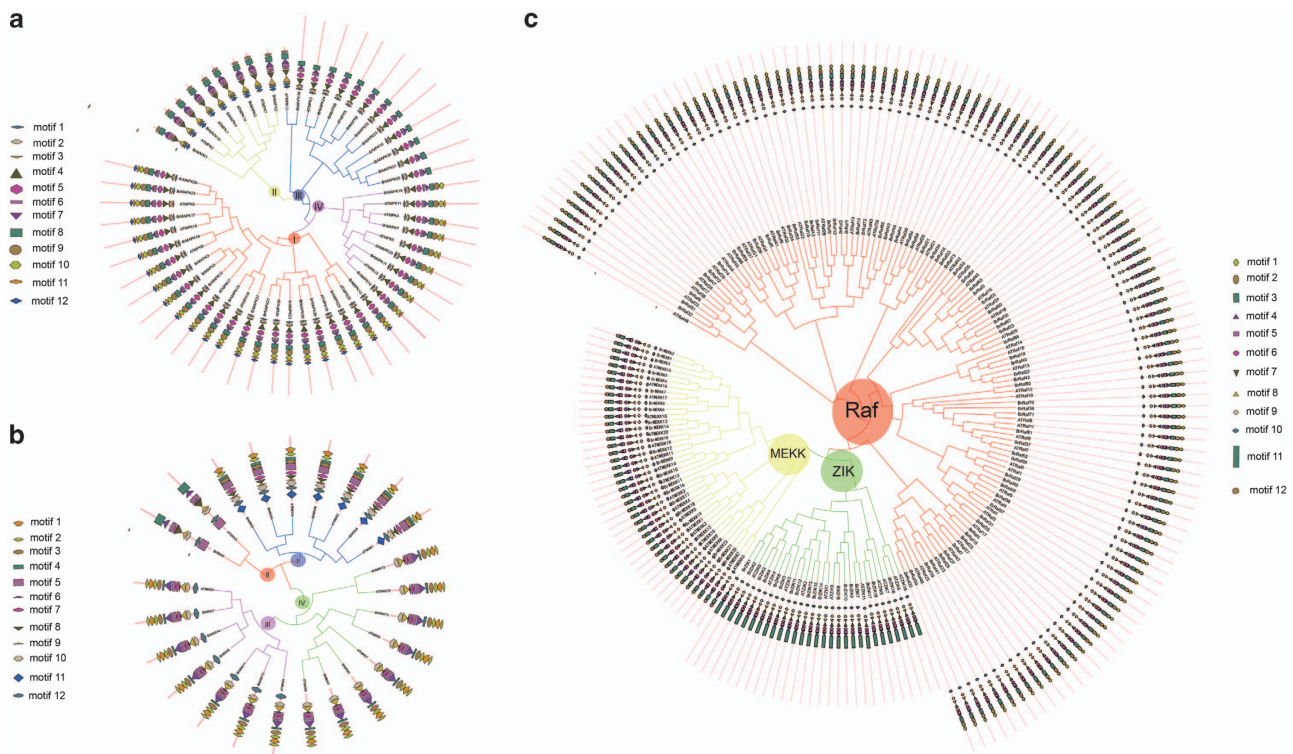


Figure 2. Schematic diagram of amino-acid motifs of MAPK cascade genes in *Arabidopsis* and *B. rapa*. (a) Phylogenetic relationships and protein motifs of MAPK genes. (b) Phylogenetic relationships and protein motifs of MAPKK genes. (c) Phylogenetic relationships and protein motifs of MAPKKK genes.

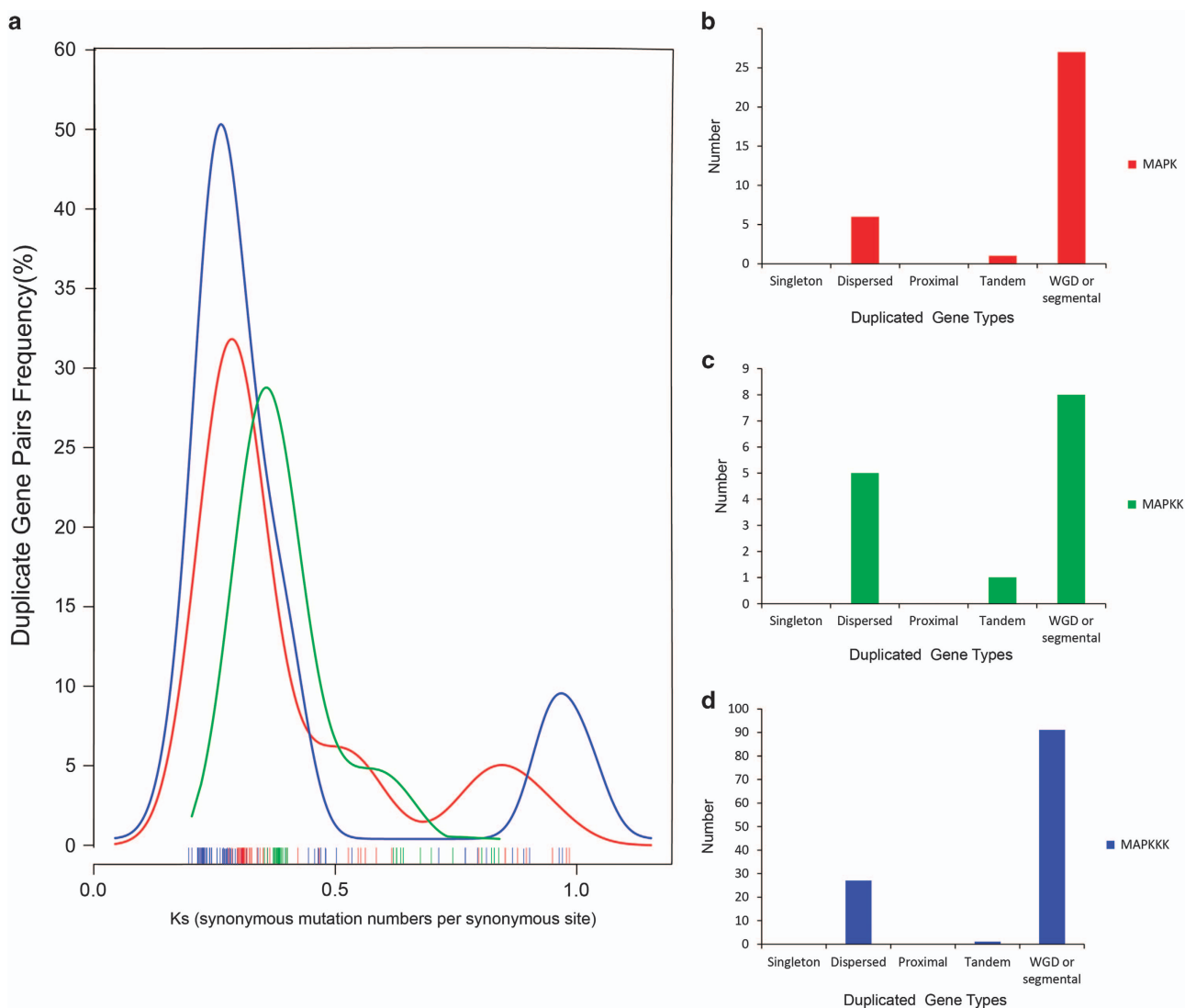


Figure 3. Pairwise comparison of the Ks values for MAPK cascade homologous genes in *B. rapa* and *A. thaliana* and their different duplicated types. (a) The distribution of Ks values for MAPK cascade genes between *B. rapa* and *A. thaliana*. (b–d) Singleton, dispersed, proximal, tandem and segmental duplicated types of MAPK cascade genes in *B. rapa*.

conserved motifs of BrMAPK proteins (Figure 2a). Eight of the 12 motifs were conserved in all the BrMAPK proteins, but the remaining 4 motifs (9–12) were not. In addition, each subfamily had its specific motif. For example, aside from the conserved motifs, MAPK proteins in classes I and II had specific motif 10 in their N-terminal region, whereas those in class IV contained motifs 8 and 9 in their C-terminal region (Figure 2). The annotated motifs from MEME revealed that eight of the 10 motifs (1–8) corresponded with the kinase domains in MAPKs (Supplementary Figure S1). Class II BrMAPK genes had no introns and all the BrMAPKK genes contained a protein kinase domain (Supplementary Figure S1).

