Review TRENDS in Plant Science Vol.11 No.4 April 2006

Ancient signals: comparative genomics of plant MAPK and MAPKK gene families

Louis-Philippe Hamel[1*], Marie-Claude Nicole[1*], Somrudee Sritubtim[2], Marie-Josée Morency[1], Margaret Ellis[2], Juergen Ehlting[2], Nathalie Beaudoin[3], Brad Barbazuk[4], Dan Klessig[5], Justin Lee[6], Greg Martin[5], John Mundy[7], Yuko Ohashi[8], Dierk Scheel[6], Jen Sheen[9], Tim Xing[10], Shuqun Zhang[11], Armand Seguin[1] and Brian E. Ellis[2]

[1]Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, 1055 du PEPS, Sainte-Foy, Québec, Canada G1V 4C7
[2]Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC, Canada V6T 1Z3
[3]Département de biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1
[4]Donald Danforth Plant Science Center, St. Louis, MO 63132, USA
[5]Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, NY 14853, USA
[6]Leibnitz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle, Germany
[7]Institute of Molecular Biology & Physiology, Øster Farimagsgade 2A, 1353 Copenhagen K, Denmark
[8]Plant Physiology Department, National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
[9]Harvard Medical School, Department of Molecular Biology, Wellman 11, Massachusetts General Hospital, Boston, MA 02114, USA
[10]Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6
[11]Department of Biochemistry, University of Missouri-Columbia, 109 Life Sciences Center, Columbia, MO 65211, USA

MAPK signal transduction modules play crucial roles in regulating many biological processes in plants, and their components are encoded by highly conserved genes. The recent availability of genome sequences for rice and poplar now makes it possible to examine how well the previously described *Arabidopsis* MAPK and MAPKK gene family structures represent the broader evolutionary situation in plants, and analysis of gene expression data for *MPK* and *MKK* genes in all three species allows further refinement of those families, based on functionality. The *Arabidopsis* MAPK nomenclature appears sufficiently robust to allow it to be usefully extended to other well-characterized plant systems.

MAPK families: conservation and diversity

Integration of the myriad cellular processes that enable eukaryotic organisms to grow and reproduce successfully requires the coordinated activity of an elaborate matrix of signal transduction proteins, within which the most prominent super-family consists of the protein kinases (PK). Within this super-family, the mitogen-activated protein kinases (MAPKs) form a distinctive and highly conserved PK sub-family. The hierarchical organization of three classes of functionally-related kinases: the MAPKs themselves, the MAPK kinases (MAPKKs), and the MAPKK kinases (MAPKKK), allows these proteins to operate as signal transmission cascades capable of efficiently amplifying, integrating and channeling information between the cellular environment and the metabolic and transcriptional response centres. Biochemical and genetic evidence points to a complex network organization in which kinases at one level can be activated by more than one upstream effector and can, in turn, act upon more than one target [1,2]. When combined with the availability of multiple related kinases at each level, and the interaction with modifying proteins such as scaffolds [3,4] and protein phosphatases [5,6], this complexity creates a remarkably versatile matrix of signaling capacities.

MAPK cascades operate at the core of eukaryotic signal transduction networks, and their component kinases have been highly conserved through evolution [7]. Sequencing of genomes as diverse as *Arabidopsis*, yeast, worm, fly and humans has revealed that the three classes of kinases involved in these cascades (MAPKKKs, MAPKKs and MAPKs) always occur as gene families. Functional analysis has demonstrated that, despite structural conservation over evolutionary time, individual gene family members often play distinct roles, although some degree of functional redundancy is also observed [8,9]. Although our knowledge of signal transduction in plants is less well developed than in some other phyla, there is considerable evidence for broad conservation of signaling modalities between plants and other eukaryotic organisms. An initial phylogenetic analysis of the gene families encoding the three classes of kinases in the *Arabidopsis* genome suggested that they have been amplified, relative to yeast or model metazoan lineages [10]. For example, *Arabidopsis* possesses genes encoding 20 MAPKs and ten

Corresponding author: Ellis, B.E. (bee@interchange.ubc.ca).
* These authors contributed equally to this work.
Available online 14 March 2006

 1360-1385/$ - see front matter doi:10.1016/j.tplants.2006.02.007

MAPKKs compared with six MAPKs and six MAPKKs in yeast and ten MAPKs and seven MAPKKs in humans. It was not known whether the expansion of these families observed in *Arabidopsis* was generally representative of higher plants or if it represents a lineage-specific pattern. The recent completion of two additional higher plant genome-sequencing efforts (*Populus* and *Oryza*) has now made it possible to address this question directly.

