



## RESEARCH PAPER

## OPEN ACCESS

## Insilico identification and characterization of abiotic stress responsive genes in *Cucumis melo* L. using bioinformatics tools

Asyia, Bibi Sadia\*, Wajia Noor, Shazia Irfan, Yasmeen Rehmat

Department of Botany, SBK Women's University, Quetta, Pakistan

**Key words:** Drought stress, Cold stress, Salt stress, Stress responsive genes, *Arabidopsis thaliana* *Cucumis melo*

<http://dx.doi.org/10.12692/ijb/19.6.173-182>

Article published on December 29, 2021

### Abstract

Plants are exposed to various abiotic stress responsive genes. The present study was designed to investigate cold, drought and salt stress responsive genes in *Cucumis melo* L. for regulation of gene expression as 101273 ESTs of *Cucumis melo* L. and 3702 ESTs of *Arabidopsis thaliana* were explored for insilico of twenty one abiotic stress responsive genes. RAB18 (96%), CSD1 (94%), HVA22A (86%) GSTF7 (73%), CDSP32 (65%), ATMRK1 (65%) MPK3 (63%), ELIP1 (59%), COR47 (56%), ERD14 (42%), DDF1 (42%) DI19 (41%), LOS4 (40%), ABCF3 (37%), P5CS1 (36%), CDPK1 (27%), DREB2A (23%), STZ (18%) and HOS10 (57%) gene was expressed in three stresses cold drought and salt. Bioinformatics tools NCBI, ORF, BLAST, CDD, and MUSCLE were used to find out homology, conserved domains then conserved analysis of nucleotides and phylogenetic analysis of *Cucumis melo* L. The current study revealed that similarly conserved domain families in *Arabidopsis thaliana* and *Cucumis melo* L. were Ras super-family (RAB18), Cu-Zn-superoxide-Disntase super-family (CSD1), TB2-DP1-HVA22 super-family (HVA22A), GST-N-Phi thioredoxin-like super family (GSTF7), TRX-COSP32 (CDSP32), PKc-like super-family (ATMRK1, (MPK3), ), Chloroa-b-bind super-family (ELIP1), Dehydrin super-family (COR47,(ERD14), AP2 super-family (DDF1), Zf-Di19 super family (DI19), DEXDc super-family (LOS4), SunT super-family (HOS10) and (ABCF3), PROB super-family (P5CS1)and Zf-C2H2-6 super-family (STZ). Phylogenetic results revealed interesting fact although genes DI19, DREB2A, CDPK1 were more closely related to in *Arabidopsis thaliana* and *Cucumis mel* L. beside this ATMRK1, ERD14, LOS4, MPK3,CDSP32, P5CS1 not showing close relations.

\*Corresponding Author: Bibi Sadia ✉ [zarmal.gul@gmail.com](mailto:zarmal.gul@gmail.com)

## Introduction

Stresses induce serious physiological and biological responses in different types of plants. These responses include stomatal closure, repression of cell growth and photosynthesis and activation of respiration (Reymond *et al.*, 2000). The number of plants when exposed to environment stress they go through physiological and biochemical adaptation (Bray. 1993). Water deficit caused by drought and high salinity has been a main selective strength in plant evolution and main factors limiting crop productivity. On the other hand, low temperature may be the most significant environmental limitation for plant distribution on land (Lytton. 2004).

In recent year remarkable change have been seen in various dominos of stress physiology. It is acknowledged that signaling pathways leading to plants stress responses are interconnected at many levels however molecular and metabolic responses observed a combination of stresses drought and heat are unique and cannot be extrapolated from plant response to the individual stress (Bohnert. 1995). Number of drought inducible genes with various functions has been identified by molecular and genomic analysis in Arabidopsis, rice and other plant; including number of transcription factors to facilitate regulate stress inducible gene expression. The products of stress inducible gene function both in the early stress response and in establishing plant stress tolerance. Various genes involved in stress tolerance were applied in genetic engineering of dehydration stress tolerance in transgenic Arabidopsis plants (Mittler. 2006).

By using cDNA array, the effect of drought and heat shock on tobacco plants coupled with physiological measurement was studied. A combination of drought and heat shock resulted in closure of stomata, suppression of photosynthesis, enhancement of respiration and increased leaf temperature. Some transcripts induced during drought (Yamaguchi and Shinizaki. 2006). Bioinformatics plays important role in today's science as the amount of data grows the demand for the tools and methods in data management, emphasis on those relevant to plant

science. These also cover some fundamental issues related to biological sequence analyzing, computational proteomics nasalization, computational metabolomic and biological data bases also explore emerging research topics in bioinformatics (Rizhsy *et al.*, 2004). Bioinformatics methods are the well-known for sequence analysis In order to find out the stress responsive genes computationally, the accessible sequences of that plant are explored (Rhee *et al.*, 2006). There are databases that have been structured for stress responsive genes in plants like Plant Stress Gene Database (Barozai and Wahid. 2012). The ESTs associated with stress responses were identified from multiple sources, based on the compiled list of stress-regulated genes documented in more than one plant species. In addition data from microarray expression profiles of possible candidate gene sequences comprising 650 from *Arabidopsis thaliana* (Prabha *et al.*, 2012).

World population is increasing by every year and there will be a need to produce more food especially crops such as rice, wheat, soy and maize. However, abiotic stresses critically threaten crop production and causes significant yield loss in large areas. Salinity can be termed as acute abiotic stress which includes all the problems due to salts quantity of sodium chloride from natural accumulation or irrigation (Stracke. 2001). Knowledge of two plants *Arabidopsis thaliana* and rice genomes has progressed during the past few years. However, *Arabidopsis thaliana* and rice have been sequenced. Our present knowledge of genetic engineering techniques also indicates that the genes of one species over to another species. Indeed, the intended uses aim to introduce new characters into an organism that otherwise would not have acquired them.

