

Genome-Wide Identification and Analysis of MAPK and MAPKK Gene Families in *Brachypodium distachyon*

Lihong Chen¹*, Wei Hu¹*, Shenglong Tan², Min Wang¹, Zhanbing Ma¹, Shiyi Zhou¹, Xiaomin Deng¹, Yang Zhang¹, Chao Huang¹, Guangxiao Yang^{1*}, Guangyuan He^{1*}

1 The Genetic Engineering International Cooperation Base of Chinese Ministry of Science and Technology, Chinese National Center of Plant Gene Research (Wuhan) HUST Part, College of Life Science and Technology, Huazhong University of Science & Technology (HUST), Wuhan, China, **2** Services Computing Technology and System Laboratory, Cluster and Grid Computing Laboratory, School of Computer Science and Technology, Huazhong University of Science & Technology (HUST), Wuhan, China

Abstract

MAPK cascades are universal signal transduction modules and play important roles in plant growth, development and in response to a variety of biotic and abiotic stresses. Although MAPKs and MAPKKs have been systematically investigated in several plant species including *Arabidopsis*, rice and poplar, no systematic analysis has been conducted in the emerging monocot model plant *Brachypodium distachyon*. In the present study, a total of 16 MAPK genes and 12 MAPKK genes were identified from *B. distachyon*. An analysis of the genomic evolution showed that both tandem and segment duplications contributed significantly to the expansion of MAPK and MAPKK families. Evolutionary relationships within subfamilies were supported by exon-intron organizations and the architectures of conserved protein motifs. Synteny analysis between *B. distachyon* and the other two plant species of rice and *Arabidopsis* showed that only one homolog of *B. distachyon* MAPKs was found in the corresponding syntenic blocks of *Arabidopsis*, while 13 homologs of *B. distachyon* MAPKs and MAPKKs were found in that of rice, which was consistent with the speciation process of the three species. In addition, several interactive protein pairs between the two families in *B. distachyon* were found through yeast two hybrid assay, whereas their orthologs of a pair in *Arabidopsis* and other plant species were not found to interact with each other. Finally, expression studies of closely related family members among *B. distachyon*, *Arabidopsis* and rice showed that even recently duplicated representatives may fulfill different functions and be involved in different signal pathways. Taken together, our data would provide a foundation for evolutionary and functional characterization of MAPK and MAPKK gene families in *B. distachyon* and other plant species to unravel their biological roles.

Citation: Chen L, Hu W, Tan S, Wang M, Ma Z, et al. (2012) Genome-Wide Identification and Analysis of MAPK and MAPKK Gene Families in *Brachypodium distachyon*. PLoS ONE 7(10): e46744. doi:10.1371/journal.pone.0046744

Editor: Keqiang Wu, National Taiwan University, Taiwan

Received: August 2, 2012; **Accepted:** September 4, 2012; **Published:** October 17, 2012

Copyright: © 2012 Chen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was financially supported by the Key Projects of International Cooperation of Chinese Ministry of Science and Technology (2009DFB20290 and 2009DFB30340, <http://www.cistc.gov.cn/englishversion/>), National Project of Oceanic and Public welfare Industry of China (201005013, <http://www.most.gov.cn/eng/programmes1/>) and the Chinese State Council Special Major Programs (2009ZX08016-001A, <http://www.most.gov.cn/eng/programmes1/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: hegy@mail.hust.edu.cn (GH); ygx@mail.hust.edu.cn (GY)

† These authors contributed equally to this work.

Introduction

The mitogen activated protein kinases (MAPK) pathway is conserved in evolution through the plant and animal kingdoms. These proteins have been implicated in diverse cellular processes including cell growth, proliferation, differentiation, survival, development and in responses to a diversity of environmental stimuli including cold, heat, reactive oxygen species, UV, drought and pathogen attack [1–4]. MAPK cascades are composed of three classes of protein kinases: MAPK kinase kinase (MAPKKKs/MEKKs), MAPK kinase (MAPKKs/MKKs) and MAPK. MAPKKKs are serine/threonine kinases phosphorylating two amino acids in the S/T-X3-5-S/T motif of the MAPKK activation loop. MAPKKs are dual-specificity kinases that activate a MAPK through double phosphorylation of the T-X-Y motif in the activation loop (T-loop). The activated MAPK leads to the phosphorylation of transcription factors and other signaling components that regulate the expression of downstream genes [5].

To date, there have been only a few MAPK cascade components that have been studied in detail in plants [4] and the most extensively studied MAPKs are MPK3, MPK4 and MPK6 in *Arabidopsis* and crop plants [5]. The *Arabidopsis* MKK1/MKK2-MPK4/MPK6 cascades have previously been shown to play important roles in the responses to salt and cold stresses and pathogen attack [6–9]. The *Arabidopsis* MKK3 participates in signaling cascades elicited by pathogen infection [10]. The *Arabidopsis* MKK4/MKK5-MPK3/MPK6 cascades play important roles in the regulation of biotic stress [11–12]. The *Arabidopsis* MPK4 and the rice OsMAPK3 negatively regulate biotic stress signaling [4,13]. In addition, several MAPK genes have been found to play an important role in the regulation of developmental processes [11,14–16]. A tobacco MAP Kinase p45^{nif4} was identified and proven to be involved in pollen maturation and germination and be activated by hydration during pollen germination [17]. The *Arabidopsis* MKK1-MPK3/MPK6 cascades are key players in stomatal development and stomatal dynamics [12,18–19]. The *Arabidopsis* MKK9-MPK3/MPK6 cascades reg-

ulate ethylene signaling and camalexin biosynthesis and may also play a role in leaf senescence [20–22]. The *Arabidopsis* MKK6-MPK4/MPK11 cascades directly regulate cytokinesis and mitosis [23–26].

So far, a large number of members of MAPK cascades from *Arabidopsis* and rice have been identified using functional genomic approach. Twenty MAPKs, 10 MAPKK and 80 MAPKKKs are present in the *Arabidopsis thaliana* genome and 15 MAPKs, 13 MAPKK and 75 MAPKKKs are present in rice [27–28]. However, no systematic investigations of MAPK and MAPKK families have been reported in *B. distachyon*. The *B. distachyon* is fast emerging as a powerful model system to study the functional genomics of the temperate grasses, cereals, and dedicated biofuel crops such as Switchgrass [29]. The attributes include small genome (~300 Mb) diploid accessions, a small physical stature, self-fertility, a short lifecycle and simple growth requirements. Recently, the *B. distachyon* genome (diploid inbred line Bd21) has been sequenced [30], which provides an unprecedented opportunity for systematic analysis of the MAPK cascades in the *B. distachyon* species.

In the present study, 16 MAPK genes and 12 MAPKK genes were identified from the Bd21 genome by database searches and they were classified according to their homology with known MAPK and MAPKK genes in *Arabidopsis* or rice. We evaluated their expansion and evolutionary mechanisms by their duplication, distribution on the chromosomes and synteny map with their orthologs in *Arabidopsis* and rice and examined their protein interaction profile by yeast two-hybrid (Y2H) assay between these two families. Subsequently, we investigated their transcript profilings in different organs and in response to different stresses by RT-PCR and that of their orthologs in *Arabidopsis* and rice using Affymetrix *Arabidopsis* and rice microarray data. Finally, we conducted the pairwise comparisons of the expression profiles of putative orthologous or paralogs pairs existing in the two family genes of *B. distachyon* and rice under different light and temperature conditions. Our results from *B. distachyon* would provide the basis for future research on their expansion and evolutionary mechanisms and the diverse signaling pathways mediated by MAPKs and MAPKKs in *B. distachyon* and are more likely to be directly translatable to agriculturally important monocots, grasses such as wheat and barley by transferring functional information between species [31].

Materials and Methods

Identification, sequence alignment and phylogenetic analyses of MAPK and MAPKK genes in *B. distachyon*

The *B. distachyon* genome assembly version 1.2 was downloaded from the website <ftp://ftp.brachypodium.org/> to construct a local protein database. Method used to identify the MAPK and MAPKK genes in *B. distachyon* was similar to that described in rice and poplar [19]. For the MAPKK gene family, the predicted proteins derived from *B. distachyon* pseudo-molecules were queried using a profile Hidden Markov Model-based search (HMMER: <http://hmmer.wustl.edu/>) with an HMM built from the ten *Arabidopsis* MAPKKs [32]. MAPKK gene models were only accepted if they displayed the consensus sequences for dual-specificity protein kinases, including the conserved aspartate and lysine residues within the active site motif (–D(L/I/V)K–), and the plant-specific phosphorylation target site motif (–S/TxxxxS/T–) within the activation loop. Similarly, the predicted *B. distachyon* proteins were queried using a profile Hidden Markov Model-based search with an HMM built from the twenty *Arabidopsis* MAPKs for the MAPK gene family. MAPK gene models were

only accepted if they contained the canonical consensus sequences for serine/threonine protein kinases, as well as an appropriately positioned activation loop (–TXY–motif). And then the predictions of MAPK and MAPKK coding sequences were verified with available EST. Finally, multiple alignments of the identified *B. distachyon* amino acid sequences of these two gene families with that of *Arabidopsis* and rice were performed by Clustal W [33] with default options and the phylogenetic trees were constructed based on the bootstrap neighbor-joining (NJ) method with a Kimura two-parameter model by MEGA [34].

Analyses of the conserved motifs, chromosomal locations, gene duplication, gene structures, promoter regions and synteny with the homologs of rice and *Arabidopsis*

We performed the MEME (Multiple Expectation maximization for Motif Elicitation) analysis on the predicted candidate proteins of MAPK and MAPKK gene families with the conditions: 1. optimum motif width was set to 6 and 50; 2. maximum number of motifs was designed to identify 10 motifs; 3. the iterative cycles were set by default [35]. The exon/intron structures and locations for the two gene families were confirmed by Blat (<http://genome.ucsc.edu/FAQ/FAQblat.html>) search using a local database containing the complete *B. distachyon* genome sequences of each chromosome. Moreover, the gene duplication events of *B. distachyon* MAPK and MAPKK genes were also investigated. We defined the gene duplication in accordance with the criteria: 1. the alignment length covered $\geq 70\%$ of the longer gene; 2. the aligned region had an identity $\geq 70\%$; 3. only one duplication event was counted for tightly linked genes [36]. A block of duplications was defined if more than one gene was involved in the duplication.

To analyze their promoter regions, the 1 kb upstream regions of the two gene families according to the position of the genes provided by the *B. distachyon* annotation information were selected and screened against the PLACE database [37]. Regulatory elements overrepresented in the dataset and known to be involved in regulation during the resistance response and under stressed conditions were selected for further analysis [38]. Among them, WBOX [sequence TGAC(C/T)] associated with the WRKY transcription factors [39], CBF (GTGAC), DRE [(G/A)CCGAC] [40] and GCC boxes associated with the ERF-type transcription factors [41] were retained for further analysis. For synteny analysis, synteny blocks between *B. distachyon* and the other two plant species of rice and *Arabidopsis* were downloaded from the Plant Genome Duplication Database [42] and those containing *B. distachyon* MAPK and MAPK genes were identified, respectively.

cDNA cloning, plasmid constructions and yeast two-hybrid assays

Six MAPK and four MAPKK genes with more EST counts support, respectively, were selected to be amplified by PCR using primers designed based on the predicted sequences (listed in Supplemental Table S1). A complete list of primers used for RT-PCR, cDNA cloning and expression studies during the course of this work were provided in supplemental Table S1. Plasmid constructions and yeast two-hybrid assay for the two family genes were performed as described by Albrecht et al. [43]. All the yeast two-hybrid assays were performed in triplicate.

Plant materials, growth conditions, stress treatments and expression profile analyses of the two family genes in *B. distachyon* by RT-PCR

The community standard diploid inbred line of *B. distachyon*, Bd21, was used for all the experiments. After germination, they were planted in greenhouse and were grown under natural light and temperature conditions. The 3-week-old seedlings were used for all treatments. For drought, salinity, cold, abscisic acid (ABA), H₂O₂ and methyl jasmonate (MeJA) treatments, seedlings were subjected to 20% polyethylene glycol (PEG), 200 mM NaCl, 4°C conditions, 100 μM ABA, 10 mM H₂O₂ and 100 μM MeJA, respectively. The leaves of all the samples for RNA extractions were collected at 3 h after treatments. The roots, stems, leaves and caryopses of more than two-month old Bd21 were collected separately for RNA isolation and used for organ-specific expression analysis.

Total RNA was isolated according to the method described in the grape [44]. For RT-PCR, the specific primers were designed according to the predicted two family gene sequences by Primer 5 software (Table S1). A *B. distachyon* β-actin gene (ID: Bradi2g24070.1), amplified with primers 5'- CCCGATGGA-CAGGTTATCACTA-3' and 5'- ATAGAGCCACCAATC-CAAACAC-3', was used as a control. The following program was used for RT-PCR: 94°C for 5 min followed by 35 cycles at 94°C for 10 s, 55–59°C for 15 s and 72°C for 30 s, followed by a 5 min extension step at 72°C. While the number of cycles of PCR for actin gene was set as 30. The experiments were repeated at least three times with independent RNA samples.

Expression analyses of the two family genes in *Arabidopsis* and rice

Affymetrix *Arabidopsis* and rice microarray data were downloaded from ArrayExpress [45], PLEXdb [46] and Rice Oligonucleotide Databases [47]. Expression intensities in different development stages are log 2-based values and then were used for the tissue expression analysis. For the abiotic stress treatments, expression patterns of all samples (at least two biological repeats) were transformed into log 2-based numbers and normalized according to the quantile method for standardization of array data. Expression of a gene (up- or down-regulated) was defined as a gene with a log 2-based ratio (Cold, drought and salt-stressed sample/mock sample) higher than 0.5 or lower than -0.5; and a significant difference in gene expression between the treated plants and the control indicated by ≤ 0.05 by paired t-test [permutations, all possible combinations; false discovery rate (FDR) correction, adjusted Bonferroni method] [48]. Hierarchical clustering of expression profiles of the two family genes in *Arabidopsis* and rice was performed using R version 2.15.1.

Pairwise comparisons of the expression profiles of putative orthologous or paralogs existing in the two family genes of *B. distachyon* and rice under different light and temperature conditions

Available expression data of the two family genes of *B. distachyon* and rice under different light and temperature conditions were downloaded from PLEXdb [46] and Rice Oligonucleotide Databases [47]. All expression data (relative amount of mRNA) were composed of thirty nine treatment points. The Pearson correlation of the expression of orthologs or paralogs pairs existing in the two family genes was calculated on the basis of the expression data under the corresponding light and temperature conditions.

Results and Discussion

Identification and annotation of the *B. distachyon* MAPK and MAPKK families

Availability of complete *B. distachyon* genome sequences has made it possible for the first time to identify all the MAPKK and MAPK family members in this plant species. In order to determine the two family genes, we performed HMM searches and identified 16 MAPK and 12 MAPKK genes, respectively, from *B. distachyon* genomes. Based on *Arabidopsis* MAPK and MAPKK nomenclature suggestions [27], each gene was named with a two-letter code corresponding to *B. distachyon* (*Bd*), followed the family designation (*MPK* or *MKK*), and finally a number (Table 1). Because there was the occurrence of alternative mRNA splicing in some genes of the two families, the following analysis was restricted to the longest protein for each gene except specific description.

