International Journal of Bioinformatics Research
ISSN: 0975–3087, E-ISSN: 0975–9115, Vol. 3, Issue 1, 2011, pp-167-177
Available online at http://www.bioinfo.in/contents.php?id=21

PREDICTION OF DOWNSTREAM INTERACTION OF TRANSCRIPTION FACTORS WITH MAPK3 IN *ARABIDOPSIS THALIANA* USING PROTEIN SEQUENCE INFORMATION

GOHAR TAJ*, SUGANDHA SHARMA, VIKRAM S. GAUR, ANIL KUMAR
Dept. of Molecular Biology & Genetic Engineering , College Of Basic Sciences & Humanities, G.B. Pant University of Agriculture & Technology, Pantnagar, Uttrakhand-263145, India.
*Corresponding Author: gohartajkhan@rediffmail.com

Received: December 13, 2010; Accepted: February 18, 2011

Abstract- Protein-Protein interactions (PPIs) are vital to most biological processes thus the identification of PPIs is of primary importance. In the present work, we endeavor to identify the downstream interaction partners of Mitogen Activated Protein Kinase3 (MAPK3) in *Arabidopsis Thaliana* using the information of protein sequences through Support Vector Machine (SVM) approach. The approach here used is supervised learning based on physiochemical properties of protein sequences through which we predict whether the MAPK3 proteins interact with downstream transcription factor proteins *viz.,* Myb, bZIP, WRKY, Myb-related proteins, AP2/EREBP, and NAC with which its interaction is almost unknown. The Myb-related transcription factor family is showing maximum interaction percentage *i.e.* 71.14% with MAPK3 while minimum interaction percentage is 21.15% which is shown by NAC transcription factor family. The interaction percentage shown by the gene loci of rest transcription factor family *i.e.* Myb, bZIP, AP2/EREBP, WRKY are 67.78%, 68.05%, 21.91% and 58.33% respectively. The results of our study clearly revealed the complexity of MAPK3 interaction with several variants of different transcription factors and the same can be verified by different methodology of wet lab experimentation for elucidating the role in various biological processes.
Key Words*:* Protein-Protein interactions (PPIs); Mitogen Activated Protein Kinase3 (MAPK3); Support Vector Machine (SVM), Transcription Factor (TF).

INTRODUCTION

Signalling through Mitogen Activated Protein Kinase (MAPK) is a fundamental and conserved process in eukaryotes and transduce external signals through protein phosphorylation [1-2]. Signal transduction is the means by which cells respond to extracellular information by interacting with other proteins. The knowledge of the interactions among proteins is essential for understanding the molecular mechanisms inside the cell [3]. In plants signal transduction are mediated by a special class of Kinases known as MAP kinases. MAP kinase pathway is one of the main phosphorylation pathway that plants use in biotic and abiotic stress resistance. It is a cascade consisting of several classes of kinases, each having a different role in signal integration and divergence. The *Arabidopsis* genome contains 23 MAPKs, 10 MAPK2 and 60 MAPK3. These kinases are involved in a variety of functions, including growth, development and responses to environmental and endogenous stimuli as well as responses to plant hormones. The role of these kinases in how plants respond to pathogen attack and other abiotic and biotic factors is currently poorly understood. This is primarly because of many MAPK-substrate interactions are very transient and unstable in order to facilitate rapid signaling. Therefore, it is difficult to identify the downstream interacting partners or the substrates of these kinases, which complicates the understanding of the development of pathogenesis in plants. Many experimental methods have been developed to study the protein-protein interactions including yeast two hybrid systems, affinity purification followed by mass spectrometry and the phage display libraries, but these methods have its own limitations and suffer from high false positive rate [4]. Therefore these limitations highlight the need of *in silico* interaction predictions.

Transcription factors regulate the expression of genes at the transcriptional level and act as transcriptional activators or repressors. Transcription factors/nuclear proteins represent the largest group of MPK targets and regulated by MAPK phosphorylation [5]. To date, a number of transcription factors including TGA, Myb, bzip, WRKY, Myb-related proteins, AP2/EREBP, homeobox have been shown to act downstream/substrates of various MAPKS [6-11, 5]. As stated earlier WRKY has been shown to act downstream of various MAPKS. WRKY transcription

factors are involved in several diverse pathways [12] and perform central functions in the regulation of developmental process, senescence and defence responses against pathogens [13-16]. WRKY genes contain at least one conserved DNA binding region, called WRKY domain, comprising the highly conserved about 60 amino acids in length WRKYGQK sequence [16]. Several biochemicals, molecular and genetic studies shows that bZIP transcription factors regulates and play role in various diverse function such as cell elongation [17], development and stress [18]. bZIP transcription factors super family consists of 75 members in Arabidopsis [19]. NAC one of the largest plant's specific transcription factor family comprises more than 100 members [20-21] and have a characteristic conserved domain at N-terminal named NAC domain [22] The NAC are known to posses diverse roles. NAC proteins bound specifically to the CATGTG motif both in vitro and in vivo. MYB factors represent a family of proteins that include the conserved MYB DNA-binding domain. Plants contain a MYB-protein subfamily that is characterised by the R2R3-type MYB domain. Remarkably, the myb-domain includes only one of the typical two or three tryptophan repeats found in other myb-like proteins [23]. AP2 and EREBPs are transcription factors that contain the AP2 DNA-binding domain. This family of transcription factor genes play a variety of roles throughout the plant life cycle like floral organ identity determination or control of leaf epidermal cell identity, to forming part of the mechanisms used by plants to respond to various types of biotic and environmental stress and constitute a multigene family [20].

