

Characterization of mitogen activated protein kinases (MAPKs) in the *Curcuma longa* expressed sequence tag database

Raj Kumar Joshi*, Basudeba Kar, Sanghamitra Nayak

Centre of Biotechnology, School of Pharmaceutical Sciences; Siksha O Anusandhan University, Bhubaneswar-751003, India; Raj Kumar Joshi - Email: rajkumar.joshi@yahoo.co.in; Phone: +91-9437684176; *Corresponding author

Received September 17, 2011; Accepted September 19, 2011; Published October 14, 2011

Abstract:

Mitogen activated protein kinase (MAPK) cascades are universal signal transduction modules that play crucial role in plant growth and development as well as biotic and abiotic stress responses. 20 and 17 MAPKs have been characterized in *Arabidopsis* and rice respectively, which are used for identification of the putative MAPKs in other higher plants. However, no MAPK gene sequences have yet been characterized for asexually reproducing plants. We describe the analysis of MAPK EST sequences from *Curcuma longa* (an asexually reproducible plant of great medicinal and economic significance). The four *Curcuma* MAPKs contains all 11 MAPK conserved domains and phosphorylation-activation motif, TEY. Phylogenetic analysis grouped them in the subgroup A and C as identified earlier for *Arabidopsis*. The *Curcuma* MAPKs identified showed high sequence homology to rice *OsMPK3*, *OsMPK4* and *OsMPK5* suggesting the presence of similar key element in signaling biotic and abiotic stress responses. Although further *in vivo* and *in vitro* analysis are required to establish the physiological role of *Curcuma* MAPKs, this study provides the base for future research on diverse signaling pathways mediated by MAPKs in *Curcuma longa* as well as other asexually reproducing plants.

Keywords: *Curcuma longa*, expressed sequence tags, MAPKs, TBLASTN

Background:

Protein phosphorylation/dephosphorylation through protein kinases is a key regulatory mechanism in modulating the activation of intracellular responses to extracellular information [1]. The mitogen-activated protein kinase (MAPK) cascades are known to be one of the major phosphorylation pathways which function downstream of sensors/receptors and regulate cellular responses to external and endogenous stimuli [2]. A typical MAPK cascade consists of three steps: a MAP kinase kinase kinase (MAPKKK) activates a particular MAP kinase kinase (MAPKK) through phosphorylation on two serine/threonine residues in a conserved S/T-X3-5-S/T motif, the activated MAPKK can in turn phosphorylate a MAPK on threonine and tyrosine residues within TEY or TDY motif located in the activation loop (T-loop) between kinase domain VII and VIII [3]. The MAPKs are known to regulate a myriad of

physiological and developmental responses such as cell growth, cell differentiation, hormone signaling, pathogen infection, wounding, drought, low temperature, high salinity etc [1]. More than 60 MAPKs have been isolated and characterized in plants, and the analysis of *Arabidopsis* genome sequences has revealed the existence of more than 20 MAPK genes [4], which suggests that the MAPK cascades in plant may be quite complex. Based on the phylogenetic analysis of amino acid sequence and phosphorylation motif, plant MAPKs have been grouped into at least four groups (A, B, C, and D) [5]. According to amino acid sequences, MAPKs share highly conserved region over the entire lengths with highest similarity in the eleven domains that are necessary for the catalytic function of serine/threonine protein kinases [5]. Above facts suggest that MAPKs play crucial roles in plant growth and development. In order to obtain more information on the

specific structural and regulatory patterns of MAPKs, it is necessary to isolate and identify more MAPK genes from different species. Most of the studies on the plant MAPK genes have been carried out in model plant species such as *Arabidopsis*, tobacco and rice. *Curcuma longa* L. (turmeric) of the family Zingiberaceae is one of the most important crops with great medicinal and economic significance. Continuous domestication of the preferred genotypes coupled with their exclusive vegetative nature seems to have eroded the genetic base of these crops and as a result, all of their cultivars available today are equally susceptible to major biotic and abiotic stresses. Moreover, turmeric is completely sterile and is propagated exclusively by vegetative means using rhizome. Characterization of MAPKs in turmeric can provide lot of information on the possible mechanism of various stress responses in asexually reproducing plants. The availability of 12,593 EST sequences from GenBank database of *Curcuma longa* provides an opportunity to the characterization of MAPK gene sequences in turmeric. We describe the identification and characterization of plant MAPKs in *Curcuma longa* EST database using known MAPK sequences.