## Materials and methods

Abiotic stress responsive genes drought, salt, cold in *Cucumis melo* L. was find out using different bioinformatics tools NCBI, ORF, BLAST, CDD, phylogeny based on homology search. Total 21 stress responsive genes of drought, cold and salt of model plant *Arabidopsis thaliana* were selected. The cold stress responsive genes were ATMRK1, MPK3, ELIP1, HOS10, COR47, ERD14, LOS4 and seven drought

stress responses genes of subject plant were RAB18, GSTF7, CDSP32, HOS10, DI19, ABCF3, DREB2A respectively the salt stress genes were CSD1, HVA22A, HOS10, DDF1, P5CS1, CDPK1. There are 101273 ESTs of *Cucumis melo* L. ([https://www.ncbi.nlm.nih.gov/genbank/dbest/dbest\\_summary/](https://www.ncbi.nlm.nih.gov/genbank/dbest/dbest_summary/)).

Description of basic steps of selected methodology is given below:

#### *Identification of gene*

Abiotic stress responsive genes that respond to different environmental stresses were identified in model plant *Arabidopsis thaliana* through literature, available public data and different websites. Methodology of Shahid *et al.*, (2012) was used for genes identification.

On bases of availability of expressed sequence tags abiotic stress responsive genes of selected plant *Cucumis melo* L. was identified as *Cucumis melo* L. belongs to cucurbitaceae family includes cucumbers, ground melons, pumpkins and squashes that include number of wild and cultivated varieties (Rhee *et al.*, 2006). Family crops fruits with flashy and usually edible just like pumpkins and based on vegetative morphology and fruit variation (Lester. 1997). These are essential source of dietary fiber, mineral and vitamin C. There is significant economic value of melons (Munger and Robinson, 1991).

#### *Fetching of genes*

Abiotic stress responsive genes that respond to different stresses condition of model plant *Arabidopsis thaliana* was fetched from publically available databases like; a complete nucleotide sequence of selected genes was fetched from NCBI. The gene name was given in NCBI after that selected RNA sequence for the gene then save the sequence in the FASTA formate (<https://www.ncbi.nlm.nih.gov/>).

#### *ORF prediction of Model plant genes*

Open reading frames of fetched genes sequences of model genes was find out using ORF finder. It provided exact coding sequences of reference genes (Din & Barozai, 2014). Genes of *Arabidopsis thaliana*

was subjected to database ORF then selected longest frames (<https://www.ncbi.nlm.nih.gov/orffinder/>).

#### *Gene's identification through homology search*

BLAST is one of the advance tools for comparative genomic approach during homology search. The fetched genes sequences of model plants was subjected to BLAST (designed by Altschul *et al.*, 1990) against the publicly available genomic and transcriptomic sequences (Expressed sequence tags) of *Cucumis melo* L. at NCBI Genebank to carryout homology search (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

#### *ORF of Cucumis melo L.*

Open reading frames of fetched genes sequences of *Cucumis melo* L. was find out using ORF finder and saved the result of longest frame noted the length of nucleotied, amino acids. (<https://www.ncbi.nlm.nih.gov/orffinder/>).

#### *CDD investigation and comparison*

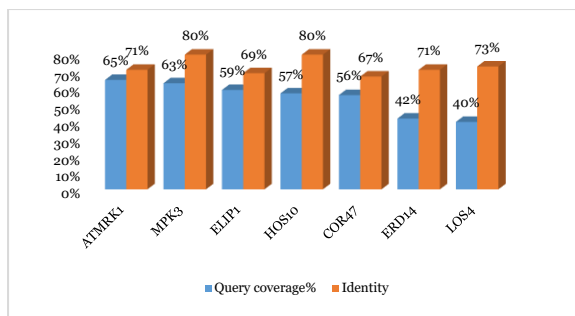
Conserved domains on reference genes of *Arabidopsis thaliana* and on predicted abiotic stress responsive genes of *Cucumis melo* L. was investigated through conserved domain database (CDD). The conserved domains was identified and compared (<https://www.ncbi.nlm.nih.gov/cdd>).

#### *MUSCLE identified genes*

For phylogenetic analysis, Cladograms was constructed using multiple sequence comparison by Log-Expectation (MUSCLE), an advance version of Clustal-w2. The gene of model plant, predicted genes of *Cucumis melo* L. and one same gene from other plants was subjected to MUSCLE. (<https://www.ebi.ac.uk/Tools/msa/muscle/>).

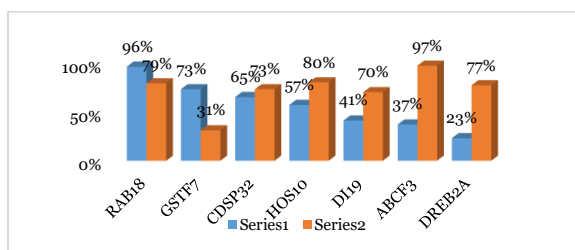
#### **Result and discussion**

Two genes MPK3 (63%), ATMRK1 (65%) query coverage came in category of highly similar. Remaining five genes COR47 (56%), HOS10 (57%) ELIP1 (59%), ERD14 (42%), LOS4 (40%) came into category of moderately similarity.