Thus it is evident that the above discussed transcription factors *viz.,* Myb, bZIP, WRKY, Myb-related proteins, AP2/EREBP, and NAC play crucial role in signal transduction, defence responses against pathogens development and stress, *etc.* in plant life cycle. And the MAPK cascade is one of the major and evolutionally conserved signaling pathways and plays key role in the regulation of stress and developmental signals in plants. Keeping in view of the above, in present study the interaction of MAPK3 is predicted with the various transcription factors *viz.,* NAC, Myb, bZIP, WRKY, Myb-related proteins, and AP2/EREBP transcription family. In this work we employed Support Vector Machine (SVM) technique for building a supervisory learning model which can be subsequently employed for identification of query sequences. Different types of features were employed as domain knowledge. In supervised learning, objects in a given collection are classified using a set of attributes, or features. The result of the classification process is a set of rules that prescribe assignments of objects to classes based solely on values of features. SVMs are a useful technique for data classification. Support vector machines consider a two-class, linearly separable classification problem. While many decision boundaries exist that are capable of separating all the training samples into two classes. Among these decision boundaries, SVMs find the one that achieves maximum margin between the two classes. The margin is defined as the distance between a planar decision surface that separates two classes and the closest training samples to the decision surface. Given a training set of instance-label pairs (x_i, y_i), i = 1……, l where x_i £ R^n and y £ $(1,-1)^l$, the SVM [24-25] requires the solution of the following optimization problem:

$\min_{w, b, \xi}$ ½ $W^T W + C\sum^{l}_{i=1} \xi_i$

Subject to $y_i (w^t \phi (x_i) + b) \geq 1 - \xi_i$,

$\xi \geq 0$

Here training vectors x_i are mapped into a higher dimensional space by the function ϕ. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. C > 0 is the penalty parameter of the error term [26].

MATERIALS AND METHODS

The approach here used is supervised learning through which we predict whether the MAPK3 proteins interact with other proteins *viz.,* Myb, bZIP, WRKY, Myb-related proteins, AP2/EREBP, and NAC with which its interaction is almost unknown. Based on examples of interacting pairs and non interacting pairs, we train a binary classifier to predict the class (interacting or non-interacting) of a set of protein sequences. The classification model for predicting protein-protein interaction (PPIs) is based on SVM.

Collection of data set, Feature extraction and Model Construction

PPI can be defined as four interaction modes: electrostatic interaction, hydrophobic interaction, steric interaction and hydrogen bond [27]. Here seven physicochemical feature groups of amino acids are selected to reflect these interaction modes and they are : *1)* Amino acid, dipeptide composition, *2)* Normalized Moreau-Broto autocorrelation, *3)* Moran autocorrelation, *4)* Geary autocorrelation , *5)* Composition, transition, distribution, *6)* Sequence order, *7)* Pseudo amino acid composition (lamda=30). Amino acid indices used in these auto-correlation descriptors are: *1)* Hydrophobicity scales, *2)* Average flexibility indices, *3)* Polarizability parameter *4)* Free energy of solution in water *5)* Residue accessible surface area in trepeptide, *6)* Residue volume, *7)* Steric parameter, *8)* Relative mutability. These features are extracted with the help of PROFEAT server (http://jing.cz3.nus.edu.sg/cgi-bin/prof/prof.cgi). The final training data set consisted of 92 protein pairs, half from the positive data set and half from the negative data set. The positive data set is formed by

the set of 46 non redundant proteins that shows interaction with MAPK3 in *Arabidopsis Thaliana* as proved by the wet lab experimentation [28] and was downloaded from TAIR (http://www.arabidopsis.org/). The negative data is formed by the artificial protein sequences by shuffling the sequences of positive data set at k-let count one. It has been demonstrated that if a sequence of one interacting pair is shuffled, then the two proteins can be deemed not to interact with each other. This is based on K-let Shuffling algorithm [29]. The positive and negative data set must meet these requirements: *(i)* the non-interacting pairs cannot appear in the whole set of interacting pairs, *(ii)* the number of negative pairs is equal to that of positive pairs. Whereas, the testing sets are composed of protein sequences of Myb, bZIP, WRKY, Myb-related proteins, AP2/EREBP, and NAC, transcription family containing 149, 72, 72, 49, 146, and 104 gene loci respectively; which are downloaded from DATF (http://datf.cbi.pku.edu.cn/). The model is generated by web server "gist-classify" [30] (Gist, version 2.2, http://svm.sdsc.edu/cgi-bin/nph-SVMsubmit.cgi) based on SVM. For model generation the SVM is trained by a labelled training set and then using the trained SVM to make predictions about the classifications of an unlabeled test set *i.e* of Myb, bZIP, WRKY, Myb-related proteins, AP2/EREBP, and NAC, transcription family is achieved.

RESULTS

There is very little work in the literature regarding prediction of MAPK3 interaction with other proteins in *Arabidopsis Thaliana*. The MAPK cascade is one of the major and evolutionary conserved signaling pathway and plays crucial role in the regulation of stress and developmental signals in plants. In this work we employed SVM based predictive modeling approach to predict the interaction of MAPK3 with Myb, bZIP, WRKY, Myb-related proteins, AP2/EREBP, and with NAC transcription factor family. In Myb transcription factor family 101 gene loci are interacting "Fig. (7)" and 48 are non-interacting out of 149 gene loci "Table (1)"; while in Myb-related transcription factor family 35 gene loci "Fig. (4)" from total of 49, showing positive interaction while rest are showing negative interaction with MAPK3 "Table(4)". Likewise in bzip transcription factor family 23 gene loci are non-interacting while rest out of 72 "Fig. (5)" are showing interaction with MAPK3 "Table (2)"; while in WRKY transcription factor family out of 72 gene loci, 42 "Fig. (2)" showed the interaction with MAPK3 and rest are showing non-interaction "Table (3)". In AP2/EREBP transcription factor family out of 146 gene loci 32 "Fig. (6)" are interacting with MAPK3 and rest are non-interacting "Table (5)". While in NAC transcription factor family from 104 gene loci 22 loci are showing interaction "Fig. (3)" while rest are non-interacting with MAPK3 "Table (6)". The Myb-related transcription factor family is showing maximum interaction percentage *i.e.* 71.14% with MAPK3 while minimum interaction percentage is 21.15% which is shown by NAC transcription factor family. While interaction percentage of Myb and bzip transcription factor family are almost equal *i.e.* 67.78% & 68.05% respectively. The interaction percentage of AP2/EREBP and NAC transcription factor family are comparatively low but similar *i.e.* 21.91% & 21.15% respectively and by WRKY family is 58.33% "Fig. (1)".

DISCUSSION

Networks of PPI provide a framework for the understanding of biological processes and can give insights into the mechanisms of diseases. Thus the understanding of biological mechanisms requires the knowledge of protein-protein interaction. MAPK is a conserved link between cell receptor and cell response and is mediated through gene expression which is regulated by transcription factors. The paper focuses on identifying different interacting partners of an important MAP kinase of Arabidopsis involved in disease signalling process. The identification of such partners will further enhance our knowledge about the mechanisms of disease resistance and help in developing novel strategies combating the pathogens. To our understanding this is the first report which has tried to identify the huge amount of downstream interacting partners of MAPK3.

