Detection of a MAPK-Like Gene in *Calotropis procera* Plant from the *De Novo* Assembled Genome Contigs of the High Throughput Sequencing Dataset

A.M. Ramadan[1,2], A.M. Shokry[1,2], N.O. Gadalla[1,3], S.M. Hassan[1,4], S. Edris[1,4], M.A. Al-Kordy[1,3], O.A. Abuzinadah[1], J.S.M. Sabir[1], S.Y.M. Alakilli[1], H.S. Al-Zahrani[1], R.M. Hussein[1,3], F.M. El-Domyati[1,4] and A. Bahieldin[1,4,*]

[1]Department of Biological Sciences, Faculty of Science, King Abdulaziz University (KAU), P.O. Box 80141, Jeddah 21589, Saudi Arabia

[2]Agricultural Genetic Engineering Research Institute (AGERI), Agriculture Research Center (ARC), Giza, Egypt

[3]Genetics and Cytology Department, Genetic Engineering and Biotechnology Division, National Research Center, Dokki, Egypt

[4]Department of Genetics, Faculty of Agriculture, Ain Shams University, Cairo, Egypt

*Corresponding Author: bahieldin55@gmail.com

Abstract: Mitogen-activated protein kinase (MAPK) cascade comprises a class of kinases in eukaryotic systems to link perception of external environmental stimuli with changes in cellular organization or gene expression. The wild plant species *Calotropis procera* (*C. procera*) has many potential applications and beneficial uses in medicine, industry and ornamental field and provides an excellent source of genes for drought-resistance and salt-tolerance. However, the biological significance of MAPKs in *C. procera* has not yet been described. In this study, we uncovered and characterized one MAPK-like gene in this medicinal plant from the *de novo* assembled genome contigs of the high throughput sequencing dataset. DNA samples were sent to Beijing Genomics Institute (BGI), Shenzhen, China for deep sequencing and dataset were provided for bioinformatics analysis. A number of GenBank accessions for MAPK protein sequences were utilized in BLAST with the recovered *de novo* assembled contigs and homology modeling was carried out using Swiss-Model, accessible via the EXPASY. Superimposition of *C. procera* MMK2-like partial sequence model on other MAPK proteins was also constructed by using RasMol and Deep-View program. The functional domains were identified from the NCBI conserved domain database (CDD) to provide insights into sequence/structure/function relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM). Then, protein structure alignment was carried out to build models of several MAPK proteins structures and compared them with the human ERK5 crystal structure to identify conserved and diverse structure domains. The results indicated that the longest assembled sequence was 647 nt length and protein sequence obtained from ORF analysis has a length of 218 deduced amino acids. Domain analysis revealed the presence of a protein kinase domain, whose function has been evolutionarily conserved from *Escherichia coli* to *Homo sapiens*. Results at different levels indicated that the PREDICTED mitogen-activated protein kinase homolog MMK2-like of *Vitis vinifera* is the most closely-related protein to *C. procera* MAPK-like protein. Theoretical 3D model for *C. procera* MAPK-like protein indicated the presence of different domains (i.e., for phosphorylation of MAP2K, participation in the interaction of MAPK with its direct upstream activator, etc.). These results support our finding of obtaining a *C. procera* sequence belonging to MAPK protein family. Also, the results proof the accuracy of our theoretical 3D modeling for *C. procera* MAPK-like protein.

[A.M. Ramadan, A.M. Shokry, N.O. Gadalla, S.M. Hassan, S. Edris, M.A. Al-Kordy, O.A. Abuzinadah, J.S.M. Sabir, S.M. Alakilli, H.S. Al-Zahrani, R.H. Hassan, F.M. El-Domyati and A. Bahieldin.**Detection of a MAPK-Like Gene in *Calotropis procera* Plant from the *De Novo* Assembled Genome Contigs of the High Throughput Sequencing Dataset**] *Life Science J* 2012;9(1s):157-166] (ISSN:1097-8135). http://www.lifesciencesite.com. 25

Keywords: Mitogen-activated protein kinase, *C. procera*, NCBI, 3Dmodeling.

1. Introduction

Plants have developed sophisticated defense mechanismsto deal with diverse unfavorable environmental factors (Somssich, 1997; Widmann *et al*., 1999). Protein kinases play a central role in cell signal transduction through phosphorylation to counteract diverse extracellular stimuli such as biotic and abiotic stresses as well as a range of developmental responses including differentiation, proliferation and cell death. One of the most commonly studied mechanisms is the mitogen-activated protein kinase (MAPK) cascade, comprising a class of protein kinases in eukaryotic systems to link perception of external stimuli with changes in cellular organization or gene expression (Widmann *et al*., 1999; Taj *et al*., 2010).

MAPK was first identified by Sturgill and Ray (1986) as microtubule-associated protein kinase. Then, a large number of genes encoding MAPK pathway components have been uncovered in several plant genomes (Mizoguchi *et al*., 1993; Seo *et al.*, 1995; Mizoguchi *et al*., 1996; Ligterink *et al*.,1997, Zhang and Klessig, 1997; Mizoguchi *et al.,* 1998; Hardie, 1999; Yang *et al*., 2001). MAPK cascades comprise a series of sub-families, i.e., MAP4K, MAP3K, MAP2K, MAPK, that are sequentially elicit to activate transcription factors, phospholipases and express specific sets of genes as a response to environmental stimuli (Lin *et al*., 1993; Jonak *et al*., 2002; Cheong *et al*., 2003; Tatebayashi *et al*., 2003; Sasabe *et al*., 2006; Swarbreck *et al*., 2008). MAPK also activates protein kinases that serve as a MAPK substrate named as MAPK-activated protein kinases (MAPKAP-kinase) found in mammalian system (Gerits, *et al*., 2008). Animal MAPK comprises three large families, *i.e.*, ERK, JNK and p38 family. While plant MAPKs also constitute a large family, for example the *Arabidopsis* genome consists of 23 MAPKs, 12 of them are ERK type, the others are plant-specific (Katuo *et al*., 2005; Cvetkovska *et al*., 2005), and no obvious JNK or p38 MAPK homolog has been identified.