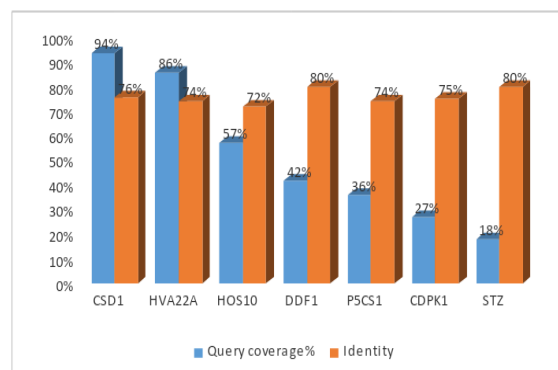
**Fig. 1.** Homology of *Cucumis melo* L. with the model gene from *A. thaliana*.

Among drought stress responsive genes, three genes showed high similarity, these were RAB18 (96%), CDSP32 (65%) and GSTF7 (73%). Three genes expressed query coverage in category of moderately similar their coverage were HOS10 (57%), DI19 (41%) and ABCF3 (37%). Single gene DREB2A (23%) came into category that was classified as less similar.



**Fig. 2.** Homology of *Cucumis melo* L. with the model gene from *A. thaliana*.

The highly query coverage similarity in salt stress responsive genes were CSD1 (94%) and HVA22A (86%). Three genes HOS10 (57%), DDF1 (42%) and P5CS1 (36%), showed moderate similarities two genes came into category Less similarity CDPK1 (27%) and STZ (18%) came into category.



**Fig. 3.** Homology of *Cucumis melo* L. with the model gene from *A. thaliana*.

*Sequence analysis of Length base pair of mRNA and amino acid*

In cold stress responsive genes in model plant and subject plant the length base pair of mRNA and amino acid were mostly similar. The frame preferred 1+ and 3+ are common in most cold stress genes. In two genes similar frame preferred which were ELIP1 (+3)

**Table 1.** ORF features for stress responsive genes in *Cucumis melo* L. and *Arabidopsis thaliana*.

SN	Gene name	Stress genes							
		mRNA length base pairs		Frame preferred		Framelength base pairs		Amino acids length	
		<i>A.thaliana</i>	<i>C.Melo</i>	<i>A.thaliana</i>	<i>C.Melo</i>	<i>A.thaliana</i>	<i>C.Melo</i>	<i>A.thaliana</i>	<i>C.Melo</i>
1	ATMRK1	2075	945	1+	3+	1035	558	344	185
2	MPK3	1858	888	3+	1+	1113	804	370	267
3	ELIP1	1035	788	3+	3+	588	579	195	192
4	HOS10	675	860	1+	3+	639	711	212	236
5	COR47	2320	727	1+	2+	798	600	265	199
6	ERD14	1178	865	1+	1+	558	714	185	237
7	LOS4	1860	790	3+	1+	1491	483	496	160
8	RAB18	1030	851	1+	2+	639	642	212	213
9	GSTF7	1213	744	3+	1+	630	648	209	215
10	CDSP32	1424	720	2+	2-	909	576	302	191
11	HOS10	675	860	1+	3+	639	711	212	236
12	DI19	1185	769	3+	3+	603	483	200	160
13	ABCF3	2493	799	2+	3+	2148	714	715	237
14	DREB2A	1651	785	3+	3+	1008	498	335	165
15	CSD1	873	704	1+	3+	459	459	152	152
16	HVA22A	977	757	2+	3+	534	546	177	181
17	HOS10	675	860	1+	3+	639	711	212	236
18	DDF1	1135	821	3+	3+	630	690	209	229
19	P5CS1	2614	812	3+	1+	2145	627	714	208
20	CDPK1	2681	770	2+	2-	1638	720	302	191
21	STZ	1303	786	3+	1+	684	657	227	218

and ERD14 (+1) respectively two genes ATMRK1 and HOS10 having similar frame in model plant (1+) and different in subject plant (+3). The COR47 gene in model plant having (+2) Frame dissimilar to all other genes.

The drought stress responsive genes DI19 and DREB2A containing similar frame (+3) and +3 frame preferred familiar in most of the genes. The length base pair of mRNA and amino acids were same. In salt stress gene CSD1 the amino acid length and frame length base pairs in *Arabidopsis thaliana* and *Cucumis melo* was similar (152) and (459) however, CDPK1 gene different frame preferred in subject plant (\_2) and model plant(+2).

#### Conserved domain

Conserved domain study of model plant *Arabidopsis thaliana* genes in *Cucumis melo* L. Study showed that all genes same conserved domains were present in *Cucumis melo* L. Irrespective of percentage query coverage and identity.

The similar conserved domains in cold stress responsive genes were Dehydrin super-family (COR47), Chloroa-b-bind super-family (ELIP1), Dehydrin (ERD14), SMT super-family (HOS10), DEXDc super-family (LOS4), PKc-like super-family (MPK3), PKc-like super-family (ATMRK1).

**Table 2.** Comparison of existence and type of conserved domain families between *Cucumis melo* L. and *Arabidopsis thaliana*.