In the present work, we predicted the downstream interaction of transcription factors with MAPK3 in *Arabidopsis Thaliana* using only the information of protein sequences through SVM. The results of our study clearly revealed the complexity of MAPK3 interaction with several variants of same or different transcription factors triggered in response to diverse upstream stimuli. The PPI networks can give insights into the mechanisms of diseases and provide a spectrum for the understanding of biological processes. Interaction networks can aid in designing signal transduction pathway and help to find the disease suppressive agents as well as uncover the key genes those are responsible for senescence and defence responses against pathogens. The same can be verified by wet lab experimentation using phosphoproteomics, Yeast2Hybrid system and affinity immuno chip based system. The results of present study further suggest to validate these results by physiochemical features. The need for bioinformatics methods to find out protein partners is being driven by the

generation of sequences at a rate far beyond our ability to carry out experimental functional analysis. Presently predictions of interaction of proteins are based on docking studies which is very computationally intensive and therefore unlikely to be able to scale to large interactions [3]. The main motivation behind this work is to perform a large-scale screening of pairs of proteins likely to interact, before validating the prediction by more expensive computational methods such as docking, or experimental methods for pairs of particular interest. The specificity, precision, and accuracy of PPI obtained in present study can also be compared to provide best prediction model. The non-interaction partners predicted in present study prompted us to look alternative MAPK proteins which are interacting with these non-interacting gene loci of various transcription factors.

Acknowledgment

Authors are grateful to sub-DIC, Bioinformatics unit at G.B. Pant University of Agriculture and Technology, Pantnagr, India for providing computational facility. This study was supported by Department of Biotechnology, Govt. of India under Programme Mode Support Project.

List of Abbreviations

Protein-Protein interactions - PPIs
Mitogen Activated Protein Kinase3 - MAPK3
Support Vector Machine – SVM
Transcription Factor - TF

References

[1] Bethkea G., Unthana T., Uhrigb J.F., Poschla Y., Gustc A.A., Scheela D. and Leea J. (2008) *PNAS* 106(19):867-872.

[2] Colcombet J. and Hirt H. (2008) *Biochem. J.* 413:217-226.

[3] Hue M., Riffle M., Vert J.P. and Noble W.S. (2010) *BMC Bioinformatics* 11:144.

[4] Deane C.M., Salwinski L., Xenaios I. and Eisenberg D. (2002) *Molecular and Cellular Proteomics* 1:349-356.

[5] Popescu S.C., Popescu G.V., Bachan S., Zhang Z., Gerstein M., Snyder M. and Kumar S.P.D. (2008) *Genes & Development* doi:10.1101/gad.1740009.

[6] Asai T., Tena G., Plotnikova J., Willmann M.R., Chiu W.L., Gomez-Gomez L., Boller T., Ausubel F.M. and Sheen J. (2002) *Nature* 415:977-83.

[7] Cheong Y.H., Moon B.C., Kim J.K., Kim C.Y., Kim M.C., Kim I.H., Park C.Y., Kim J.C., Park B.O., Koo S.C., Yoon H.W., Chung W.S., Lim C.O., Lee S.Y. and Cho M.J. (2003) *Plant Physiol* 132:1961–1972.

[8] Ekerngren S.K., Liu Y., Schiff M., Kumar S.P.D. and Martin G.B. (2003) *Plant Journal* 36: 905-917.

[9] Kim C.Y. and Zhang S. (2004) *Plant J.* 38:142-151.

[10] Liu Y. and Zhang S. (2004) *Plant Cell* 16: 3386–3399.

[11] Koo S.C., Moon B.C., Kim J.K, Kim C.Y., Sung S.J., Kim M.C., Cho M.J. and Cheong Y.H. (2009) *Biochemical and Biophysical Research Communications* pp. 365–370.

[12] Ulker B. and Somssich I.E. (2004) *Curr Opin in Plant Biol* 7(5):491-498.

[13] Rushton P.J. and Somssich I.E. (1998) *Curr Opin in Plant Biol*. 1:311-315.

[14] Rushton P.J., Torres J.T., Parniske M., Wernert P., Hahlbrock K. and Somssich I.E. (1996) *EMBO (Eur. Mol. Biol. Organ.) J.* 15:5690-5700.

[15] Singh K.B., Foley R.C. and Onate-Sanchez L. (2002) *Curr Opin in Plant Biol* 5(5):430-436.

[16] Eulgem T., Rushton P.J., Robatzek S. and Somssich I.E. (2000) *Trends in Plant Sci* 5(5):199-206.

[17] Fukazawa J., Sakai T., Ishida S., Yamaguchi I. and Kamiya Y. (2000) *Plant cell* 12: 901-915.

[18] Jakoby M., Weisshaar B., Droge-Laser W., Carbajosa J.V., Tiedeman J. and Kroj, T. (2002) *Trends in Plant Science* 7:106-111.

[19] Pitzschkea A., Djameia A., Teigea M., and Hirta H. (2009) *PNAS* doi:10.1073/pnas.0905599106.

[20] Riechmann J.L., Heard J., Martin G., Reuber L., Jiang C.Z., Keddie j., Adam L., Pineda O., Ratcliffe O.J., Samaha R.R., Creelman R., Pilgrim M., Broun P., Zhang, J.Z., Ghandehari D., Sherman B.K. and Yu G.L. (2000) *Science* 290(5499):2105-2110.

[21] Olsen A.N., Ernst H.A., Leggio L.L. and Skriver K. (2005) *Trends Plant Sci* 10:79–87.

[22] Aida M., Ishida T., Fukaki H., Fujisawa H. and Tasaka M. (1997) *Plant Cell* 9:841–857.

[23] Kirik V. and Baumlein H. (1996) *Pub Med Central PMID*: 8996094.

[24] Boser E., Guyon I. and Vapnik V. (1992) *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM Press* pp. 144-152.

[25] Cortes C., and Vapnik V. (1995) *Machine Learning* 20:273-297.

[26] Hsu C.W., Chang C.C. and Lin C.J. (2003) *A Practical Guide to Support Vector Classification.*

[27] Guo Y., Yu L., Wen Z. and Li M. (2008) *Nucleic Acids Research* 36(9).

[28] Feilner T., Hultschig C., Lee J., Meyer S., Immink R.G.H., Koenig A., Possling A., Seitz H., Beveridge A., Scheel D., Cahill D.J., Lehrach H., Kreutzberger J. and Kersten B. (2005) *Molecular & Cellular Proteomics* 4:1558-1568.