The *Calotropis procera* (*C. procera*) of the family Ascelpiadaceae, a drought-resistant, salt-tolerant wild plant species locally known as "Oshar" with the English name of "Giant", is an evergreen poisonous shrub. Through its wind- and animal-dispersed seeds, it quickly becomes established as a weed along degraded roadsides, lagoon edges and in overgrazed native pastures. It has a preference for areas of abandoned cultivation especially sandy soils with low rainfall (Francis, 2003; Orwa *et al*., 2009). *C. procera* is native to west and east Africa, and south Asia, while naturalized in Australia, Center and South America, and the Caribbean island (Rahman & Wilcock, 1991; Brandes, 2005; Orwa *et al*., 2009).

Although *C. procera* plant is toxic, it has many potential applications and beneficial uses. In medicine, it is both poisonous and health-giver in much the same way as digitalis. The aqueous extract of *C. procera* (latex) inhibits cellular infiltration and affords protection against development of neoplastic changes in transgenic mouse model of hepatocellular carcinoma (Choedon *et al*., 2006). The root extractof *C. procera* has protective activity against carbon tetrachloride-induced liver damage (Basu *et al*., 1992). *C. procera* latex is also reported to possess interesting activities such as the ability to combat diarrhea or retard insect larval development (Kumar *etal*., 2001; Morsy *et al*., 2001). Chloroform extract of roots has been reported to possess anti-inflammatory activity (Basu and Chaudhuri, 1991; Kumar & Basu, 1994). Aqueous extract of the flowers was found to exhibit analgesic, antipyretic and anti-inflammatory activity (Mascolo *et al*., 1988). The alcoholic extracts from different parts were found to possess antimicrobial and spermicidal activity (Qureshi *et al*., 1991; Kishore *et al*., 1997). It has also been proven to have anti-fungal properties and can be used effectively in fungal diseases of the skin such as athlete’s foot and ringworm (Kuta, 2008). Laticifer proteins (LP) recovered from the latex of this medicinal plantare targets for DNA topoisomerase I that triggers apoptosis in cancer cell lines (Soares *et al*., 2007). Also, *C. procera* has tannins, latex, rubber and a dye that are used in industrial practices (Orwa *et al*., 2009). *C. procera* is a potential plant for bioenergy and biofuel production in semi arid regions (Garg & Kumar, 2011). In ornamental field, *C. procera* is occasionally grown as an ornamental in dry or coastal areas because it is handsome, of a convenient size, and is easy to propagate and manage. It is recommended as a host plant for butterflies. As *C. procera* is beneficial for human, it provides an excellent source of genes for drought-resistance and salt-tolerance. However, the biological significance of MAPKs in *C. procera* has not yet been described.

In this study, we uncovered and characterized a MAPK-like gene in this medicinal plant from the *de novo* assembled genome contigs of the high throughput sequencing dataset.

2. Materials and Methods

Isolation of nuclear DNA

Extraction of total DNA was performed using the modified procedure of Gawel and Jarret (1991). Three samples of leaf discs of *C. procera* were frozen in liquid nitrogen (approximately 50 mg tissue each) were collected from upper leaves. To remove RNA contamination, RNase A (10 mg/ml, Sigma, USA) was added to the DNA samples and incubated at 37oC for 30 min. Estimation of the DNA concentration in different samples was done by measuring optical density at 260 nm according to the equation: DNA concentration (ug/ml) = OD260 X 50x dilution factor. DNA samples were sent to Beijing Genomics Institute (BGI), Shenzhen, China for deep sequencing and dataset were provided for analysis.

Sequence filtering and bioinformatic analysis

The raw sequence data were obtained using the Illumina python pipeline v. 1.3. For obtained libraries, only high quality reads (quality >20) were retained. Then, *de novo* assembly of the obtained short (pair- and single-end) read dataset was performed using assembler Velvet (Zerbino & Birney, 2008) followed by creation of putative unique transcript (PUTs) with a combination of different k-mer lengths and expected coverage. In total, the yielded EST assemblies from Velvet were merged into *Arabidopsis thaliana* MPK4 accession number NM_116367, where the identity of sequences was over 95% and 40 bp overlapping.

Basic local alignment search tool (BLAST)

The BLAST finds regions of local similarity between sequences. The program compares nucleotide or protein of deduced amino acids to sequence databases, and calculates the statistical significance of matches based on pair-wise alignment method. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families (http://www.ncbi.nlm.nih.gov/BLAST).

AlignX and ClastalW

ClustalW (Higgins & Sharp, 1988) is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms. AlignX® Module: Rapid Multiple Sequence Alignment With Minimal Preparation AlignX® uses a modified Clustal W algorithm to generate multiple sequence alignments of either protein or nucleic acid sequences for similarity comparisons and for annotation. The power of AlignX® is that it maintains annotated features within the alignment for easy visualization and localization of regions of interest.

Determination of phylogenetic relationships

The neighbor joining method was used to build a tree where the evolutionary rates are free to differ in different lineages. To evaluate the reliability of the inferred trees, CLC Genomics Workbench was used to allow the option of doing a bootstrap analysis. A bootstrap value is attached to each branch, and this value is a measure of the confidence in this branch.

Utilized nucleotide sequence accession numbers

The GenBank accession numbers for MAPK Protein sequences data reported utilized in this work are shown in Table 1.

The 3D homology modeling

Homology modeling was carried out using Swiss-Model, protein modeling server, accessible via the EXPASY (http://www.expasy.org/). Superimposition of *C. procera* MMK2-like partial sequence model on other MAPK proteins was constructed by using RasMol (http://www.umass.edu/microbio/rasmol/) and Deep-View program (http://spdbv.vital-it.ch/). The functional domains were identified from the NCBI conserved domain database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), which uses 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM).

Table 1. Accession number, description of the gene and organism, whose gene was isolated.

