## **BIOINFORMATION** Discovery at the interface of physical and biological sciences

open access

www.bioinformation.net

Volume 6(7)

Web Server

# MAP Kinase analyser: A tool for plant kinase and substrate analysis

Saranga Dhar Samantaray<sup>1</sup>, Parvendra Singh<sup>1</sup>, Bonthla Venkata Suresh<sup>2</sup>, Sugandha Sharma<sup>2</sup>, Gohar Taj<sup>2</sup>\*, Anil Kumar<sup>2</sup>

<sup>1</sup>Department of Computer Science & Technology, College of Basic Sciences & Humanities, G.B. Pant University of Agriculture & Technology, Pantnagar, Uttrakhand-263145, India; <sup>2</sup>Department of Molecular Biology & Genetic Engineering, College of Basic Sciences & Humanities, G.B. Pant University of Agriculture & Technology, Pantnagar, Uttrakhand-263145, India; Gohar Taj - Email: gohartajkhan@rediffmail.com; \*Corresponding author

Received May 28, 2011; Accepted May 30, 2011; Published June 23, 2011

#### Abstract:

MAPK (Mitogen Activated Protein Kinase) is a Ser/Thr kinase, which plays a crucial role in plant growth and development, transferring the extra cellular stimuli into intracellular response *etc.* Manual identification of these MAPK in the plant genome is tedious and time taking process. There are number of online servers which predict the P-site (phosphorylation site), find the motifs and domain but there is no specific tool which can identify all them together. In order to identify the P-Site, phosphorylation site consensus sequences and domain of the MAPK in plant genome, we developed a tool, MAP Kinase analyzer. MAP kinase analyzer take protein sequence as input in the fasta format and the output of tool includes: 1) The prediction of the phosphorylation site *viz.*, Serine (S), Threonine (T), and Tyrosine (Y), Contex, Position, Score and phosphorylating kinase as well as the graphical output; 2) Phosphorylation site consensus sequence pattern for different kinases and 3) Domain information about the MAPK's. The MAP kinase analyzer tool and supplementary files can be downloaded from http://www.bioinfogbpuat/mapk\_OWN\_1/.

#### **Background:**

MAPK are special class of kinases which are activated by various growth factors, differentiating factors, M-phase phosphorylation cascade reactions and involved in biotic and abiotic stress signaling pathways [1]. They play a key role in the transmission of external signals such as mitogens, hormones and different stresses. A number of prediction servers are available over the World Wide Web. These servers facilitate the prediction like GPS: a comprehensive www server for phosphorylation sites prediction; PPSP: prediction of PKspecific phosphorylation site with Bayesian decision theory; PhoScan: Prediction of kinase-specific phosphorylation sites with sequence features by a log-odds ratio approach; MEME Suite: Motif-based sequence analysis tools; FANMOD: a tool for fast network Protein Consensus Sequence Motif detection; SMART (a Simple Modular Architecture Research Tool); d-Omix: a mixer of generic protein domain analysis tool; Phospho.ELM etc. Manual identification of MAPKs are tedious and time consuming, in order to identify MAPK there are no specific tools as yet, which predicts altogether the phosporylation site (P-Site), P-site consensus sequence or pattern and domain. Thus we developed a tool MAP kinase analyzer, which solves most of the above faced problems.

#### Methodology:

#### Datasets:

For training, the P-sites, positive datasets were obtained from Phospho.ELM database [2], which contains 2540 substrate proteins from different species covering 4799 S, 974 T and 1433 Y sites. To remove redundant fragments within the datasets, the initial datasets were filtered using a 30% sequence identity. The negative (*i.e.* non-phosphorylation sites,  $N_S$ ,  $N_T$  and  $N_Y$ ) were

ISSN 0973-2063 (online) 0973-8894 (print) Bioinformation 6(7): 286-287 (2011) obtained from these 2540 protein sequences and represented all S, T and Y residues that were not reported as being phosphorylated in Phospho.ELM database. For the phosphorylation site consensus sequence and domain analysis, the primary data is retrieved from TAIR [3], NCBI [4] and UniProt databases [5], and after retrieval these data are manually as well as with the help of tool like Multalin [6], optimized.