To assess which gene in this study had expression support, we compared the predicted genes against available ESTs (<http://www.brachypodium.org/>). Because many members of these two families are similar to one another, only top matches were considered, after applying a high match stringency of at least 95% of nucleotide identity between EST and genomic sequences (Table S2). At this threshold, 7 MAPKK genes and 14 MAPK genes in this study had EST supports, representing 87.5% and 58.3% of predicted MAPKK and MAPK genes in the study, respectively. Four MAPKK genes (six proteins generated by alternative splicing) and six MAPK genes with more EST counts support were selected to be amplified by PCR for the following protein interactive analysis. All the cDNAs of our cloned genes of the two families have been submitted to GenBank and their accession numbers in GenBank were shown in Table S2.

Multiple alignments, phylogenetic and domain analyses of MAPK and MAPKK families

The phylogenetic relationship of the MAPK family proteins in *B. distachyon* was examined by multiple sequence alignment of their conserved protein kinase, which spans approx. about 300 amino acids. The alignment analysis based on T-loop motif sequences of MAPK family from three species divided them into four groups (Figure 1). As shown in Figure 1, the sequences in the T-loop motif were highly conserved. A comparison with the protein kinase motifs of the MAPK representative proteins from *Arabidopsis* and rice could result in a better separation of the different groups and subgroups. Through phylogenetic analyses of the conserved motifs of the three species, we found that the BdMAPKs could also be classified into four groups corresponding to the groups A, B, C and D in *Arabidopsis* (Figure 2a) [49]. Those *B. distachyon* and rice MAPKs that harbored a -TEY- signature in their activation loop clustered with their *Arabidopsis* homologs except BdMAPK11 in which the Met-Glu-Tyr (-MEY-) replaced the -TEY- signature; while the large clade of group D MAPKs all displayed a distinctive -TDY- rather than -TEY- signature (Figure 1 and Figure 2a). This case that “MEY” replaced the “TEY” signature was also observed in tomato but not observed in rice [50], even though the relationship of the two monocots species *B. distachyon* and rice are evolutionarily more closely than that of *B. distachyon* and dicots tomato.

Moreover, four MAPK genes (*BdMPK3*, *BdMPK4*, *BdMPK6* and *BdMPK11*) belonging to group A and B were found to have a CD domain in their C-terminal region, which functioned as a binding site of MAPKKs and conserved as a DxxDE(P)xC motif in the evolution from the multiple sequence alignment results (Figure 2b). Group C and D did not possess such a CD domain, but group D had a relatively long C-terminal region. The results were

Table 1. Nomenclature for MAPKs and MAPKKs in *Arabidopsis*, *Brachypodium* and *Oryza*.

Family	Bd gene name	Bd gene model	At gene name [*]	At gene code [*]	Os gene name [*]	Os gene code [*]
MAPK	<i>BdMPK3</i>	Bradi1g65810	<i>AtMPK1</i>	At1g10210	<i>OsMPK3</i>	Os03g17700
	<i>BdMPK4</i>	Bradi3g32000	<i>AtMPK2</i>	At1g59580	<i>OsMPK4</i>	Os10g38950
	<i>BdMPK6</i>	Bradi1g49100	<i>AtMPK3</i>	At3g45640	<i>OsMPK6</i>	Os06g06090
	<i>BdMPK7-1</i>	Bradi1g34030	<i>AtMPK4</i>	At4g01370	<i>OsMPK7</i>	Os06g48590
	<i>BdMPK7-2</i>	Bradi4g24912	<i>AtMPK5</i>	At4g11330	<i>OsMPK14</i>	Os02g05480
	<i>BdMPK11</i>	Bradi3g16560	<i>AtMPK6</i>	At2g43790	<i>OsMPK16</i>	Os11g17080
	<i>BdMPK14</i>	Bradi3g03780	<i>AtMPK7</i>	At2g18170	<i>OsMPK17-1</i>	Os06g49430
	<i>BdMPK16</i>	Bradi2g36470	<i>AtMPK8</i>	At1g18150	<i>OsMPK17-2</i>	Os02g04230
	<i>BdMPK17</i>	Bradi1g34700	<i>AtMPK9</i>	At3g18040	<i>OsMPK20-1</i>	Os01g43910
	<i>BdMPK20-1</i>	Bradi2g44350	<i>AtMPK10</i>	At3g59790	<i>OsMPK20-2</i>	Os05g50560
	<i>BdMPK20-2</i>	Bradi2g15317	<i>AtMPK11</i>	At1g01560	<i>OsMPK20-3</i>	Os06g26340
	<i>BdMPK20-3</i>	Bradi1g41780	<i>AtMPK12</i>	At2g46070	<i>OsMPK20-4</i>	Os01g47530
	<i>BdMPK20-4</i>	Bradi2g45870	<i>AtMPK13</i>	At1g07880	<i>OsMPK20-5</i>	Os05g49140
	<i>BdMPK20-5</i>	Bradi2g16337	<i>AtMPK14</i>	At4g36450	<i>OsMPK21-1</i>	Os05g50120
	<i>BdMPK21-1</i>	Bradi2g15620	<i>AtMPK15</i>	At1g73670	<i>OsMPK21-2</i>	Os01g45620
	<i>BdMPK21-2</i>	Bradi2g45010	<i>AtMPK16</i>	At5g19010		
			<i>AtMPK17</i>	At2g01450		
			<i>AtMPK18</i>	At1g53510		
			<i>AtMPK19</i>	At3g14720		
			<i>AtMPK20</i>	At2g42880		
MAPKK	<i>BdMKK1</i>	Bradi1g51000	<i>AtMKK1</i>	At4g26070	<i>OsMKK1</i>	Os06g05520
	<i>BdMKK3-1</i>	Bradi4g39490	<i>AtMKK2</i>	At4g29810	<i>OsMKK3</i>	Os06g27890
	<i>BdMKK3-2</i>	Bradi1g41860	<i>AtMKK3</i>	At5g40440	<i>OsMKK4</i>	Os02g54600
	<i>BdMKK3-3</i>	Bradi3g11260	<i>AtMKK4</i>	At1g51660	<i>OsMKK5</i>	Os06g09180
	<i>BdMKK4</i>	Bradi3g53650	<i>AtMKK5</i>	At3g21220	<i>OsMKK6</i>	Os01g32660
	<i>BdMKK5</i>	Bradi1g46880	<i>AtMKK6</i>	At5g56580	<i>OsMKK10-1</i>	Os02g46760
	<i>BdMKK6</i>	Bradi1g75150	<i>AtMKK7</i>	At1g18350	<i>OsMKK10-2</i>	Os03g12390
	<i>BdMKK10-1</i>	Bradi1g11525	<i>AtMKK8</i>	At3g06230	<i>OsMKK10-3</i>	Os03g50550
	<i>BdMKK10-2</i>	Bradi1g69400	<i>AtMKK9</i>	At1g73500		
	<i>BdMKK10-3</i>	Bradi1g10800	<i>AtMKK10</i>	At1g32320		
	<i>BdMKK10-4</i>	Bradi1g10770				
	<i>BdMKK10-5</i>	Bradi1g10790				

*From Hamel et al [19]. At: *Arabidopsis thaliana*, Os: *Oryza sativa*, Bd: *B. distachyon*. doi:10.1371/journal.pone.0046744.t001

consistent with that of MEME results (Figure 2c). A total of ten motifs were identified from the BdMPK proteins. Nine of ten motifs corresponded with the eleven domains (I–XI) that are found in *B. distachyon* serine/threonine protein kinases (Figure S1). From the MEME results, we could also find that all the members of large group D contained the extra motif 9 besides their common motifs shared with other subgroup A, B and C (Figure 2c). The last subdomain 11 of serine/threonine protein kinases was corresponding with the motif 7, while no subdomains corresponded with the motif 9, which only existed in the group D and followed behind the motif 7. Whether the difference between the large group D and other subgroups A, B and C might have some effect on their biological function needs to be further studied in the future. However, so far relatively little work has been done to identify the function of MAPKs in the large group D.

For the *B. distachyon* MAPKKs, similar work was done as the MAPK gene family. Animal MAPKKs had the S/T-X5-S/T motif as the phosphorylation site which was also observed in some

fungi and many plant MAPKKK kinases (MEKKKs). Comparison analysis between the amino acid sequences of *B. distachyon*, *Arabidopsis* and rice MAPKK S/T-X5-S/T motifs was shown in Figure 3. The results revealed that the activation loop motif of four members of the ‘Group D’ (*OsMKK4*, *OsMKK5*, *BdMKK4* and *BdMKK5*) and all the members of ‘Group C’ was either absent or located 3–5 residues downstream of the canonical position (Figure 3), especially the *BdMKK10-1*, *BdMKK10-3*, *BdMKK10-4*, *BdMKK10-5* and *OsMKK10-3* which were more divergent in the S/T-X5-S/T motif compared with others. More interestingly, all these four genes had no EST support and the expression profiles of their orthologous in *Arabidopsis* and rice were also not detected by RT-PCR in the two species [19]. In order to find the convincing evidence to determine whether these genes with untypical domain were expressed in a certain condition, they were subjected to RT-PCR and microarray data analysis in the following section.

Previous study showed that MAPKK proteins from the same subgroup or clade tend to cluster together. However, the

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Group A																														
BdMPK6	C	D	F	G	L	A	R	T	T	.	S	E	T	D	F	M	T	E	Y	V	V	T	R	W	Y	R
OsMPK6	C	D	F	G	L	A	R	T	T	.	S	E	T	D	F	M	T	E	Y	V	V	T	R	W	Y	R
AtMPK6	C	D	F	G	L	A	R	V	T	.	S	E	S	D	F	M	T	E	Y	V	V	T	R	W	Y	R
BdMPK3	C	D	F	G	L	A	R	P	S	.	S	E	S	D	M	M	T	E	Y	V	V	T	R	W	Y	R
OsMPK3	C	D	F	G	L	A	R	P	S	.	S	E	S	D	M	M	T	E	Y	V	V	T	R	W	Y	R
Group B																														
BdMPK4	A	D	F	G	L	A	R	T	T	.	S	E	T	D	L	M	T	E	Y	V	V	T	R	W	Y	R
OsMPK4	A	D	F	G	L	A	R	T	T	.	T	E	T	D	L	M	T	E	Y	V	V	T	R	W	Y	R
BdMPK11	G	D	F	G	L	A	R	T	T	.	T	E	T	D	F	M	M	E	Y	V	V	T	R	W	Y	R
AtMPK11	G	D	F	G	L	A	R	T	K	.	S	E	T	D	F	M	T	E	Y	V	V	T	R	W	Y	R
AtMPK4	G	D	F	G	L	A	R	T	K	.	S	E	T	D	F	M	T	E	Y	V	V	T	R	W	Y	R
Group C																														
BdMPK7-1	C	D	F	G	L	A	R	T	N	N	T	K	G	Q	F	M	T	E	Y	V	V	T	R	W	Y	R
BdMPK7-2	C	D	F	G	L	A	R	T	N	N	T	K	G	Q	F	M	T	E	Y	V	V	T	C	L	Y	R
OsMPK7	C	D	F	G	L	A	R	T	N	N	T	K	G	Q	F	M	T	E	Y	V	V	T	R	W	Y	R
BdMPK14	C	D	F	G	L	A	R	T	N	S	S	K	G	Q	F	M	T	E	Y	V	V	T	R	W	Y	R
OsMPK14	C	D	F	G	L	A	R	T	N	S	S	K	G	Q	F	M	T	E	Y	V	V	T	R	W	Y	R
AtMPK2	C	D	F	G	L	A	R	T	S	N	T	K	G	Q	F	M	T	E	Y	V	V	T	R	W	Y	R
AtMPK14	C	D	F	G	L	A	R	T	.	.	.	Y	E	Q	F	M	T	E	Y	V	V	T	R	W	Y	R
Group D																														
BdMPK20-3	C	D	F	G	L	A	R	V	A	F	S	D	.	A	P	T	T	V	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK20-3	C	D	F	G	L	A	R	V	A	F	T	D	.	A	P	T	T	V	F	W	T	D	Y	V	A	T	R	W	Y	R
AtMPK20	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	T	I	F	W	T	D	Y	V	A	T	R	W	Y	R
BdMPK20-2	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	T	V	F	W	T	D	Y	V	A	T	R	W	Y	R
BdMPK20-1	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	T	I	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK20-1	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	T	I	F	W	T	D	Y	V	A	T	R	W	Y	R
BdMPK20-5	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	T	V	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK20-5	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	T	V	F	W	T	D	Y	V	A	T	R	W	Y	R
BdMPK20-4	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	T	V	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK20-4	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	T	V	F	W	T	D	Y	V	A	T	R	W	Y	R
AtMPK19	C	D	F	G	L	A	R	V	S	F	N	D	.	T	P	T	T	V	F	W	T	D	Y	V	A	T	R	W	Y	R
BdMPK16	C	D	F	G	L	A	R	V	A	I	S	D	.	T	P	T	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK16	C	D	F	G	L	A	R	V	A	F	S	D	.	T	P	T	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
AtMPK16	C	D	F	G	L	A	R	V	A	F	N	D	.	T	P	T	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
BdMPK17	C	D	F	G	L	A	R	V	S	F	N	D	.	A	P	T	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK17-1	C	D	F	G	L	A	R	A	S	F	N	D	.	A	P	S	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK17-2	C	D	F	G	L	A	R	V	A	F	N	D	.	S	P	S	T	I	F	W	T	D	Y	V	A	T	R	W	Y	R
AtMPK15	C	D	F	G	L	A	R	V	S	F	N	D	.	A	P	T	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
BdMPK21-1	C	D	F	G	L	A	R	V	S	F	N	D	G	A	P	S	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK21-1	C	D	F	G	L	A	R	V	S	F	N	D	.	T	P	S	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
BdMPK21-2	C	D	F	G	L	A	R	V	S	F	N	D	.	T	P	S	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
OsMPK21-2	C	D	F	G	L	A	R	V	S	F	D	D	.	T	P	S	A	I	F	W	T	D	Y	V	A	T	R	W	Y	R
AtMPK17	C	D	L	G	L	A	R	V	S	F	T	D	.	S	P	S	A	V	F	W	T	D	Y	V	A	T	R	W	Y	R

Figure 1. Alignment of multiple *B. distachyon* MAPK, selected *Arabidopsis* and rice MAPK domain amino acid sequences. Alignment was performed using Clustal W. The conserved T-loop amino acid signature is highlighted in yellow, and gaps are marked with dot lines. doi:10.1371/journal.pone.0046744.g001

evolutionary analysis of MAPKK genes in the three species could not be clarified for all clades and revealed some interesting observations (Figure 4a). For example, most MAPKK genes were placed in generally well-resolved clades that each contained the members from both monocots and eudicots; while the MKK7-9 clade was not the case, for which no *B. distachyon* and rice orthologs could be identified for the AtMKK7-9. The different evolutionary

patterns of MKK7-9 clade in *Arabidopsis*, *B. distachyon* and rice might occur after their divergence. Compared with MKK10 clade from *Arabidopsis* (1) and rice (3), *B. distachyon* MKK10 clade was the most expanded with five genes. And MKK10 clade also had the most members in the three species compared with other clades. More interestingly, the relative evolution distance between BdMKK10-2 and other members of its same BdMKK10 clade,