[29] Jiang M., Anderson J., Gillespie J. and Mayne M. (2008) *BMC Bioinformatics* 9:192.

[30] Pavlidis P., Wapinski I. and Noble W.S. (2004) *Bioinformatics* 20(4):586-587.

Table 1- Myb transcription factor family genes showing interaction and non-interaction with MAPK3.

Gene loci Id Showing Interaction	101	AT1G06180.1,AT1G08810.1,AT1GO9540.1,ATAT1G09770.1,AT1G14350.1,AT1G16490.1,AT1G17950.1,AT1G18570.1,AT1G18710.1,AT1G22640.1,AT1G25340.1,AT1G26780.1,AT1G34670.1,AT1G35515.1,AT1G48000.1,AT1G49010.1,AT1G56160.1,AT1G56650.1,AT1G57560.1,AT1G63910.1,AT1G66230.1,AT1G66370.1,AT1G66390.1,AT1G68320.1,AT1G69560.1,AT1G73410.1,AT1G74080.1,AT1G74430.1,AT1G74650.1,AT1G79180.1,AT2G02820.1,AT2G16720.1,AT2G23290.1,AT2G25230.1,AT2G26950.1,AT2G26960.1,AT2G31180.1,AT2G32460.1,AT2G36890.1,AT2G37630.1,AT2G38090.1,AT239880.1,AT2G47190.1,AT2G47460.1,AT4G05100.1,AT4G09450.1,AT4G09460.1,AT4G13480.1,AT4G17785.1,AT4G18770.1,AT4G21440.1,AT4G22680.1AT4G25560.1,AT4G26930.1,AT4G28110.1,AT4G32730.1,AT4G33450.1,AT4G34990.1,AT4G37260.1,AT4G37780.1,AT4G38620.1,AT5G01200.1,AT5G02320.1,AT5G05790.1,AT5G06100.1,AT5G06110.1,AT5G07690.1,AT5G07700.1,AT5G08520.1,AT5G10280.1,AT5G11050.1,AT5G11510.1,AT5G12870.1,AT5G14340.1,AT5G1531`0.1,AT5G16600.1,AT5G16770.1,AT5G17800.1,AT5G23000.1,AT5G23650.1AT5G26660.1,AT5G35550.1,AT5G39700.1,AT5G40330.1,AT5G40360.1,AT5G40430.1,AT5G41020.1,AT5G45420.1,AT5G49330.1,AT5G49620.1,AT5G54230.1,AT5G55020.1,AT5G56110.1,AT5G57620.1,AT5G58850.1,AT5G59780.1,AT5G60890.1,AT5G62320.1,AT5G62470.1,AT5G65230.1,AT5G65790.1
Gene loci Id Showing Non-Interaction	48	AT1G66380.1,AT3G01140.1,AT3G01530.1,AT3G02940.1,AT3G06490.1,AT3G08500.1,AT3G09230.1,AT3G09370.1,AT3G10580.1,AT3G10590.1,AT3G10595.1,AT3G11280.1,AT3G11440.1,AT3G11450.1,AT3G12720.1,AT3G12820.1,AT3G13540.1,AT3G13890.1,AT3G18100.1,AT3G23250.1,AT3G24310.1,AT3G27785.1,AT3G27810.1,AT3G27920.1,AT3G28470.1,AT3G28910.1,AT3G29020.1,AT3G30210.1,AT3G46130.1,AT3G47600.1,AT3G48920.1,AT3G49690.1AT3G50060.1,AT3G52250.1,AT3G53200.1,AT3G55730.1,AT3G60460.1,AT3G61250.1,AT3G62610.1,AT4G00540.1,AT4G01680.1,AT5G04760.1,AT5G14750.1,AT5G40350.1,AT5G52260.1,AT5G52600.1,AT5G58900.1,AT5G67300.1

Table 2. bZIP transcription factor family genes showing interaction and non-interaction with MAPK3.

Gene loci Id Showing Interaction	49	AT1G03970.1,AT1G06070.1,AT1G06850.1AT1G08320.1,AT1G13600.1,AT1G19490.1,AT1G22070.1,AT1G32150.1,AT1G35490.1,AT1G42990.1,AT1G43700.1,AT1G45249.1,AT1G49720.1,AT1G58110.1,AT1G59530.1,AT1G68640.1,AT1G68880.1,AT1G75390.1,AT1G77920.1,AT2G04038.1,AT2G12900.1,AT2G12940.1,AT2G13150.1,AT2G16770.1,AT2G18160.1,AT2G21230.1,AT2G21235.1,AT2G22850.1,AT2G31370.1,AT2G35530.1,AT2G36270.1,AT2G40620.1,AT2G40950.1,AT2G41070.1,AT2G42380.1,AT2G46270.1,AT3G44460.1,AT3G51960.1,AT3G62420.1,AT5G08141.1,AT5G15830.1,AT5G24800.1,AT5G28770.1,AT5G38800.1,AT5G42910.1,AT5G44080.1,AT5G49450.1,AT5G60830.1,AT5G65210.1
Gene loci Id Showing Non-Interaction	23	AT3G10800.1,AT3G12250.1,AT3G17609.1,AT3G19290.1,AT3G30530.1,AT3G49760.1,AT3G54620.1,AT3G56660.1,AT3G56850.1,AT3G58120.1,AT4G01120.1,AT4G02640.1,AT4G34000.1,AT4G34590.1,AT4G35040.1,AT4G36730.1,AT4G37730.1,AT4G38900.1,AT5G06839.1,AT5G06950.1,AT5G06960.1,AT5G10030.1,AT5G11260.1

*Table **3-** WRKY transcription factor family genes showing interaction and non-interaction with MAPK3.*

Gene loci Id Showing Interaction	42	AT1G13960.1,AT1G18860.1,AT1G29280.1,AT1G29860.1,AT1G30650.1,AT1G55600.1,AT1G62300.1,AT1G64000.1,AT1G66550.1,AT1G66560.1,AT1G66600.1,AT1G68150.1,AT1G69310.1,AT1G69810.1,AT1G80590.1,AT1G80840.1,AT2G03340.1,AT2G4880.1,AT2G21900.1,AT2G23320.1,AT2G24570.1,AT2G25000.1,AT2G30250.1,AT2G30590.1,AT2G34830.1,AT2G37260.1,AT2G38470.1,AT2G40740.1,AT2G40750.1,AT2G44745.1,AT2G46130.1AT2G46400.1,AT2G47260.1,AT3G01080.1,AT3G01970.1,AT3G04670.1,AT3G56400.1,AT3G58710.1,AT3G62340.1,AT4G01250.1,AT4G01720.1,AT5G45270.1

Gene loci Id Showing Non-Interaction	30	AT4G04450.1,AT4G11070.1,AT4G12020.1,AT4G18170.1,AT4G22070.1,AT4G23550.1,AT4G23810.1,AT4G24240.1,AT4G26440.1,AT4G26640.1,AT4G30935.1,AT4G31550.1,AT4G31800.1,AT4G39410.1,AT5G01900.1,AT5G07100.1,AT5G13080.1,AT5G15130.1,AT5G22570.1,AT5G24110.1,AT5G26170.1,AT5G28650.1,AT5G41570.1,AT5G43290.1,AT5G45050.1,AT5G46350.1,AT5G49520.1,AT5G52830.1,AT5G56270.1

Table 4- Myb-related proteins transcription factor family genes showing interaction and non-interaction with MAPK3.