Accession no.	Description	Organism latin name
NP_001117210	mitogen-activated protein kinase 11	*Arabidopsis thaliana*
NP_001233897	mitogen-activated protein kinase 7	*Solanum lycopersicum*
NP_001234660	mitogen-activated protein kinase 6	*Solanum lycopersicum*
XP_002874986	mitogen-activated protein kinase 4	*Arabidopsis thaliana*
NP_563631	mitogen-activated protein kinase 11	*Arabidopsis thaliana*
XP_002278860	PREDICTED: mitogen-activated protein kinase homolog MMK2	*Vitis vinifera*
XP_002284710	PREDICTED: mitogen-activated protein kinase homolog NTF6	*Vitis vinifera*
XP_002874986	mitogen-activated protein kinase 4	*Arabidopsis lyrata subsp. Lyrata*
XP_002892056	mitogen-activated protein kinase 11	*Arabidopsis lyrata subsp. Lyrata*
XP_003525376	PREDICTED: mitogen-activated protein kinase 4-like	*Glycine max*
XP_003532933	PREDICTED: mitogen-activated protein kinase 2-like	*Glycine max*
XP_003534546	PREDICTED: mitogen-activated protein kinase homolog MMK2-like	*Glycine max*
XP_003548645	PREDICTED: mitogen-activated protein kinase homolog MMK2-like	*Glycine max*
XP_003573472	PREDICTED: mitogen-activated protein kinase 2-like isoform 1	*Brachypodium distachyon*
XP_003573473	PREDICTED: mitogen-activated protein kinase 2-like isoform 2	*Brachypodium distachyon*
XP_003574247	PREDICTED: mitogen-activated protein kinase 6-like	*Brachypodium distachyon*
XP_003611065	mitogen-activated protein kinase	*Medicago truncatula*
XP_003622463	mitogen-activated protein kinase	*Medicago truncatula*
XP_003624049	mitogen-activated protein kinase	*Medicago truncatula*
XP_003633959	PREDICTED: mitogen-activated protein kinase homolog MMK2-like	*Vitis vinifera*
XP_002276158	PREDICTED: mitogen-activated protein kinase 4	*Vitis vinifera*
XP_002279719	PREDICTED: mitogen-activated protein kinase 16-like	*Vitis vinifera*
XP_002283794	PREDICTED: mitogen-activated protein kinase 20-like	*Vitis vinifera*
XP_002284377	PREDICTED: mitogen-activated protein kinase 9-like	*Vitis vinifera*
XP_002284807	PREDICTED: mitogen-activated protein kinase 3	*Vitis vinifera*
XP_002285641	PREDICTED: mitogen-activated protein kinase 19	*Vitis vinifera*
XP_003634202	PREDICTED: mitogen-activated protein kinase 9-like	*Vitis vinifera*

Structure alignment

Protein 3D structure comparison is a challenging task that depends on the alignment algorithm, the similarity measure, and the fractions of the protein structures considered for the pairwise structure alignment (Kolodny *et al.*, 2005). DaliLite was proven to be very accurate structural alignment method on representative datasets (Hou *et al.*, 2002; Day *et al.*, 2003; Barthel *et al.*, 2007). Models of several MAPK proteins structure were built and compared with the human ERK5 crystal structure, genbank accession number AAA81381. The protein model and ERK5 3D-structure were applied to pairwise comparison of protein structures using DaliLite program server at EBI https://www.ebi.ac.uk/Tools/dalilite/ (Holm *et al.*, 2008) and their alignments were used to identify conserved and diverse structure domains. Root mean square deviation (RMSD) which measures the average distance between the backbone of superimposed proteins was measured using DaliLite according to the following formula:

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2}$$

NGS sequence

Whole-RNAseq, paired-end short-sequence reads for *C. procera* were generated using the Illumina Genome Analyser IIx (GAIIx) according to manufacturer's instructions (Illumina, San Diego, CA). Assemblies were mapped to *Arabidopsis thaliana* MPK4 accession number NM_116367 using SAOP (Li *et al.*, 2009). The number of reads aligned was 1073 with average coverage of 64.34. The length of consensus sequence equal 855 nt. The longest sequence with 647 nt length with high quality was used for further investigation. ORF analysis showed a partial length ORF within the first 402 nt as shown in figure1.

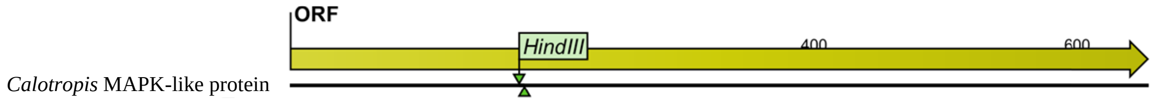

Figure1. ORF analysis for the obtained MAPK sequence. This Sequence was characterized by the presence of *Hin*dIII site at ~200 nt length.

3. Results and Discussion

The protein sequence obtained from ORF analysis with a length of 218 deduced amino acids was analyzed against the pfam database (http://pfam.sanger.ac.uk/) to allocate protein domains. Domain analysis revealed the presence of a protein kinase domain (accession number PF0069) as shown in Figure 2, whose function has been evolutionarily conserved from *Escherichia coli* to *Homo sapiens*. Protein kinases play a role in a multitude of cellular processes, including division, proliferation, apoptosis, stress tolerance and differentiation (Manning *et al.*, 2002) Phosphorylation usually results in a functional change of the target protein by changing enzyme activity, cellular location, or association with other proteins (Stout*et al.*, 2004).

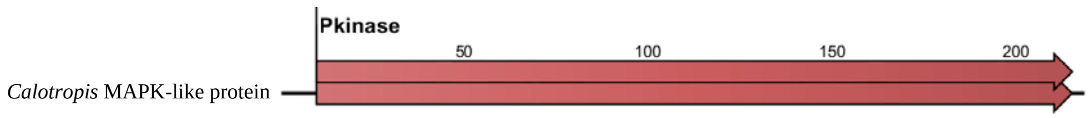

Figure2. Protein kinase domain of the deduced amino acid sequence of the obtained MAPK-like protein as analyzed by pfam database.

BLAST analysis

BLAST (either protein-protein BLAST or BLASTp) was performed to identify sequence similarity with homologous proteins from other organisms to the obtained *C. procera* MAPK-like protein (http://blast.ncbi.nlm.nih.gov/). The interpretation of the score and sequence similarity from BLAST searching eventually led to the identification of putative or homologous protein sequences. Results for the most closely-related protein to *C. procera* MAPK-like protein indicated that the PREDICTED mitogen-activated protein kinase homolog MMK2-like of *Vitis vinifera* has the lowest e-value (1e-149). These results indicate that the speculated *C. procera* MAPK-like protein can be a member of MMK2 protein family.