SN	Gene name	Stress genes	
		Conserved domain family	
		<i>A.thaliana</i>	<i>Cucumis melo. L</i>
1	COR47	Dehydrin super-family	Dehydrin superfam
2	ELIP1	Chloroa-b-bind super-family	Chloroa-b-bind super-family
3	ERD14	Dehydrin	Dehydrin
4	HOS10	SMT super-family	SMT super-family
5	LOS4	DEXDc super-family	DEXDc super-family
6	MPK3	PKc-like super-family	PKc-like super-family
7	ATMRK1	PKc-like super-family	PKc-like super-family
8	ABCF3	SunT super-family	SunT super-family
9	CDSP32	TRX-COSP32	TRX-COSP32
10	DI19	Zf-Di19 super family	Zf-Di19 super family
11	DREB2A	AP2 super family	AP2 super family
12	HOS10	SMT-super family	SMT-super family
13	RAB18	Ras super-family	Ras super-family
14	GSTF7	GST-N-Phi thioredoxin-like super family	GST-N-Phi thioredoxin-like super family
15	CDPK1	STKs-CAMK PKc- like super family	PKc- like super family
16	CSD1	Cu-Zn-superoxide-Disntase super-family	Cu-Zn-superoxide-Disntase super-family
17	DDF1	AP2 super-family	AP2 super-famiy
18	HOS10	SMT super-family	SMT super-family
19	HVA22A	TB2-DP1-HVA22 super-family	TB2-DP1-HVA22 super-family
20	P5CS1	PROB super-family	PROB super-family
21	STZ	Zf-C2H2-6 super-family	Zf-C2H2-6 super-family

Respectively the similar conserved domains in drought stress responsive genes were SunT super-family (ABCF3), TRX-COSP32 (CDSP32), Zf-Di19 super family (DI19), AP2 super family (DREB2A), SMT-super family (HOS10), Ras super-family (RAB18), GST-N-Phi thioredoxin-like super family (GSTF7).The similar conserved domains in drought stress responsive genes were STKs-CAMK PKc- like super family (CDPK1), Cu-Zn-superoxide-Disntase super-family (CSD1), AP2 super-family (DDF1), TB2-

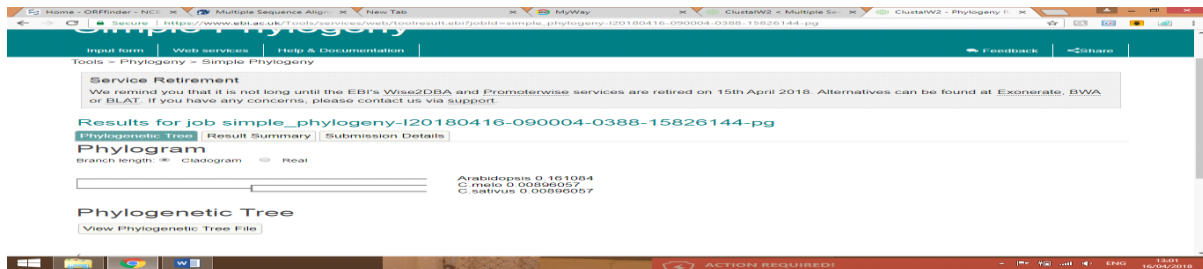
DP1-HVA22 super-family (HVA22A), PROB super-family (P5CS1), Zf-C2H2-6 super-family (STZ).

#### Phylogenetic analysis of genes

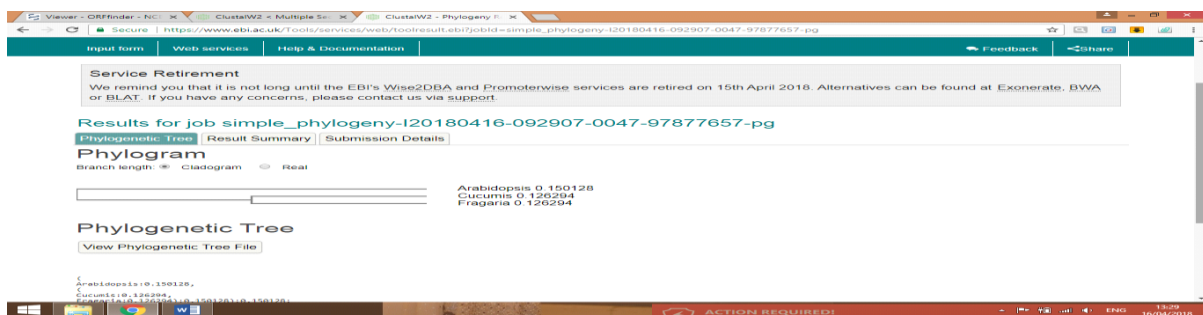
Plant genes mainly stress related genes are frequently clustered in gene families consisting members with unnecessary functions. The genomic information accessible in public databases has made study of the origin and expansion of gene families possible. Over the course of evolution plant genomes

have undergone large scale duplication and rearrangement during which some duplicated genes gain new functions, some persist maintaining their ancestral functions, while some others entirely lose functionality (Demuth & Hahn, 2009) Cladogram was constructed comparing three homolog of abiotic stress responsive genes one from model plant (*Arabidopsis thaliana*), second from subject plant (*Cucumis melo* L.) Third from dicot and monocot

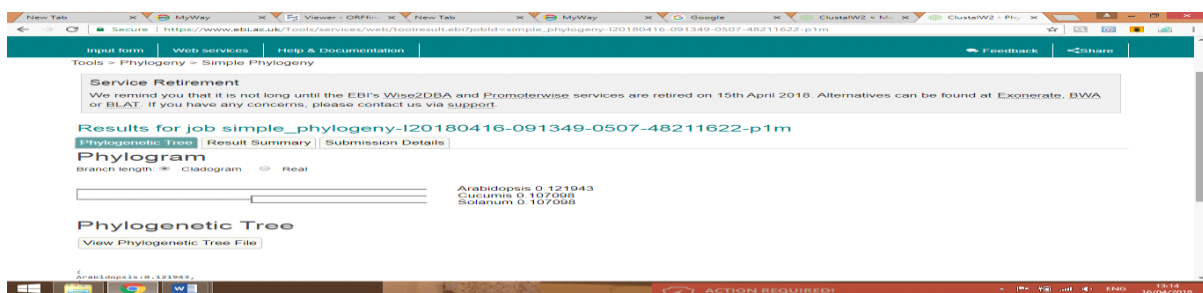
plants. In abiotic stress responsive genes DI19, DREB2A, CDPK1 were more closely related to in *Arabidopsis thaliana* and *Cucumis melo*. In ATMRK1 genes *Cucumis melo* L. *Cucumis sativus*, ERD14 gene *Arabidopsis Thaliana* to *Raphanu*, LOS4 gene *Cucumis melo* L. to *Fragaria*, MPK3 and CDSP32 genes *Cucumis melo* L. to *Solanum* and P5CS1 *Arabidopsis Thaliana* to *Brassica*. showing close relation.