#### Algorithm:

In the case of P-site prediction we designed a neural network for the prediction and genetic algorithm for training this neural network which can be called as genetic algorithm neural network (GANN) [7]. The GANN uses GA (Genetic Algorithm) to optimize the connection weights of the ANN (Artificial Neural Network) over the training dataset. In our GANN model, the number of input nodes is equal to the dimensionality of feature vector, *i.e* 24. The neural network uses a sigmoid function to provide a continuous activation function. GANN is used to construct a P-site predictor with the following configuration: (1) the maximum generation number-1000; (2) the threshold of the fitness value-0.7; (3) population size-100; (4) crossover probability -0.9. In case of Psite consensus pattern and domain search for the MAP kinases, firstly MSA (Multiple Sequence Alignment) has been done using Multalin tool. From MSA, we detected 11 conserved sub-domains which are present in all plant MAPKs. Among all sub-domains, 8 sub-domains contain the MAP kinase activation loop ((TE/D) Y) and the MAPK's signature sequence (T(E/D) YxxTRWYRAPEL) [8]. These sub-domains are optimized and included in the tool for prediction. This tool is developed on NetBeans suite with Java language [9].

### **BIOINFORMATION**

	equence A	nalysis		G.S.Part	University of Agri. & Tech.						
		V.	-	19	II ZS	Phosp Total au 70 Your Seq 6272	phorylation Site Pa mber of amino acid is: ence:	rediction	2		
NIV.	THE T	A A	H.			Output:	GAYGIYESVLÜTETNELV	AMERTA			Graphical Output
TIL			200			Site	Context	Position	Score	) P	hosphorylating Group
				-		5	FRAMETHOGOFENDEGSUFEIT	5 / 21	0.525167	j	Tyrosine Motif
						5	HEEGEFENDFERLFEITSCHAPP	20	1.789-045	Terr	ninal Methylation in Bacterial Piles
Gelender	Welc	ome to MA	PK Sequ	ience A	nalysis	-5	STOPCSLEDTSKIRPPIPICAG	25	0.5685577	17	Tyroskie Motif
ator 2011		-				т	12 INNEGGOQUITOFRAVE		0.51250035		No Netch
	3 (Mai	18	a new of network	Sand method	to policing printing	17.11	MINTOGO CYTOFPAYETHOO	8 . 9	0.538175	)	Tyrosine Motif
1 2 2 4 5 5	18		INDONESI.		and the second second second	TIL	GGGQYTDFPANETHGGQF159DH	16 16	0.6992743	)	No Match
8 8 10 11 12 13	1 1 2	5.4				Υ	ANTGGGQTTDFPAVETHGG		.0.73889635	)	Tyrosine Motif
1. 1. 1. 1. 1. 2.	a 13		ric mbook to to	rpet we Ju	rightEuX database, using	Y I	PARETHOGOFENDFGSLFEITS	K 24 -	0.6031892	]	Tyrosine Motif
Phosphorylati	ion Site Cons	concue So		1,005	-	Domai	oressensemented		0.6751746	)	Tyrodise Motif
Phosphorylati Total number of and Input Sequence: 400 #STGGGYTOFFA IFFIGEGATORY Phosphorylation Site	on Sile Cons o acid is: .vetmodqfistbi .svedtetmeevas Coosemus Sog. Fo	sensus See	quence	raity. 25		Y	DPGGIFETSSCREPHERGEAU in Information her of amino acid is: sence: D00000000000000000 00 y TDFPAYETNOGGF 00 GY TDFPAYETNOGGF 00 GY TDFPAYETNOGGF 00 GY TDFPAYETNOGGF 00 GY TDFPAYETNOGGF 00 GY TDFPAYETNOGCF 00 GY TDFF 00 G		0.6751746 1.7 E1 75 K A NAP ON M R Q F 50 V 1 R Q F 50 V 1 R Q F 50 V 1 M C 1 M N K M K 1 T K E M K R 1 T K E	7 K.P.P. III D.A.K. III D.A.K. IIII D.A.K. IIIII D.A.K. IIIIII D.A.K. IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	Tyroshe Motif
Phosphorylation Total number of analo Input Sequence: arr arr arr brocket of contained arr arr arr phosphorylation Site	on Sile Cons wardd In: 	sensus Sei Föströrtsi und: Consensus Sequence Found	quence start	End	Pictorial Representation	у <b>Domai</b> Total num 270 Input Sequ матория как с как с к к к к к к к к к к к к к	DPGSIFETSSCREPPERCAL In Information due of amilino acid is: armsce: DSS acid biological SS acid biolo	1 5 Y 0 1 F 6 6 6 C Y AMK 8 F 9 Y 9 P 1 6 7 F 5 Y 0 7 F 5 Y	- 0.6751746 L F E F T5 K A N AF D N M R Q 5 D N H S E N D F M T M R T S E N F M D A R T Y R D C E Q I K E I	1 M P P M D A K 5 M V 5 M V 5 M V 5 C P P Q Q H 1 Y Q Q H 1 Y Q	Tyroshe Motif
Phosphorylation Total number of amin Input Sequence: ATO #TGGGQTTDFFA IFFGRGATGIVE Phosphorylation Site Bictinyl-Lysive	on Site Cons to acid in: (vctmccqristo) (svctrctmctra) Consentuut Soq. Fo Pattern [AMjinkim/*	sensus Sei rostrortsi and: Consensus Sequence Found	quence start	End 68	Pictorial Representation	у <b>Domain</b> 200 Тотаl num 200 Тария Segn матос 1 каза каз	DPGSIFETSSCREPPERCAT		1.0.4751746 L.F.E.F.T.S.K.I A.N.A.F.O.N.I.I R.G.K.V.F.H.T R.G.K.V.F.H.T R.G.K.V.H.H. N.R.R.T.Y.E. D.C.E.Q.I.K.C.I D.C.E.Q.I.K.C.I	1 M D A K 1 M D A K 2 M D A K 2 K V V 2 K V V 2 K V V 2 K V V 2 K V V 4 K V V V 4 K V V V 4 K V V V V V 4 K V V V V V V V 4 K V V V V V V V V V V V V V V V V V V	Tyroshe Motif