The BrMAPKK genes were clustered into four distinct gene classes (I, II, III and IV) (Figure 2b), and varied in kinase domain presence or absence, adenosine triphosphate (ATP) binding site and serine/threonine protein kinase active site according to phylogenetic analysis (Supplementary Figure S1). The BrMAPKK class III and IV subfamily genes have no introns, except for BrMAPKK9, which had one intron (Supplementary Figure S1). The BrMAPKKK gene family in *B. rapa* was clustered into three subfamilies (MEKK, RAF and ZIK) (Figure 2c). All the BrMAPKKK proteins also had a protein kinase domain (Supplementary Figure S1c).

Duplication and Ks analysis for MAPK cascade genes in *A. thaliana* and *B. rapa*

We reconstructed 24 conserved ancestral genomic blocks (GBs) according to a previous report.⁵⁷ The color-coding of these blocks was based on their positions in a proposed ancestral karyotype (AK1–8).^{57,58} To investigate the BrMAPK cascade genes of chromosomal location, all of them were mapped onto ten chromosomes in *B. rapa*, except *BrZIK4*, which is located on Scaffold000203 (Supplementary Figure S2). Among the ten sets of chromosomes in *B. rapa*, chromosome 09 contained more BrMAPK cascade genes (30 genes) than any of the other chromosomes. Among the LF, MF1 and MF2 subgenomes, LF contained more MAPK cascade genes than the MF1 and MF2 subgenomes. The most BrMAPKKK genes (18%) clustered in the AK8 region, whereas for BrMAPK and BrMAPKK genes, the most (20%) clustered in AK1 region (Supplementary Figure S2).

Gene duplication was the driving force for gene families evolution.⁵⁹ We used the MCScanX software to detect the MAPK cascade genes duplication type. All of the BrMAPK cascade genes were traced to one of five duplication types (singleton, WGD,

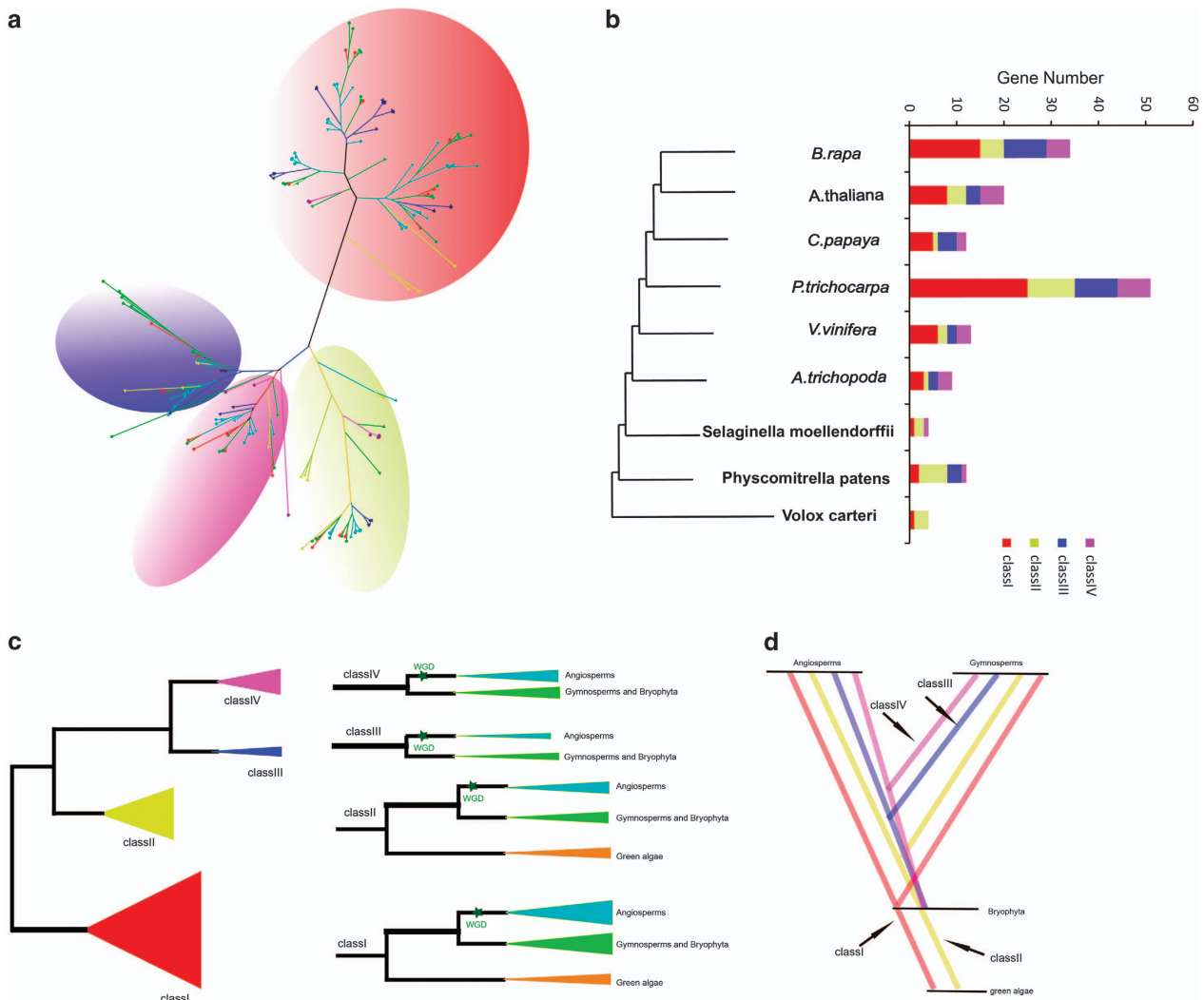


Figure 4. The duplication origin of MAPK genes in plants. **(a)** Phylogenetic relationships among MAPK genes. **(b)** Comparison of numbers of MAPK genes in representative species. **(c)** The phylogenetic relationships of MAPK genes. The green stars denote the inferred whole-genome duplication in angiosperms. **(d)** Schematic representation of the duplication history of MAPKs in plants. The red, yellow, blue and violet lines indicate class I, II, III and IV MAPK genes, respectively.

tandem, proximal or dispersed duplication types) (Figure 3). Twenty-seven BrMAPK, 8 BrMAPKK and 91 BrMAPKKK genes in *B. rapa* were duplicated from WGD or segmental events compared to only 1 BrMAPK, 1 BrMAPKK and 1 BrMAPKK genes from tandem duplication and 6 BrMAPK, 5 BrMAPKK and 27 BrMAPKK genes from dispersed duplication (Figure 3). The expansion of the BrMAPK cascade gene family was derived by the WGD or segmental duplication and dispersed gene duplication in *B. rapa*. The BrMAPKKK gene Ks values ranged from 0.2 to 0.5 and focused on about 0.25 (~8.5 Myr), whereas the BrMAPKK genes ranged from 0.4 to 0.7 and had a mean of about 0.40 (~12.5 Myr; Figure 3a and Supplementary Table S3). The BrMAPK gene Ks values ranged from 0.25 to 0.5 and centered ~0.3 (~10.5 Myr). The divergence time for the BrMAPKKK duplicated gene pairs was 8.49 MYA (Million years ago), which indicates that their divergence occurred during the *Brassica* triplication event (5–9 MYA). The BrMAPK and BrMAPKK duplicated gene happened 10 to 16.6 MYA and occurred during the divergence of *B. rapa* from *Arabidopsis* (9.6–16.1 MYA).

Expansion and evolution of MAPK genes in plants

We performed a phylogenetic tree of MAPK genes with additional eight key species in the tree of plant life to survey the evolutionary

relationships of the MAPK gene family in plants. We selected four angiosperms *Vitis vinifera*, *P. trichocarpa* and *C. papaya* because these species did not undergo α and β duplications, and *A. trichopoda* is a basal angiosperm that did not undergo the γ duplication event.^{60–62} A total of 204 MAPK genes were identified from the nine plant species (Supplementary Table S4; Figure 4a). All the MAPK genes segregated into four distinct classes (classes I, II, III and IV) with 100% bootstrap support (Supplementary Figure S3). The nucleotide distance between class I subfamily with class II, III and IV subfamilies was higher than those between MAPK classes II and III subfamilies, class II and IV subfamily, and classes III and IV subfamilies (Supplementary Figure S4). These findings indicated that class I had a distant phylogenetic relationship with the other three subfamilies, which is consistent with the results from the phylogenetic trees (Figure 4).