Methodology:

Protein sequences of reported plant MAPKs were used to query the *Curcuma longa* expressed sequence tag (EST) database with the TBLASTN algorithm [6]. The complete list of sequences used as baits includes *OsMPK1*, *OsMPK2*, *OsMPK3*, *OsMPK5* and *OsMPK7* from rice [7], *AtMPK2*, *AtMPK3*, *AtMPK4* and *AtMPK8* from *Arabidopsis thaliana* [5], *ZmMPK5*, *ZmMPK6*, *ZmMPK7* and *ZmSIMK* from *Zea mays* [8] and MAP kinase gene *TaMPK3* from *Triticum aestivum* [9]. All turmeric sequences used during this work were obtained from *Curcuma longa* EST database. We have mined 12,593 EST sequences consisting of two tissue libraries of 6870 rhizomes (DY395309-DY388440) and 5723 leaves (DY388439-DY382717). The EST sequences were screened against the UniVec database from NCBI (<ftp://ftp.ncbi.nih.gov/pub/UniVec/>) for detecting vector and adapter sequences by using the program Cross_Match. CAP3 program was used to assemble the EST sequence into contigs for creating a non-redundant dataset. The program TBLASTN was used to perform reverse alignment on *Curcuma longa* contigs. All MAPK clusters found in *Curcuma* database were translated to obtain their putative protein sequences. The Open Reading Frames (ORFs) for each searched contig was predicted using the ExPASy Translate Tool (bo.expasy.org/tools/dna.html). Protein sequences obtained were used in a second round of TBLASTN search against the non-redundant protein database at the National Center for Biotechnology Information (NCBI) to identify their closest homologues. Additional domains were detected using the Prosite (<http://bo.expasy.org/prosite>) and Pfam (<http://www.sanger.ac.uk/Software/Pfam/search.shtml>) prediction programs. Multiple alignments of proteins deduced and bait sequences were performed using the ClustalX program [10]. The phylogenetic tree was constructed by the neighbor-joining method [11] using the NJ algorithm implemented in the Molecular Evolutionary Genetics Analysis (MEGA) software package version 2.1 with the Poisson correction [12]. For construction of the phylogenetic tree the confidence levels for the nodes were determined with 1000 replications using the internal branch test [13].

Results and Discussion:

Extensive bioinformatics analysis resulted in 4 MAPK sequences retrieved through TBLASTN alignment of 4035 *Curcuma longa* contig sequences (Table 1 see supplementary material). All the four MAPK sequences showed significant hit (score >500, e-value equal to 0.0) with average length of 367 amino acids and found to have homologous sequences in the NCBI GenBank database. Genome database search and analysis was performed to identify members of the MAPK gene family in many plants including *Arabidopsis*, rice, poplar and grapevine [5, 14, and 15]. Alignment of the four *Curcuma* MAPK sequences with MAPKs from other plants indicated that they contained all 11 conserved sub-domains. The TEY motif was also found located in the activation loop between kinase domain VII and VIII. The C-terminus of all the four sequences was also characterized by the presence of a common docking (CD) domain that functions as the binding site of MAPKKs. Domain scan using the PROSITE and pFAM analysis showed the presence of MAPK signature domain, ATP binding domain and serine/threonine protein kinase active site (Figure 1). This data confirmed that the four *Curcuma* contigs sequences belong to the MAPK family. A phylogenetic tree was constructed by alignment of the amino acid sequences using the ClustalX method to investigate the molecular evolution and phylogenetic relationship between *Curcuma* MAPK sequences and other MAPKs. The phylogenetic tree displayed that the *Curcuma* MAPKs were placed in two subgroups (A and C) as identified earlier for *Arabidopsis* based on sequence similarity (figure 2). Earlier reports also suggest that MAPK sequences with TEY signature in their activation loop have been clustered with their *Arabidopsis* homologs in the previously defined A, B and C clades [15]. CL.CON 2468 and CL.CON 1447 showed highest similarity with *OsMPK5* in group-A. It has been documented that MAPK members in the group A are implicated in signaling biotic and abiotic stress responses [16]. *OsMPK5* has been demonstrated to be a key element for managing defense response against blast disease [17]. Likewise, CL.CON 3107 and CL.CON 3756 were placed in the group C with *OsMPK3* and *ZmMPK7*, which are transcriptionally regulated by various abiotic stresses [18]. CL.CON.3756 also showed highest homology with *OsMPK4*, as obtained from BLASTp analysis. This is understood from the fact that both *OsMPK3* and *OsMPK4* exhibit 91% amino acid identity even though located in different chromosomes and are clustered in the group C. The *Curcuma* MAPKs in each group showed 99% amino acid identity with each other. This suggests that they are meant for either same function or are represented as replicative forms of the same gene. The *Curcuma* MAPK sequences identified in this study belong to the intron-harboring members with CL.CON 2468 and CL.CON 1447 consisting 4 and 7 introns in their open reading frames respectively. However, CL.CON 3107 and CL.CON 3756 consisted of only a single intron each suggesting the occurrence of gene loss among the members of subgroup C. Two grapevine MAPKs, *VvMPK1* and *VvMPK2* are also found to be clustered in subgroup C and possesses single introns [19]. Although the physiological function of the *Curcuma* MAPK sequences has not yet been investigated, the high level of homology compared with the known MAPKs indicate that they may have similar function with *OsMPK5*, *OsMPK4* and *OsMPK3*. However, actual function of the *Curcuma* MAPKs identified in this study cannot be ascertained