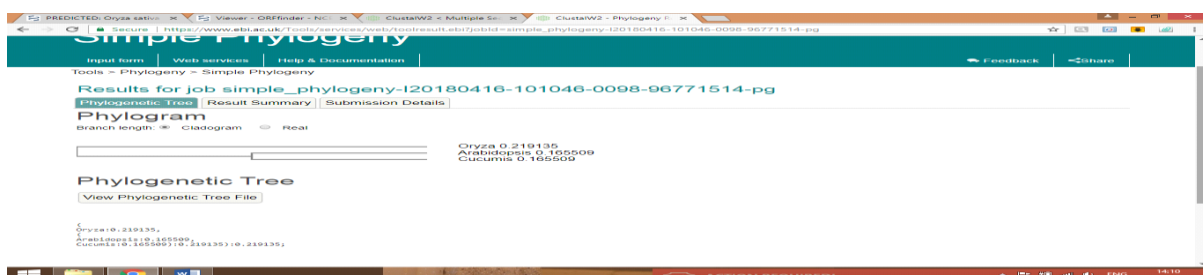
**Fig. 1.** Cladogram of *Arabidopsis Thaliana*, *Cucumis melo* L. *Cucumis sativus* for ATMRK1.



**Fig. 2.** Cladogram of *Arabidopsis. Thaliana*, *Cucumis melo* L. and *Fragaria vesca* for LOS4 gene.



**Fig. 3.** Cladogram of *Arabidopsis. Thaliana*, *Cucumis melo* L. and *Solanum lycopersicum* for MPK3 gene.



**Fig. 4.** Cladogram of *Arabidopsis. Thaliana*, *Cucumis melo* L. and *Oryza sativa* for DI19 gene.



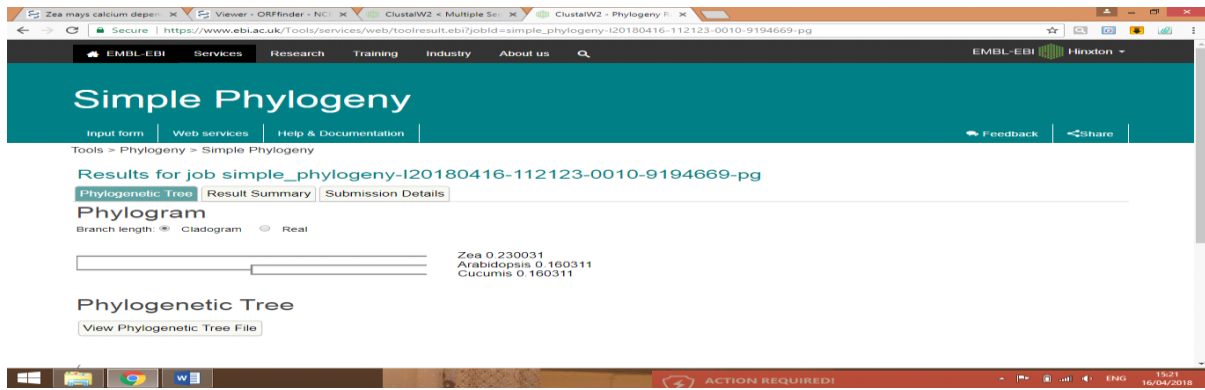


Fig. 5. Cladogram of *Arabidopsis thaliana*, *Cucumis melo* L. and *Zea mays* for CDPK1 gene.