Multi-sequence alignment (MSA) and phylogenetic analysis

The best BLAST search hits were used to perform multi-sequence alignment. This resulted in 20 sequences originating from 7 different species. An alignment of the 21 sequences was obtained by gap open penalty of 10 and gap extension penalty of one. Sequences with more than 85% identity with the obtained *C. procera* MAPK-like protein were used (Table 3 & Figure 3). The results also show that the closest sequence to the obtained *C. procera* MAPK-like protein is *Vitis vinifera* PREDICTED: mitogen-activated protein kinase homolog MMK2-like with accession number XP_003633959. These results support the obtained BLAST results. MSA results were used to perform phylogenetic tree for the 21 proteins and results (Figure 4) were similar to those of previous analyses.

Table 2: Accession number for each protein, description, organism name and the calculated e-value of homologous proteins to *C. procera*MAPK-like protein identified using BLASTP search.

Accession	Description	Latin name	e-value
XP_003633959	PREDICTED: mitogen-activated protein kinase homolog MMK2-like	*Vitis vinifera*	1e-149
XP_003622463	Mitogen-activated protein kinase	*Medicago truncatula*	5e-146
XP_002874986	mitogen-activated protein kinase 4	*Arabidopsis lyrata subsp. Lyrata*	7e-145
NP_192046	mitogen-activated protein kinase 4	*Arabidopsis thaliana*	5e-145
XP_003624049	Mitogen-activated protein kinase	*Medicago truncatula*	2e-145
XP_002278860	PREDICTED: mitogen-activated protein kinase homolog MMK2	*Vitis vinifera*	2e-144
XP_003534546	PREDICTED: mitogen-activated protein kinase homolog MMK2-like	*Glycine max*	5e-143
XP_003548645	PREDICTED: mitogen-activated protein kinase homolog MMK2-like [Glycine max].	*Glycine max*	2e-142
NP_001233897	mitogen-activated protein kinase 7 [Solanum lycopersicum].	*Solanum lycopersicum*	3e-141
XP_003611065	Mitogen-activated protein kinase [Medicago truncatula].	*Medicago truncatula*	6e-139
XP_003574247	PREDICTED: mitogen-activated protein kinase 6-like [Brachypodium distachyon].	*Brachypodium distachyon*	5e-138
NP_001234660	mitogen-activated protein kinase 6 [Solanum lycopersicum].	*Solanum lycopersicum*	3e-138
XP_003532933	PREDICTED: mitogen-activated protein kinase 2-like [Glycine max].	*Glycine max*	2e-137
XP_003573472	PREDICTED: mitogen-activated protein kinase 2-like isoform 1 [Brachypodium distachyon].	*Brachypodium distachyon*	2e-137
XP_003573473	PREDICTED: mitogen-activated protein kinase 2-like isoform 2 [Brachypodium distachyon].	*Brachypodium distachyon*	1e-137
XP_002892056	mitogen-activated protein kinase 11 [Arabidopsis lyrata subsp. lyrata].	*Arabidopsis lyrata subsp. Lyrata*	2e-136
NP_563631	mitogen-activated protein kinase 11 [Arabidopsis thaliana].	*Arabidopsis thaliana*	1e-136
XP_003525376	PREDICTED: mitogen-activated protein kinase 4-like [Glycine max].	*Glycine max*	1e-136
NP_001117210	mitogen-activated protein kinase 11 [Arabidopsis thaliana].	*Arabidopsis thaliana*	7e-135
XP_002284710	PREDICTED: mitogen-activated protein kinase homolog NTF6 [Vitis vinifera].	*Vitis vinifera*	1e-135

Table 3: Pairwise alignment between each hit MAPK sequence as compared to the obtained sequence of *C. procera* MAPK-like protein.

Accession	Gaps	Differences	Distance	Identity%	Identities
XP_003633959	0	14	0.07	93.58	204
XP_003548645	0	23	0.11	89.45	195
XP_003534546	0	21	0.10	90.37	197
XP_003624049	0	21	0.10	90.37	197
XP_003622463	0	20	0.10	90.83	198
XP_003532933	0	32	0.16	85.32	186
XP_003525376	0	32	0.16	85.32	186
XP_002278860	0	19	0.09	91.28	199
NP_192046	0	22	0.11	89.91	196
XP_002874986	0	22	0.11	89.91	196
NP_001117210	0	30	0.15	86.24	188
XP_002892056	0	29	0.14	86.7	189
NP_001233897	0	27	0.13	87.61	191
XP_003573472	0	32	0.16	85.32	186
XP_003573473	0	32	0.16	85.32	186
NP_001234660	0	28	0.14	87.16	190
XP_003611065	0	30	0.15	86.24	188
XP_003574247	0	31	0.15	85.78	187
NP_563631	0	30	0.15	86.24	188
XP_002284710	0	30	0.15	86.24	188

