J Genet Resour 2019;5(2): 72-82 *Homepage: http://sc.journals.umz.ac.ir/*

RESEARCH ARTICLE *DOI: 10.22080/jgr.2019.2341*

J G R

Functional Annotation of Two *Hypothetical* Proteins Reveals Valuable Proteins Involved in Response to Salinity: An *in silico* Approach

Kiarash Jamshidi Goharrizi[1], Farzane Amirmahani[2*], Foad Fatehi[3], Maryam Nazari[4] and Sayyed Saeed Moosavi[4]

[1] Department of Plant Breeding, Yazd Branch, Islamic Azad University, Yazd, Iran.
[2] Department of Biology, Faculty of Sciences, University of Isfahan, Isfahan, Iran.
[3] Department of Agriculture, Payame Noor University, Tehran, Iran.
[4] Department of Agronomy and Plant Breeding, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran.

ARTICLEINFO

Article history:
Received 07 April 2019
Accepted 20 June 2019
Available online 08 July 2019

Keywords:
Hypothetical protein
UCB-1 pistachio rootstock
Response to salinity
Reverse transcriptase enzyme

Corresponding author:
✉ F. Amirmahani
Farzanemahani@yahoo.com

Print & Online ISSN:
p-ISSN 2423-4257
e-ISSN 2588-2589

ABSTRACT

Through the exponential development in the specification of sequences and structures of proteins by genome sequencing and structural genomics approaches, there is a growing demand for valid bioinformatics methods to define these proteins function. In this study, our objective is to identify the function of unknown proteins from *UCB-1 pistachio* rootstock and specify their classification using bioinformatics tools. In previous research, we recognized 5 *HPs* in proteomic profile of the University of California at Berkeley I *pistachio* rootstock leaf under salinity stress. Two of them had 2.95 and 2.29-fold up-regulation under salinity stress. In this study, the probable function and characterization of these *HPs* were recognized using different statistical methods and programs. According to our analyses, these *HPs* have similarities with reverse transcriptase enzyme as well as helicase enzyme and some responsive proteins to salt stress. These observations suggest a close relationship between the overexpression of these enzymes and plant responses to salinity stress. These stress-responsive proteins could provide a novel plant defense strategy in response to salinity.

Please cite this paper as: Jamshidi Goharrizi K, Amirmahani F, Fatehi F, Nazari M, Moosavi S. Functional annotation of two *Hypothetical* proteins reveals valuable proteins involved in response to salinity: An *in silico* approach. *J Genet Resour* 5(2):72-82. doi: 10.22080/jgr.2019.2341

Introduction

Pistachio, with at least 11 species and as the most important nut all over the world, belongs to the *Anacardiaceae* family (Moazzzam Jazi *et al.,* 2017). *P. integerrima, P. atlantica,* and *P. vera* are three of the most important *pistachio* species that *P. vera* is used as cultivated species, while *P. integerrima, P. atlantica* is used as rootstock for *P. vera* cultivation (Moazzzam Jazi *et al.,* 2017). Hybrid *UCB-1 pistachio* rootstock was produced with a cross between *P. atlantica* female × *P. integerrima* male and by controlled pollination in the University of California at Berkeley (Ahmad *et al.,* 2005). *UCB-1* has better salinity tolerance (Ahmad *et al.,* 2005; Ferguson *et al.,* 2002; Akbari *et al.,* 2018), better yields (Ferguson *et al.,* 2002), greater *Verticillium* tolerance (Morgan *et al.,* 1992) and better cold-resistance (Epstein *et al.,* 2004) than other *pistachio* rootstocks.

When plants are exposed to salinity stress, agricultural fertility and grow of them are limited (Schwarz *et al.,* 2010). The yield of *pistachio*, which is classified as a salt-tolerant plant, is significantly reduced by salinity stress (Moazzam Jazi *et al.,* 2016; Jamshidi *et al.,* 2019). In plants, physiological, biochemical and molecular reactions are the controllers of tolerance to salinity stress (Mostek *et al.,* 2015; Kiarash *et al.,* 2018). Plant proteomic analysis is a useful technique for a better understanding of the plant's mechanisms in reaction to abiotic

stresses such as drought and salinity (Barkla Bronwyn *et al.,* 2013). In plants and under various tensions, the proteomic analysis provides essential information about the expression of the proteins (Barkla Bronwyn *et al.,* 2013).