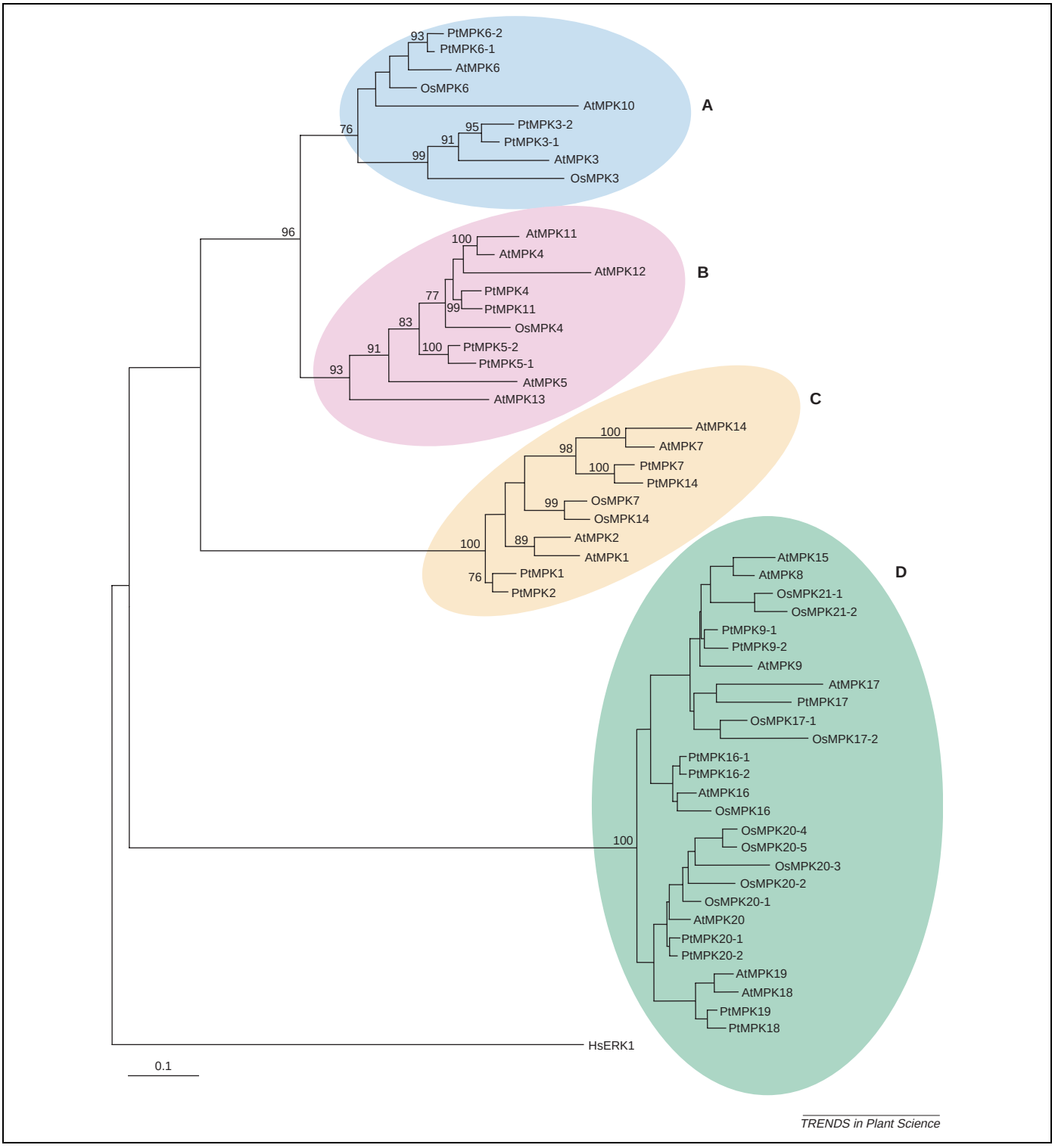

Figure 1. Phylogenetic relationships of *Arabidopsis*, poplar and rice MPK genes. The *Populus* genome assembly version1.0 (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) was searched using the 20 *Arabidopsis* MPK amino acid sequences as direct queries. The predicted *Oryza* peptide set derived from TIGR rice pseudomolecules (Version 3.0) (http://www.tigr.org/tdb/e2k1/osa1/data_download.shtml) was queried using a profile Hidden Markov Model-based search (HMMER http://hmmer.wustl.edu/) with an HMM built from the 20 *Arabidopsis* MPKs. These searches retrieved 21 full-length poplar MPK homologs and 15 rice MPK homologs. MPK 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. The protein kinase domains of each sequence were aligned with ClustalX (1.81), using HsERK1 as an outgroup, and adjusted manually. The following modified alignment parameters were used: Pairwise alignment – Gap opening, 35.0, Gap extension, 0.75; Multiple alignment – Gap opening, 15.0, Gap extension, 0.30. The resulting alignments in Phylip format were submitted to *PHYML online* (http://atgc.lirmm.fr/phyml/) to generate a maximum likelihood bootstrapped tree. To identify the species of origin for each MPK, a species acronym is included before the protein name: At, *Arabidopsis thaliana*; Hs, *Homo sapiens*; Os, *Oryza sativa*; Pt, *Populus trichocarpa*.