Phosphorylati Total number of analn N Input Sequence: aN martocograpse iiFickCardive Phosphorylation Site Bactoyl Lysine Termina Methylation in Bacterial Pilm	Image:	Sensus Sen IFGSLFEITSI INSEA Sequence Found ABC .COLFE	quence Start 22	End 53	Pictorial Representation	Y           Domain           200           Input Sequence           abstract Sequence           abstrand           abstrand	DPGSIFETSSCREPPHORCAN In Information ther of amilino acid is: armsce: DDGSIFETSSCREPPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGSIFETSSCREPHORCAN DDGS		0.4751746 LFEITSK AHAFONK RGFINT GLEVIN RGLVIN RGLVIN AHAFONK RHITKI DEEQIKI Start	r K P P H D A K I S T E S A H I E C F P Q L P H Q L P H Q L P H Q L I Y Q End	Pictorial Representation
Phosphorylati Total number of and Total number of and Total number of and Total number of and N Input Sequence: ATO MATCOOUTOFFA IFFICACOUTOFFA IFFI	in         Site Cons           in         Site Cons           in         Site Cons           in         Site Cons           Site Cons         Site Cons           Site Cons         Site Cons           Site Cons         Site Cons           Consensus Seq. For         Site Cons           [AM]MGUN*         [AM]MGUN*           [G[W]]r[St][[T]         [AG_SWWP]E           G.G., G[AV, FWWP]         G.G., G[AV, FWWP]	sensus Sea rostreitsi net: Consensus Sequence Found AMK COLFE	guence Start 53	End 52	Pictorial Representation	Y           Domcai           270           Input Sequence           e0012100           MOTOL           1 × 1 0           1 ×	DPGGLFETSSCHEPHERGEGAN In Information her of amino acid is: armor: DDDGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOGGT DGGTDFFAYETHOG DGGTDFFAYETHO		LFEITSK ANAFOMM ASTROMMA SCHOLMAN SCHOL	C R.P.P. M.D.A.K. S.S.T.K. S.S.T.K. S.S.T.K. Q.L.P.M. Q.L.P.M. Q.L.P.M. M.I.Y.Q. End 33 90	Pictorial Representation
Phosphorylati Total number of analn Total number of analn Trends State Terminal Methylation Site Discretise Phose Trends Methylation Site Discretise Phose Tyrosine Motif	m     m	Semsus Sen FGSLFEITSI MARIA Sequence Found ABC GRCATCI YTDF	quence Start 65 23 47 3	End 48 31 32 12	Pictorial Representation	Y           Domain           270           Tagant Sequence           edb22500           NN T 0           1 N 0           C 0 0           T 1 N 0           C 0 0           T 0 0           C 0 1           C 0 0           T 0 0           C 0 1           Domain           1           2           5	Depositerrescrependenciar           Information           Aber of ampino acid is:           assisce:           DDDStructure           DOST DF PAYSTHOOD TO           DAGY DS TO STAND           DE MONTONIC STAND           DE	S Y OL F G S     C Y ANK F G      C	0.4751746	2 APP 8 D 2 A 5 S T 2 5 S T	Pictorial Representation
Phosphorylatio Total number of amin Total number of amin Input Sequence: arro #STGGGQTTDTFA ITFIGEGATOTO Phosphorylation Sile Biothyl-Lynkie Terniau Methyl Plans Disucleotide Binding Proteins Tyrosine Motif Tyrosine Motif		sensus Sei rostrontsi and: Consensus Sequence Found AMK Coller GREATCH	guence Start 63 23 63 23 24	End 48 33 32 28	Pictorial Representation	Y           Domain           208           Input Seq           e002200           0.00 200 <td>DPGSLFETTSSCREPPERCACK           in Information           her of amplies acid is:           according to the part of the optimized of the part of the part of the optimized of the the part of the optimized of the part of</td> <td></td> <td>L F E J TS K A N AF D N H K G K V J H S C V J K N R R T V C D C C Q I C V Start 7 44 83 93 94 94 94 94 94 94 94 94 94 94 94 94 94</td> <td>r R.P.P. R.D.A.K. I S.T.G. S.A.K.I S.T.G. Q.A.L.N. MI Y.Q. End 300 1227 1255</td> <td>Pictorial Representation</td>	DPGSLFETTSSCREPPERCACK           in Information           her of amplies acid is:           according to the part of the optimized of the part of the part of the optimized of the the part of the optimized of the part of		L F E J TS K A N AF D N H K G K V J H S C V J K N R R T V C D C C Q I C V Start 7 44 83 93 94 94 94 94 94 94 94 94 94 94 94 94 94	r R.P.P. R.D.A.K. I S.T.G. S.A.K.I S.T.G. Q.A.L.N. MI Y.Q. End 300 1227 1255	Pictorial Representation
Phosphorylatio Total number of amin Tyroine Motif Tyroine		Sensus Sen FGSLFEITSI MERTA Sequence Found ANK GREATER TDF 1999	guenco Start 5 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	End 31 22 24 44	Pictorial Representation	Y           Domain           200           Input Seg           a00/2016           N = 7 do           N = 7 do           T = 0 do	DPGSLFEITSSCHEPPERCAT           in Information           her of amilino acid is:           armino acid is:		L F E I TS R I A NAF O HU R G K V I M S E N F V E V D E C Q I K C I Start V 44 1 13 120 120 120	End 10 10 10 10 10 10 10 10 10 10	Pictorial Representation