Gene loci Id Showing Interaction	35	AT1G01060.1,AT1GO1380.1,AT1G01520.1,AT1G06910.1,AT1G09710.1,AT1G15720.1,AT1G17460.1,AT1G17520.1,AT1G18330.1,AT1G18960.1,AT1G19000.1,AT1G49950.1,AT1G58220.1,AT1G70000.1,AT1G71030.1,AT1G72650.1,AT1G72740.1,AT1G74840.1,AT2G30420.1,AT2G30432.1,AT2G36960.1,AT2G42150.1,AT2G44430.1,AT2G46410.1,AT2G46830.1,AT2G47210.1,AT3G05380.1,AT3G09600.1,,AT1G72650.1,AT3G10113.1,AT3G16350.1,AT3G21440.1,AT3G49850.1,AT3G53790.1,AT4G01060.1
Gene loci Id Showing Non-Interaction	14	AT2G13960.1,AT4G01280.1,AT4G39160.1,AT4G39250.1,AT5G02840.1,AT5G17300.1,AT5G27610.1,AT5G37260.1,AT5G47390.1,AT5G52660.1,AT5G53200.1,AT5G58340.1,AT5G61620.1,AT5G67580.1

Table 5- AP2/EREBP transcription factor family genes showing interaction and non-interaction with MAPK3.

Gene loci Id Showing Interaction	32	AT1G01250.1,AT1G03800.1,AT1G04370.1,AT1G06160.1,AT1G12610.1,AT1G12630.1AT1G15360.1,AT1G16060.1,AT1G19210.1,AT1G21910.1,AT1G22190.1,AT1G22810.1,AT1G22985.1,AT1G24590.1,AT1G25470.1,AT1G25560.1,AT1G28160.1,AT1G28360.1,AT1G36060.1,AT1G44830.1,AT1G46768.1,AT1G50640.1,AT1G50680.1,AT1G51120.1,AT1G51190.1,AT1G53170.1AT1G53910.1AT1G63030.1,AT1G64380.1,AT1G68840.1,AT1G71130.1,AT1G68550.1
Gene loci Id Showing Non-Interaction	114	AT1G12890.1,AT1G12980.1,AT1G13260.1,AT1G28370.1,AT1G33760.1,AT1G43160.1,AT1G49120.1,AT1G71450.1,AT1G71520.1,AT1G72360.1,AT1G72570.1,AT1G74930.1,AT1G75490.1,AT1G77200.1,AT1G77640.1,AT1G78080.1,AT1G79700.1,AT1G80580.1,AT2G20350.1,AT2G20880.1,AT2G22200.1,AT2G23340.1,AT2G25820.1,AT2G28550.1,AT2G31230.1,AT2G33710.1,AT2G35700.1,AT2G36450.1,AT2G38340.1,AT2G39250.1,AT2G40220.1,AT2G40340.1,AT2G40350.1,AT2G41710.1,AT2G44840.1,AT2G44940.1,AT2G46310.1,AT2G47520.1,AT3G11020.1,AT3G14230.1,AT3G15210.1,AT3G16280.1,AT3G16770.1,AT3G20310.1,AT3G20840.1,AT3G23220.1,AT3G23230.1,AT3G23240.1,AT3G25730.1,AT3G25890.1,AT3G50260.1,AT3G54320.1,AT3G54990.1,AT3G57600.1,AT3G60490.1,AT3G61630.1,AT4G06746.1,AT4G11140.1,AT4G13040.1,AT4G13620.1,AT4G16750.1,AT4G17490.1,AT4G17500.1,AT4G18450.1,AT4G23750.1,AT4G25470.1,AT4G25480.1,AT4G25490.1,AT4G27950.1,AT4G28140.1,AT4G31060.1`,AT4G32800.1,AT4G34410.1,AT4G36900.1,AT4G36920.1,AT4G37750.1,AT4G39780.1,AT5G05410.1`,AT5G07310.1,AT5G07580.1,AT5G10510.1,AT5G11190.1,AT5G11590.1,AT5G13330.1,AT5G13910.1,AT5G17430.1,AT5G18450.1,AT5G18560.1,AT5G19790.1,AT5G21960.1,AT5G25190.1,AT5G25390.1,AT5G25810.1,AT5G43410.1,AT5G44210.1,AT5G47220.1,AT5G47230.1,AT5G50080.1,AT5G51190.1,AT5G51990.1,AT5G52020.1,AT5G53290.1,AT5G57390.1,AT5G60120.1,AT5G61590.1,AT5G61600.1,AT5G61890.1,AT5G64750.1,AT5G65130.1,AT5G65510.1,AT5G67000.1,AT5G67010.1,AT5G67180.1,AT5G67190.1

Table 6-NAC Transcription Factor Family genes showing interaction and non-interaction with MAPK3.