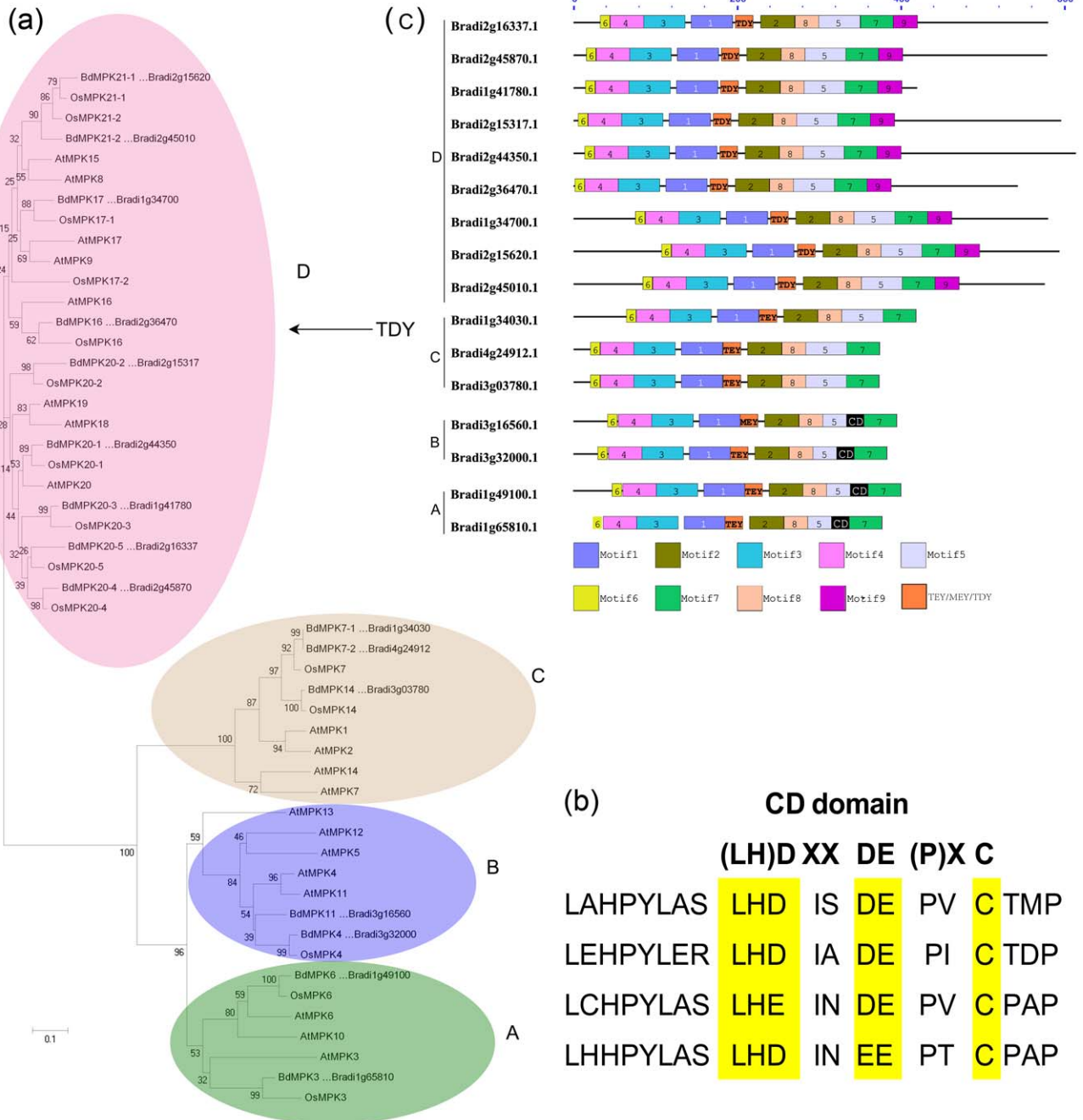


Figure 2. Phylogenetic and domain analyses of MAPKs. (a) Phylogenetic relationships of *B. distachyon*, *Arabidopsis* and rice MAPK genes. The phylogenetic tree was constructed by NJ (Neighbor-joining) method using the MEGA 4 program. Different color patternings indicate different gene clusters or superclusters. (b) Comparison of CD domain among four *B. distachyon* MAPKs (c) Schematic diagram of amino acid motifs of *B. distachyon* MAPKs from different groups. Motif analysis was performed using MEME 4.0 software as described in the methods. The black solid line represents the corresponding BdMAPK and its length. The different-colored boxes represent different motifs and their position in each BdMAPK sequence. A detailed motif introduction for all the *B. distachyon* MAPKs is shown in Figure S1. doi:10.1371/journal.pone.0046744.g002

was much farther than that between each other of BdMCK10-1, BdMCK10-3 and BdMCK10-4 in the phylogenetic tree, which may be due to the reason that BdMCK10-2 evolved prior to other members in the evolutionary timescale of life.

The conserved motifs of MAPKK family proteins in *B. distachyon* were also investigated using MEME as that of MAPK family, and a schematic overview of the identified motifs was shown in

Figure 4b. In the BdMCK10 clade, except that the closely related paralogs BdMCK10-1 and BdMCK10-3 had an extra motif 10 in the N-terminal of their sequence, all other members shared nine conserved motifs with the same organization. However, the motif 10 of other clades was not in the N-terminal of the sequence and all the members of them did not contain the motif 7 except BdMCK4 and BdMCK5 compared with that of BdMCK10

Group A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50			
AtMKK1	I	I	H	R	D	L	K	P	S	N	L	L	I	N	H	R	G	E	V	K	I	T	D	F	G	V	S	K	I	L	T	S	T	S	S	L	A	N	S	F	V	G	T	Y	P	Y			
AtMKK2	I	I	H	R	D	L	K	P	S	N	L	L	I	N	H	R	G	E	V	K	I	T	D	F	G	V	S	T	V	M	T	N	T	A	G	L	A	N	T	F	V	G	T	Y	N	Y			
OsMKK1	I	I	H	R	D	L	K	P	S	N	L	L	I	N	H	M	G	E	V	K	I	S	D	F	G	V	S	A	I	I	A	S	S	S	A	Q	R	D	T	F	T	G	T	Y	N	Y			
BdMKK1	V	I	H	R	D	L	K	P	S	N	L	L	I	N	H	R	G	E	V	K	I	S	D	F	G	V	S	A	I	I	A	S	S	S	A	Q	R	D	T	F	T	G	T	Y	N	Y			
OsMKK6	V	I	H	R	D	I	K	P	S	N	L	L	V	N	R	K	G	E	V	K	I	T	D	F	G	V	S	A	V	L	A	S	S	M	G	Q	R	D	T	F	V	G	T	Y	N	Y			
BdMKK6	V	I	H	R	D	I	K	P	S	N	L	L	V	N	H	K	G	E	V	K	I	T	D	F	G	V	S	A	V	L	A	S	S	I	G	Q	R	D	T	F	V	G	T	Y	N	Y			
AtMKK6	V	I	H	R	D	I	K	P	S	N	L	L	V	N	H	K	G	E	V	K	I	S	D	F	G	V	S	A	S	L	A	S	S	M	G	Q	R	D	T	F	V	G	T	Y	N	Y			
Group B																																																					
BdMKK3-1	L	V	H	R	D	I	K	P	A	N	M	L	V	N	L	K	G	E	A	K	I	T	D	F	G	V	S	A	G	L	D	N	T	M	A	M	C	A	T	F	V	G	T	V	T	Y			
BdMKK3-2	L	V	H	R	D	I	K	P	A	N	M	L	V	N	L	K	G	E	A	K	I	T	D	F	G	V	S	A	G	L	D	N	T	M	A	M	C	A	T	F	V	G	T	V	T	Y			
OsMKK3	L	V	H	R	D	L	K	P	A	N	L	L	V	N	L	K	G	E	A	K	I	T	D	F	G	V	S	A	G	L	D	N	T	M	A	M	C	A	T	F	V	G	T	V	T	Y			
BdMKK3-3	L	V	H	R	D	I	K	P	A	N	L	L	L	N	L	K	G	D	T	K	I	T	D	F	G	V	T	S	G	L	H	D	S	I	D	M	C	A	T	F	L	G	S	V	T	Y			
AtMKK3	L	V	H	R	D	I	K	P	A	N	L	L	I	N	L	K	G	E	P	K	I	T	D	F	G	I	S	A	G	L	E	N	S	M	A	M	C	A	T	F	V	G	T	V	T	Y			
Group C																																																					
AtMKK10	I	V	H	G	D	I	K	P	S	N	L	L	I	N	K	K	G	E	V	K	I	A	D	F	G	A	S	R	I	V	A	G	G	D	Y	G	S	N	G	T	C	A	Y		
OsMKK10-1	V	A	H	G	D	V	K	P	S	N	L	L	V	G	H	R	G	E	I	K	I	A	D	F	G	A	S	R	V	V	T	G	R	D	E	A	H	H	Q	S	P	G	T	W	A	Y			
OsMKK10-2	I	V	H	R	D	L	K	P	S	N	L	L	L	G	A	D	G	E	V	K	I	A	D	F	G	A	G	K	V	L	R	R	R	L	D	P	C	A	S	Y	V	G	T	A	A	Y			
BdMKK10-2	V	V	H	R	D	L	K	P	S	N	L	L	L	G	S	H	G	E	V	K	I	A	D	F	G	A	G	K	V	L	R	R	R	L	D	P	C	A	S	Y	V	G	T	A	A	Y			
BdMKK10-1	I	V	H	L	D	V	K	P	G	N	L	L	V	S	T	G	G	E	V	K	I	A	D	F	G	I	A	R	V	L	P	R	A	G	G	D	D	V	R	C	T	A	Y	A	G	T	A	A	Y
BdMKK10-3	I	V	H	L	D	I	K	P	G	N	L	L	V	S	T	G	G	E	V	K	V	A	D	F	G	I	A	K	V	L	P	R	A	G	A	D	D	A	R	C	K	S	Y	A	G	T	A	A	Y
OsMKK10-3	V	V	H	R	D	I	K	P	G	N	L	L	V	S	V	D	G	E	V	K	I	A	D	F	G	I	A	K	V	V	P	P	R	R	G	G	E	H	R	A	A	Y	E	Y	E	G	T	A	A	Y			
BdMKK10-4	V	A	H	R	D	I	K	P	A	N	L	L	L	S	A	A	G	E	V	K	I	A	D	F	G	I	A	K	V	V	V	S	G	A	G	G	R	A	R	A	L	A	Y	E	G	T	V	A	Y
BdMKK10-5	V	V	H	R	D	I	K	P	A	N	L	L	V	S	G	A	G	G	V	K	V	A	D	F	G	I	A	M	V	L	P	S	R	A	G	G	E	R	C	A	A	Y	E	G	T	V	A	Y	
Group D																																																					
AtMKK7	I	V	H	R	D	I	K	P	A	N	L	L	L	N	S	R	N	E	V	K	I	A	D	F	G	V	S	K	I	I	T	R	S	L	D	Y	C	N	S	Y	V	G	T	C	A	Y			
AtMKK9	I	V	H	R	D	I	K	P	A	N	L	L	L	N	S	K	N	E	V	K	I	A	D	F	G	V	S	K	I	L	V	R	S	L	D	S	C	N	S	Y	V	G	T	C	A	Y			
AtMKK8	I	V	H	R	D	I	K	P	A	N	L	L	R	S	S	K	E	E	V	K	I	A	D	F	G	V	S	K	I	V	V	R	S	L	N	K	C	N	S	F	V	G	T	F	A	Y			
AtMKK5	I	V	H	R	D	I	K	P	S	N	L	L	I	N	S	A	K	N	V	K	I	A	D	F	G	V	S	R	I	L	A	Q	T	M	D	P	C	N	S	S	V	G	T	I	A	Y			
AtMKK4	I	V	H	R	D	I	K	P	S	N	L	L	I	N	S	A	K	N	V	K	I	A	D	F	G	V	S	R	I	L	A	Q	T	M	D	P	C	N	S	S	V	G	T	I	A	Y			
OsMKK4	I	V	H	R	D	I	K	P	S	N	L	L	I	D	S	G	R	R	V	K	I	A	D	F	G	V	G	R	I	L	N	Q	T	M	D	P	C	N	S	S	V	G	T	I	A	Y			
BdMKK4	I	V	H	R	D	I	K	P	S	N	L	L	I	D	S	G	R	R	V	K	I	A	D	F	G	V	G	R	I	L	N	Q	T	M	D	P	C	N	S	S	V	G	T	I	A	Y			
OsMKK5	I	V	H	R	D	I	K	P	S	N	L	L	I	D	S	A	R	R	V	K	I	A	D	F	G	V	G	R	I	L	N	Q	T	M	D	P	C	N	S	S	V	G	T	I	A	Y			
BdMKK5	I	V	H	R	D	I	K	P	S	N	L	L	I	D	S	A	R	R	V	K	I	A	D	F	G	V	G	R	I	L	N	Q	T	M	D	P	C	N	S	S	V	G	T	I	A	Y			

Figure 3. Alignment of multiple *B. distachyon* MAPKK, selected *Arabidopsis* and rice MAPKK domain amino acid sequences. Alignment was performed using Clustal W. The conserved S/T-X5-S/T motif and active site D(I/L/V)K motif were highlighted in yellow. The genes with incomplete activation loop motif in which the 5'- or 3'- S/T residue is either absent or located 3–5 residues downstream of the canonical position were highlighted in grey. And gaps are marked with dot lines.
doi:10.1371/journal.pone.0046744.g003

clade. In addition, we also found that all the members of BdMKK3 clade had long terminal sequences both at the 5' and 3' terminal besides sharing the conserved motifs with the same organization (Figure 4b). And the genes from *B. distachyon* tend to be clustered more closely with that from rice than that from *Arabidopsis* by comparison with the phylogenetic analysis of the two family genes from different species, which may be the reason that *Arabidopsis* and Poaceae split far before the separation of *B. distachyon* and rice.

Genomic distribution and gene duplication

Figure 5 showed the locations of the MAPK and MAPKK genes on the 5 chromosomes of *B. distachyon*. They were separately located on each chromosome individually or in clusters, and their distribution was non-random. Their chromosomal distribution pattern revealed that certain chromosomes and chromosomal regions had a relatively high density of MAPK or MAPKK genes. For example, none of these two gene family members was located on chromosome 5. Seventy five percent of MAPKK genes were located on the chromosome 1 and the remaining members were respectively located on chromosome 3 and 4. For the MAPK family, the majority of members were respectively located on the chromosome 1, 2 and 3 except BdMPK7-2. Based on Holub's definition of a gene cluster [51], it is a region that contains four or

more genes within 200 kb or less. In this study, we found only a gene cluster with four members in the MAPKK family using this criteria and no gene cluster was found in the MAPK family (Figure 5). However, if we relaxed the clustering criteria with at least two genes in a cluster, 69% of MAPK members occurred in clusters but it did not have some effect on the number of MAPKK gene clusters (Figure 5).