Figure 1: A Snapshots of the MAP kinase analyzer showing homepage, P-site prediction, P-site consensus sequence pattern and domain prediction of MAPK.

#### User interface:

In the case of P-site prediction, when user inputs a sequence in fasta format, the sequence is then processed by the neural network and result is displayed in the form of numeric scores corresponding to all S, T and Y present in the sequence and also highlight them. The user can opt for a threshold value at which user wants to see the P-sites. In the case of P-site consensus sequence pattern or motif and domain analysis for MAPK, a file is created containing all the patterns that have to be searched for in the given sequence. The sequence is then processed and the motif as well as domains present in the sequence is displayed to the user in text as well as graphical representation form.

#### Utility:

A novel tool for P-site, P-site consensus sequence pattern and for the domain prediction of MAPK has been developed (**Figure 1**). We have given a graphical and user friendly interface, so the tool is easy to use. Through this tool we can identify whether the given protein sequence is a MAP kinase or not on the basis of presence of the specific MAPK domain, in addition, we can also identify the possible kinase substrate by the analysis of P-site consensus sequence pattern, which consequently gives an idea about the functioning of the protein.

#### **Conclusion:**

Performance evaluation with dataset and database variants clearly indicates that MAP kinase analyzer has significantly high accuracy in terms of specificity and sensitivity. To the best of our knowledge MAP kinase analyzer is the first ever tool which identifies the P-Site, phosphorylation site consensus sequences and domain of the MAPK in plant genome altogether.

#### Acknowledgments:

Authors are grateful to Sub-DIC, Bioinformatics unit at G.B. Pant University of Agriculture and Technology, Pantnagar, India for providing computational facility. This study was supported by Department of Biotechnology, Govt. of India under Programme Mode Support Project. Authors also acknowledge Payal Agarwal, Hitesh Arora, Abhishek Choudhary, and Manesh Pal for their kind support in the development of tool.

#### **References:**

- [1] Jonak et al. Cell Mol Life Sci. 1999 55: 204 [PMID: 10188583]
- [2] http://phospho.elm.eu.org/
- [3] http://www.arabidopsis.org
- [4] http://blast.ncbi.nlm.nih.gov
- [5] http://www.uniprot.org
- [6] http://www.multalin.toulouse.inra.fr/multalin/multalin.html
- [7] http://www.cbs.dtu.dk/services/NetPhos/
- [8] Nadarajah K & Sidek HM. Asian J Plant Sci. 2010 9: 1
- [9] http://www.java.com

#### Edited by P Kangueane

#### Citation: Samantaray et al. Bioinformation 6(7): 286-287 (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.