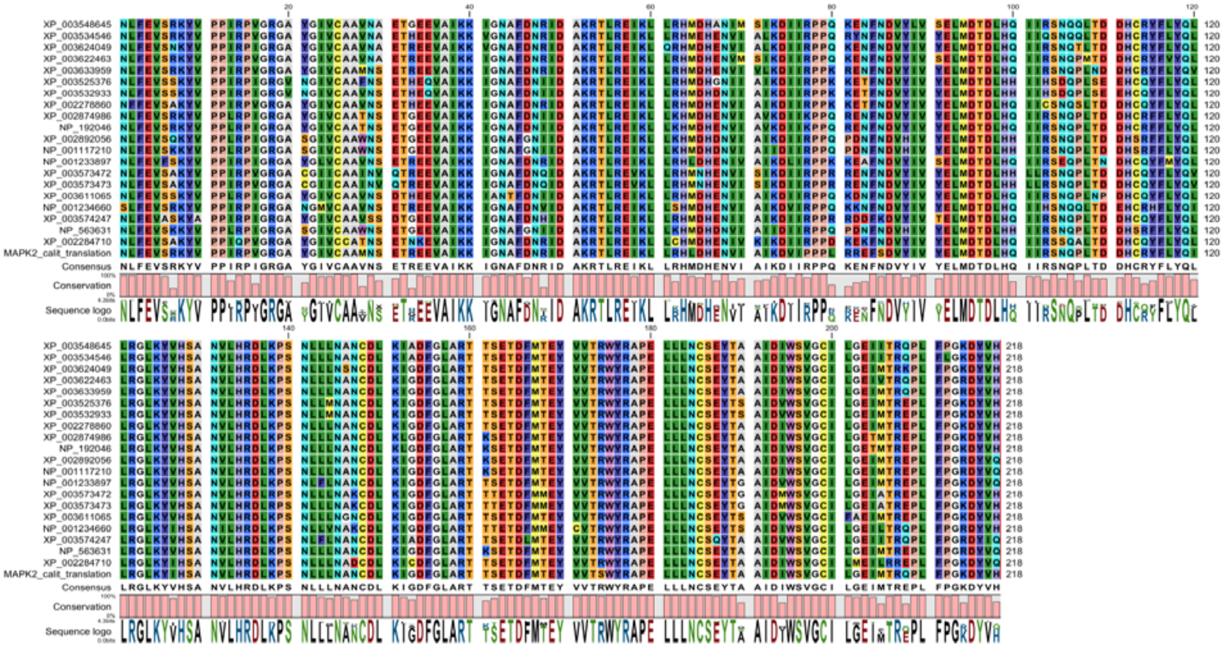

Figure 3. Multi-sequence alignment of the 20 MAPK sequences with the obtained *C. procera* MAPK-like protein sequence.

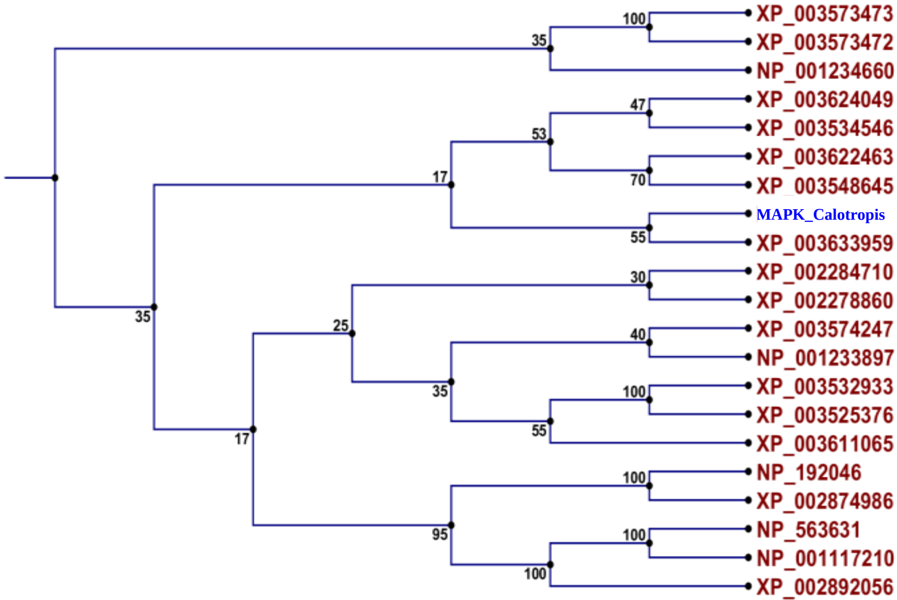

Figure 4. Phylogenetic analysis of 20 MAPK proteins and *C. procera* MAPK-like protein.

3D structure modeling

MAPK signaling efficiency and specificity can be achieved through specialized docking motifs present in components of the cascade (Figure 5). Based on structural alignment, a theoretical 3D model for *C. procera* MAPK-like protein was created, corresponding to residues 1-220 of the primary structure (Figure 6). The predicted model was created using the Swiss-Model, protein modeling server. The overall dimensions of the model are 61.707Å X 55.313Å X 43.264Å.TXY, including the phosphorylation site of TEY (residues Thr168-Glu169-Tyr170) for activation by MAP2K, the D-domain (also referred to as the D site, δdomain, or DEJL domain) consisting of a core of basic residues Lys138-Pro139 upstream a hydrophobic patch (Lys/Arg-Lys/Arg-Xaa2-6-Φ-X-Φ, where Φ is a hydrophobic residue, such as Leu, Iso or Val) as described by Dalby (1998) (residues 143-152). As described by Ferrell (1999), this domain also participates in the interaction of MAPK with its direct upstream activator (residues 156-167). Also, there are three hyper-variant regions (A = RKYV, B = LRRE & C = GLARTTSETDFM) scattered in the molecule.

a.

```
NLFEVSRKYVPPIRPVGRGAYGIVCAAMNSETREEVAIKKIGNAFDNRIDAKRTLREIKLLRHLDHENVIAIKDVIPPPLRREFSDV
YIVYELMDTDLHQIIRSNQPLTDDHCRYFLYQILRGLKYIHSANVLHRDLKPSNLLLNANCDLKIGDFGLARTTSETDFMTEYVVTS
WYRAPELLLNCSEYTAAIDIWSVGCILGEIMTRQPLFPGKDYVH
```

b.

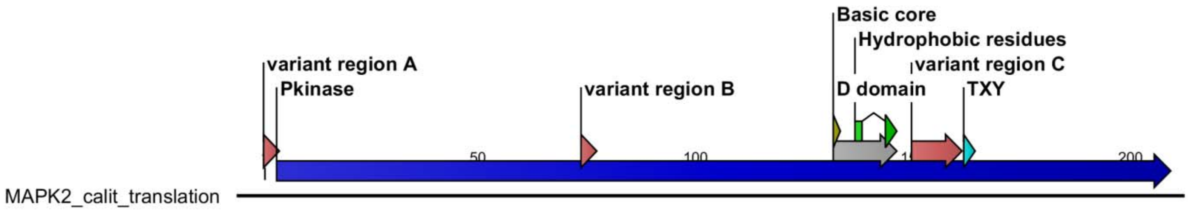

Figure 5. Identified sequence (a) and motifs (b) in the obtained *Calotropis* MAP2K-like protein sequence. Protein kinase domain (dark blue), TXY Phosphorylation site: TEY (light blue), D domain: Underlined (gray), Hydrophobic residues within D-Domain (Green), Basic residues within Basic core (Tan), Hyper variant region A: RKYV (red), Hyper variant region B: LRRE (red), Hyper variant region C: GLARTTSETDFM (red).