We further determined the tandem and segmental duplication events of these two gene families on the five *B. distachyon* chromosomes. As shown in Figure 5, two MAPK gene clusters (The names of each gene cluster was labeled in the same color except the black color) containing 6 tandem duplicated genes were both located on chromosome 2 and one MAPKK gene cluster with four members were located on chromosome 1. Intriguingly, the two gene clusters of MAPK family as well as *BdMPK16*, *BdMPK17* and *BdMPK20-3* accounting for 56.25% of MAPK genes were also involved in the segmental duplication event and formed a supercluster, which showed that these members may be originated from chromosomal segment duplication and subsequent divergence under the selective pressure of external environment. The duplicate chromosomal segments of these two families were connected with different color dotted lines. Although several paralogs such as *BdMPK3* and *BdMPK6*, *BdMPK4* and *BdMPK11* shared high similarity of amino acid sequences, they

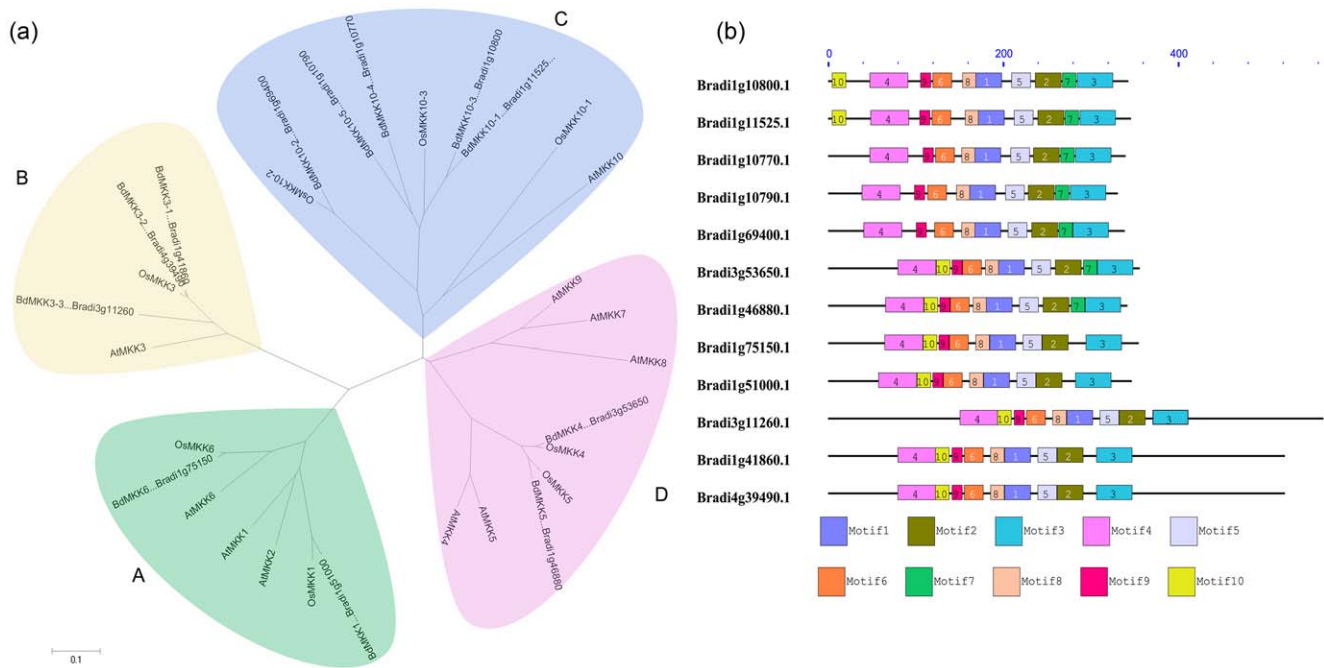


Figure 4. Phylogenetic and domain analyses of MAPKKs. (a) Phylogenetic relationships of *B. distachyon*, *Arabidopsis* and rice MAPKK genes. The phylogenetic tree was constructed by NJ (Neighbor-joining) method using the MEGA 4 program. Different color patterns indicate different gene clusters or superclusters. (b) Schematic diagram of amino acid motifs of *B. distachyon* MAPKKs. The black solid line represents the corresponding BdMAPKK and its length. The different-colored boxes represent different motifs and their position in each BdMAPKK sequence. doi:10.1371/journal.pone.0046744.g004

were far from each other on the chromosome and therefore were not considered as tandem duplicated genes. More interestingly, *BdMPK7-1* and *BdMPK17* were tightly linked on chromosome 1, but they only shared 44.35% similarity of amino acids sequences and were also not considered as tandem duplicated genes, which may be evolved from different progenitor genes. The result was consistent with that of phylogenetic and domain analysis mentioned above and the two genes were located on different subgroups of the MAPK phylogenetic tree. *BdMPK7-1* and *BdMPK17* were in the subgroup C and D, respectively (Figure 2). However, *BdMPK7-1*, *BdMPK7-2* and *BdMPK14* located on different chromosomes had higher similarity in their amino acids sequence than that of other paralogs, reached up to more than 90% similarity between each other and happened to cluster together in the phylogenetic tree, which showed that these genes were possibly originated from recent inter-chromosomal gene duplication event. In addition, we also found that although the *BdMCK10-2* and the gene cluster consisting of other members of *BdMCK10* subgroup were all located on the same chromosome, they were far from each other in the physical distance, which was consistent with the result of MAPKK phylogenetic and multiple alignment analysis. Taken together, a total of 87.5% of MAPK genes and 66.7% of MAPKK genes were the results of either tandem duplication or segmental duplication event, which suggested that the two duplication patterns played a major role in the expansion of the two families in *B. distachyon*.

Analyses of the MAPK and MAPKK gene structures and their promoter regions

To obtain some insight into the gene structures of the MAPK and MAPKK family genes, their exon/intron organizations were analyzed. It was found that all the members of MAPK family had introns varying from three to eleven, while 58.3% of the total

MAPKK genes had no intron. The MAPK and MAPKK genes could be divided into three and two subgroups based on their exon/intron structures, respectively (Figure 6). The gene structures of MAPK family were more divergent than that of MAPKK family. Some losses or gains of exons were also identified during the evolution of these two family genes at the 5' end or 3' terminal, and even in the middle of their sequences such as *BdMPK7-1*, *BdMCK1* and *BdMPK6*. Nonetheless, the exon/intron structures of each gene cluster originated from tandem or segmental duplication event in the two gene families tended to share similar structure organizations except tiny difference. One such example was *BdMPK4* and *BdMPK11*. These findings suggested these two family genes preserved a relatively constant exon-intron composition in each subgroup of them during the evolution of the *B. distachyon* genome. Moreover, we examined the *B. distachyon* MAPKs and MAPKKs for the occurrence of alternative mRNA splicing. The cDNA sequences obtained in our study were compared with all available full-length cDNA and EST sequence data. The investigations identified two alternatively spliced mRNAs for each *BdMCK1* and *BdMCK6*, respectively. The differential splicing events occurring in the two genes resulted in the difference in the C-terminal part of their translated protein (Figure S2). And the differentially spliced transcript forms (*BdMCK6.1* and *BdMCK6.2*) of *BdMCK6* were exemplarily selected for the following interaction and expression analysis in order to assess whether they have similar function.

In addition, we also analyzed the promoter sequences in 1 kb region upstream of the predicted MAPK and MAPKK encoding genes as described in *M. truncatula* [52]. Four regulatory elements including the WBOX cassettes, CBF boxes, DRE boxes and the GCC motif implicated in either response to pathogens or plant stress were identified as being overrepresented in the 1 kb region upstream of the two gene families. The analysis results showed that

Chromosome Location Information

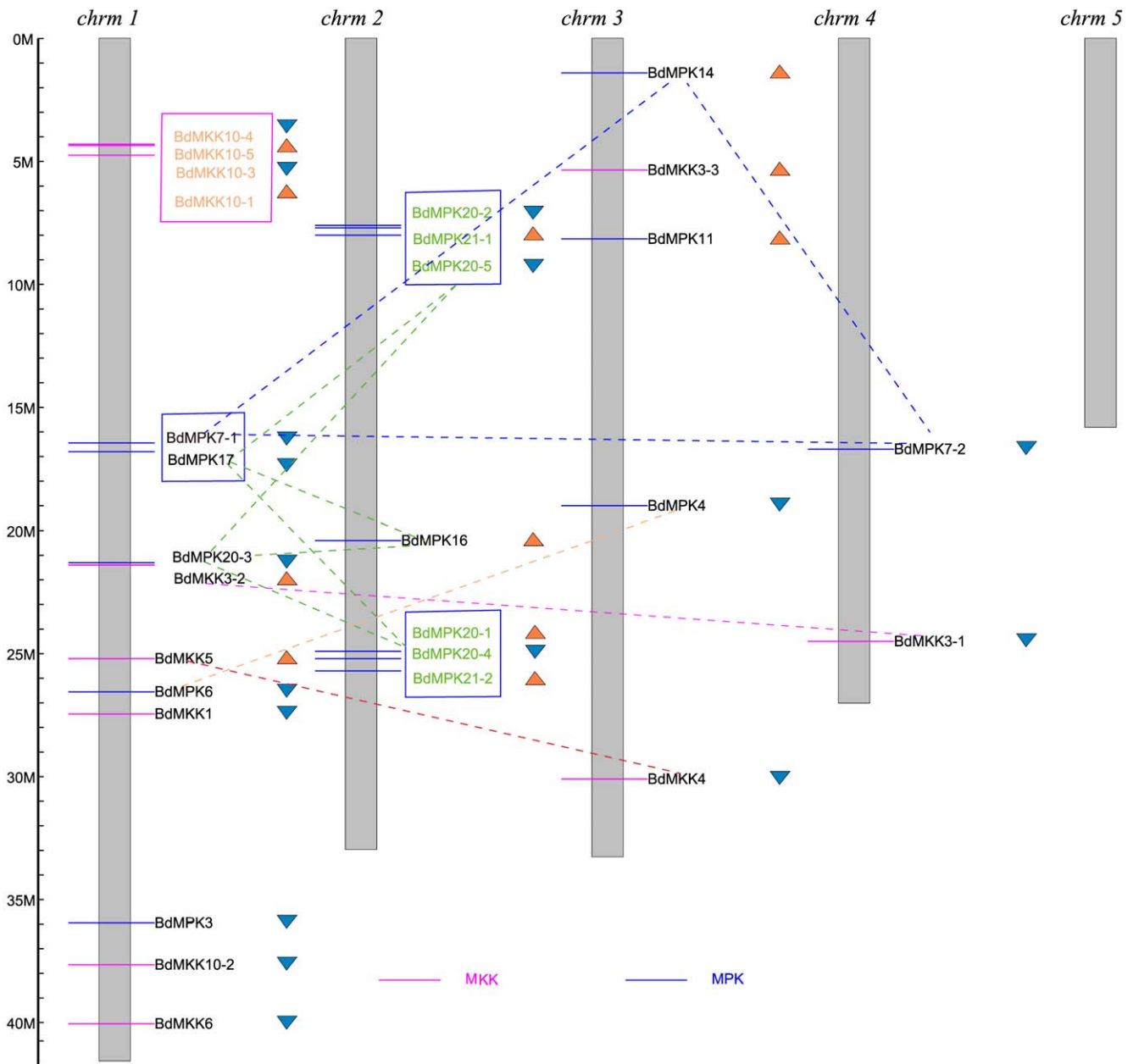


Figure 5. Chromosomal locations of *B. distachyon* MAPK (in blue) and MAPKK (in purple) genes. The MAPK and MAPKK gene clusters were separately indicated with blue and purple rectangles. Different color dotted lines connect the MAPK and MAPKK genes present on duplicate chromosomal segments. The names of each tandem duplicated gene cluster of the two families were indicated with the same color (except the names in black). The orange and blue triangles indicate the upward and downward direction of transcription, respectively.
doi:10.1371/journal.pone.0046744.g005

WBOX element existed in all the members of the two families, averaging 7.38 per MAPK gene and 8.15 per MAPKK gene, respectively (Table S2). And 62.5% of MAPK genes and 75% MAPKK genes contained at least 6 predicted WBOXs. In contrast, the average numbers of other element types were 1.75 (CBF), 2.06 (GCC) and 1.06 (DRE) for MAPK family and 0.92 (CBF), 2.25 (GCC) and 0.5 (DRE) for MAPKK family, respectively. Of the predicted MAPK genes, 50% contained the four multiple boxes, while there was only 33.3% of MAPKK genes including these four multiple boxes. Moreover, we see no clear

evidence of a correlation between the arrangement of these promoter cassettes (WBOX, CBF, DRE and GCC) and *in silico* expression via EST counts (Table S2).