Information on transcriptional regulation of individual *Arabidopsis* MAPK components has also been limited to reports of differences in transcription of a few specific MAPK genes in different organs, tissue and/or cell-types, or expression changes induced by extracellular stimuli [11,12]. Analysis of the family structure and expression of the full complement of MAPK and MAPKK genes in two eudicot species (*Populus trichocarpa* and *Arabidopsis thaliana*), and comparison of the eudicot data with those derived from a monocot (*Oryza sativa*), provides initial insights into the degree of structural and functional conservation of these two important signal transduction gene families in higher plants.

MPKs

Previous analysis of the *Arabidopsis* genome identified 20 MAPK (MPK) genes and ten MAPKK (MKK) genes, and also assigned a systematic nomenclature to these sequences [10]. Because only limited information was available at that time concerning the expression or biological roles of these 30 genes in *Arabidopsis*, no comprehensive attempt had been made to assess their biological functionality. Exhaustive searches of the current poplar and rice genome sequence databases has revealed 21 putative poplar MPK gene models and 15 models for rice, indicating that the two eudicot MPK gene families are similar in size, whereas the rice family is substantially smaller. Phylogenetic analysis places representatives of the poplar and rice MPK homologs into each of the four clades that had been identified earlier in *Arabidopsis* [10] (Figure 1). Those poplar and rice MPKs that possess a -TEY- signature in their activation loop cluster with their *Arabidopsis* homologs in the previously defined A, B and C clades. However, the inclusion of the two additional taxa now allows better definition of the large clade of 'Group D' MPKs, whose members all display a distinctive -TDY-, rather than -TEY-, signature. Clade D can now be resolved into three sub-groups, each of which contains both eudicot and monocot members (Figure 2). The plant MPK gene family structure thus reflects an ancient pattern of diversification that was already established before the evolutionary divergence of the monocots and eudicots.

For all three species, apparently orthologous genes can be readily identified in some groups of MPKs, such as those defined by AtMPK3, AtMPK6, AtMPK7 and AtMPK14. This robust pattern encourages us to propose that the systematic MPK nomenclature adopted earlier [10] for *Arabidopsis* MPKs could be extended to the poplar MPKs (PtMPK), and to the rice MPKs (OsMPK). Although not perfect, such a model should help to avoid further development of the confusing trivial nomenclature that already marks the rice MPK gene family (Table 1). In the case of poplar, no MAPK gene names have yet appeared in the literature.

The strong evolutionary conservation of MPK sequences also suggests that orthologous numbering of family members would not be inappropriate in some cases, recognizing that such numbering is driven strictly by predicted evolutionary relationships and is not based on evidence of conserved biological function. We therefore propose that orthologous numbering of *PtMPK* and *OsMPK* genes be adopted, based on the originally assigned AtMPK numbers, where these relationships appear clear. Inevitably, paralogous relationships within the *Arabidopsis* MPK gene family were sometimes not captured in the original AtMPK numbering (e.g. paralogs MPK7 and MPK14), but our improved understanding of the MPK evolutionary relationships now makes it possible to assign paralogous nomenclature to poplar and rice genes for which a single *Arabidopsis* ortholog exists (e.g. *PtMPK20-1* and *PtMPK20-2* versus *AtMPK20*).

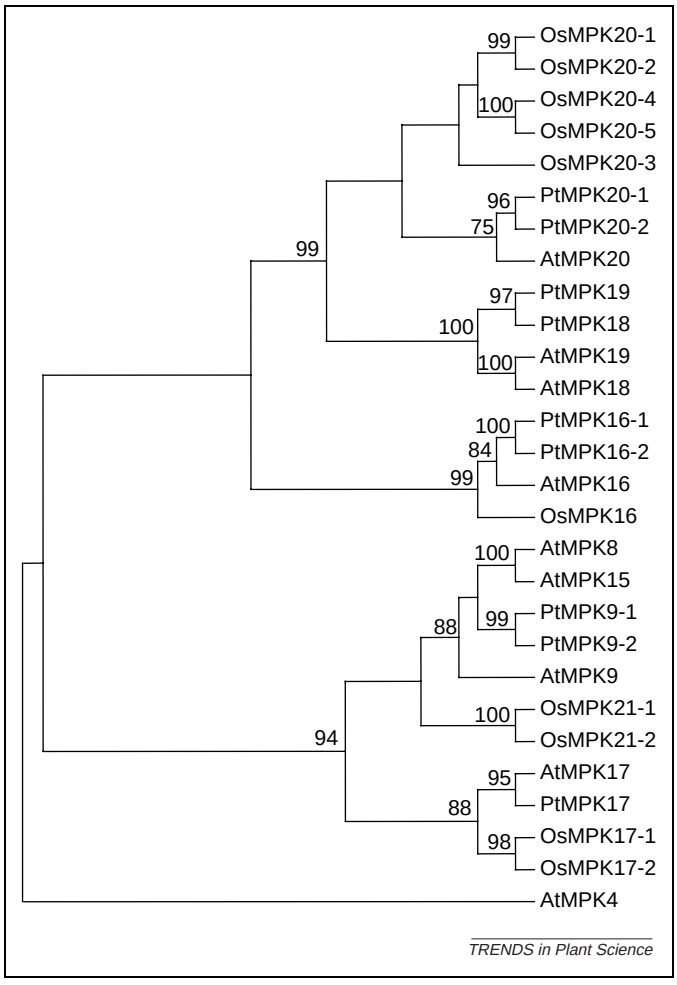

Figure 2. Clade D MPK gene phylogeny. To obtain better resolution within clade D, a separate alignment was carried out using full-length amino acid sequences, and the same alignment parameters as in Figure 1, to generate a *PHYML* bootstrapped tree. *PHYML* default values were used except for 100 bootstraps, JTT substitution model, and four substitution rate categories. Only bootstrap scores >70 are indicated. The species acronyms are as in Figure 1.