due to the fact that the *Curcuma* cDNA library that contributed the *Curcuma* EST sequences were derived out of the leaf, stem and root tissue of the plant without being exposed to any stress conditions. Hence, the MAPKs characterized in this study must be of constitutive nature.

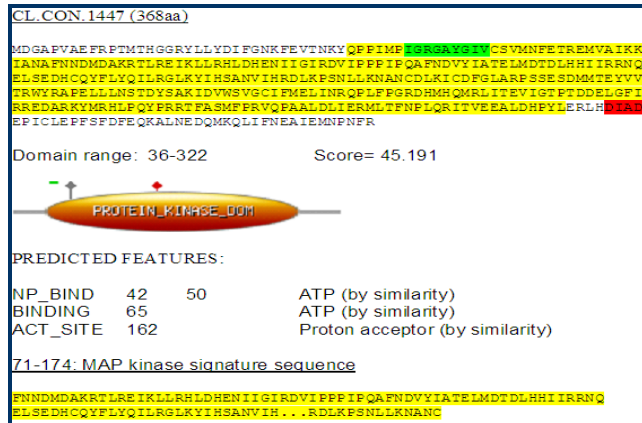


Figure 1: Domain scan result for *Curcuma longa* MAPK sequence CL.CON. 1447. A distinct protein kinase domain was found with MAP kinase signature sequence, ATP binding site and serine/threonine protein kinase active site.

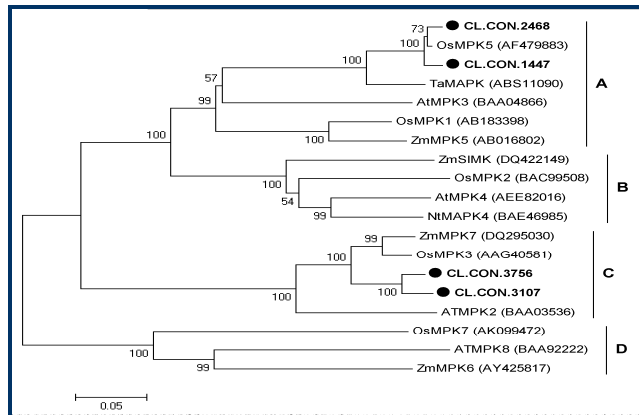


Figure 2: Phylogenetic tree of MAPKs from *Curcuma longa* and other higher plants. Letters A-D on the right indicates different groups of MAPKs. The protein sequences of the MAPKs used for construction of the tree are listed in the GenBank database under the following accession numbers: *OsMPK5* (AF479883), *TaMAPK* (ABS11090), *AtMPK3* (BAA04866), *OsMPK1* (AB183398), *ZmMPK5* (AB016802), *ZmSIMK* (DQ422149), *OsMPK2* (BAC99508), *AtMPK4* (AEE82016), *NtMAPK4* (BAE46985), *ZmMPK7* (DQ295030), *OsMPK3* (AAG40581), *AtMPK2* (BAA03536), *OsMPK7* (AK099472), *AtMPK8* (BAA92222), *ZmMPK6* (AY425817).