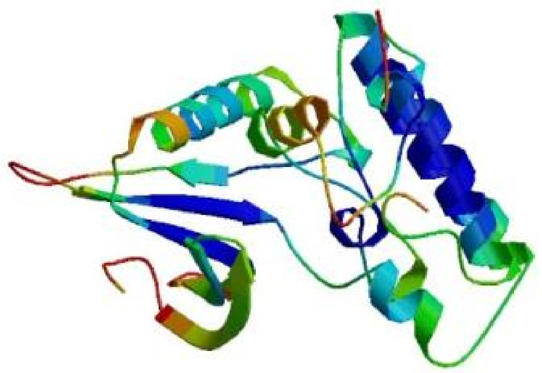

Figure 6. Theoretical 3D model for *C. procera* MAPK-like protein.

Structure alignment

We applied DaliLite, on nine protein 3D structures. Eight of the nine 3D models were created based on structural alignment using Swiss-Model while one structure related to human ERK5 crystal structure was downloaded from protein database. ERK5 is the closest homologous protein sequence with available 3D structure to the obtained *C. procera* MAPK, however, ERK5 is a human MAPK also known as MAPK7. MAPK7 is proposed to play a role in the pathology of cancer (Lochhead *et al*., 2012). To proof the accuracy of our theoretical 3D model of *C. procera* MAPK-like protein, we used DaliLite to computes optimal and suboptimal structural alignments between ERK5 and the theoretical 3D model of *C. procera* MAPK-like protein. The resulting superimposed figure is shown in Figure 7 with Z-score of 28.2, number of equivalent residues of 208 and RMSD of 0.5. The figure shows that 3D model of *C. procera* MAPK amino acids (yellow) has almost the same coordinates of ERK5 (gray) except in three positions A (residues 7-10), B (residues 80-83) and C (residues 156-167), which are hyper-variant. Region C is located upstream the TEY dual phosphorylation motif within the activation loop, which participates in the interaction of MAPK with its direct upstream activator MAP2K. This region greatly varies from one MAPK to another as it is responsible of the specificity of MAPK to dock with its activating protein (Gray et al., 2001). These results support or finding of obtaining a *C. procera* sequence belonging to MAPK protein family. Also, the results proof the accuracy of our theoretical 3D modeling for *C. procera* MAPK-like protein.

Further analysis on the predicted 3D structures of seven published *V. vinifera* MAPK protein sequences in the Genbank as compared to the obtained *C. procera* sequence belonging to MAPK protein family was done (Table 4 & Figure 8). Results of RMDS and Z-score showed that the lowest RMSD (0.4) was computed when comparing the 3D MMK2 model of *V. vinifera* versus that of *C. procera* MAPK-like protein, which is low enough to support that *C. procera* MAPK-like protein is MMK2-like protein.

Table 4: Results of RMDS and Z-score for all compared 3D structures with *C. procera* MAPK-like protein.

Organism	MAPK type	Z-score	Number of equivalent residues	RMSD
V. vinifera	MMK2 or MAPK1	30.7	216	0.4
Homo sapiens	MAPK7	28.2	208	0.5
V. vinifera	MAPK 16	27.7	207	0.5
V. vinifera	MAPK 20	27.6	206	0.5
V. vinifera	MAPK 4	28.9	214	0.9
V. vinifera	MAPK 19	28.5	215	1.1
V. vinifera	MAPK 3	28.4	213	1.2
V. vinifera	MAPK 9	27.6	210	1.2

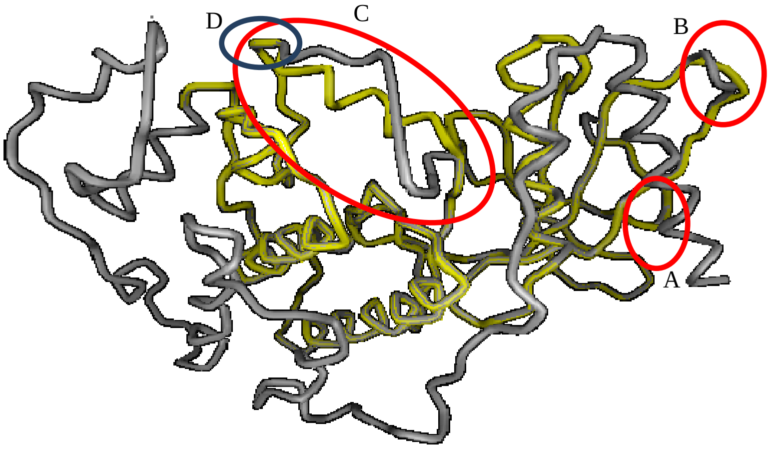

Figure 7. Superimposed figure between ERK5 and *C. procera* MAPK. A & B = first and second hyper-variant regions, C = third variant region and activated loop domain, D = TEY motif.