Gene duplication is a prominent feature within both the eudicot and monocot MPK gene families. Ancient duplication events can be detected that might represent previously described whole genome duplications in *Arabidopsis* [13], poplar [14] and rice [15]. One early duplication event appears to have generated the basal split between the plant TEY MPKs (Groups A, B and C) and the TDY clade, groups that have since both expanded and remained monophyletic. It is interesting that the *Chlamydomonas* genome encodes five clearly recognizable MPKs, three of which carry -TEY- activation loop signatures and the other two contain -TDY- (data not shown), indicating that this

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

	Pt gene	Pt gene model	At gene	At gene code	Os gene	Os gene code	Previous rice gene names
MPK	PtMPK1	grail3.0003026801	AtMPK1	At1g10210	OsMPK3	Os03g17700	OsMAP1; OsMAPK2; OsMRSMK2; OsBIMK1
	PtMPK2	fgenesh4_pm.C_LG_V000601	AtMPK2	At1g59580	OsMPK4	Os10g38950	
	PtMPK3-1	estExt_fgenesh4_pm.C_LG_IX0462	AtMPK3	At3g45640	OsMPK6	Os06g06090	
	PtMPK3-2	fgenesh4_pm.C_LG_I000779	AtMPK4	At4g01370	OsMPK7	Os06g48590	OsMAPK4, OsMAP-kinase2; OsMSRMK3
	PtMPK4	estExt_fgenesh4_pg.C_LG_XIV0319	AtMPK5	At4g11330	OsMPK14	Os02g05480	OsMAPK3
	PtMPK5-1	fgenesh4_pm.C_LG_I000354	AtMPK6	At2g43790	OsMPK16	Os11g17080	
	PtMPK5-2	fgenesh4_pm.C_LG_III000431	AtMPK7	At2g18170	OsMPK17-1	Os06g49430	OsBWMK1
	PtMPK6-1	estExt_fgenesh4_pm.C_LG_VII0025	AtMPK8	At1g18150	OsMPK17-2	Os02g04230	OsRMAPK2
	PtMPK6-2	estExt_Genewise1_v1.C_LG_XVII0005	AtMPK9	At3g18040	OsMPK20-1	Os01g43910	
	PtMPK7	fgenesh4_pm.C_scaffold_57000068	AtMPK10	At3g59790	OsMPK20-2	Os05g50560	
	PtMPK9-1	estExt_fgenesh4_pm.C_LG_XII0156	AtMPK11	At1g01560	OsMPK20-3	Os06g26340	
	PtMPK9-2	estExt_fgenesh4_pm.C_LG_XV0119	AtMPK12	At2g46070	OsMPK20-4	Os01g47530	OsWJUMK1; OsMPKG1
	PtMPK11	estExt_fgenesh4_pg.C_LG_II1481	AtMPK13	At1g07880	OsMPK20-5	Os05g49140	
	PtMPK14	grail3.0019023301	AtMPK14	At4g36450	OsMPK21-1	Os05g50120	
	PtMPK16-1	eugene3.00081905	AtMPK15	At1g73670	OsMPK21-2	Os01g45620	OsMPKG2
	PtMPK16-2	estExt_Genewise1_v1.C_LG_X4145	AtMPK16	At5g19010			
	PtMPK17	estExt_Genewise1_v1.C_LG_X6513	AtMPK17	At2g01450			
	PtMPK18	estExt_fgenesh4_pg.C_1680031	AtMPK18	At1g53510			
	PtMPK19	estExt_fgenesh4_pg.C_LG_I3001	AtMPK19	At3g14720			
	PtMPK20-1	gw1.V.609.1	AtMPK20	At2g42880			
	PtMPK20-2	estExt_fgenesh4_pm.C_LG_II0282					
MKK	PtMKK2-1	eugene3.00180195	AtMKK1	At4g26070	OsMKK1	Os06g05520	OsMEK2
	PtMKK2-2	estExt_fgenesh4_pm.C_LG_VI0403	AtMKK2	At4g29810	OsMKK3	Os06g27890	
	PtMKK3	estExt_fgenesh4_pg.C_LG_I2149	AtMKK3	At5g40440	OsMKK4	Os02g54600	
	PtMKK4	gw1.X.3215.1	AtMKK4	At1g51660	OsMKK5	Os06g09180	
	PtMKK5	eugene3.00080074	AtMKK5	At3g21220	OsMKK6	Os01g32660	OsMEK1
	PtMKK6	estExt_Genewise1_v1.C_1450125	AtMKK6	At5g56580	OsMKK10-1	Os02g46760	
	PtMKK7	gw1.122.164.1	AtMKK7	At1g18350	OsMKK10-2	Os03g12390	
	PtMKK9	gw1.XII.274.1	AtMKK8	At3g06230	OsMKK10-3	Os03g50550	
	PtMKK10	eugene3.00290225	AtMKK9	At1g73500			
	PtMKK11-1	fgenesh4_pg.C_LG_VIII001629	AtMKK10	At1g32320			
	PtMKK11-2	gw1.X.5118.1					

divergence was already present in the MPK family possessed by the common ancestor of the Chlorophyta and the Embryophyta.