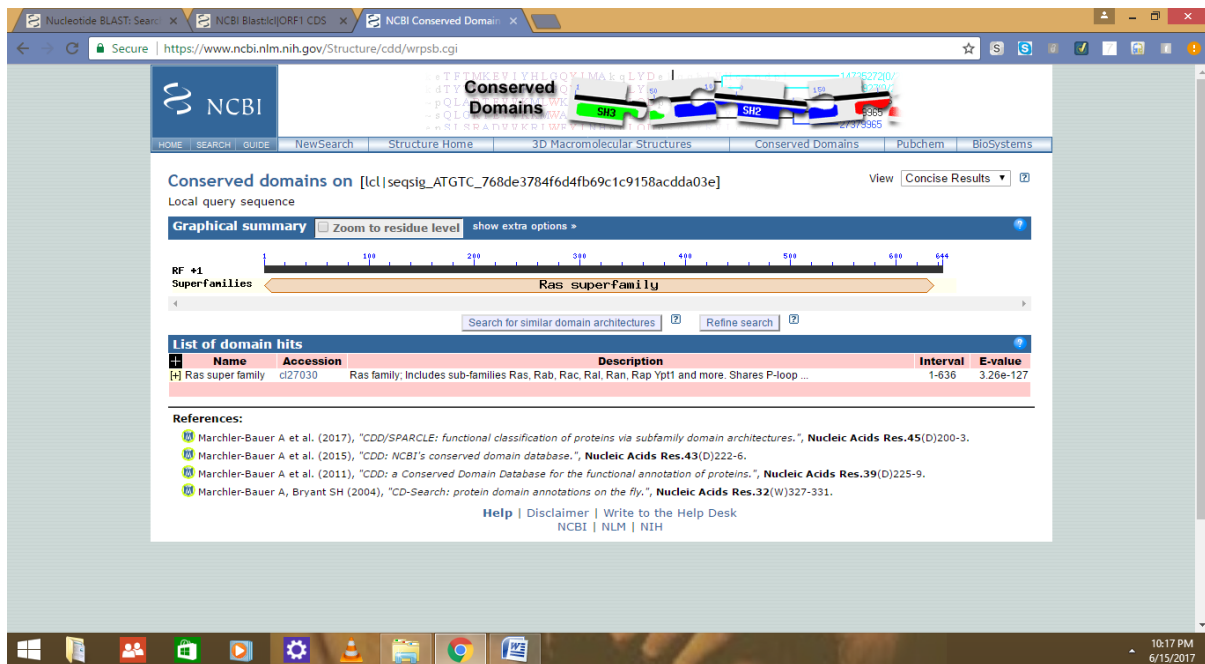


Fig. 6. Conserved domain of *Cucumis melo* L. (RAB18).

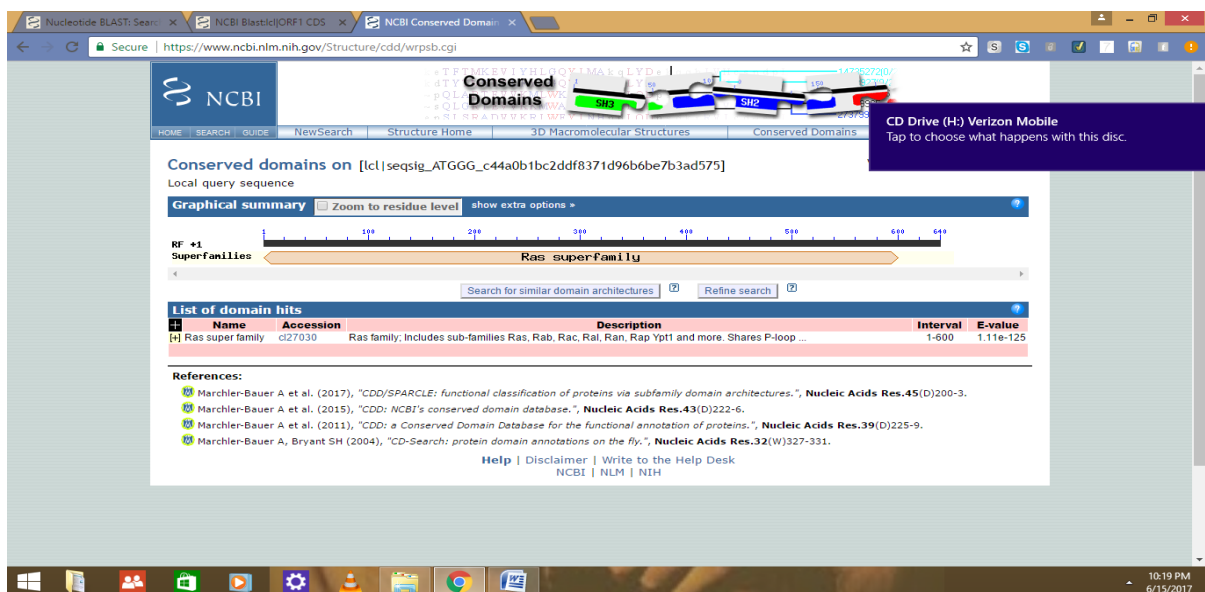


Fig. 7. Conserved domain of *Arabidopsis thaliana* (RAB18).

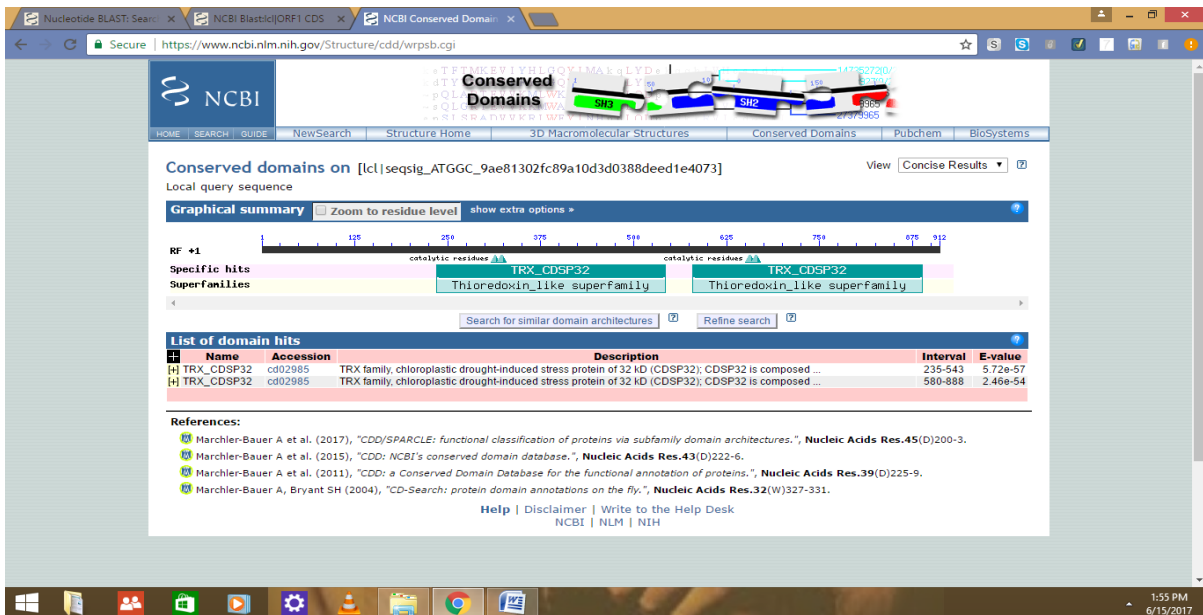


Fig. 8. Conserved domain of *Cucumis melo* L.(CDSP32).