The gene numbers for the four classes in the MAPK gene family were marked in the nine plant species (Figure 4b). Only one class I and three class II subfamilies of MAPK genes were found in *Volvox carteri*, suggesting that MAPK gene expansion happened after the divergence of green algae. All these 9 species had conserved numbers (1–5) of class II MAPK genes, except *P. trichocarpa*, which had 10 MAPK genes. Based on the phylogenetic relationships, the

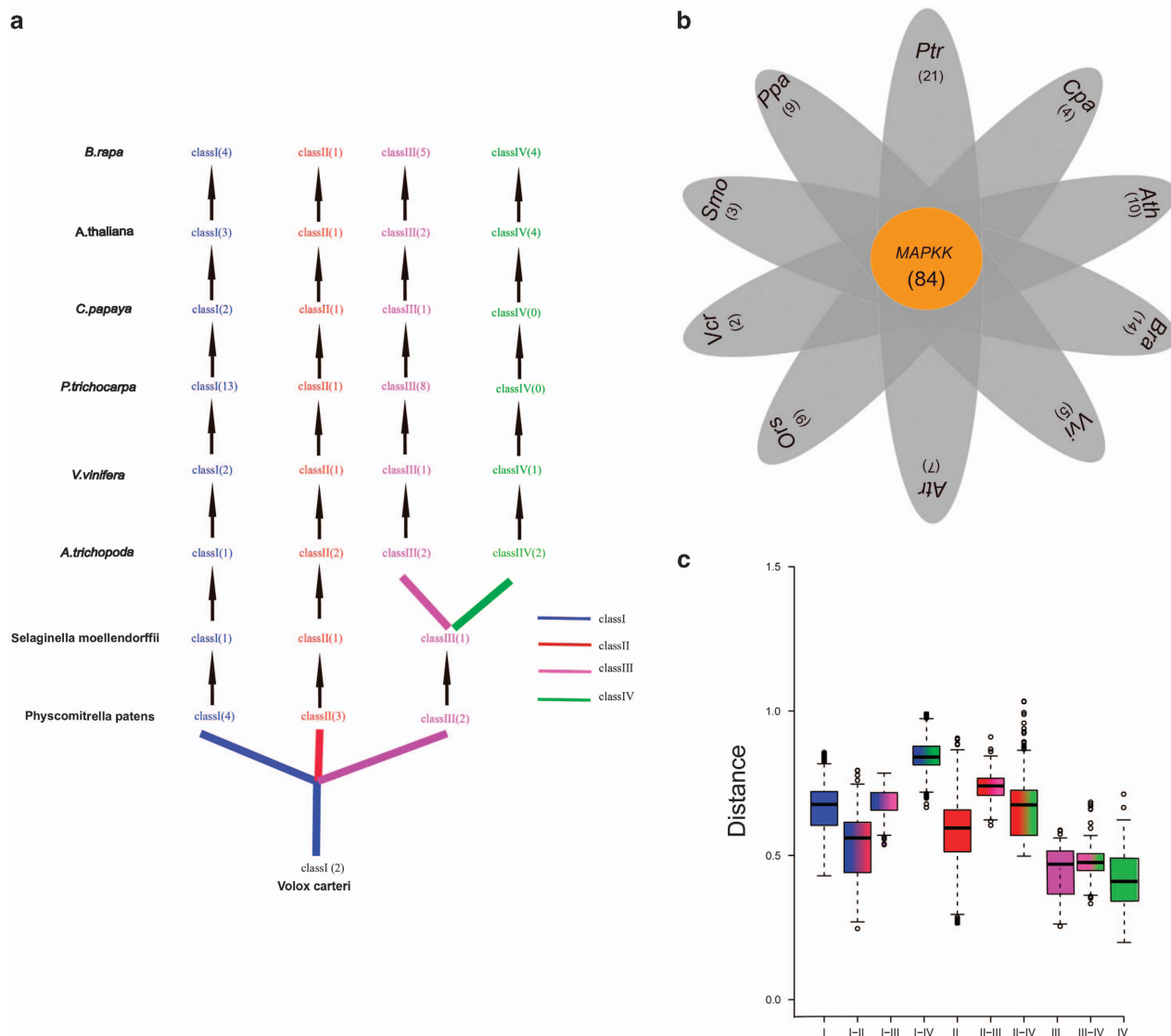


Figure 5. The analysis of MAPKK gene evolution. (a) Comparison of numbers of MAPKK genes in the phylogenetic tree representing plant species. (b) Venn diagram shows the number of common gene families and genes in 10 plants. (c) Genetic distance among the different classes of MAPKK genes.

class II subfamily of MAPK genes originated from green algae, whereas classes III and IV subfamilies originated from land plants (Figure 4c). Class II subfamily of MAPK genes diverged from Bryophyta class III and IV subfamilies, and that the rapid expansion of classes II, III and IV subfamilies may have primarily led to MAPK gene expansion in angiosperm plants. The evolution of MAPK genes was re-constructed and suggests that classes III and IV subfamilies originated from the class II subfamily, which was the initial group, and that the class III subfamily separated from the class IV subfamily in plants classified in the Bryophyta (Figure 4d).

Expansion and evolution of MAPKK genes in plants

Based on the phylogenetic trees for all the MAPKKs, we tried to reconstruct the expansion history of the MAPKK families. A total of 84 MAPKK genes were identified in the nine species, although only two MAPKK genes were detected in *V. carteri*. According to the phylogenetic tree topologies, the MAPKK genes could be grouped into four classes (classes I-IV) (Supplementary Figure S5). The MAPKK genes from *P. patens* and *S. moellendorffii* were clustered into only three MAPKK subfamilies included classes I, II and III, whereas all MAPKK subfamilies were present in the

angiosperms, which expanded rapidly, and was accompanied by several WGD (Supplementary Figure S6). The number of MAPKK genes in most angiosperms was greater than that in the other non-angiosperm species (*V. carteri* and *P. patens*), and was most likely due to WGDs, which leads to gene expansion (Figure 5). The number of MAPKK genes in class II was a highly conserved in the all species except *A. trichopoda* (2), *P. pteridophyta* (3) and *V. carteri* (no class II MAPKK genes were identified; Figures 5a and b). One class I MAPKK gene subfamily was found in *V. carteri* and suggested that MAPKK gene expansion occurred in land plants. However, no class II MAPKK genes were detected in *V. carteri* that indicated the class II gene subfamily may be unique to land plants.

The four MAPKK classes have similar domain architectures and motif compositions (Supplementary Figure S7). In addition to the domain compositions identified by pfam, 10 common motifs embedded in MAPKK gene domains were identified by MEME for all of the MAPKK genes. Class I subfamily of MAPKK genes had motif 8. Although the two MAPKK genes in *V. carteri* do not have motif 8, all the other species have similar motif compositions. From algae to angiosperm, the MAPKK gene family has highly conserved domains and motifs.