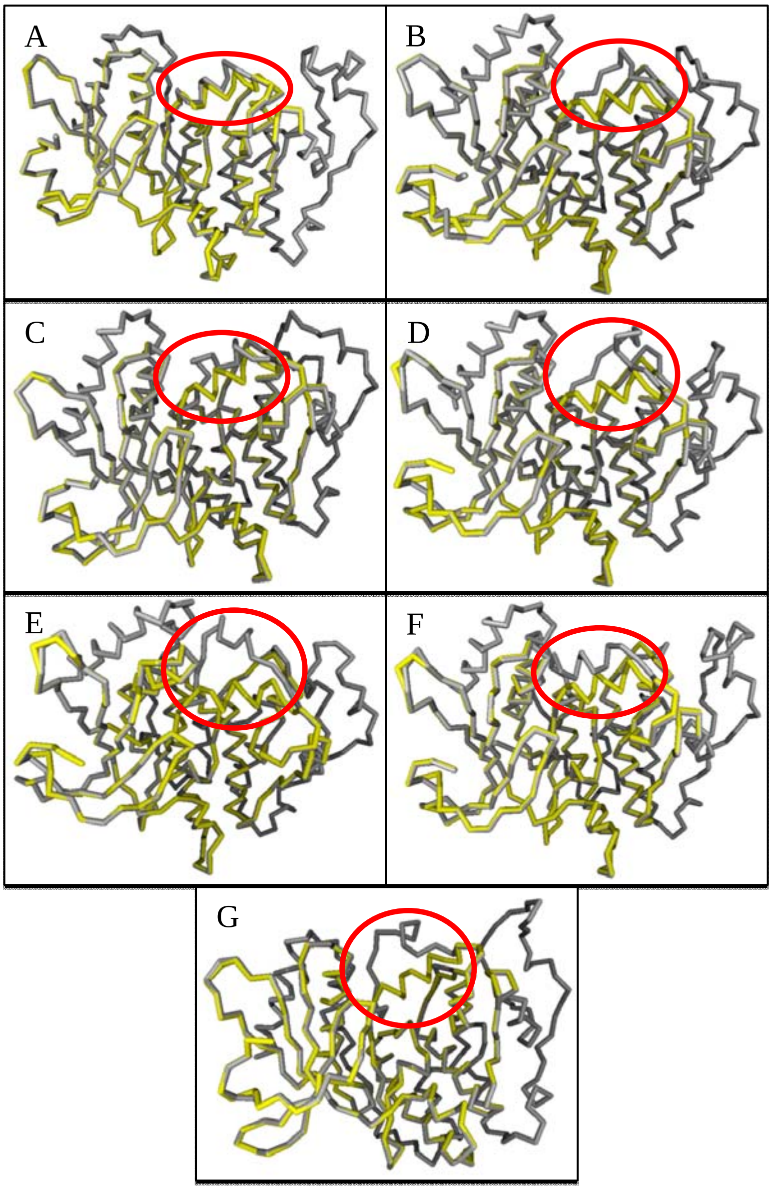

Figure 8. Superimposed figure between *C. procera* MAPK protein and each *V. vinifera* MAPK with annotation for activated loop domain. A: MMK2 or MAPK1, B: MAPK 3, C: MAPK 4, D: MAPK 9, E: MAPK 16, F: MAPK 19 and G: MAPK 20.

Acknowledgments

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under Grant no. 2-3-1432/HiCi. The authors, therefore, acknowledge with thanks DSR technical and financial support.

4. References

Barthel D, Hirst JD, Blazewicz J, Burke EK, Krasnogor N (2007). A decision support system for protein (structure) comparison, knowledge, similarity and information. BMC Bioinformatics, 8:416.

Basu A, Chaudhuri AK (1991). Preliminary studies on the anti-inflammatory and analgesic activities of *Calotropis procera* root extract. J. Ethnopharmacol, 31:319.

Basu A, Sen T, Ray RN, Nag Chaudhuri AK (1992). Hepatoprotective effects of *C. proceraprocera* root extract on experimental liver damage in animals. Fitoterapia, 63:507-514.

Brandes D (2005). *C. proceraprocera* on Fuerteventura. Working Group for Vegetation Ecology, Institute of Plant Biology, TechnicalUniversity Braunschweig, Germany. Available at:http://www.biblio.tu-bs.de/geobot/fuerte.html

Cheong YH, Moon BC, Kim JK, Kim CY, Kim MC, Kim IH, Park CY, Kim JC, Park BO, Koo SC, Yoon HW, Chung WS, Lim CO, Lee SY, Cho MJ (2003). BWMK1, a rice mitogen-activated protein kinase, locates in the nucleus and mediates pathogenesis related gene expression by activation of a transcription factor. Plant Physiol, 132:1961-1972.

Choedon T, Mathan G, Arya S, Kumar VL, Kumar V (2006). Anticancer and cytotoxic properties of the latex of *C. procera procera* in a transgenic mouse model of hepatocellular carcinoma. World J Gastroenterol, 12:2517-2522.

Cvetkovska M, Ramptitsch C, Bykova N, Xing T (2005). Genomic analysis of MAPK Cascade in Arabidopsis defense responses. Plant Mol Biol Rep, 23:331-343.

Dalby KN, Morrice N, Caudwell FB, Avruch J, Cohen P (1998). Identification of regulatory phosphorylation sites in mitogen-activated protein kinase (MAPK)-activated protein kinase-1a/p90rsk that are inducible by MAPK. J Biol Chem, 273:1496-1505.

Day R, Beck DAC, Armen RS, Daggett V (2003). A consensus view of fold space: Combining SCOP, CATH, and the Dali Domain Dictionary. Protein Sci, 12:2150-2160.

Ferrell Jr JE (1999). Building a cellular switch: more lessons from a good egg. Bioessays 21:866-870.

Francis JK (2003). *Calotropis procera*. U.S. Department of Agriculture, Forest Service, International Institute of Tropical Forestry, Puerto Rico.

Garg J, Kumar A (2011). Hydrocarbon from plants as renewable source of energy. Bioherald, 1: 31-35.

Gawel NJ, Jarret RL (1991). A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. Plant Mol Biol Rep, 9: 262-266.

Gerits N, Shiryaev A, Kostenko S, Klenow H, Shiryaeva O, Johannessen M, Moens U (2008). The transcriptional regulation and cell-specific expression of the MAPK-activated protein kinase mk5. Cell Mol Biol Letter, 14:548-574.

Gray P, Fred R, Tara BG, Bing-E Xu, Mahesh K, Kevin B, Melanie HC (2001). Mitogen-activated protein (MAP) kinase pathways: Regulation and physiological functions. Endocrine Rev, 22(2): 153-183.

Hardie, DG (1999). Plant protein serine/threonine kinases: classification and functional. Ann rev Plant Physiol Plant Mol Biol, 50:97-131.

Higgins DG, Sharp PM (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene, 73(1):237-244.

Hou J, Sims GE, Zhang C, Kim SH (2002). A global representation of the protein fold space. Proc Natl Acad Sci,100:2386-2390.

Jonak C, Okresz L, Bogrw L, Hirt H (2002). Complexity, cross talk and integration of plant MAP kinase signaling. Curr Opin Plant Biol, 5:415-424.