More recent duplications are also obvious, some having occurred before the monocot–eudicot divergence (e.g. within the MPK20 group), and others apparently having taken place after that event (e.g. within the MPK3 group). Most striking is the large number of recent MPK gene duplications in the poplar genome (10) relative to the number in *Arabidopsis* (5) and rice (4). This pattern is consistent with other evidence of extensive segmental duplications in the poplar genome (G. Tuskan, personal communication), and it also indicates that the approximate correspondence in MPK family size between *Arabidopsis* and poplar is not necessarily indicative of fully synonymous family structures. In some cases, two poplar paralogs can be identified for a single *AtMPK* gene (e.g. *PtMPK16-1* and *PtMPK16-2* versus *AtMPK16*), whereas in others, recently duplicated poplar genes (e.g. *PtMPK5-1* and *PtMPK5-2*) have an evolutionary relationship that is different from that of the related *Arabidopsis* homologs (*AtMPK5* and *AtMPK13*). Moreover, both products of the ancestral gene duplication events that gave rise to *AtMPK10* and *AtMPK13* appear to have been uniquely retained in *Arabidopsis* but not in the poplar or rice lineages.

One of the more striking duplication patterns is the remarkable lineage-specific amplification of the *OsMPK20* gene set, whose five closely related members are scattered over three linkage groups in the rice genome. Not only is this expansion unmatched in the two eudicot genomes (one *AtMPK20* ortholog and two *PtMPK20* candidates) but it also emphasizes that the rice MPK gene family is likely to be functionally even smaller than its total MPK gene count would suggest.

One metric for functionality of MPK gene family members is detection of expression. High confidence ESTs could be identified for the *OsMPK20-1* and *OsMPK20-4* genes, and the rice MPSS database also provided convincing evidence of expression for *OsMPK20-1, OsMPK20-3, OsMPK20-4* and *OsMPK20-5* (see Supplementary material). Based on this pattern, it appears that loss of function in the *OsMPK20-2* gene might already be underway, even though the sequence integrity of the encoded open reading frame has apparently been retained.

Although the various expression assays yielded contrasting results sometimes, they collectively reveal that 19 of the 20 *Arabidopsis* MPKs are expressed at relatively low levels in most tissues (see Supplementary material). One of the more actively transcribed members of this family is AtMPK3. Rapid induction of both expression and