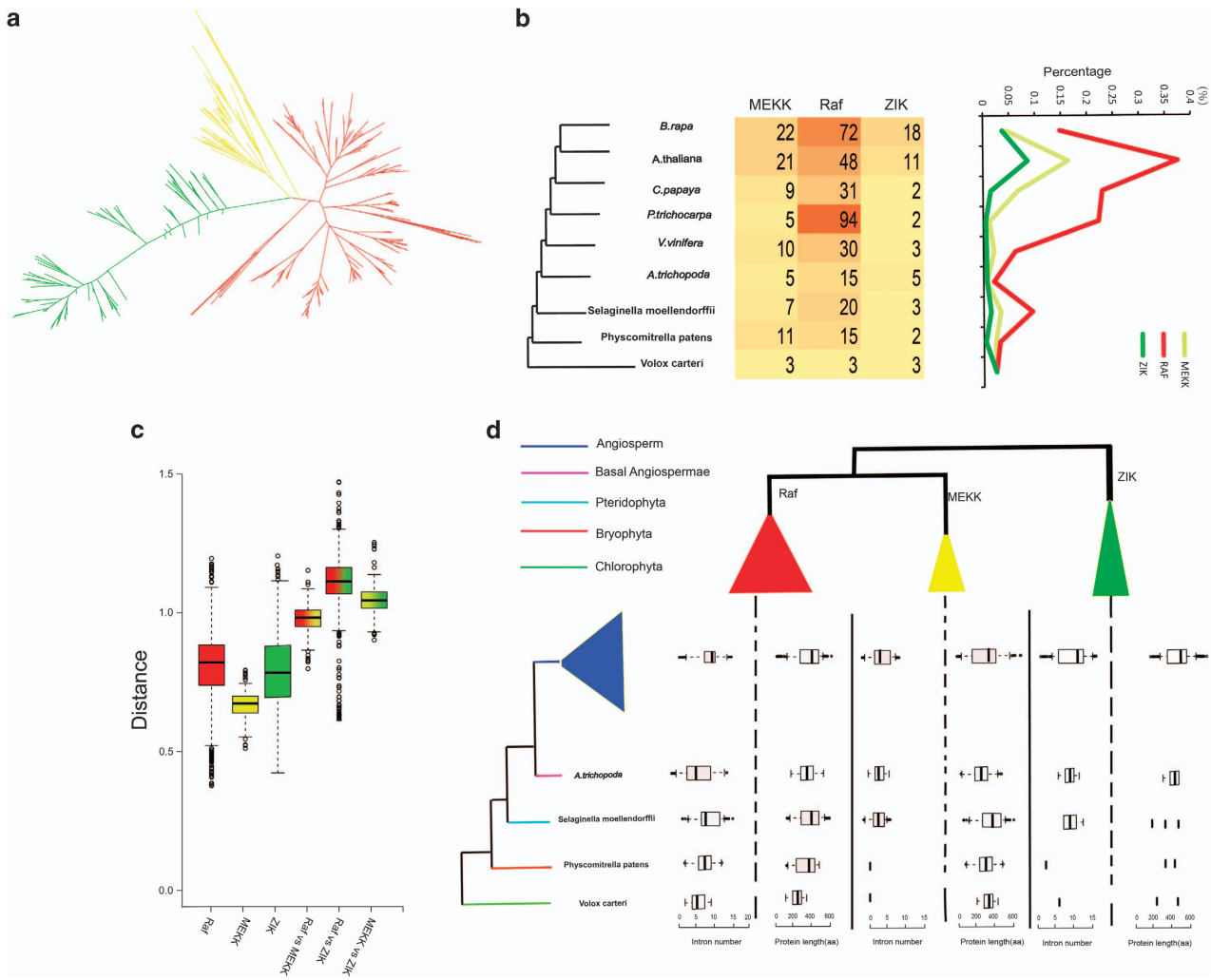


Figure 6. The analysis of MAPK cascade genes evolution. **(a)** Phylogenetic relationships among MAPK cascade genes. **(b)** The percentage and numbers of MAPK cascade genes in representative species. **(c)** Genetic distance among the different clades of MAPK cascade genes. **(d)** Distribution of structural characteristics of MAPK cascade genes in plant kingdom.

Phylogenetic trees were constructed in the same nine species to investigate the evolution of the MAPK cascade genes (Supplementary Figure S6). Genetic distances were analyzed to examine the relationships among the four classes (Figure 5c). The genetic distances for class III subfamily versus class IV subfamily were smaller than class III versus class II subfamily, and class III subfamily versus class I subfamily, which indicated that class III subfamily had a closer relationship with class IV subfamily. Moreover, the genetic distances for classes I, II, and III subfamilies were similar, whereas that of subfamily IV was the smallest, which suggested that the degree of sequence divergence for class IV subfamily was lower than the other subfamilies (Figure 5c). The evolutionary history of MAPKs was surmised that class I subfamily was the initial class from which MAPKs originated. Additionally, the class IV subfamily separated from class III subfamily in angiosperm plants.

Expansion and evolution of MAPK cascade genes in plants

We constructed a phylogenetic tree of the MAPK cascade genes that clustered into three subfamilies (MEKK, ZIK and Raf) (Figure 6a, Supplementary Figure S8). The MEKK, ZIK and Raf subfamily were detected in *V. carteri* (Figure 6b) that suggested they all originated from the green algae. Genetic distances were analyzed to further

determine the relationships among the three subfamilies (Figure 6c). The genetic distances between MEKK subfamily and the Raf subfamily were smaller than MEKK subfamily versus other subfamilies (Figure 6c). These results indicated that MEKK subfamily has a close relationship with Raf subfamily and that MEKK and Raf subfamilies may share a common evolutionary origin.

Analysis of MAPKs family size in nine plants indicated that it underwent rapid expansion during evolution, and further expansion in the Brassicaceae (Figure 6b). Phylogenetic trees were constructed for each species and the MAPK cascade family divided into three classes (MEKK, ZIK and Raf) (Supplementary Figure S8). The WGD had important role in the expansion and evolution of gene families in plant genomes, the rapidly increased gene numbers of the MAPK cascade family in *B. rapa* (18) and *A. thaliana* (11) (Figure 6b) gave a good example for that. Notably, the number of MAPKs in *P. trichocarpa* and *B. rapa* is more than other species, which was caused by salicoid duplication and Brassica-specific WGD events. During expansion of the MAPK cascade gene family, the Raf subfamily played a major role in this process.

We compared gene structure characteristics between algae and higher plants to better understand the evolutionary characteristics of MAPK cascade genes in the plant kingdom (Figure 6d). The gene structure of Raf was quite simple in lower species with most genes