Gene loci Id Showing Interaction	22	AT1G02210.1,AT1G02220.1,AT1GO2230.1,AT1G02250.1,AT1G19040.1,AT1G26870.1,AT1G28470.1,AT1G32870.1,AT1G33060.1,AT1G33280.1,AT1G34180.1,AT1G34190.1,AT1G52890.1,AT1G54330.1,AT1G56010.1,AT1G60280.1,AT1G60350.1,AT1G61110.1,AT1G65910.1,AT2G02450.1,AT2G17040.1,AT2G18060.1
Gene loci Id Showing Non-Interaction	82	AT1G01010.1,AT1G01720.1,AT1G03490.1,AT1G12260.1,AT1G25580.1,AT1G32510.1,AT1G32770.1,AT1G52880.1,AT1G60300.1,AT1G60340.1,AT1G60380.1,AT1G62700.1,AT1G64105.1,AT1G69490.1,AT1G77450.1,AT1G79580.1,AT2G24430.1,AT2G27300.1,AT2G33480.1,AT2G43000.1,AT2G46770.1,AT3G01600.1,AT3G03200.1,AT3G04060.1,AT3G04070.1,AT3G04420.1,AT3G04430.1,AT3G10480.1,AT3G10490.1,AT3G10490.2,AT3G10500.1,AT3G15170.1,AT3G15500.1,AT3G15510.1,AT3G17730.1,AT3G18400.1,AT3G29035.1,AT3G44290.1,AT3G44350.1,AT3G49530.1,AT3G55210.1,AT3G56520.1,AT3G56530.1,AT3G56560.1,AT3G61910.1,AT4G01520.1,AT4G01540.1,AT4G01550.1,AT4G10350.1,AT4G17980.1,AT4G28500.1,AT4G28530.1,AT4G29230.1,AT4G35580.1,AT4G36160.1,AT5G04400.1,AT5G04410.1,AT5G07680.1,AT5G08790.1,AT5G09330.1,AT5G13180.1,AT5G14000.1,AT5G14490.1,AT5G17260.1,AT5G18270.1,AT5G18300.1,AT5G22290.1,AT5G22380.1,AT5G39610.1,AT5G39690.1,AT5G39820.1,AT5G41090.1,AT5G46590.1,AT5G50820.1,AT5G53950.1,AT5G56620.1,AT5G61430.1,AT5G62380.1,AT5G63790.1,AT5G64060.1,AT5G64530.1,AT5G66300.1

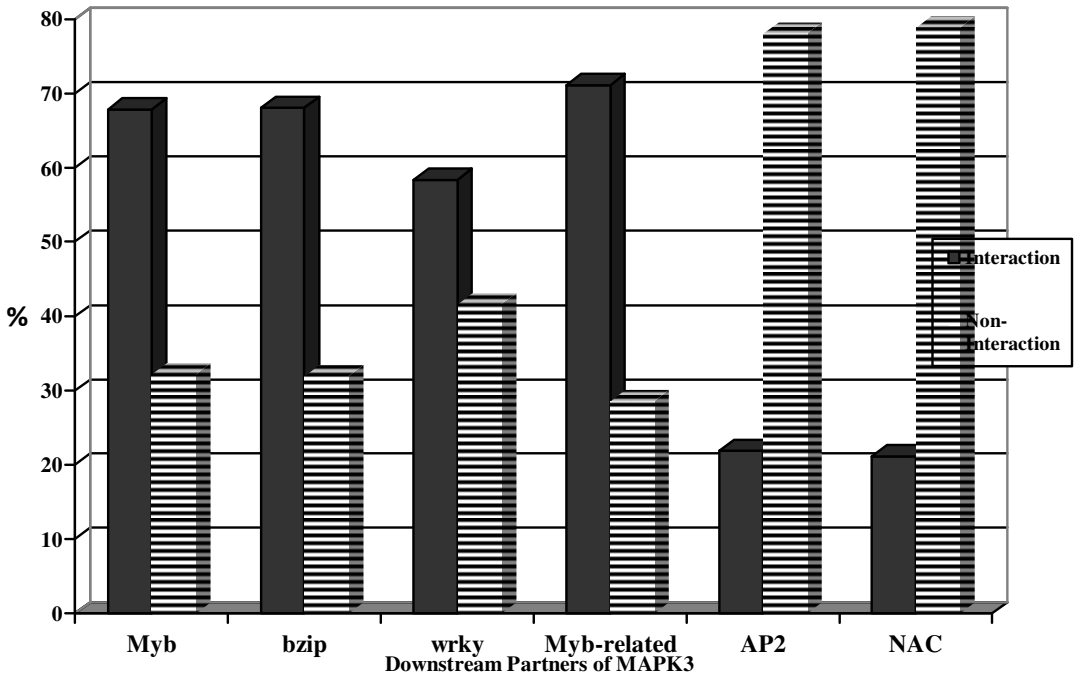

Fig: 1 Percentage of different transcription factor showing interaction and non-interaction with MAPK3.

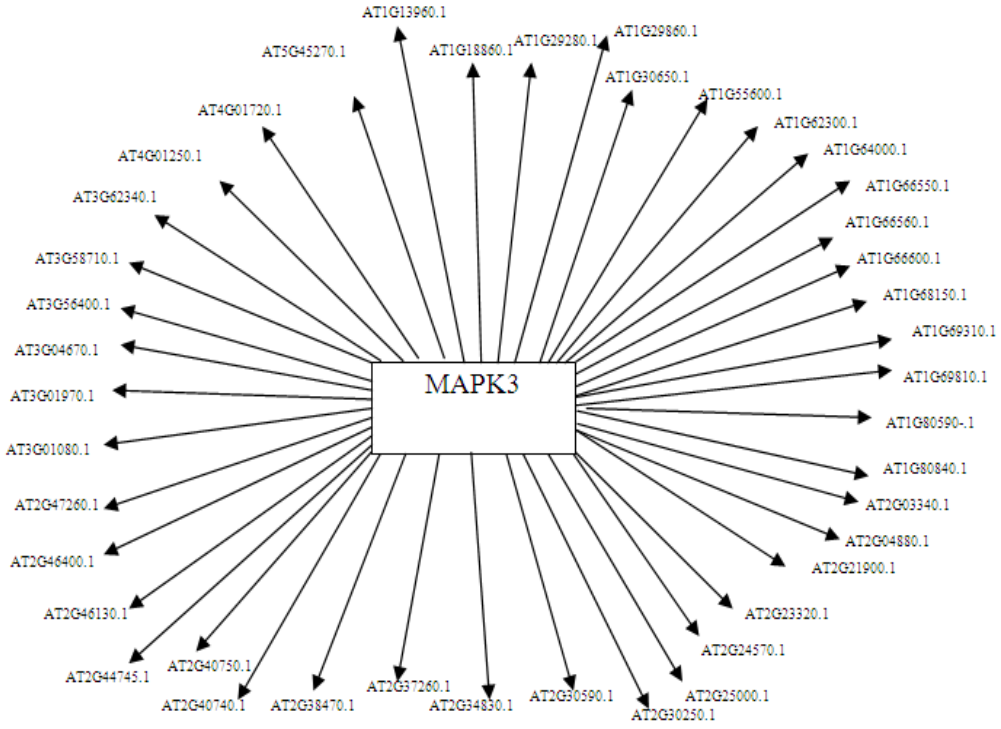

Fig.2: Interaction of WRKY transcription factor with MAPK3.