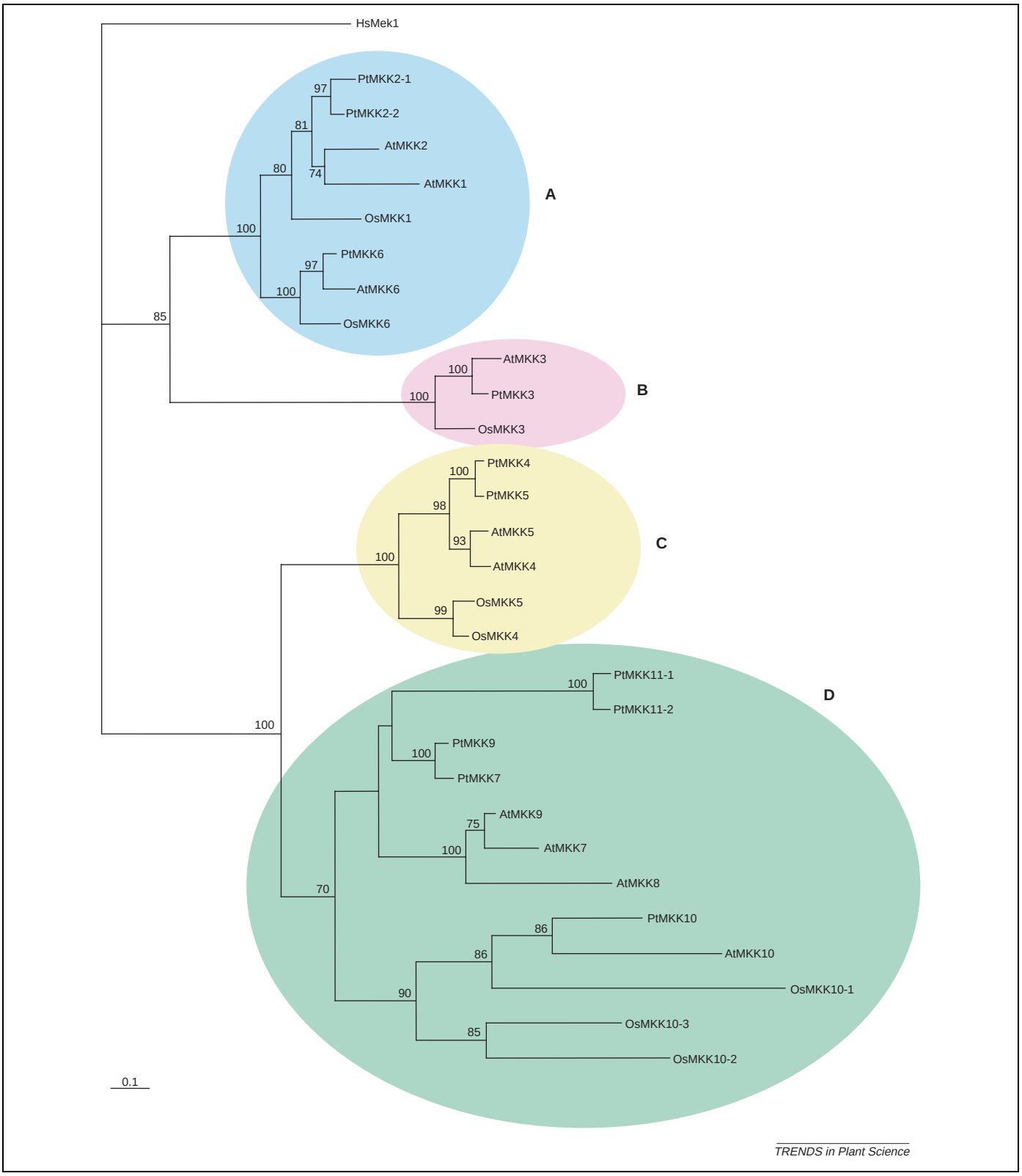

Figure 3. Phylogenetic relationships of *Arabidopsis*, poplar and rice MKK genes. The *Populus* genome assembly version 1.0 (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) was searched using the ten *Arabidopsis* MKK amino acid sequences as direct queries. The predicted *Oryza* peptide set derived from TIGR rice pseudomolecules (Version 3.0) (http://www.tigr.org/tdb/e2k1/osa1/data_download.shtml) was queried using a profile Hidden Markov Model-based search (HMMER http://hmmer.wustl.edu/) with an HMM built from the ten *Arabidopsis* MKKs. These searches retrieved nine full-length poplar MKK homologs, and eight rice MKK homologs. MKK 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/T**xxxxx**S/T**-, within the activation loop. The protein kinase domains of these sequences were aligned with ClustalX (1.81), using HsMek1 as an outgroup. The following modified alignment parameters were used: Pairwise alignment – Gap opening, 35.0, Gap extension, 0.75; Multiple alignment – Gap opening, 15.0, Gap extension, 0.30. The resulting alignments in Phylip format were submitted to *PHYML online* (http://atgc.lirmm.fr/phyml/) to generate bootstrapped trees. Default values were used except for 100 bootstraps, JTT substitution model, and four substitution rate categories. Only bootstrap scores >70 are shown. To identify the species of origin for each MKK, a species acronym is included before the protein name: At, *Arabidopsis thaliana*; Hs, *Homo sapiens*; Os, *Oryza sativa*; Pt, *Populus trichocarpa*.

post-translational activation of AtMPK3 and its orthologs in other herbaceous species have been observed in response to a wide range of stresses, including wounding, ozone, elicitor treatment and Avr-R gene interaction during the hypersensitive response [11,16,17]. By contrast, little *AtMPK10* expression could be detected in any of the three assay systems, suggesting that this gene family member is either expressed primarily in situations not examined in these assays, or that the gene is becoming non-functional, even though the encoded protein still displays all the elements thought to be necessary for MPK activity. The AtMPK10 sequence is also characterized by a relatively large divergence from other members of the MPK6 cluster (Figure 1).

Almost all the poplar and rice MPK genes are also expressed at modest levels in all tissues scored, but in poplar the highest expression was usually shown by *PtMPK17,* a member of the Group 'D' MPKs, whereas the *PtMPK3* genes were relatively weakly expressed. Neither poplar nor rice appears to possess a MPK orthologous to *AtMPK10*. In light of the recent duplication of several of the poplar MPK genes, and evidence that duplicated eukaryotic genes often undergo rapid differential evolution of regulatory patterns [18,19], it is of particular interest to compare the expression profiles of those duplicates. In each tissue assayed, some *PtMPK* duplicates (3-1 and 3-2; 5-1 and 5-2) consistently showed differential patterns of expression, which is suggestive of evolving sub-functionalization, whereas other duplicated genes (16-1 and 16-2; 20-1 and 20-2) showed similar levels of expression for each member (see Supplementary material). Two rice MPK genes, *OsMPK4* and *OsMPK20-2,* yielded no MPSS signal in the cDNA libraries surveyed, although the apparent absence of *OsMPK4* expression might be an artefact of the MPSS assay because high confidence *OsMPK4* ESTs can be identified in the databases. In addition, the poplar and *Arabidopsis MPK4* orthologs are actively transcribed in most tissues.