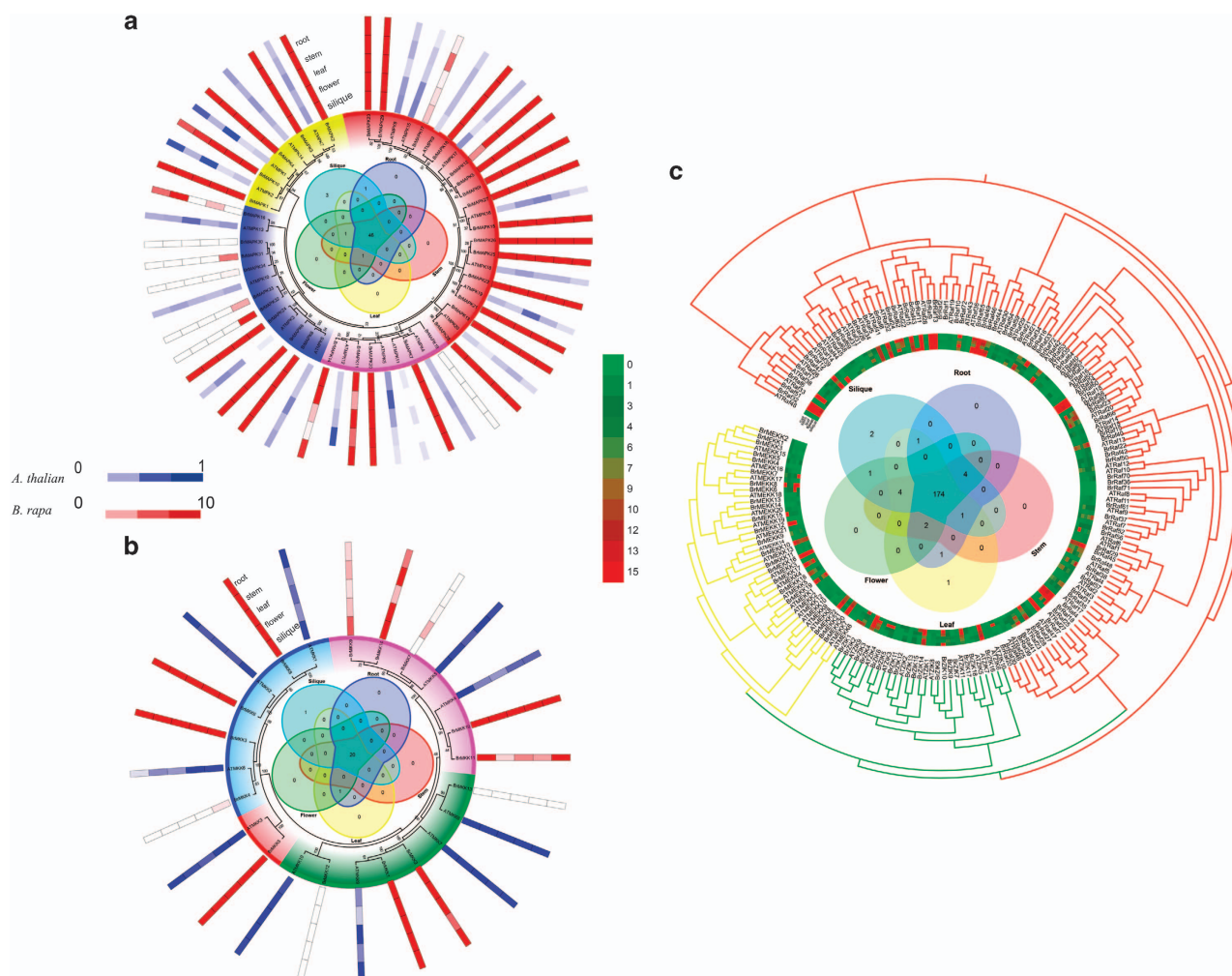


Figure 7. Expression pattern analysis of MAPK cascade genes in *A. thaliana* and *B. rapa*. Heat map representation and hierarchical clustering of MAPK cascade genes ((a) MAPK, (b) MAPKK, (c) MAPKKK) in root, stem, leaf, flower and silique. Venn diagram depicting the distribution of shared expression of the MAPK cascade genes among five *Brassica rapa* tissues.

having just five introns, whereas the number of introns increased to about 10 in the angiosperms (Figure 6d). The length of the protein has been stable at about 386 amino acids. MEKKs have no introns in *V. carteri* and *P. peridophyta* (Figure 6d).

Comparative expression patterns of the MAPK cascade genes in different tissues from *B. rapa* and *A. thaliana*

The MAPK cascade gene expression patterns between *A. thaliana* and *B. rapa* were investigated in roots, stems, leaves, flowers and siliques (callus only was studied in *B. rapa* and mature pollen studied in *A. thaliana*; Figure 7 and Supplementary Tables S5 and S6). Gene expression of BrMAPKs was detected in the roots, stems, leaves and flowers (28 BrMAPKs, 82%) and siliques (31 BrMAPKs, 91%; Figure 7a). All AtMAPKs showed expression (mean-normalized value > 1) in all tissues (Figure 7a). Gene expression of BrMAPKK and AtMAPKK was detected in at least one of the tissues, except BrMAPKK13 and BrMAPKK14 in which no expression was found in any of the tissues. Among the 196 BrMAPKKKs (including 84 AtMAPKKKs and 112 BrMAPKKKs), 5 (*BrRaf36*, *BrRaf70*, *BrRaf63*, *BrZIK11*, and *BrRaf71*) were not expressed, and 2 (*BrMEKK7* and *BrRaf21*) were expressed only in the flower tissue (Figure 7c and Supplementary Figure S9). Additionally, a total of 174 MAPKKKs had high expression levels (FPKM value > 10) in the five tissues. We also analyzed MAPK cascade genes for tissue-specific expression profiles in five tissues; some BrMAPK cascade

genes were highly expressed (FPKM value > 10) in flowers, including *BrMAPK32*, *BrMAPKK12*, and *BrMEKK7*, whereas they had no or low expression in the other tissues (Figures 7a and c).

MAPK cascade genes in the five tissues on the phylogenetic tree were selected to survey divergent functions of the homologous genes (Figure 7). AtMAPKKs Class IV subfamily showed high expression (mean-normalized value > 1) in the five tissues, except *BrMAPKK13* and *BrMAPKK14* were not expressed. Class III AtMAPKs also showed expression in all five tissues, except *BrMAPK30* and *BrMAPK34* were not expressed, and *BrMAPK31*, *BrMAPK32* and *BrMAPK33* were expressed only in the roots. These results indicated that class IV MAPKK and class III MAPK genes may have lost some of their functions after duplication events according to the expression patterns.

Cis-elements and interaction network analysis among MAPK cascade proteins in *B. rapa*

To analyze the *cis*-regulatory elements, we used the Plant *Cis*-acting Regulatory DNA Elements (PLACE) online tool to identify MAPK cascade genes in *B. rapa*. A total of 11 common *cis*-regulatory elements were detected in the promoter regions of the MAPK cascade genes and were highly conserved among all of the MAPK cascade genes studied in *B. rapa* (Figure 8). Three common *cis*-regulatory elements, ARBE, the CGTCA-motif, and the GARE-motif, were responsive to plant hormones, including ABA, JA and

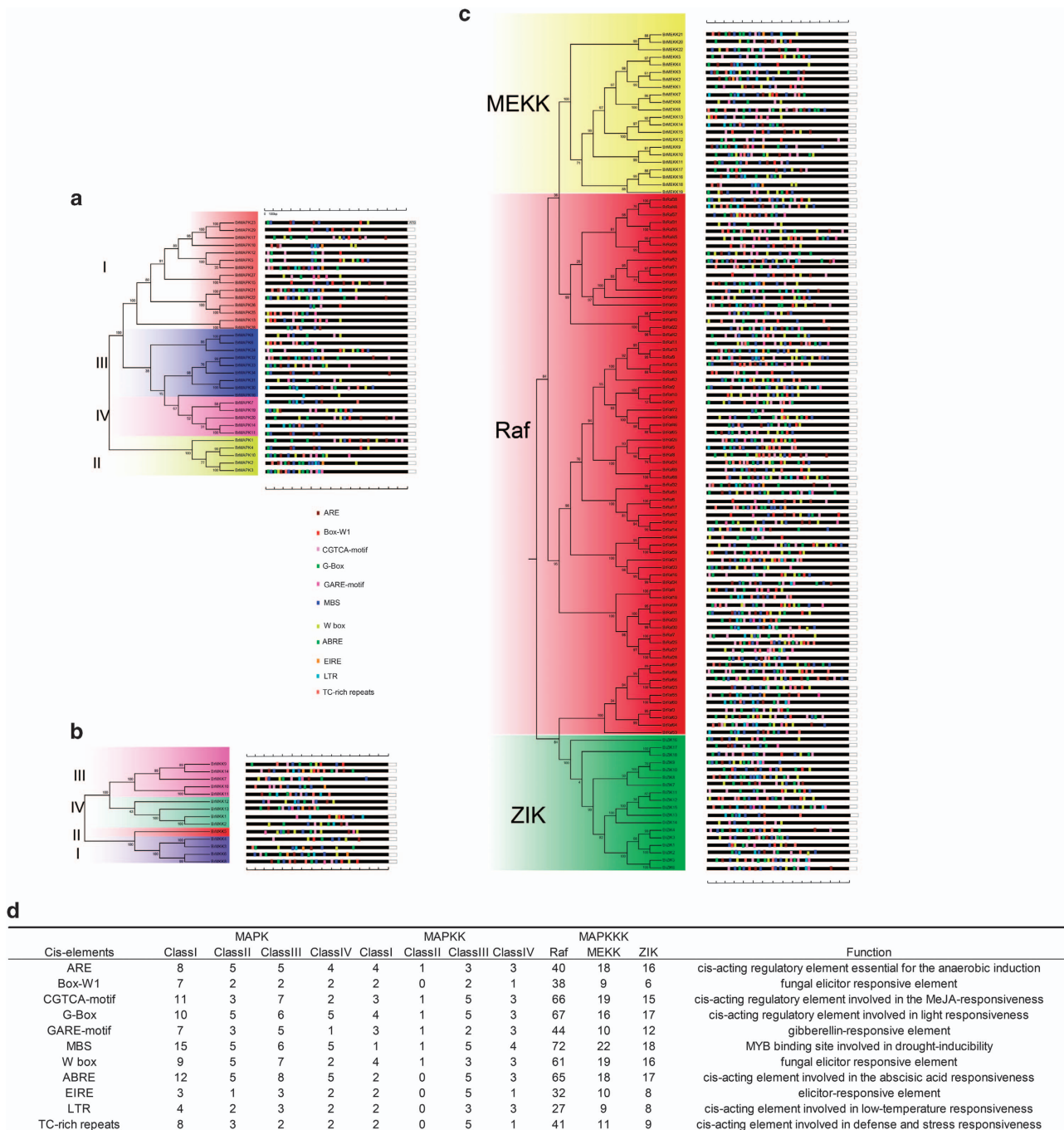


Figure 8. The *cis*-elements of the promoter regions of *B. rapa* MAPK cascade genes. The micro-segments in different colors were the putative elements sequences. The description of the ten *cis*-elements are in brackets.