Synteny analyses of MAPKs and MAPKKs between *B. distachyon* and other two plant species of rice and *Arabidopsis*

To further explore the origin and evolutionary process of *B. distachyon* MAPK and MAPKK genes, we analyzed the compar-

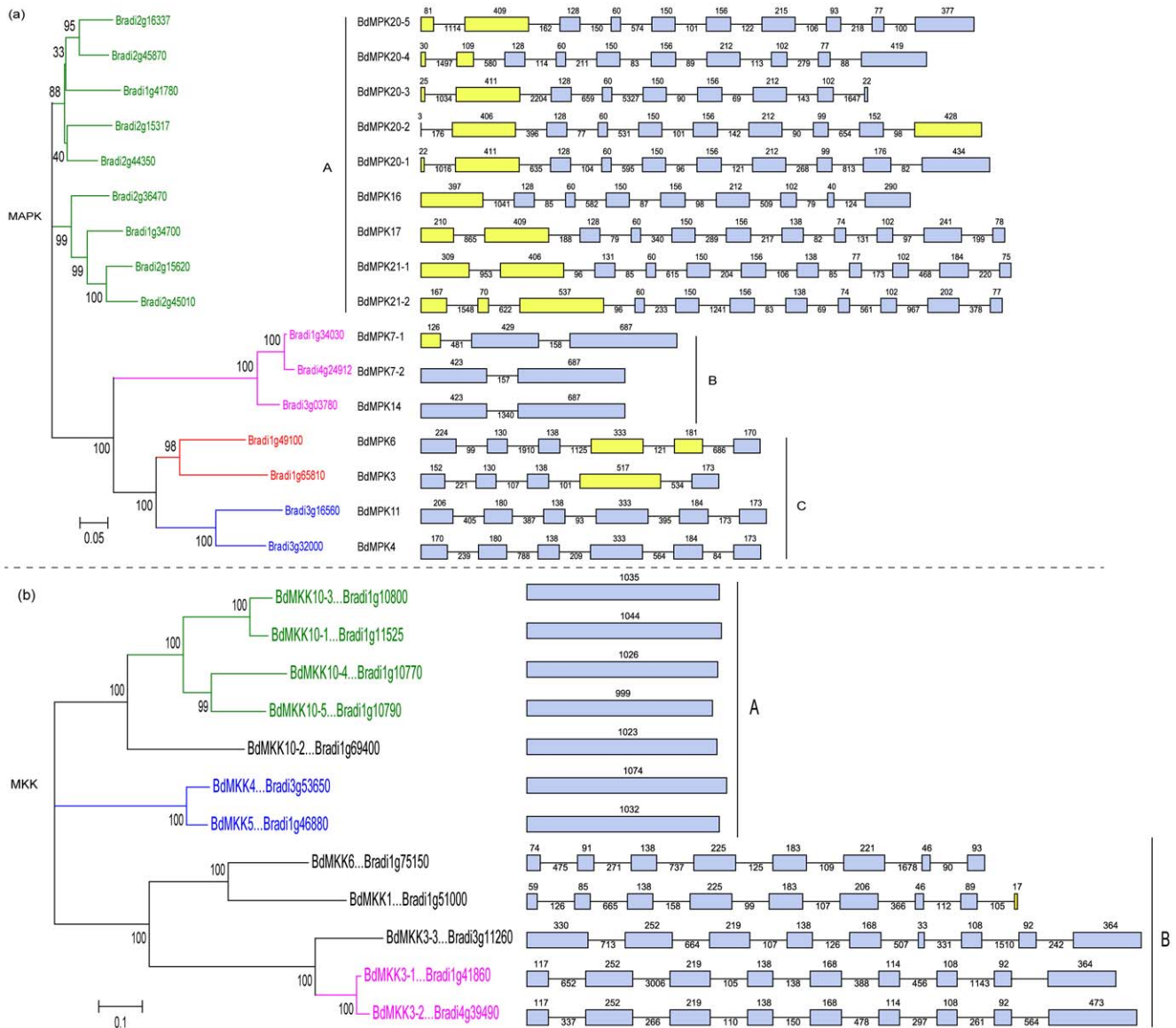


Figure 6. Phylogenetic analysis (Left) and exon-intron structures (Right) of *B. distachyon* MAPKs (a) and MAPKKs (b). The gene names of each gene cluster or supercluster originated from tandem or segmental duplication event were indicated with the same color in the phylogenetic tree except the black color. Only coding exons, represented by blue or yellow boxes, were drawn to scale. The black solid lines connecting two exons represented introns. Exons with different structures among the two family genes in the same subgroups were marked in yellow. Numbers above the exons or below the introns of each gene structure represented the size of exon and intron, respectively.
doi:10.1371/journal.pone.0046744.g006

ative synteny map among *B. distachyon*, rice and *Arabidopsis* genomes. *Arabidopsis* and rice are two most important eudicot and monocot model plant species and the functions of many *Arabidopsis* and rice genes have been well characterized. Genomic comparison is a quick way to transfer genomic knowledge acquired in one taxon to a less-studied taxon [44,53]. Thus, through comparative genomics analysis we could confidently infer the functions of *B. distachyon* MAPK and MAPKK genes based on their *Arabidopsis* and rice homologs.

Through large-scale synteny analysis between *B. distachyon* and the other two species of rice and *Arabidopsis*, 50% of MAPK genes and 42% of MAPKK genes from *B. distachyon* were identified to be synteny with their orthologs of rice genome, while there was only one gene (*BdMPK4*) in these two families found being synteny with *AtMPK11* of *Arabidopsis* (Figure 7). The result was consistent with

the current understanding of plant evolutionary history. From Figure 7, we also found that their homologs of the tightly linked genes in the *B. distachyon* chromosome were still tightly linked in the rice chromosome with identical or opposite orientation such as the pair of *BdMPK20-1/BdMPK20-4* and *OsMPK20-1/OsMPK20-4*, the pair of *BdMKK5/BdMPK6/BdMKK1* and *OsMKK1/OsMPK6/OsMKK5*, and the pair of *BdMPK3/BdMKK10-2* and *OsMKK10-2/OsMPK3*. Such instance could be explained as random translocations and insertion events or thought of that the progenitor species had donor region (a tightly linked cluster of related or unrelated sequences in one part of the genome) and the *B. distachyon* or rice accepted this region in the process of species evolution. More challenging for synteny interpretation were cases where duplicated *B. distachyon* MAPK or MAPKK genes yet corresponded to one or two rice homologs and the closest homologs in the phylogenetic

tree were not the same with that of syntenic analysis such as *BdMPK7-1/BdMPK14-OsMPK14* and *BdMCK4/BdMCK5-OsMCK4/OsMCK5* (Figure 7). Of course, there were also some MAPK and MAPKK genes of *B. distachyon* that could not mapped to any syntenic blocks with the other two species. However, we could not conclude that the two families from *B. distachyon*, rice and *Arabidopsis* did not share a common ancestor. That may be the reason that their genomes have undergone multiple rounds of significant chromosomal rearrangement and fusions, followed by selective gene loss, which can severely obscure the identification of chromosomal syntenic blocks during the speciation of *B. distachyon*, rice and *Arabidopsis* [44].

Analyses of the interactions between MAPK and MAPKK family members

MAPK cascades are conserved signaling modules found in all eukaryotic cells, including plants, fungi and animals and they play a remarkably important role in plant signaling of a variety of abiotic and biotic stresses. However, so far only a few *Arabidopsis* MAPK cascade components and their interaction have been studied in detail. In order to unravel the interactions between the last two kinases within the MAPK cascade in *B. distachyon*, we studied their interactions between several genes of the two families

using yeast two-hybrid assay. Four MAPKK genes (five proteins generated from alternative splicing) and six MAPK genes with more EST counts support were cloned into the respective DNA-binding domain and GAL4 activation domain plasmids. After transformation into the yeast strain YH2Gold, interaction was monitored by growth on interaction-selecting media lacking His. As shown in Figure 8, *BdMPKK3-1* exhibited a significant interaction with *BdMPK7-1* and weak interaction with *BdMPK14*, while its sister gene *BdMPKK3-2* derived from gene duplicate event did not show any affinity toward the pair of duplicated *BdMPK7-1/BdMPK14* and other MAPKK genes. The interactive partners between *BdMPKK3-1* and *BdMPK7-1/BdMPK14* were also found in their homologs of *Arabidopsis* which were activated by Pathogen and JA [5]. In addition, the two isoforms (*BdMAKK6.1* and *BdMAKK6.2*) of *BdMCK6* generated by alternative splicing were also cloned into the DNA-binding domain plasmids in order to investigate whether they had similar function. From Figure 8, we can find that they indeed interacted with the same gene *BdMPK6* but had different affinity, which may be due to tiny difference of their amino sequences. Therefore, the exact structural features determining the specificity of complex formation remain to be uncovered. Similarly, the two isoforms also interacted only with *BdMPK6* and did not interact with the sister gene *BdMPK3* of

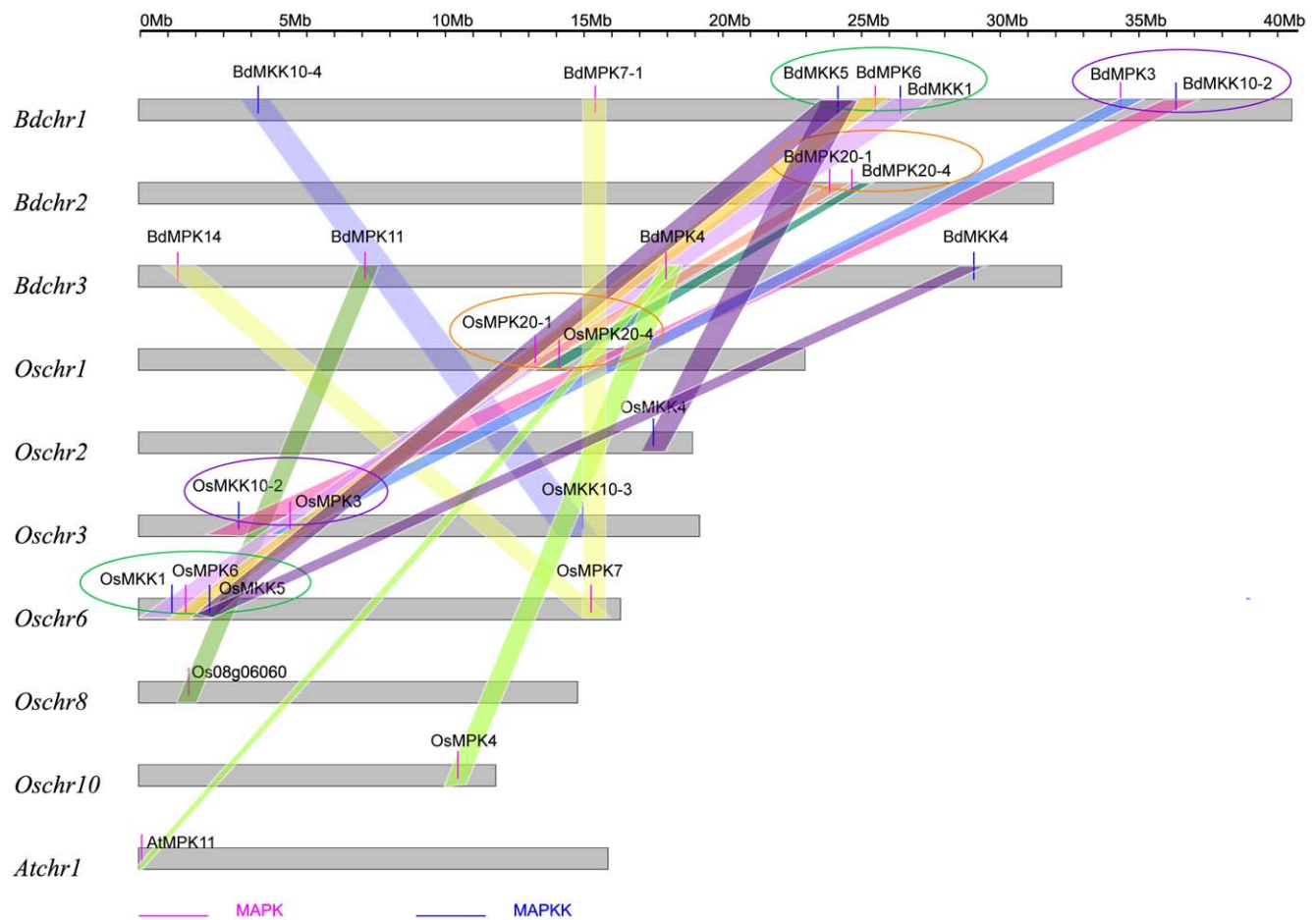


Figure 7. Syntenic analyses of MAPKs and MAPKKs between *B. distachyon* and the other two plant species of rice and *Arabidopsis*. To identify the species of origin for each chromosome, a species acronym is included before the chromosome name: At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Bd, *B. distachyon*. Different color bars connect syntenic regions between *B. distachyon* and the other two plant species of *Arabidopsis* and rice chromosomes. The region of *B. distachyon* tightly linked cluster likely to be syntenic with that of the rice tightly linked cluster by translocation, insertion or other events was indicated with the ellipse in the same color. doi:10.1371/journal.pone.0046744.g007

BdMPK6 as the case of *BdMPK3-1*. So far, the interactive partner of *BdMCK6/BdMPK6* has not been found in their homologs of *Arabidopsis* and other plants. That is maybe the reason that there were few reports on the MAPKK family, especially on the MKK6 gene. More interestingly, the homologs of *BdMPK3/BdMPK6* in *Arabidopsis* are closely related genes that show a high level of functional redundancy [4–5] and always accompany together in the previous reported signaling pathway. However, the two duplicated genes of *BdMPK3/BdMPK6* in *B. distachyon* were not the case from the results of interactive yeast two-hybrid interaction assay, which indicated that sequence similarity and evolutionary history were not sufficient to predict MAPKK-MAPK interactions; their interaction profiles found in one model plant did not always exist in other model plants and also suggested that even a high conservation of certain MAPKKs or MAPKs does not necessarily represent the functional redundancy.

Expression profiles of MAPK and MAPKK genes in *B. distachyon*, *Arabidopsis* and rice under normal grow and various abiotic stress conditions

The expression profiles of *B. distachyon* MAPK and MAPKK genes. To observe the organ-specificity of the two family members, we confirmed their transcript level in four different organs: roots, stems, leaves and caryopses. For MAPK family, the vast majority of its members were expressed in all the detected organs except the *BdMPK21-1* and *BdMPK21-2* (Figure 9). *BdMPK20-3* had higher expression level than other members of MAPK family in all the tested corresponding organs. The expression of *BdMPK6* was induced in stem, leaf and caryopses, while its closely duplicated sister gene *BdMPK3* was not the case. Similar result was observed in the duplicated pair of *BdMPK7-1/BdMPK7-2* and other duplicated gene pairs which both existed in MAPK and MAPKK families. These results showed that although the duplicated genes had higher similarities in amino acid, we can't speculate that they should have similar function or were involved the same signaling pathway. Therefore their expression

intensity and/or organ-specificity varied depending on their specific function but not on the sequence similarity. In addition, we increased the mRNA amount of each organ and the cycles of RT-PCR in order to investigate whether *BdMPK21-1* and *BdMPK21-2* that were not detected in the same experimental condition as other members had trace expression in the undetected organs. However, we still could not detect the expression of *BdMPK21-1* in roots and stems or the expression of *BdMPK21-2* in all the four organs.

Compared with MAPK family, the expression intensity of each MAPKK member varied slightly greater in different organs than that of each MAPK member. For example, 81.25% of MAPK members had relatively uniform high level expression under our four detected tissues, while there were only 33.33% of MAPKK genes with similar expression level. Fifty percent of MAPKK genes were induced or up-regulated in caryopses compared with that in roots, whereas only 18.75% MAPK genes were the case and most of them had no significant difference. From these results, it is interesting to speculate that most of MAPK genes with relative uniform high expression levels in all the *B. distachyon* detected organs might play key roles in plant development and half of MAPKK genes induced in caryopses may represent their unique functions in this developmental stage. However, more researches are needed to determine the functions of the two family genes through other biological experimental technology.