Katou S, Karita E, Yamakawa H, Seo S, Mitsuhara I, Kuchitsu K, Ohashi Y (2005). Catalytic activation of the plant MAPK phosphatase NtMKP1 by its physiological substrate salicylcic acid-induced protein kinase but not by Calmodulins. J Biol Chem, 280:59569-59581.

Kishore N, Chopra AK, Khan O (1997). Antimicrobial properties of *C. procera procera* Ait. in different seasons. A study *in vitro*. Biol Mem, 23:53.

Kolodny R, Koehl P, Levitt M (2005). Comprehensive evaluation of protein structure alignment methods. J Mol Biol, 346:1173-1188.

Kumar S, Dewan S, Sangraula H, Kumar VL (2001). Anti-diarrhoeal activity of the latex of *C. procera procera*. J Ethnopharmaco, 76:115-118.

Kumar VL, Basu N (1994). Anti-inflammatory activity of the latex of *C. procera procera*. J Ethnopharmaco, 44:123-125.

Kuta FA (2008). Anti-fugal Effect of *C. procera procera* stem bark on *Epidermophyton flocosum* and *Trichphyton gypseum*. African J Biotech,7(13):2116-2118.

Holm L, Kääriäinen S, Rosenström P, Schenkel A (2008). Searching protein structure databases with DaliLite v.3. Bioinformatics, 24(23):2780-2781.

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics, 25(15):1966-1967.

Ligterink W, Kroj T, Zurnieden U, Hirt H, Scheel D (1997). Receptor mediated activation of MAP Kinase in pathogen defense of plants. Science, 276:2054-2057.

Lin LL, Wartmann M, Lin AY, Knopf JL, Seth A, Davis RJ (1993). cPLA2 is phosphorylated and activated by MAP kinase. Cell, 72:269-278.

Lochhead PA, Gilley R, Cook SJ (2012). ERK5 and its role in tumour development. Biochem Soc Trans, 40(1):251-256.

Manning G, Plowman GD, Hunter T, Sudarsanam S (2002). Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci, 27 (10):514-520.

Mascolo N, Sharma R, Jain SC, Capasso F (1988). Ethnopharmacology of *Calotropis procera* flowers. J Ethnopharmacol, 22(2):211-221.

Mizoguchi T, Hayashida N, Yamaguchi-shinozaki K, Kamada H, Shinozaki K (1993). AtMPKs: a gene family of MAPK in *Arabidopsis thaliana*. FEBS Let, 336:440-444.

Mizoguchi T, Ichimura K, Irei K (1998). Identification of a possible MAPK cascade in *Arabidopsis thaliana* based on pairwise yeast two hybrid analysis and functional test of yeast mutants. FEBS Let, 437:56-60.

Mizoguchi T, Irie K, Hirayama T, Hayashida N, Yamaguchi-Shinozaki K, Matsumoto K, Shinozaki K (1996). A gene encoding a mitogen-activated protein kinase kinase kinase is induced simultaneously with genes for a mitogen-activated protein kinase and an S6 ribosomal protein kinase by touch, cold and water stress in *Arabidopsis thaliana*. Proc Natl Acad Sci USA, 93:765-759.

Morsy, TA, Rahem MA, Allam KA (2001). Control of *Musa domestica* third instar larvae by the latex of *C. procera procera* (Family: Asclepiadaceae). J Egy Soc Parasitology, 31:107-110.

Orwa C, Mutua A, Kindt R, Jamnadass R, Simons A (2009). Agroforestree Database:a tree reference and selection guide version 4.0 (http://www.worldagroforestry.org/af/treedb/).

Qureshi MA, Qureshi NM, Arshad R, Begum R (1991). A Study on the antisperm activity in extracts from different parts of *C. procera - procera*.Pak. J Zool, 23:161.

Rahman MA, Wilcock CC (1991). A taxonomic revision of *C. procera*(Asclepiadaceae). Nordic J Botany, 11(3):301-308.

Sasabe M, Soyano T, Takahashi Y, Sonobe S, Igarashi H, Itoh TJ, Hidaka M, Machida Y (2006). Phosphorylation of NtMAP65-1 by a MAP kinase downregulates its activity of microtubule bundling and stimulates progression of cytokinesis of tobacco cells. Genes Dev, 20:1004-1014.

Seo S, Okamoto N, Seto H, Ishizuka K, Sano H, Ohashi Y (1995). Tobacco MAP Kinase- a possible mediator in wound signal transduction pathways. Science, 270:1988-1992.

Soares de Oliveira J, Pereira Bezerra D, Teixeira de Freitas CD, Delano Barreto Marinho Filho J, Odorico de Moraes M, Pessoa C, Costa-Lotufo LV, Ramos MV (2007). *In vitro* cytotoxicity against different human cancer cell lines of laticifer proteins of *Calotropis procera* (Ait.) R. Br. Toxicol In Vitro, 21(8):1563-1573.

Somssich IE (1997). MAP kinases and plant defense. Trends Plant Sci, 2:406-408.

Stout TJ, Foster PG, Matthews DJ (2004) High-throughput structural biology in drug discovery: protein kinases. Curr Pharm Des, 10 (10):1069-1082.

Sturgill TW, Ray LB (1986). Muscle proteins related to microtubule associated protein-2 are substrates for an insulinstimulatable kinase. Biochem Biophys Res Commun, 134:565-571.

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008). The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res, 36:1009-1014.

Taj G, Agarwal P, Grant M, Kumar A (2010).MAPK machinery in plants: recognition and response to different stresses through multiple signal transduction pathways. Plant Signal Behav, 5(11):1370-1378.

Tatebayashi K, Takekawa M, Saito H (2003). A docking site determining specificity of Pbs2 MAPKK for Ssk2/Ssk22 MAPKKKs in the yeast HOG pathway. EMBO J, 22:3624-3634.

Widmann C, Gibson S, Jarpe MB, Johnson GL (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. Physiol Rev, 79:143-180.

Yang KY, Liu Y, Zhang S (2001). Activation of a Mitogenactivated protein kinase pathway is involved in disease resistance in tobacco. Proc Natl Acad Sci USA, 98:741-746.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res, 18(5):821-829.

Zhang SQ, Klessig DF (1997). Salicyclic acid activates 48 KDa MAP kinase in tobacco. Plant Cell, 9:809-824.