GA, and suggested that they could affect MAPK cascade gene expression levels in *B. rapa*. Some common *cis*-regulatory elements responsive to both abiotic and biotic stresses were fungal elicitor-responsive elements (W-box and Box-W1), a light-responsive element (G-Box), low-temperature responsiveness (LTR), defense and stress responsiveness (TC-rich repeats), and a drought-responsive element (MBS). The diverse regulatory elements indicated the importance of MAPK cascade genes in stress tolerance. There were two more elements including a regulatory element essential for anaerobic induction (ARE) and an elicitor-responsive element (EIRE). In addition, we statistics the number of genes that contained *cis*-regulatory in each subfamily (Figure 8d). In the MAPK, most genes contained more MBS *cis*-regulatory

elements than other *cis*-regulatory elements. Only the class II subfamily of MAPKK had six *cis*-regulatory elements. In MAPKKK, most genes also contained *cis*-regulatory elements of MBS, which is similar to MAPK genes. These results indicated that MAPK cascade genes were responsive to various stresses, which may be due to the binding of corresponding *cis*-elements that regulate the expression of MAPK cascade genes.

The interaction network of BrMAPK cascade proteins was constructed by STRING software (Supplementary Figure S10). The lines stand for the positive correlations that were retrieved from STRING database corresponding to the protein interactions. The interaction network contained 42 nodes and 167 pairs of interacting genes. It showed a very complicated correlation

among the BrMAPK cascade proteins (Supplementary Figure S10). This phenomenon indicated that MAPK cascade genes are involved in many fundamental mechanisms and regulate many downstream factors and/or are regulated by many upstream genes.

DISCUSSION

MAPK cascade genes have been shown to be involved in growth, development and responses to a variety of stress stimuli. The MAPK signaling pathway is one of the most important and conserved ubiquitous modules for signal transduction in plants^{14,20,21,63,64} (Supplementary Figure S11).

If genes products are involved in complex regulatory networks, via different metabolic pathways or transcriptional regulation, they are likely to be retained.^{41,65} In this study, we identified 34 BrMAPK, 15BrMAPKK and 112 BrMAPKKK genes in the *B. rapa* genome that displayed high retention after the WGD. More MAPK (59%) and MAPKKK (46%) genes retained two or three copies in *B. rapa* than MAPKK (40%) genes. By analyzing the duplicated types, we found more segmental duplication in BrMAPKs (79.4%) and BrMAPKKKs (81.3%) than in BrMAPKKs (53.3%). BrMAPKs and BrMAPKKKs that participate in development, stress and defense, and play particularly important roles in signaling pathways. This result is agreement with the gene dosage hypothesis.⁶⁶ Comparison of nucleotide distances showed that BrMAPKs and BrMAPKKKs diverged approximately 8.5 MYA, earlier than BrMAPKKs that separated at 12.5 MYA.

During evolutionary, angiosperm genomes have experienced polyploidization events, which have led to the MAPK cascade gene expansion, followed by divergence by (1) subfunctionalization, (2) neofunctionalization, or (3) nonfunctionalization.^{62,67–69} These fates have provided choices for duplicated genes to gain functional diversification.

We determined the evolutionary histories of the MAPK cascade genes through comparison analyses that examined expansion, phylogenetic relationships, characteristic structures and nucleotide distances. The MAPK class II subfamily had a close relationship with the classes III and IV subfamilies. We constructed a MAPK evolutionary model whereby classes III and IV subfamilies might have originated from class II, which was the ancestral group, followed by class III subfamily separating from class IV subfamily in Bryophyta (Figure 4). At present, MAPK cascade gene family members have been reported from a few plant species, including

rice (*Oryza sativa*),⁷⁰ Arabidopsis (*A. thaliana*),⁷¹ maize (*Zea mays*),⁷² Brachypodium distachyon,⁷³ canola (*B. napus*),⁷⁴ apple (*Malus domestica*)⁷⁵ and grape (*V. vinifera*).⁷⁶ MAPK genes had conserved amino acids sequences that reflects their common evolutionary ancestry.⁷⁷ Before the divergence of the monocots and eudicots, the plant MPK gene family structure already was established from an ancient pattern of diversification.⁷⁷ MAPK cascades play central roles in the eukaryotic signal transduction networks, and they have highly conserved component kinases during evolution.⁷⁸

In flowering plants, MAPK genes number was twice that of MAPKKs.³⁷ The MAPKK gene family has retained highly conserved domains and motifs from algae to angiosperms (Supplementary Figure S7). In their evolutionary, class II and III subfamilies of MAPKK genes might have originated from class I subfamily, which was the ancestral class, followed by class IV subfamily separating from class III subfamily in angiosperms. The MAPKK genes in *Arabidopsis* and *B. rapa* possessed similar gene complements and experienced several recent gene duplication events. After comparing the MAPKK genes in *V. vinifera* and *B. rapa*, we found that *B. rapa* had a relatively large number of MAPKK genes compared to *V. vinifera*. The reason might be that *V. vinifera* had not undergone a α and β duplication events. The MAPKK gene family formed four distinct subfamilies in the phylogenetic tree, which is consistent with the result for *B. rapa* and *A. thaliana*.³⁷

In the three subfamilies in the MAPKKK family, MEKK has a closer relationship to Raf than MEKK to ZIK, and plant MEKKs and Rafs may share a common evolutionary origin. Moreover, the increase of Rafs played a major role in the expansion of the MAPKKK gene family. An evolutionary history of MAPK cascade gene family was inferred in the plant kingdom; all three gene families exist in the genomes from Bryophyta to Angiospermae (Figure 9, Supplementary Figure S12).

Many of the MAPK pathway members have been characterized according to their response to various environmental stimuli in plants.⁶³ The MEKK1-MPK4 cascade plays an important role in ROS (reactive oxygen species) metabolism,⁷⁹ and MKK1-MPK6 cascade is involved in H₂O₂ metabolism.³³ A TIPK (*Trichoderma*-induced MAPK) was reported to response to pathogen,⁸⁰ whereas CsNMAPK regulates NO₃⁻ stress,²⁷ and salt stress.²⁸ Plant MAPKs play roles in different hormones pathways,⁸¹ such as MKK3/MPK6 participate in jasmonic acid (JA) signaling.⁸² MAPKs were involved in abscisic acid (ABA) signaling pathway.³³ To better understand the possible transcriptional regulation functions of MAPK cascade

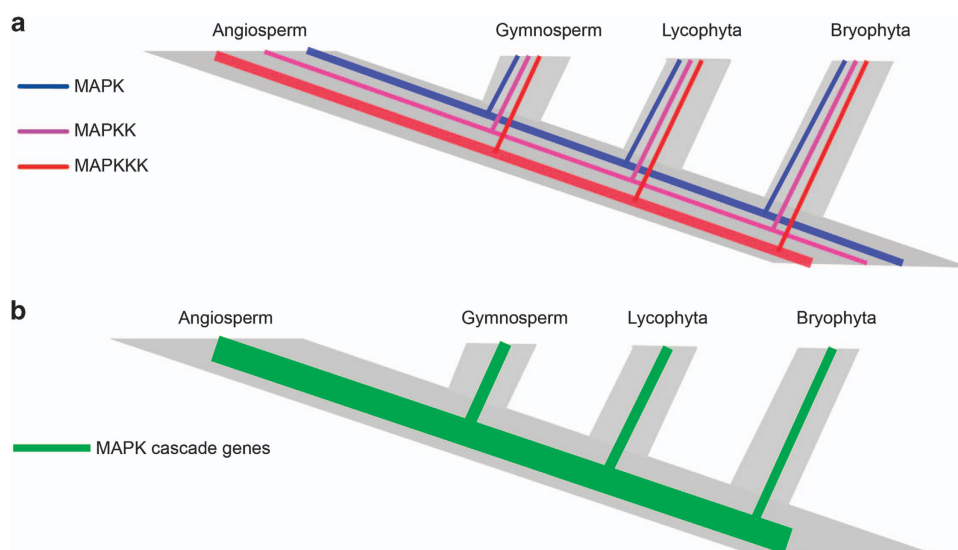


Figure 9. Evolutionary history of the MAPK cascade genes in the plant kingdom.

genes in *B. rapa*, we identified common conserved *cis*-regulatory elements in the promoter regions of MAPK cascade genes. We also found that 11 common *cis*-elements in all the MAPK cascade genes of *B. rapa* were responsive to plant hormones and environmental stresses.