Evidence is accumulating that MAPK and MAPKK proteins are involved in responding to various abiotic stresses. Thus, we used RT-PCR method to examine the expression of the two family genes in response to six different abiotic stresses or hormone treatments: drought, salinity, cold, ABA, MeJA and H₂O₂. From the results we found that each member of the two family genes showed differential expressions in response to at least two treatments except for *BdMPK21-2* that did not show any detectable expression in all treatments as that in the organs mentioned above, which indicated that either this gene is expressed in other organs or under other stress treatments, expressed with low level or it is a pseudogene (Figure 9, Figure 10). And 61% of the two family genes were

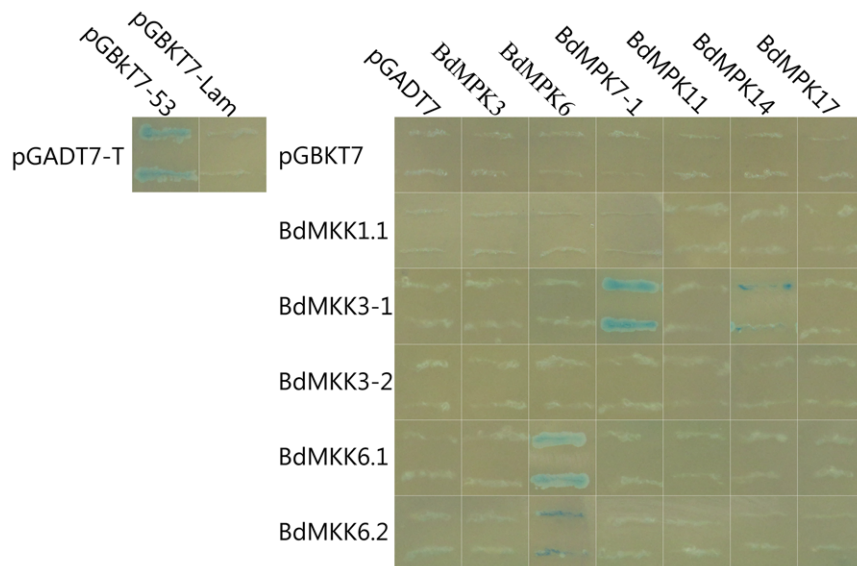


Figure 8. Comparative yeast two-hybrid interaction analyses of six *BdMPKs* with five *BdMKKs*. The yeast strain YH2Gold containing the indicated plasmid combinations was grown on nutritional selective medium minus His, Leu, Trp and plus X- α -Gal (SD/-His-Leu-Trp/X- α -Gal, TDO/X). Positive or negative control on the left corner showed the interaction between SV40 large T-antigen and murine p53 or SV40 large T-antigen and human lamin C (Lam) examined on TDO/X, respectively. doi:10.1371/journal.pone.0046744.g008

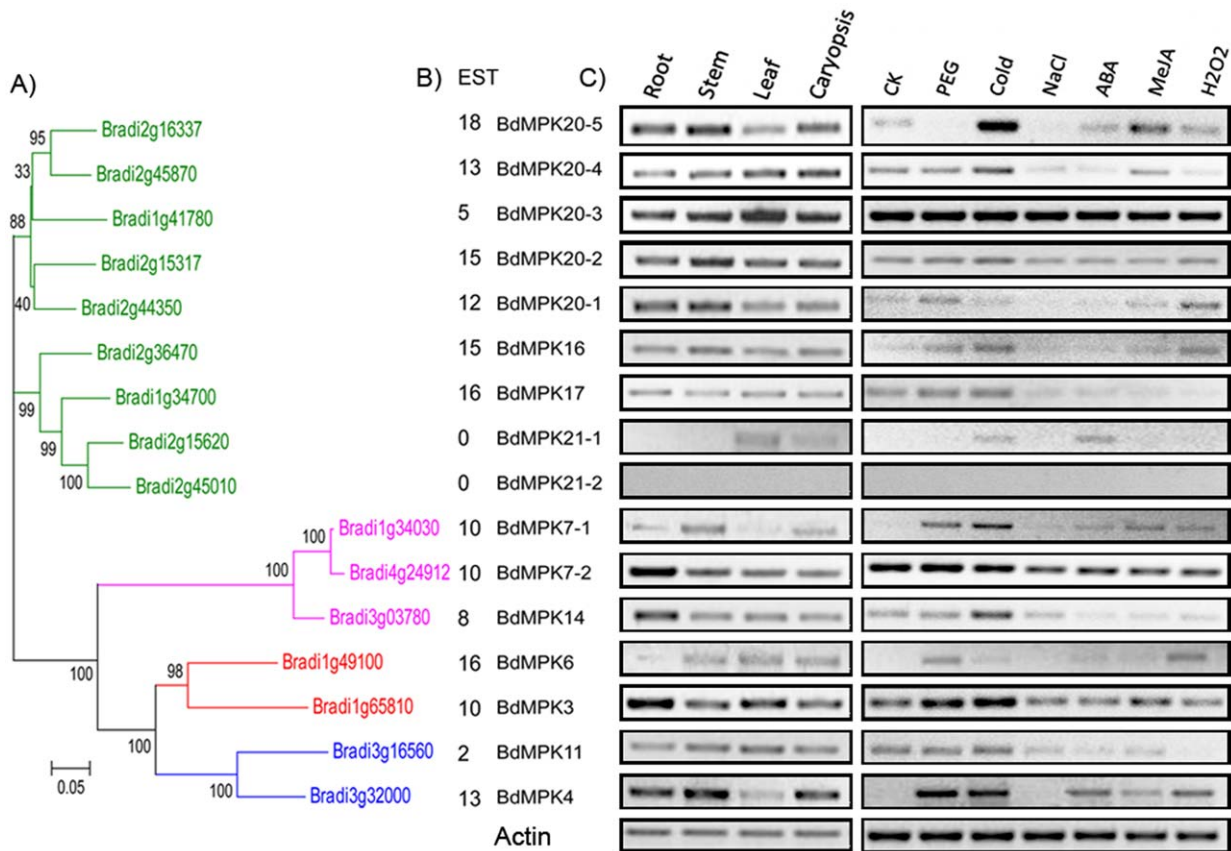


Figure 9. Phylogenetic (A), EST (B) and RT-PCR (C) analyses of *Brachypodium* MAPK genes. The gene names of each gene cluster originated from tandem or segmental duplication event were indicated with the same color. doi:10.1371/journal.pone.0046744.g009

induced or constitutively expressed in all the six treatments, implying their important roles in the stress tolerance. Of course, there were some genes that were up-regulated and down-regulated in one treatment but not in all the treatments. For example, 43.75% of MAPK family genes were up-regulated and only one gene was down-regulated compared with the control in the PEG treatment, while 66.7% of MAPKK genes were up-regulated and none of them were down-regulated. More than fifty six percent of MAPK family genes were up-regulated and none of them was down-regulated compared with the control in the cold treatment, while 16.7% of MAPKK genes were up-regulated and one of them was down-regulated. Taken together, the majority of the two family genes were expressed or induced under the stress conditions of drought, salt and cold, ABA, MeJA and H₂O₂ respectively, implying that these members might be putative regulators in response to various abiotic stress or hormone treatments. Moreover, we also found that the genes of the two families with EST support were expressed with higher level or in a much wider range of libraries, including those constructed from various stress conditions and organs than that with no EST support. However, there was no clear evidence of a correlation between the expression intensity by RT-PCR and the EST counts of each gene from the two families. For example, the four genes *BdMCK10-1*, *BdMCK10-3*, *BdMCK10-4* and *BdMCK10-4* derived from tandem duplication event had no EST support, but varying expression intensity was detected in some tissues or stress conditions, which showed that the expression patterns of the two family genes were highly variable and were not strongly associated with their EST counts or sequence similarity.

The expression profiles of *Arabidopsis* MAPK and MAPKK genes. To observe the organ-specificity and the responses to stress conditions of the two family members in *Arabidopsis*, a total of 4 experiments including 201 hybridizations from the Affymetrix *Arabidopsis* genome array were obtained. All the genes of the two families corresponding to 30 probe sets were identified in the 4 experiments. Detailed expression of the two family genes under different developmental stages was provided in Table S3. Heatmap representation of expression profiles of these two family genes during *Arabidopsis* development was shown in Figure 11. For the MAPK family genes, *AtMPK5* and *AtMPK11* had higher expression in rosette, flower, and seeds than that of other organs. *AtMPK1*, *AtMPK10*, *AtMPK12*, *AtMPK13* and *AtMPK14* were expressed with low abundance in nearly all the organs, whereas *AtMPK3*, *AtMPK4*, *AtMPK6*, *AtMPK17* and *AtMPK20* were constitutively expressed with high abundance in nearly all the organs (Figure 11a). More interestingly, this case was consistent with the expression profiles of their homologs in *B. distachyon*. For example, *BdMPK3*, *BdMPK4*, *BdMPK6*, *BdMPK17* and all the members of *BdMPK20* clade were constitutively expressed with high level in nearly all the *B. distachyon* detected organs. For the MAPKK family genes, *AtMKK2*, *AtMKK5* and *AtMKK9* clustered together in the heatmap with high expression in the majority of organs, while the clustered *AtMKK7*, *AtMKK8* and *AtMKK10* had low expression in nearly all the organs. The expression level of other members varied depending on different organs.

In addition, we also identified the expression profiles of the two family genes under cold, drought and salt stress conditions through

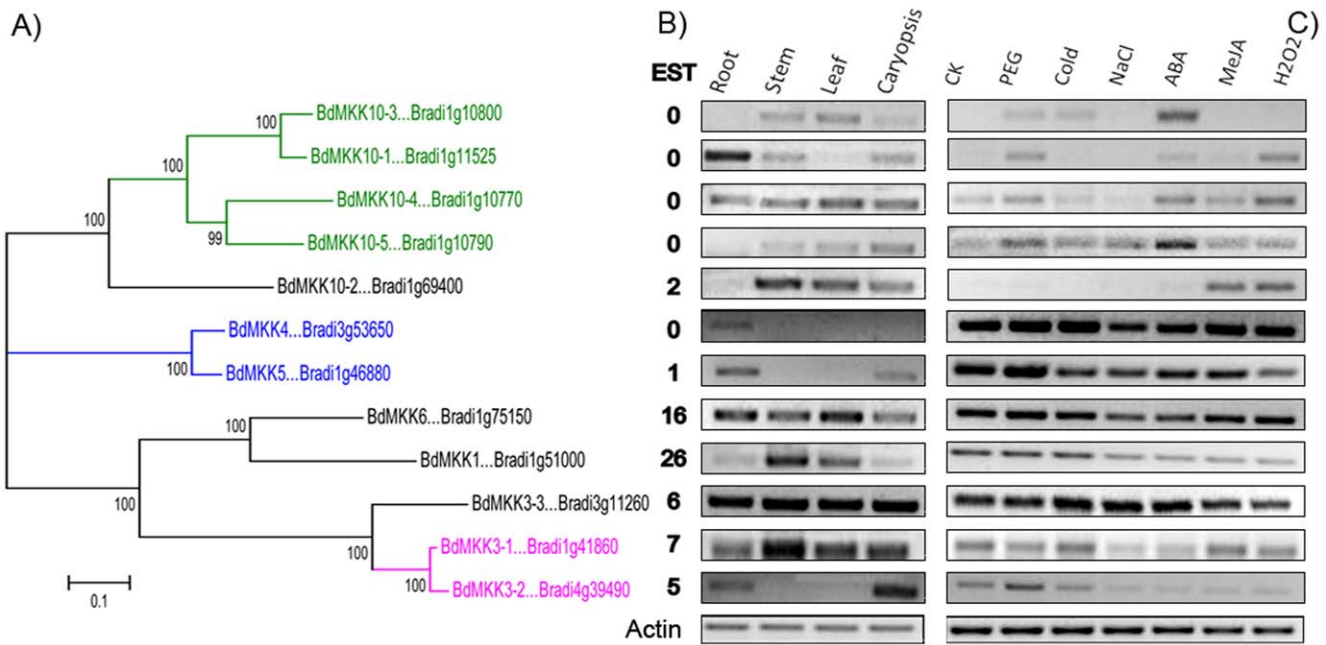


Figure 10. Phylogenetic (A), EST (B) and RT-PCR (C) analyses of MAPKK genes. The gene names of each gene cluster originated from tandem or segmental duplication event were indicated with the same color (except the black color). doi:10.1371/journal.pone.0046744.g010

analysis of publicly available microarray data sets. Details of the experimental analytic result were provided in Table S4. A total of 5 MAPK genes (*AtMPK3*, *AtMPK5*, *AtMPK7*, *AtMPK11* and *AtMPK19*) were significantly up-regulated and 4 MAPK genes (*AtMPK7*, *AtMPK8*, *AtMPK16* and *AtMPK20*) were significantly down-regulated under the cold or salt stress conditions according to the strict criteria of up-regulated or down-regulated genes described in the method section. Among the 5 up-regulated genes, *AtMPK3* was identified under the cold treatment, which was consistent with previous report that the expression of *AtMEKK1* and *AtMPK3* could be induced by cold [54]. Actually, *AtMPK3* was one of the best-characterized members in the *Arabidopsis* MAPKs and involved in many signal pathways activated by a diversity of stimuli including abiotic stresses, pathogens and oxidative stress. For the MAPKK family, four genes were up-regulated or down-regulated under one or two stress conditions and *AtMKK9* was up-regulated under both the cold and salt treatments. However, none of MAPK and MAPKK members was identified to be significantly up-regulated or down-regulated under the drought treatment. Besides these genes in the two families which showed significant changes in response to the stress conditions, there are also lots of genes showing tiny changes under different stress treatments as provided in Table S4.

The expression profiles of rice MAPK and MAPKK genes. Similarly, the expression profiles of the two family genes in rice under different developmental stages and the three stress conditions mentioned above were also examined as that in *Arabidopsis* using the available microarray database. A total of five experiments including 70 hybridizations from the Affymetrix rice genome array were obtained. Twenty two genes of the two families corresponding to 27 probe sets were identified in the five experiments and there was only one representative MAPKK genes whose corresponding probe was not found. Detailed expression profiles of the two family genes under different developmental stages were provided in Table S5. Heatmap representation of expression profiles of these two family genes

during rice development was shown in Figure 12. There were almost 60% of MAPK genes clustered together in the heatmap with high expression in nearly all the organs except *OsMPK20-5* in anther and *OsMPK3* in ovary and shoot apical meristem (SAM). The expression profiles of these genes happened to be consistent with that of their homologs in *B. distachyon* and *Arabidopsis* except *AtMPK7* and *AtMPK14*, especially the *AtMPK14* which not only had low expression in all the detected organs but also had no significant fluorescent signal in all the stress conditions. More interestingly, the expression profiles of two paralogs *OsMPK17-1* and *OsMPK17-2* originated from tandem duplication event were quite different under the majority of rice developmental stages, which may reveal that after duplication one daughter gene retained the ancestral function while the other acquired new functions [55]. However, the expression profiles of duplicated paralogs *OsMPK21-1* and *OsMPK21-2* showed highly similar expression intensities, which may indicate subfunctionalization in the course of evolution. Besides this case, there were also several cases existing in the homologs of MAPKK family. For example, *OsMKK1* and *OsMKK6* clustered together in the heatmap were constitutively expressed with high expression in all the tissues and their homologs in *B. distachyon* also had similar expression profiles. *OsMKK10-1* had low expression in all the tissues and its closest related gene *AtMKK10* in the phylogenetic tree showed similar expression profile.