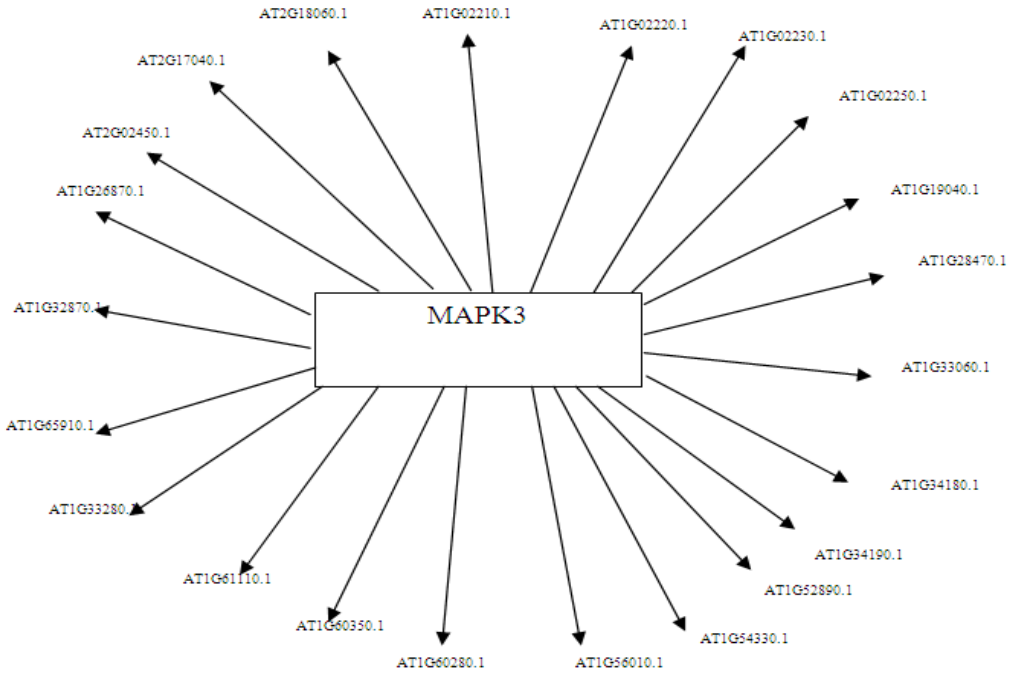

Fig.3: Interaction of NAC transcription factor with MAPK3.

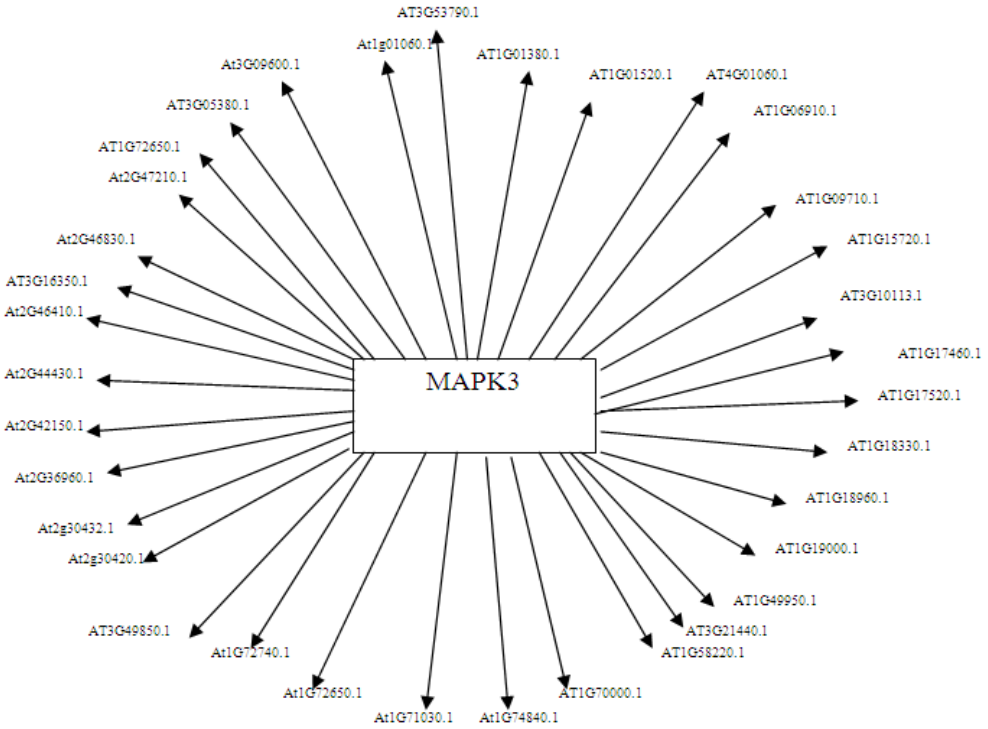

Fig.4: Interaction of Myb-related transcription factor with MAPK3.