CONCLUSIONS

In summary, genome-wide analysis of MAPK cascade gene families was first performed in *B. rapa*, and 34 BrMAPK, 14 BrMAPKK and 112 BrMAPKKK genes were identified. We analyzed the characteristic structures, evolutionary patterns, gene duplication, expression divergence and *cis*-regulatory elements of MAPK cascade genes in plants. Comparative evolutionary analysis from eight other plants was also conducted in the tree of plant life. Our study provides new insights into the evolution of the MAPK cascade gene family in *B. rapa* and the plant kingdom in general. We further confirmed the significant regulation of the MAPK cascade genes in *B. rapa* during growth, development and stress tolerance. Our study provides useful information regarding the MAPK cascade genes functional divergence and conservation, and also aids in understanding the effects of polyploidy during evolution with regards to MAPK cascade genes.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the key project of National Natural Science Foundation of China (No. 31330067), the National Natural Science Foundation of China (No. 31471886), and National vegetable industry technology system (CARS-23-A-06), and Jiangsu modern agriculture (vegetable) industrial technology system (SXGC[2017] 273).

AUTHOR CONTRIBUTIONS

PW conceived and designed the experimental design. YL and WW contributed to the experimental work. XH contributed to the interpretation of the results and coordinated the study. All the authors read and approved the final manuscript.

REFERENCES

- Cardinale F, Meskiene I, Ouaked F, Hirt H. Convergence and divergence of stress-induced mitogen-activated protein kinase signaling pathways at the level of two distinct mitogen-activated protein kinase cascades. *Plant Cell* 2002; **14**: 703–711.
- Garrington TP, Johnson GL. Organization and regulation of mitogen-activated protein kinase signaling pathways. *Curr Opin Cell Biol* 1999; **11**: 211–218.
- Bethke G *et al.* Flg22 regulates the release of an ethylene response factor substrate from MAP kinase 6 in *Arabidopsis thaliana* via ethylene signaling. *Proc Natl Acad Sci USA* 2009; **106**: 8067–8072.
- Fiil BK, Petersen K, Petersen M, Mundy J. Gene regulation by MAP kinase cascades. *Curr Opin Plant Biol* 2009; **12**: 615–621.
- Stefan M, Lynn UZ, Geetha G, Stefan B, Baldwin IT. MAPK-dependent JA and SA signalling in *Nicotiana attenuata* affects plant growth and fitness during competition with conspecifics. *BMC Plant Biol* 2012; **12**: 213–213.
- Popescu SC, Popescu GV, Bachan S *et al.* MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes Dev* 2009; **23**: 80–92.
- Sasabe M, Soyano T, Takahashi Y *et al.* Phosphorylation of NtMAP65-1 by a MAP kinase down-regulates its activity of microtubule bundling and stimulates progression of cytokinesis of tobacco cells. *Genes Dev* 2006; **20**: 1004–1014.
- Shen H, Liu C, Zhang Y *et al.* OsWRKY30 is activated by MAP kinases to confer drought tolerance in rice. *Plant Mol Biol* 2012; **80**: 241–253.
- Yoo SD, Cho YH, Tena G, Xiong L, Sheen J. Dual control of nuclear EIN3 by bifurcate MAPK cascades in C₂H₄ signalling. *Nature* 2008; **451**: 789–795.
- Fryxell KJ. The coevolution of gene family trees. *Trends Genet* 1996; **12**: 364–369.
- Ligterink W, Hirt H. Mitogen-activated protein [MAP] kinase pathways in plants: versatile signaling tools. *Int Rev Cytol* 2001; **201**: 209–275.

- Bögre L. Complexity, cross talk and integration of plant MAP kinase signalling. *Curr Opin Plant Biol* 2002; **5**: 415–424.
- Hirt H. A new blueprint for plant pathogen resistance. *Nat Biotechnol* 2002; **20**: 450–451.
- Seo S, Okamoto M, Seto H, Ishizuka K, Sano H, Ohashi Y. Tobacco MAP kinase: a possible mediator in wound signal transduction pathways. *Science* 1995; **270**: 1988–1992.
- Bögre L, Hirt H. Wounding induces the rapid and transient activation of a specific MAP kinase pathway. *Plant Cell* 1997; **9**: 75–83.
- Mizoguchi T, Irie K, Hirayama T *et al.* A gene encoding a mitogen-activated protein kinase kinase is induced simultaneously with genes for a mitogen-activated protein kinase and an S6 ribosomal protein kinase by touch, cold, and water stress in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 1996; **93**: 765–769.
- Morris PC, Ellis BE, Ecker JR *et al.* Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci* 2002; **7**: 301–308.
- Morris PC, Ellis BE, Ecker JR *et al.* *Arabidopsis* MAP kinase 4 negatively regulates systemic acquired resistance. *Cell* 2000; **103**: 1111–1120.
- Kovtun Y, Wanling C, Tena G, Sheen J. Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc Natl Acad Sci USA* 2000; **97**: 2940–2945.
- Ichimura K, Mizoguchi T, Yoshida R, Yuasa T, Shinozaki K. Various abiotic stresses rapidly activate *Arabidopsis* MAP kinases ATMPK4 and ATMPK6. *Plant J Cell Mol Biol* 2000; **24**: 655–665.
- Mikolajczyk M, Awotunde OS, Muszyńska G, Klessig DF, Dobrowolska G. Osmotic stress induces rapid activation of a salicylic acid-induced protein kinase and a homolog of protein kinase ASK1 in tobacco cells. *Plant Cell* 2000; **12**: 165–178.
- Schaffer R, Wisman E. Microarray Analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell* 2001; **13**: 113–123.
- Ren D, Yang H, Zhang S. Cell Death Mediated by MAPK Is associated with hydrogen peroxide production in *Arabidopsis*. *J Biol Chem* 2002; **277**: 559–565.
- Xu J, Li Y, Wang Y *et al.* Activation of MAPK kinase 9 induces ethylene and camalexin biosynthesis and enhances sensitivity to salt stress in *Arabidopsis*. *J Biol Chem* 2008; **283**: 26996–27006.
- Matsuoka D, Nanmori T, Sato KI, Fukami Y, Kikkawa U, Yasuda T. Activation of AtMEK1, an *Arabidopsis* mitogen-activated protein kinase kinase, in vitro and in vivo: analysis of active mutants expressed in *E. coli* and generation of the active form in stress response in seedlings. *Plant J* 2002; **29**: 637–647.
- Teige M, Scheikl E, Eulgem T *et al.* The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Mol Cell* 2004; **15**: 141–152.
- Xu H, Wang X, Sun X, Shi Q, Yang F, Du D. Molecular cloning and characterization of a cucumber MAP kinase gene in response to excess NO₃⁻ and other abiotic stresses. *Sci Hort* 2008; **117**: 1–8.
- Xu H, Sun X, Wang X, Shi Q, Yang X, Yang F. Involvement of a cucumber MAPK gene (CSNMAPK) in positive regulation of ROS scavenging and osmotic adjustment under salt stress. *Sci Hort* 2011; **127**: 488–493.
- Ning J, Li X, Hicks L, Xiong L. A Raf-like MAPKKK gene DSM1 mediates drought resistance through reactive oxygen species scavenging in rice. *Plant Physiol* 2010; **152**: 876–890.
- Shou H, Bordallo P, Wang K. Expression of the Nicotiana protein kinase (NPK1) enhanced drought tolerance in transgenic maize. *J Exp Bot* 2004; **55**: 1013–1019.
- Berri S, Colcombet J. Constitutively active mitogen-activated protein kinase versions reveal functions of *Arabidopsis* MPK4 in pathogen defense signaling. *Plant Cell* 2012; **24**: 4281–4293.
- Gao M, Liu J, Bi D *et al.* MEK1, MKK1/MKK2 and MPK4 function together in a mitogen-activated protein kinase cascade to regulate innate immunity in plants. *Cell Res* 2008; **18**: 1190–1198.
- Xing Y, Jia W, Zhang J. AtMKK1 mediates ABA-induced CAT1 expression and H₂O₂ production via AtMPK6-coupled signaling in *Arabidopsis*. *Plant J* 2008; **54**: 440–451.
- Asai T, Tena G, Plotnikova J *et al.* MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* 2002; **415**: 977–983.
- Beck M, Komis G, Ziemann A, Menzel D, Samaj J. Mitogen-activated protein kinase 4 is involved in the regulation of mitotic and cytokinetic microtubule transitions in *Arabidopsis thaliana*. *New Phytol* 2011; **189**: 1069–1083.
- Takahashi Y, Soyano T, Kosetsu K, Sasabe M, Machida Y. kinesin, ANP MAPKKs and MKK6/ANQ MAPKK, which phosphorylates and activates MPK4 MAPK, constitute a pathway that is required for cytokinesis in *Arabidopsis thaliana*. *Plant Cell Physiol* 2010; **51**: 1766–1776.
- Dóczi R, Okrészl L, Romero AE, Paccanaro A, Bögre L. Exploring the evolutionary path of plant MAPK networks. *Trends Plant Sci* 2012; **17**: 518–525.
- Feng K, Liu F, Zou J *et al.* Genome-wide identification, evolution, and co-expression network analysis of mitogen-activated protein kinase kinases in *Brachypodium distachyon*. *Front Plant Sci* 2016; **7**: 1400.