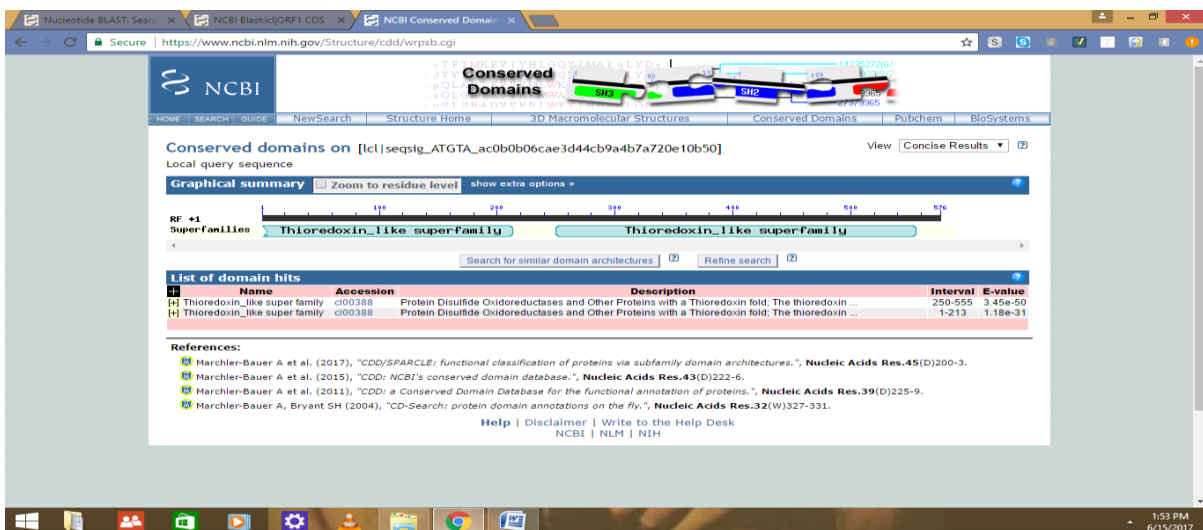


Fig. 9. Conserved domain of *Arabidopsis thaliana*(CDSP32).

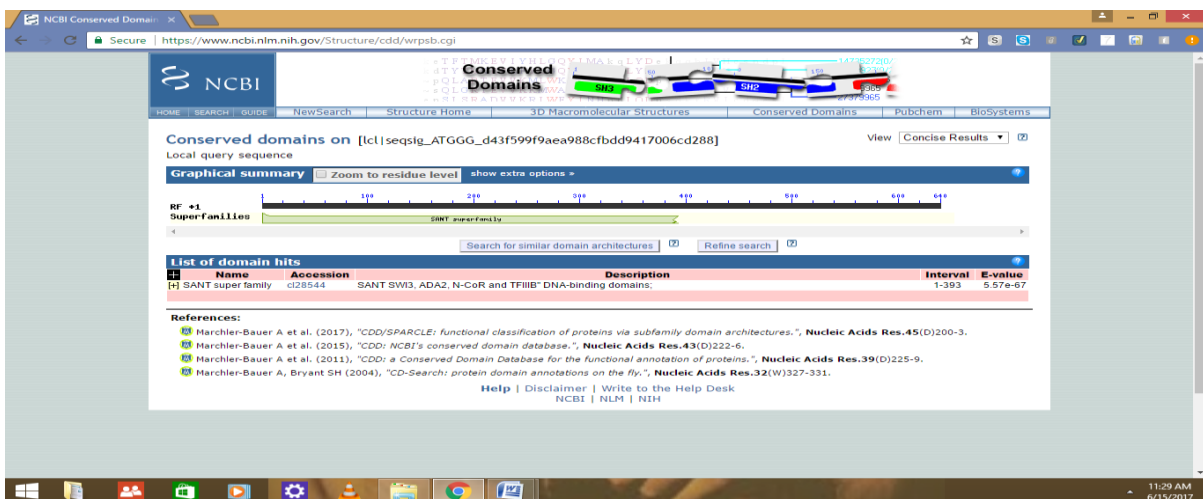
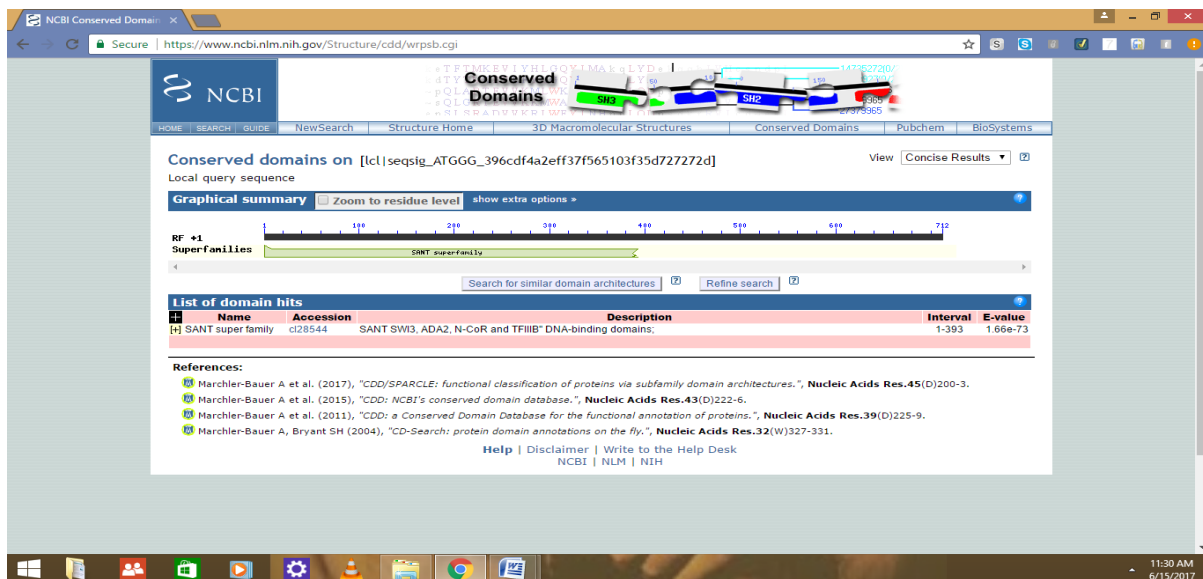