Conclusion:

Turmeric, also known as the ‘golden spice’, is one of the most important herbs in tropical and sub-tropical countries and valued worldwide as a spice, food preservative, and in traditional medicine. It consisted of about 110 species which are

basically sterile and propagates exclusively by vegetative means. This indicates that turmeric genotypes may provide invaluable tools for studying and characterizing various genes that are expressed in asexually reproducing plant. In the present study, we identified four *Curcuma* MAPK sequences from the EST database of the plant. The predicted amino acid sequences of the *Curcuma* MAPKs showed high similarity in the eleven kinase subdomains that are necessary for the catalytic function of the serine/threonine protein kinase. Moreover, they were categorized in subgroup A and C exhibiting high similarity with MAPK sequences from rice and citing their involvement in biotic or abiotic stress responses. Further biochemical and functional analysis of the characterized MAPK sequences are necessary to establish their physiological role in *Curcuma longa*. The analysis will act as a framework for further dissection of the MAPK signaling network in *Curcuma* as well as other asexually reproducing plants.

Acknowledgement:

The authors are grateful to Dr. Manoj Ranjan Nayak, President, Siksha O Anusandhan University for his encouragement and support.

References:

- [1] Jonak C *et al.* *Cell Mol Life Sci.* 1999 **55**: 204 [PMID: 10188583]
- [2] Hirt H. *Trends Plant Sci.* 1997 **2**:131
- [3] Jonak C *et al.* *Curr Opin Plant Biol* 2002 **5**: 415 [PMID: 12183180]
- [4] Asai T *et al.* *Nature* 2002 **415**: 977 [PMID: 11875555]
- [5] Ichimura K *et al.* *Trends Plant Sci.* 2002 **7**: 301 [PMID: 12119167]
- [6] Altschul SF *et al.* *Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694]
- [7] Reyna NS & Yang Y, *Mol Plant Microbe Interact* 2006 **19**: 530 [PMID: 16673940]
- [8] Wu T *et al.* *Mol Biol Rep.* 2011 **38**: 3967 [PMID: 21120617]
- [9] Rudd JJ *et al.* *Plant Physiol.* 2008 **147**: 802 [PMID: 18441220]
- [10] Thomson JD *et al.* *Nucleic Acids Res.* 1997 **25**: 4876 [PMID: 9396791]
- [11] Saitou N & Nei M, *et al.* *Mol Biol Evol.* 1987 **4**: 406 [PMID:3447015]
- [12] Kumar S *et al.* *Bioinformatics.* 2001 **17**: 1244 [PMID: 11751241]
- [13] Sitnikova T *et al.* *Mol Biol Evol.* 1995 **12**: 319 [PMID: 7700156]
- [14] Nicole M-C *et al.* *BMC Genomics.* .2006 **7**: 223 [PMID: 16945144]
- [15] Hamel LP *et al.* *Trends Plant Sci.* 2006 **11**: 192 [PMID: 16537113]
- [16] Nadarajah K *et al.* *Asian Jour Plant Sci* 2010 **9**:1
- [17] Song F *et al.* *Planta.* 2002 **215**: 997 [PMID: 12355160]
- [18] Agrawal GK *et al.* *Biochem Biophys Res Commun.* 2003 **302**: 171 [PMID: 12604328]
- [19] Hyun TK *et al.* *Genes & Genomics.* 2010 **32**: 275

Edited by P Kanguane

Citation: Joshi *et al.* *Bioinformation* 7(4): 180-183 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: A dataset of *Curcuma longa* MAPK sequences from GenBank EST database

<i>Curcuma longa</i> contig	Size (AA)	T- loop	Homologous sequence	NCBI A/C	E.value	Group
CL.CON.1447	368	TEY	MAP kinase <i>OsBIMK1</i> (<i>Oryza sativa</i> Indica group)	AAK01710	0.0	A
CL.CON.2468	366	TEY	Mitogen activated protein kinase 1, (<i>Oryza sativa</i> indica cultivar group) <i>OsMPK1</i>	AB183398	0.0	A
CL.CON.3107	370	TEY	<i>Oryza sativa</i> (indica cultivar-group) mitogen activated protein kinase 3 (<i>OsMPK3</i>)	DQ826424	0.0	C
CL.CON.3756	369	TEY	<i>Oryza sativa</i> MAPK4 protein	AJ251330	0.0	C