- 39 Caroline D, Matton DP. Genome-wide analysis of MAPKKs shows expansion and evolution of a new MEKK class involved in solanaceous species sexual reproduction. *BMC Genomics* 2015; **16**: 1037.
- 40 Moghe GD, Shiu SH. The causes and molecular consequences of polyploidy in flowering plants. *Ann N Y Acad Sci* 2014; **1320**: 16–34.
- 41 Birchler JA, Veitia RA. The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 2007; **19**: 395–402.
- 42 Wang X, Wang H, Wang J et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 2011; **43**: 1035–1039.
- 43 Bhattacharya D, Medlin L. Algal phylogeny and the origin of land plants. *Plant Physiol* 1998; **116**: 9–15.
- 44 Rensing SA, Lang D, Zimmer AD et al. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 2008; **319**: 64–69.
- 45 Banks JA, Nishiyama T, Hasebe T et al. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* 2011; **332**: 960–963.
- 46 Kawahara Y, de la Bastide M, Hamilton JP et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 2013; **6**: 4.
- 47 Goodstein DM, Shu S, Howson R et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 2012; **40**: D1178–D1186.
- 48 Wang Y, Tang H, DeBarry JD et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 2012; **40**: e49.
- 49 Krzywinski M, Schein J, Birol I et al. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009; **19**: 1639–1645.
- 50 Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P. iPath2. 0: interactive pathway explorer. *Nucleic Acids Res* 2011; **39**: W412–W415.
- 51 Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**: 1792–1797.
- 52 Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 2006; **4**: 259–263.
- 53 Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comput Graph Stat* 1996; **5**: 299–314.
- 54 Koch MA, Haubold B, Mitchell-Olds T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (*Brassicaceae*). *Mol Biol Evol* 2000; **17**: 1483–1498.
- 55 Tong C, Wang X, Yu J et al. Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. *BMC Genomics* 2013; **14**: 689.
- 56 Schmid M, Davison TS, Henz SR et al. A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 2005; **37**: 501–506.
- 57 Cheng F, Mandáková T, Wu J, Xie Q, Lysak MA, Wang X. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *Plant Cell* 2013; **25**: 1541–1554.
- 58 Schranz ME, Lysak MA, Mitchell-Olds T. The ABC's of comparative genomics in the *Brassicaceae*: building blocks of crucifer genomes. *Trends Plant Sci* 2006; **11**: 535–542.
- 59 Maher C, Stein L, Ware D. Evolution of *Arabidopsis* microRNA families through duplication events. *Genome Res* 2006; **16**: 510–519.
- 60 Lee T-H, Tang H, Wang X, Paterson AH. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res* 2013; **41**: D1152–D1158.
- 61 Albert VA, Barbazuk WB, Der JP et al. The *Amborella* genome and the evolution of flowering plants. *Science* 2013; **342**: 1241089.
- 62 Jiao Y, Wickett NJ, Ayyampalayam S et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* 2011; **473**: 97–100.
- 63 Pitzschke A, Schikora A, Hirt H. MAPK cascade signalling networks in plant defence. *Curr Opin Plant Biol* 2009; **12**: 421–426.
- 64 Yu Q, Chen C, Du D et al. Reprogramming of a defense signaling pathway in rough lemon and sweet orange is a critical element of the early response to 'Candidatus Liberibacter asiaticus'. *Hort Res* 2017; **4**: 17063.
- 65 Aad G, Abajyan T, Abbott B et al. Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC. *Physics Letters B* 2012; **716**: 1–29.
- 66 Birchler JA, Veitia RA. The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 2007; **19**: 395–402.
- 67 Ohno S. *Evolution by Gene Duplication*. Springer Science & Business Media: Berlin Heidelberg. 2013.
- 68 Edger PP, Pires JC. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res* 2009; **17**: 699–717.
- 69 Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* 2010; **11**: 97.
- 70 Reyna NS, Yang Y. Molecular Analysis of the rice MAP kinase gene family in relation to magnaporthe grisea infection. *Mol Plant Microbe Interact* 2006; **19**: 530–540.
- 71 Group M. Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci* 2002; **7**: 301.
- 72 Liu Y, Zhang D, Wang L, Li D. Genome-wide analysis of mitogen-activated protein kinase gene family in maize. *Plant Mol Biol Reporter* 2013; **31**: 1446–1460.
- 73 Chen L, Zhang Y, He Y et al. Genome-wide identification and analysis of MAPK and MAPKK gene families in *Brachypodium distachyon*. *PLoS ONE* 2012; **7**: e46744.
- 74 Liang W, Bo Y, Yu BJ et al. Identification and analysis of MKK and MPK gene families in canola (*Brassica napus* L.). *BMC Genomics* 2013; **14**: 392.
- 75 Zhang S, Xu R, Luo X, Jiang Z, Shu H. Genome-wide identification and expression analysis of MAPK and MAPKK gene family in *Malus domestica*. *Gene* 2013; **531**: 377–387.
- 76 Wang G, Lovato A, Polverari A et al. Genome-wide identification and analysis of mitogen activated protein kinase kinase gene family in grapevine (*Vitis vinifera*). *BMC Plant Biol* 2014; **14**: 219.
- 77 Hamel LP, Nicole MC, Sritubtim S et al. Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci* 2006; **11**: 192–198.
- 78 Widmann C, Gibson S, Jarpe MB, Johnson GL. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol Rev* 1999; **79**: 143–180.
- 79 Nakagami H, Soukupová H, Schikora A, Zárský V, Hirt H. A Mitogen-activated protein kinase kinase mediates reactive oxygen species homeostasis in *Arabidopsis*. *J Biol Chem* 2006; **281**: 38697–38704.
- 80 Shores M, Gal-On A, Leibman D, Chet I. Characterization of a mitogen-activated protein kinase gene from cucumber required for trichoderma-conferred plant resistance. *Plant Physiol* 2006; **142**: 1169–1179.
- 81 Rodríguez MCS, Petersen M, Mundy J. Mitogen-activated protein kinase signaling in plants. *Annu Rev Plant Biol* 2010; **61**: 621–649.
- 82 Takahashi F, Yoshida R, Ichimura K et al. The mitogen-activated protein kinase cascade MKK3-MPK6 is an important part of the jasmonate signal transduction pathway in *Arabidopsis*. *Plant Cell* 2007; **19**: 805–818.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

Supplementary Information for this article can be found on the *Horticulture Research* website (<http://www.nature.com/hortres>)