MKKs

The *Arabidopsis* genome was originally notated as possessing ten members of the MKK gene family [10], but closer examination of those sequences reveals that *AtMKK10* lacks a properly constructed activation loop target site, which raises the question of its biological functionality. In addition to the thirteen poplar and rice MKK gene models possessing fully canonical motif structures [8], an additional six MKK genes (three each in poplar and rice) are deficient in one or more motif elements. Phylogenetic analysis of the MKK gene families in the three species places the MKK genes in generally well-resolved clades that each contain members from both monocots and eudicots, with the exception of the MKK7-9 clade, for which no rice ortholog can be identified (Figure 3). A five-member sister group to the MKK7-9 clade can be identified that includes *AtMKK10*, a poplar ortholog (putative *PtMKK10*) and three paralogous rice sequences (putative *OsMKK10-1*, *OsMKK10-2* and *OsMKK10-3*), but each of these gene models possesses an incomplete activation loop motif in which the 5′-S or T residue is either absent or located 3–5 residues upstream of the canonical position. Although these five genes share a common evolutionary history, the differences in the amino acid sequence of the MKK activation loop of *PtMKK10* and that of the monocot ortholog *OsMKK10* are fewer than the differences seen between the same region of *PtMKK10* and *AtMKK10*, even though the two eudicot species are evolutionarily more closely related (data not shown). In the active site region, all five genes retain the residues believed to be essential for kinase function but it is not clear whether the corresponding gene products would possess MKK activity in the context of typical MAPK cascades. Alternatively, if they were expressed, they might represent neo-functionalized versions of a duplicated canonical MKK gene, acting perhaps as non-catalytic scaffolding proteins [4]. A similar situation exists with PtMKK11-1 and PtMKK11-2. Expression analysis suggests that the issue of their functionality might be largely moot because we found no convincing evidence in the databases for transcription of *AtMKK10*, *PtMKK10*, *PtMKK11-1, PtMKK11-2*, *OsMKK10-1* or *OsMKK10-3*. Interestingly, *OsMKK10-2* is expressed in several tissues, and corresponding ESTs have been found for rice and maize.

A less complex case involves *AtMKK8*, a gene whose encoded protein possesses all the canonical MKK motifs, but for which no evidence of expression could be found in any of the three assay systems. Again, it might be diagnostic that neither the poplar nor the rice genome sequence yielded a clear ortholog of *AtMKK8,* suggesting that an ancestral gene duplication event occurring after the divergence of the monocots and eudicots could have yielded the precursors of the AtMKK7-8-9 and the PtMKK7-9-11 clades. Since that time, both PtMKK11-1 and PtMKK11-2, and AtMKK8 might have drifted toward a non-functional state, which would be consistent with the large evolutionary distances in the phylogenetic reconstruction (Figure 3). As a result, the functional *Arabidopsis* MKK gene family is likely to consist of just eight members, as does the poplar family.

One of the more interesting MKK clades consists of the MKK3 sequences. These genes are unusual in their possession of a 3′ extension encoding a NTF ('nuclear transfer factor') domain. Although stand-alone NTF proteins are found encoded in other eukaryotic genomes, including *Arabidopsis*, the combination in plants of a MAPKK and a NTF within a single gene product appears to be unique among eukaryotic taxa. No biological functions have yet been associated in plants with either MKK3, or the NTF domain, but the MKK3 genes are actively transcribed in all three species analyzed in this study (see Supplementary material). Interestingly, the *Chlamydomonas* genome encodes a single MKK and this MKK belongs to the MKK3 structural class, including the 3′-NTF domain, indicating that this chimeric arrangement has had a long and successful evolutionary history in the lineage of photosynthetic eukaryotes.

Conclusions

The apparent expansion of the MAPK gene families observed in the *Arabidopsis* genome initially suggested that higher plants had exploited MAPK signaling

versatility in support of the unique developmental and environmental needs associated with evolution of the Embryophyta, but the present analysis indicates that this expansion might be more apparent than real. At the MKK level, both *Arabidopsis* and poplar possess similar gene complements, and several of those represent the products of recent gene duplication events. The role of duplications in amplifying the MPK family is even more striking, particularly in poplar, where virtually all the PtMPK family members exist as recently duplicated pairs. Overall, the rice MPK and MKK gene families display less evidence of such recent duplication activity, and as a result both families are distinctly smaller in this monocot than in the eudicot taxa. Steven Maere *et al.* [20] recently demonstrated that gene families in certain classes of eukaryotic genes, including the protein kinases, are more likely to have undergone amplification through whole genome or large segmental duplication events than through local duplications. Consistent with this, few of the MPK or MKK gene family members occur as tandem duplicates within the three genomes examined here.