Fig. 10. Conserved domain of *Cucumis melo* L. (HOS10).





**Fig. 11.** Conserved domain of *Arabidopsis thaliana* (HOS10).

## Conclusion

In silico identification of abiotic stress responsive genes in *Cucumis melo* L. established interesting results overall results of bioinformatics investigation of twenty one abiotic responsive genes. The cold stress responsive genes were ATMRK1, MPK3, ELIP1, HOS10, COR47, ERD14, LOS4 and seven drought stress responses genes of subject plant (*Cucumis melo* L.) were RAB18, GSTF7, CDSP32, HOS10, DI19, ABCF3, DREB2A respectively the salt stress genes were RAB18, GSTF7, CDSP32, HOS10, DI19, ABCF3, DREB2A. The genes with highest query coverage, identity and high percentage identity on conservation analysis was shown by genes RAB18, CSD1, HVA22A, GSTF7, and almost all genes ATMRK1, MPK3, CDSP32, ELIP1, HOS10, COR47, ERD14, ABCF3, DREB2A, DI19, P5CS1, CDPK1, STZ expressed similar conserved domains and showed low query coverage.

Abiotic stress responsive genes (cold, drought, salt) in *Cucumis melo* L. The gene HOS10 expressed similar Query coverage (57%) but the identity varies in cold stress genes (80%) in salt stress gene (72%). Similar conserved domain SMT super-family was noticed in cold, drought and salt. In model plant *Arabidopsis thaliana* and *Cucumis melo* L. genes CDPK1, ATMRK1, MPK3 showed PKc-like super-family and COR47, ERD14 expressed Dehydrin super-family. For phylogenetic analysis Cladogram was constructed to

Comparing three homolog of abiotic stress responsive genes one from *Arabidopsis thaliana* second from *Cucumis melo* L. third from dicot and monocot plants. *Arabidopsis thaliana* and *Cucumis melo* L. DI19, DREB2A, CDPK1 genes were more closely related.

## References

- Adams CF, Richardson M.** 1981. Nutritive value of foods. USDA Home and Garden.
- Barozai MYK, Wahid AH.** 2012. In silico identification and characterization cumulative abiotic stress responding genes in Potato (*Solanum tuberosum* L.). *Pak. J. Bot* **44**, 57-69
- Bohnert HJ, Nelson DE, Jensen RG.** 1995. Adaptations to environmental stresses. The plant cell **7(7)**, 1099.
- Bray EA.** 1993. Molecular responses to water deficit. *Plant physiology* **103(4)**, 1035
- Lester G.** 1997. Melon (*Cucumis melo* L.) fruit nutritional quality and health functionality. *Hort Technology* **7(3)**, 222-227.
- Lytton J.** 2004. The cation/Ca<sup>2+</sup> exchanger superfamily: phylogenetic analysis and structural implications. *Molecular biology and evolution* **21(9)**, 1692-170

**Ma Y, Qin F, Tran LSP.** 2012. Contribution of genomics to gene discovery in plant abiotic stress responses. *Molecular plant* **5(6)**, 1176-1178.

**Mittler R.** 2006. Abiotic stress, the field environment and stress combination. *Trends in Plant Science* **11(1)**, 15-19.

**Munger HM, Robinson RW.** 1991. Nomenclature of Cucumis melo Cucurbit Genet.

**Prabha R, Ghosh I, Singh DP.** 2011. Plant Stress Gene Database: A collection of plant genes responding to stress condition. *ARPJ. Sci. Technol* **1**, 28-31.

**Reymond P, Weber H, Damond M, Farmer EE.** 2000. Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. *The Plant Cell* **12(5)**, 707-719.

**Rhee SY, Dickerson J, Xu D.** 2006. Bioinformatics and its applications in plant biology. *Annu. Rev. Plant Biol* **57**, 335-360.

**Rhee SY, Dickerson J, Xu D.** 2006. Bioinformatics and its applications in plant biology. *Annu. Rev. Plant Biol* **57**, 335-360.

**Rizhsky L, Liang H, Shuman J, Shulaev V, Davletova S, Mittler R.** 2004. When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. *Plant Physiology* **134(4)**, 1683-1696.

**Stracke R, Werber M, Weisshaar B.** 2001. The R2R3-MYB gene family in Arabidopsis thaliana. *Current opinion in plant biology* **4(5)**, 447-456.

**Yamaguchi K, Shinozaki K.** 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol* **57**, 781-803.