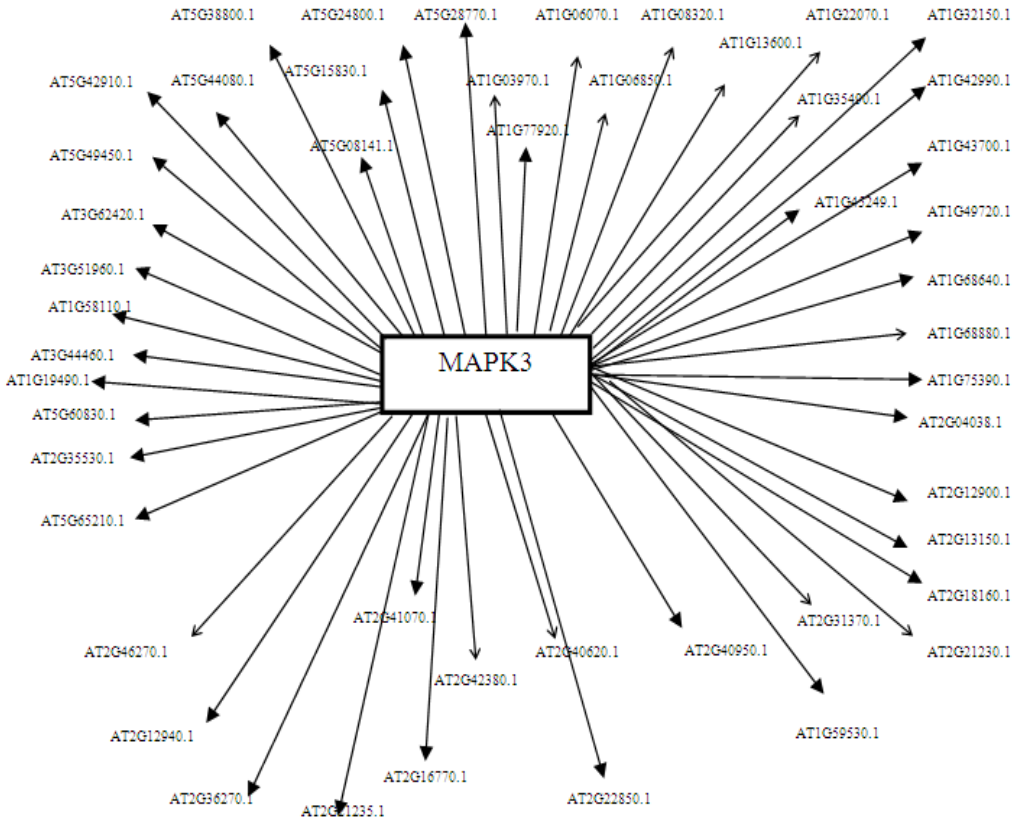

Fig.5: Interaction of bZIP transcription factor with MAPK3.

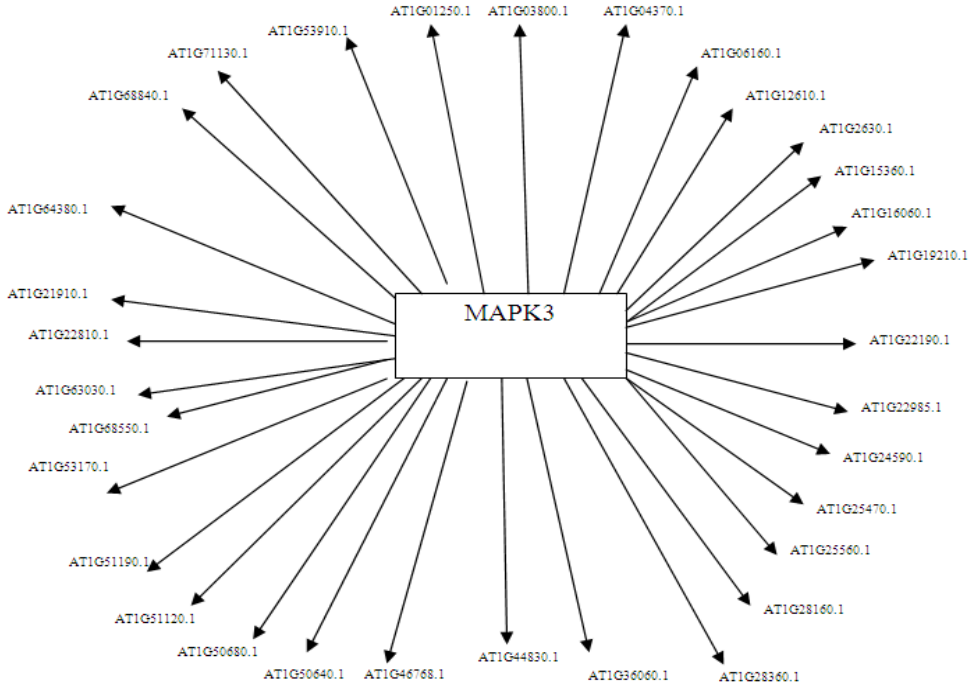

Fig.6: Interaction of AP2 transcription factor with MAPK3.

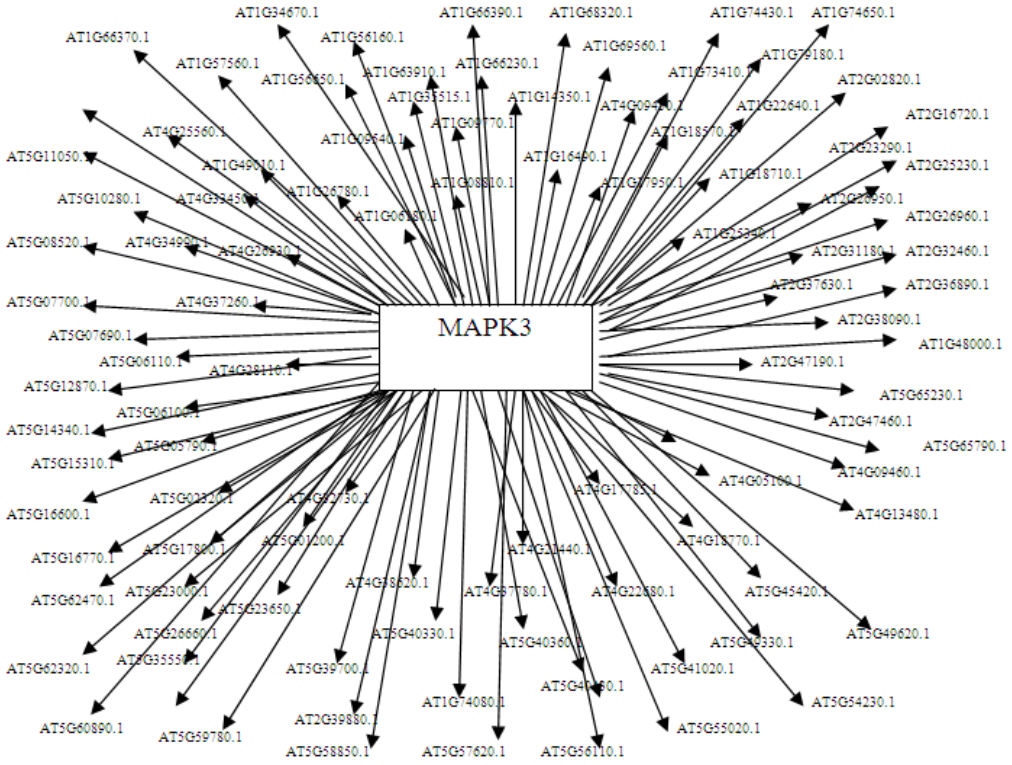

Fig.7: Interaction of Myb transcription factor with MAPK3.