When using microarray data to investigate the response of the two family genes to abiotic stresses, we found that 80% of *OsMPK* genes and 50% of *OsMPKK* genes were up-regulated or down-regulated under one or more stress conditions (Table S4). For example, *OsMPK17-1* and *OsMKK4* were both up-regulated under the three stress conditions, while the duplicated pair of *OsMPK20-2* and *OsMPK20-3* were both down-regulated. However, we also found that some closely related genes of the two families in the three species with similar expression profiles under different developmental stages did not show similar expression profiles

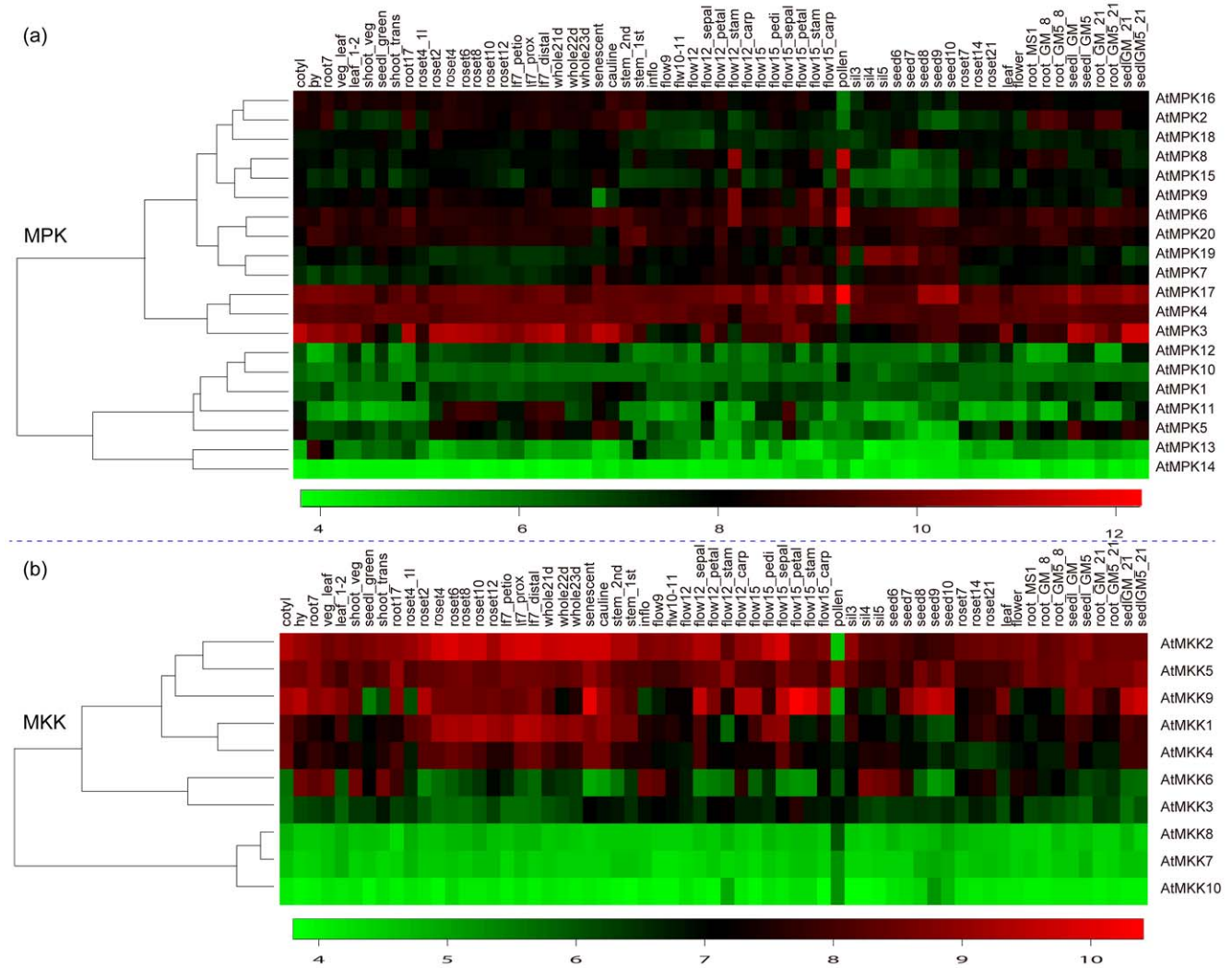


Figure 11. Expression profiles of 20 MAPK genes (a) and 10 MAPKK (b) genes in different *Arabidopsis* tissues. Details of the experimental conditions are provided in Table S3. Log₂ based value was used to create the heatmap. Difference in gene expression changes is shown in color as the scale.

doi:10.1371/journal.pone.0046744.g011

when they were under the three stress conditions and *vice versa* (Table S4).

Pairwise comparisons of the expression profiles of putative orthologous or paralogs pairs existing in the two family genes of *B. distachyon* and rice under different light and temperature conditions

The expression profiles of the two family genes of *B. distachyon* and rice under different light and temperature conditions were shown in Figure S3 and Figure S4. Most of them had comparative uniform expression profiles, while the expression intensity of *BdMPK14*, *BdMPK20-3*, *BdMPK3-3*, *BdMkk4* and *OsMPK20-1*, *OsMPK20-2*, *OsMPK20-5*, *OsMkk4* varied comparatively greater than that of others. In order to investigate whether the expression profiles of these duplicated orthologs or paralogs pairs at the terminal nodes of phylogenetic tree were correlated in the two plant species, we compared their expression profiles of them under different light and temperature conditions (Table 2). The results showed that there were only three sets of data whose correlation coefficients were greater than 0.5, indicating a positive correlation

between these orthologs pairs under different light conditions such as the pair of *BdMPK20-3* and *OsMPK20-3* (Figure 13A), while there was a set of data, whose correlation coefficients were not lower than -0.5 but may have certain negative correlation such as the pair of *BdMPK20-4* and *OsMPK20-4* (Figure 13B). In addition, we also found that some sets of data whose correlation coefficients were very small and had no clear positive or negative correlation under the different light and temperature conditions such as the pair of *BdMkk4* and *OsMkk4* (Figure 13C). The comparison analysis results indicated that the duplicated gene pair may remain parts of its ancestral function or showed divergent function under different stress conditions in the evolutionary process. However, the reasons for expression divergence of the duplicated genes in many gene families remain unclear. Previous study showed that the long-term evolutionary fate of duplicated orthologs or paralogs pairs would be determined by the functions of the duplicated genes. There are four types of functional differentiation following gene duplication: pseudogenization, conservation of gene function, subfunctionalization and neofunctionalization [55]. Many duplicated genes may be lost from the genome after the duplication

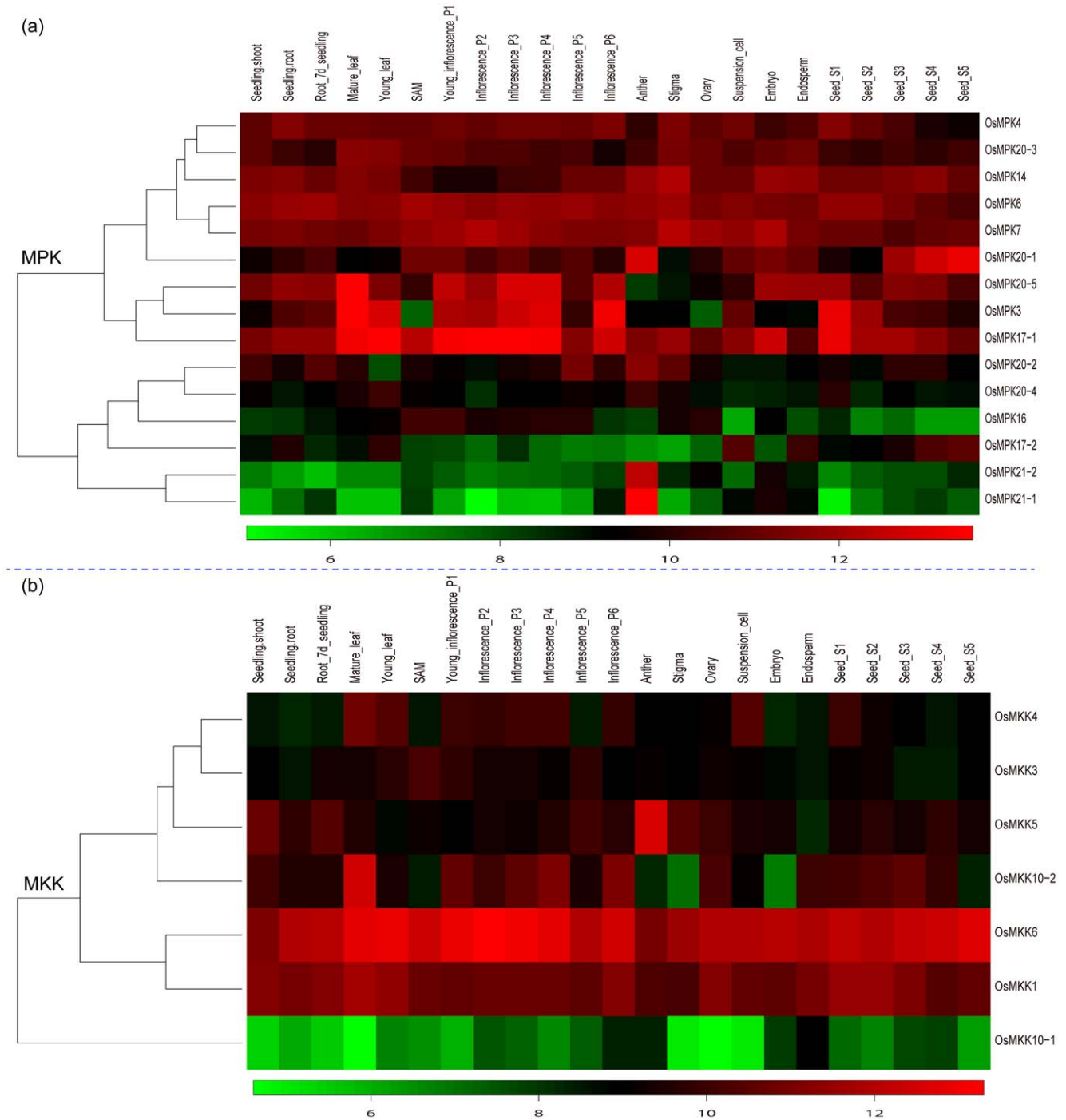


Figure 12. Expression profiles of 15 MAPK genes (a) and 8 MAPKK (b) genes in different rice tissues. Details of the experimental conditions are provided in Table S4. Log₂ based value was used to create the heatmap. Difference in gene expression changes is shown in color as the scale.

doi:10.1371/journal.pone.0046744.g012

events, and neofunctionalization and subfunctionalization are the major factors for the retention of new genes. Our analytic results of the duplicated gene pairs were consistent with this viewpoint. Besides these, the expression patterns of the two isoforms (BdMKK6.1 and BdMKK6.2) derived from alternative splicing event were also analyzed. More interestingly, the expression of one isoform BdMKK6.1 with longer protein sequence was identified in all the detected tissues and stress conditions and another one

BdMKK6.2 was just opposite by traditional RT-PCR. When using the microarray data under different light conditions, we found that the isoform BdMKK6.1 happened to be expressed with higher level than BdMKK6.2. However, the correlation coefficient of their expression profiles was very low (0.181). It may be the reason that they were not involved in this light pathway but played important roles in other signal pathways or that they indeed had some difference in their functions.

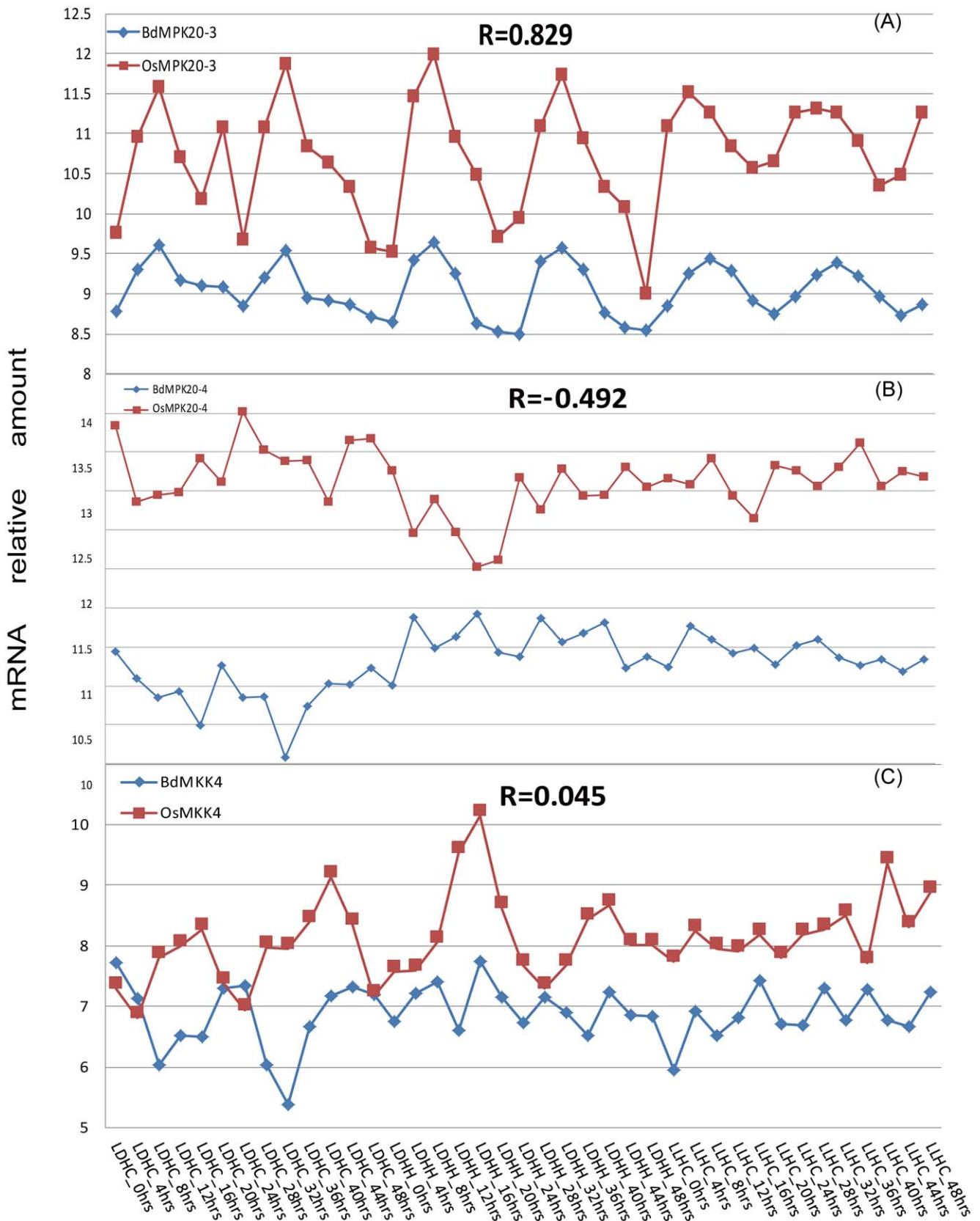


Figure 13. Pairwise comparisons of the expression profiles of putative orthologs between rice and *B. distachyon* MAPKs and MAPKKs. The mRNA relative amount represents the log₂ based value of expression intensity and the expression profiles of the two family genes in rice and *B. distachyon* were provided in Table S6. The treatment time (h) under the particular light and temperature condition is presented on the x-axis. R indicates the correlation coefficient of the expression profiles between orthologs or paralogs pairs under the corresponding light and

temperature treatments. The positive (A), negative (B) and no obvious correlation (C) were detected from our identified 17 pairs of putative orthologs or paralogs in rice and *B. distachyon* MAPKs and MAPKKs.
doi:10.1371/journal.pone.0046744.g013

Conclusion

Out of many signaling pathways involved in biotic and abiotic stress response in plants, MAPK cascade is one of the major pathways. This signaling module links external stimuli with several cellular responses and is evolutionary conserved among eukaryotic organisms [2,56]. In plants, MAP kinases are represented by multigene families and are involved in efficient transmission of specific stimuli and also involved in the regulation of the antioxidant defense system in response to stress signaling [54]. So far, significant progress has been made toward the identification and characterization of MAPK and MAPKK gene families in several model plants, while no systematic analysis for the two family genes was reported in *B. distachyon*. In this present study, we carried out a genome-wide survey of the two gene families in *B. distachyon* and characterized them on the bases of structural diversity, phylogenetic relationship, conserved protein motif, chromosomal location, gene duplication, exon/intron organization, promoter region, interaction between them and expression profiles. In addition, we also analyzed the expression patterns of the two family genes in *Arabidopsis* and rice under different developmental stages as well as under various stress conditions and comparatively analyzed the expression profiles of putative orthologous or paralogs pairs existing in the two family genes of *B. distachyon* and rice under different light and temperature conditions. Therefore, our genomic and bioinformatics analysis of the two family genes and proteins presented in this work would provide an important foundation for the further functional dissection of the last two kinases of MAPK cascade involving in the important signaling pathways in different organs and under different stress conditions. Beside these, overexpression, knock-

down or mutagenesis, and promoter analysis of selected members of the two families are underway in our laboratory so that we can accurately determine molecular pathways in the MAPK and MAPKK gene families.