Despite many large-scale phenotypic screens in *Arabidopsis* and rice, few mutants have been identified in MPK or MKK genes. The extensive gene duplication seen in these families would appear to have the potential to buffer the effects of such genetic lesions, but the isolation of a severely compromised *Arabidopsis mpk4* mutant [12] indicates that the presence of the closely related *MPK11* gene in the *mpk4* background is not sufficient to suppress the *mpk4* deficiency phenotype. Similarly, MPK3 and MPK6 (and their orthologs in other species) are closely related, and are often both post-translationally activated in response to stress. However, the *MPK6* gene does not respond transcriptionally to stress, whereas *MPK3* expression is rapidly induced [17,21]. In addition, *in vitro* assay of recombinant AtMPK3 and AtMPK6 against an array of 1690 *Arabidopsis* proteins revealed that only 26 of 87 identified substrates could be phosphorylated by both kinases [9]. Thus, a significant level of sub-functionalization is likely to have evolved within some or all the MPK and MKK gene pairs. However, such variation could be difficult to define without using more fine-grained analyses, such as promoter–reporter fusions. For example, this approach has revealed that the expression of *AtMPK12* in aerial tissues of *Arabidopsis* is largely restricted to the guard cells (S. Sritubtim and B.E. Ellis, unpublished)

The growing interest in plant MAPK modules and their cellular functions should help to fill in the many gaps in our current knowledge of their place in plant biology. Of particular importance will be the development of more and better reagents capable of distinguishing between the many plant MAPkinases and their partners. To date, attention has been largely focused on a small subset of the possible players, sometimes using tools (e.g. in-gel kinase assays) whose specificity and relative efficiency in detecting the full range of plant gene products are not well characterized. Similarly, the sequencing of additional plant genomes, particularly of more basal taxa, such as *Physcomitrella* and *Selaginella*, should provide crucial new insights into the evolutionary origins and functionalization of the MAPK signaling matrix we see in modern plants.

Acknowledgements

Financial support from the Natural Sciences and Engineering Research Council of Canada and from Genome Canada is gratefully acknowledged.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tplants.2006.02.007

References

1 Cardinale, F. *et al*. (2002) Convergence and divergence of stress-induced mitogen-activated protein kinase signaling pathways at the level of two distinct mitogen-activated protein kinase kinases. *Plant Cell* 14, 703–711
2 Garrington, T. and Johnson, G. (1999) Organization and regulation of mitogen-activated protein kinase signaling pathways. *Curr. Opin. Cell Biol.* 11, 211–218
3 Levchenko, A. *et al*. (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5818–5823
4 Whitmarsh, A.J. and Davis, R.J. (1998) Structural organization of MAPkinase signaling modules by scaffold proteins in yeast and mammals. *Trends Biochem. Sci.* 23, 481–485
5 Mapes, J. and Ota, I. (2004) Nbp2 targets the Ptc1-type 2C Ser/Thr phosphatase to the HOG MAPK pathway. *EMBO J.* 23, 302–311
6 Stephen, M.K. (2000) Protein phosphatases and the regulation of mitogen-activated-protein kinase signaling. *Curr. Opin. Cell Biol.* 12, 186–192
7 Widmann, C. *et al*. (1999) Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol. Rev.* 79, 143–180
8 Jonak, C. *et al*. (2002) Complexity, cross talk and integration of plant MAP kinase signaling. *Curr. Opin. Plant Biol.* 5, 415–424
9 Feilner, T. *et al*. (2005) High through-put identification of potential *Arabidopsis* MAP kinase substrates. *Mol. Cell. Proteomics* 4, 1558–1568
10 MAPK Group. (2002) Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci.* 7, 301–308
11 Ahlfors, R. *et al*. (2004) Stress hormone-independent activation and nuclear translocation of mitogen-activated protein kinases in *Arabidopsis thaliana* during ozone exposure. *Plant J.* 40, 512–522
12 Petersen, M. *et al*. (2000) *Arabidopsis* map kinase 4 negatively regulates systemic acquired resistance. *Cell* 103, 1111–1120
13 Ermolaeva, M. *et al*. (2003) The age of the *Arabidopsis thaliana* genome duplication. *Plant Mol. Biol.* 51, 859–866
14 Sterck, L. *et al*. (2005) EST data suggest that poplar is an ancient polyploid. *New Phytol.* 167, 165–170
15 Yu, J. *et al*. (2005) The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.* 3, e38
16 Nakagami, H. *et al*. (2005) Emerging MAP kinase pathways in plant stress signaling. *Trends Plant Sci.* 10, 339–346
17 Zhang, S. *et al*. (2000) Multiple levels of tobacco WIPK activation during the induction of cell death by fungal elicitins. *Plant J.* 23, 339–347
18 Blanc, G. and Wolfe, K.H. (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16, 1679–1691
19 Gu, X. *et al*. (2005) Rapid evolution of expression and regulatory divergences after yeast gene duplication. *Proc. Natl. Acad. Sci. U. S. A.* 102, 707–712
20 Maere, S. *et al*. (2005) Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5454–5459
21 Liu, Y. *et al*. (2003) Interaction between two mitogen-activated protein kinases during tobacco defense signaling. *Plant J.* 34, 149–160