Supporting Information

Figure S1 A detailed motif and subdomain introductions for *Brachypodium* MAPKs.

(TIF)

Figure S2 Schematic depictions of alternatively spliced BdMCKK1 and BdMCKK6 genes.

(TIF)

Figure S3 Expression profiles of BdMPKs (a) and BdMCKKs (b) under different light and temperature conditions. Details of the experimental conditions are provided in Table S6. Log₂ based value was used to create the heatmap. Difference in gene expression changes is shown in color as the scale.

(TIF)

Figure S4 Expression profiles of OsMPKs (a) and OsMCKKs (b) under different light and temperature conditions. Details of the experimental conditions are provided in Table S6. Log₂ based value was used to create the heatmap. Difference in gene expression changes is shown in color as the scale.

(TIF)

Table S1 The primer sequences used for RT-PCR amplification and gene cloning of the two family genes.

(XLSX)

Table S2 Annotation of the *Brachypodium distachyon* MAPK and MAPKK gene families.

(XLSX)

Table S3 The expression profiles (log₂ based value) of the two family genes of *Arabidopsis* under different developmental stages.

(XLSX)

Table S4 The expression profiles of the two family genes of *Arabidopsis*, *Brachypodium* and rice under three different abiotic stress conditions.

(XLSX)

Table S5 The expression profiles (log₂ based value) of the two family genes of rice under different developmental stages.

(XLSX)

Table S6 The expression profiles (log₂ based value) of the two family genes of *Brachypodium* and rice under different and temperature conditions.

(XLSX)

Author Contributions

Conceived and designed the experiments: LC WH ST. Performed the experiments: LC WH ZM MW SZ YZ XD CH. Analyzed the data: LC ST WH. Wrote the paper: LC. Provided guidance on the whole study: LC HW GY GH.

Table 2. Pearson correlation coefficients for the expression profiles of orthologs or paralogs pairs.

Family	Gene	Gene	Similarity ^a	Correlation coefficient
MAPK	<i>BdMPK3</i>	<i>OsMPK3</i>	91.9%	-0.113
	<i>BdMPK4</i>	<i>OsMPK4</i>	94.4%	0.113
	<i>BdMPK6</i>	<i>OsMPK6</i>	94.5%	0.594
	<i>BdMPK14</i>	<i>OsMPK14</i>	96.2%	0.311
	<i>BdMPK16</i>	<i>OsMPK16</i>	81.3%	-0.345
	<i>BdMPK17</i>	<i>OsMPK17-1</i>	86.5%	-0.097
	<i>BdMPK20-1</i>	<i>OsMPK20-1</i>	90.4%	0.276
	<i>BdMPK20-3</i>	<i>OsMPK20-3</i>	65.6%	0.829
	<i>BdMPK20-4</i>	<i>OsMPK20-4</i>	86.8%	-0.492
	<i>BdMPK21-1</i>	<i>OsMPK21-1</i>	81.9%	0.075
MAPKK	<i>BdMCKK1</i>	<i>OsMCKK1</i>	87.2%	-0.289
	<i>BdMCKK3-1</i>	<i>BdMCKK3-2</i>	90.2%	0.263
	<i>BdMCKK4</i>	<i>OsMCKK4</i>	90.8%	0.045
	<i>BdMCKK5</i>	<i>OsMCKK5</i>	82.9%	0.396
	<i>BdMCKK6</i>	<i>OsMCKK6</i>	93.3%	0.204
	<i>BdMCKK10-2</i>	<i>OsMCKK10-2</i>	84.8%	-0.066
	<i>BdMCKK10-4</i>	<i>BdMCKK10-5</i>	69.9%	0.657

^arepresents the similarity of amino acid.

doi:10.1371/journal.pone.0046744.t002

References

- Romeis T (2001) Protein kinases in the plant defence response. *Curr Opin Plant Biol* 4: 407–414.
- Tena G, Asai T, Chiu WL, Sheen J (2001) Plant mitogen-activated protein kinase signaling cascades. *Curr Opin Plant Biol* 4: 392–400.
- Zhang S, Klessig DF (2001) MAPK cascades in plant defense signaling. *Trends Plant Sci* 6: 520–527.
- Pitzschke A, Schikora A, Hirt H (2009) MAPK cascade signalling networks in plant defence. *Curr Opin Plant Biol* 12: 421–426.
- Colcombet J, Hirt H (2008) *Arabidopsis* MAPKs: a complex signalling network involved in multiple biological processes. *Biochem J* 413: 217–226.
- Teige M, Scheikl E, Eulgem T, Doczi R, Ichimura K, et al. (2004) The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Mol Cell* 15: 141–152.
- Gao M, Liu J, Bi D, Zhang Z, Cheng F, et al. (2008) MEKK1, MKK1/MKK2 and MPK4 function together in a mitogen-activated protein kinase cascade to regulate innate immunity in plants. *Cell Res* 18: 1190–1198.
- Qiu JL, Zhou L, Yun BW, Nielsen HB, Fil BK, et al. (2008b) *Arabidopsis* mitogen-activated protein kinase kinases MKK1 and MKK2 have overlapping functions in defense signaling mediated by MEKK1, MPK4, and MKS1. *Plant Physiol* 148: 212–222.
- Chen T, Zhu H, Ke D, Cai K, Wang C, et al. (2012) A MAP kinase kinase interacts with SymRK and regulates nodule organogenesis in *Lotus japonicus*. *Plant Cell* 24: 823–838.
- Takahashi F, Yoshida R, Ichimura K, Mizoguchi T, Seo S, et al. (2007) The mitogen-activated protein kinase cascade MKK3-MPK6 is an important part of the jasmonate signal transduction pathway in *Arabidopsis*. *Plant Cell* 19: 805–818.
- Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, et al. (2002) MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* 415: 977–983.
- Wang H, Ngwenyama N, Liu Y, Walker JC, Zhang S (2007) Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in *Arabidopsis*. *Plant Cell* 19: 63–73.
- Lee SK, Kim BG, Kwon TR, Jeong MJ, Park SR, et al. (2011) Overexpression of the mitogen-activated protein kinase gene OsMAPK33 enhances sensitivity to salt stress in rice (*Oryza sativa* L.). *J Biosci* 36: 139–151.
- Boudsocq M, Lauriere C (2005) Osmotic signaling in plants: multiple pathways mediated by emerging kinase families. *Plant Physiol* 138: 1185–1194.
- Meszáros T, Helfer A, Hatzimasoura E, Magyar Z, Serazetdinova L, et al. (2006) The *Arabidopsis* MAP kinase kinase MKK1 participates in defence responses to the bacterial elicitor flagellin. *Plant J* 48: 485–498.
- Qiu JL, Fil BK, Petersen K, Nielsen HB, Botanga CJ, et al. (2008a) *Arabidopsis* MAP kinase 4 regulates gene expression through transcription factor release in the nucleus. *EMBO J* 27: 2214–2221.
- Wilson C, Voronin V, Touraev A, Vicente O, Heberle-Bors E (1997) A developmentally regulated MAP kinase activated by hydration in tobacco pollen. *Plant Cell* 9: 2093–2100.
- Gudesblat GE, Iusem ND, Morris PC (2007) Guard cell-specific inhibition of *Arabidopsis* MPK3 expression causes abnormal stomatal responses to abscisic acid and hydrogen peroxide. *New Phytol* 173: 713–721.
- Hamel LP, Nicole MC, Sritubtim S, Morency MJ, Ellis M, et al. (2006) Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci* 11: 192–198.
- Xu J, Li Y, Wang Y, Liu H, Lei L, et al. (2008) Activation of MAPK kinase 9 induces ethylene and camalexin biosynthesis and enhances sensitivity to salt stress in *Arabidopsis*. *J Biol Chem* 283: 26996–27006.
- Yoo SD, Sheen J (2008) MAPK signaling in plant hormone ethylene signal transduction. *Plant Signal Behav* 3: 848–849.
- Zhou C, Cai Z, Guo Y, Gan S (2009) An *Arabidopsis* mitogen-activated protein kinase cascade, MKK9-MPK6, plays a role in leaf senescence. *Plant Physiol* 150: 167–177.
- Beck M, Komis G, Muller J, Menzel D, Samaj J (2010) *Arabidopsis* homologs of nucleus- and phragmoplast-localized kinase 2 and 3 and mitogen-activated protein kinase 4 are essential for microtubule organization. *Plant Cell* 22: 755–771.
- Beck M, Komis G, Ziemann A, Menzel D, Samaj J (2010) Mitogen-activated protein kinase 4 is involved in the regulation of mitotic and cytokinetic microtubule transitions in *Arabidopsis thaliana*. *New Phytol* 189: 1069–1083.
- Zeng Q, Chen JG, Ellis BE (2011) AtMPK4 is required for male-specific meiotic cytokinesis in *Arabidopsis*. *Plant J* 67: 895–906.
- Takahashi Y, Soyano T, Kosetsu K, Sasabe M, Machida Y (2010) HINKEL kinesin, ANP MAPKKKs and MKK6/ANQ MAPKK, which phosphorylates and activates MPK4 MAPK, constitute a pathway that is required for cytokinesis in *Arabidopsis thaliana*. *Plant Cell Physiol* 51: 1766–1776.
- MAPK Group (2002) Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci* 7: 301–308.
- Rao KP, Richa T, Kumar K, Raghuram B, Sinha AK (2010) *In silico* analysis reveals 75 members of mitogen-activated protein kinase kinase gene family in rice. *DNA Res* 17: 139–153.
- Bevan MW, Garvin DF, Vogel JP (2010) *Brachypodium distachyon* genomics for sustainable food and fuel production. *Curr Opin Biotechnol* 21: 211–217.
- IBI (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463: 763–768.
- Kumar S, Mohan A, Balyan HS, Gupta PK (2009) Orthology between genomes of *Brachypodium*, wheat and rice. *BMC Res Notes* 2: 93.
- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39: W29–37.
- Thompson JD, Higgins DG, Gibson TJ (1994) Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
- Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* 3: 21–29.
- Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH (2002) Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol Biol Evol* 19: 256–262.
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27: 297–300.
- Jang CS, Kamps TL, Skinner DN, Schulze SR, Vencill WK, et al. (2006) Functional classification, genomic organization, putatively cis-acting regulatory elements, and relationship to quantitative trait loci, of sorghum genes with rhizome-enriched expression. *Plant Physiol* 142: 1148–1159.
- Dong J, Chen C, Chen Z (2003) Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response. *Plant Mol Biol* 51: 21–37.
- Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K, et al. (2006) Dual function of an *Arabidopsis* transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc Natl Acad Sci U S A* 103: 18822–18827.
- Ohme-Takagi M, Suzuki K, Shinshi H (2000) Regulation of ethylene-induced transcription of defense genes. *Plant Cell Physiol* 41: 1187–1192.
- Tang H, Wang X, Bowers JE, Ming R, Alam M, et al. (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res* 18: 1944–1954.
- Albrecht V, Ritz O, Linder S, Harter K, Kudla J (2001) The NAF domain defines a novel protein-protein interaction module conserved in Ca²⁺-regulated kinases. *EMBO J* 20: 1051–1063.
- Zhang Y, Mao L, Wang H, Brocker C, Yin X, et al. (2012) Genome-wide identification and analysis of grape aldehyde dehydrogenase (ALDH) gene superfamily. *PLoS One* 7: e32153.
- Parkinson H, Kapushesky M, Kolesnikov N, Rustici G, Shojatalab M, et al. (2009) ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res* 37: D868–872.
- Wise RA, Teeter JG, Jensen RL, England RD, Schwartz PF, et al. (2007) Standardization of the single-breath diffusing capacity in a multicenter clinical trial. *Chest* 132: 1191–1197.
- Jung KH, Dardick C, Bartley LE, Cao P, Phetsom J, et al. (2008) Refinement of light-responsive transcript lists using rice oligonucleotide arrays: evaluation of gene-redundancy. *PLoS One* 3: e3337.
- Sharoni AM, Nuruzzaman M, Satoh K, Shimizu T, Kondoh H, et al. (2010) Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. *Plant Cell Physiol* 52: 344–360.
- Ichimura K, Tena G, Henery Y, Champion A, Kreis M, et al. (2002) Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci* 7: 301–308.
- Kong F, Wang J, Cheng L, Liu S, Wu J, et al. (2012) Genome-wide analysis of the mitogen-activated protein kinase gene family in *Solanum lycopersicum*. *Gene* 499: 108–120.
- Holub E (2001) Arms race is ancient history in *Arabidopsis*, the wildflower. *Nat Rev Genet* 2: 516–527.
- Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, et al. (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol* 146: 5–21.
- Lyons E, Pedersen B, Kane J, Alam M, Ming R, et al. (2008) Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol* 148: 1772–1781.
- Sinha AK, Jaggi M, Raghuram B, Tuteja N (2011) Mitogen-activated protein kinase signaling in plants under abiotic stress. *Plant Signal Behav* 6: 196–203.
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.
- Jonak C, Okresz L, Bogre L, Hirt H (2002) Complexity, cross talk and integration of plant MAP kinase signalling. *Curr Opin Plant Biol* 5: 415–424.