Hypothetical proteins (HPs) or unknown proteins are proteins that have not been linked to known genes (Bharat Siva Varma *et al.,* 2015), have not been characterized (Bharat Siva Varma *et al.,* 2015), and whose existence has been predicted (Galperin 2001). Domains of *HPs* have known proteins which have not any recognized structural or functional domain (Bharat Siva Varma *et al.,* 2015). A big section of the mammalian proteome is demonstrated by *HPs.* Domain homology searches are able to predict the function of *HPs* with different confidence levels (Srinivasan *et al.,* 2015). *HPs* have conserved domains and it is possible that these conserved domains be compared with the known family domains. In this regard, *HPs* could be categorized and determined into specific protein families by homology modeling (Srinivasan *et al.,* 2015).

Many computational tools are used to predict protein function in *HPs*. This has been attained from information derived from phylogenetic analysis, sequence similarity, active site residue similarity, conserved domains, protein-ligand interactions, protein-protein interaction, motifs, gene expression profiles and phosphorylation regions (Bharat Siva Varma *et al.,* 2015). The purpose of this study is to identify the function of unknown proteins from *UCB-1 pistachio* rootstock and specify their classification using bioinformatics tools. These data can be applied to use these proteins as beneficial targets for the engineering of the plant varieties tolerant to salt stress conditions.

Materials and Methods

Selection of the hypothetical protein

In our previous study on changes in the proteomic profile of *UCB-1 pistachio* rootstock, 5 *HPs* under NaCl stress were reported. Two of these proteins were up-regulated 2.95 and 2.29-fold under salinity stress (Jamshidi Goharrizi *et al.,* 2019). The results of the *NCBI BLAST* search tool showed sequence similarities with uncharacterized proteins named *LOC105974023* from *Erythranthe guttata* and *LOC109219975* from *Nicotiana attenuata*. Therefore, these proteins were selected for further analysis. The overall methodology for the characterization of HPs is presented in Fig.1.

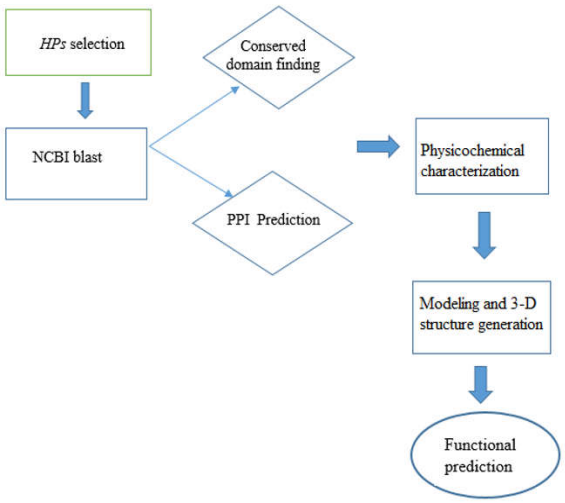

Fig.1 The overall methodology for the characterization of *HPs*.

The conserved domains analysis

Domains are often determined as (sequence or structure) units, which can be thought of as separate functional and/or structural protein units. In the course of molecular evolution, it is suggested that domains may have been used as building blocks and have faced recombination to regulate protein functions (Galperin 2001). A domain or fold might also demonstrate a higher conservancy degree as compared with the whole sequence (Lesk and Chothia 1986). In this study, five bioinformatics tools including *CDD-BLAST (Conserved Domain Database-Basic Local Alignment Search Tool), PFAM, HmmScan, SMART (Simple Modular Architecture Research Tool),* and *SCANPROSITE* were applied. These tools are able to find the specified conserved domains in the desired proteins' sequences and then can help in the putative proteins' classification in a specific protein family.

Physicochemical characterization

ProtParam tool *(http://web.expasy.org/protparam/protparam-doc.html)* in the *ExPASy* server was used to evaluate the physicochemical properties of *HPs* (Gasteiger *et al.,* 2005). Theoretical pI, molecular weight, the total number of negative

and positive residues, amino acid composition, aliphatic index, *GRAVY* or the grand average of hydropathicity, extinction coefficient, and instability index are computed by ProtParam tool (Gasteiger *et al.,* 2005). A protein aliphatic index is an indicator that is described as the comparative content occupied by aliphatic side fetter amino acids. The *GRAVY* value for a protein or peptide is computed as the aggregation of the hydropathy amounts of all of the amino acids distributed by the number of rests in the sequence. The amount of light that a protein absorbs at a determined wavelength is shown by the extinction coefficient. Instability index is an indicator for calculating protein stability in a tube. An instability index >40 is anticipated to be unstable and a value <40 is anticipated to be stable (Gasteiger *et al.,* 2005).

Sequence analysis

To calculate sequence similarity, *BLAST* or *Basic Local Alignment Search Tool* is frequently used (Altschul *et al.,* 1990; Amirmahani and Jamshidi Goharrizi 2018). The *FASTA* sequences of the selected proteins were the inquiry sequences, and analogous proteins in various databases were investigated for using the *BLASTP* program. *BLASTP* is applied to discover analogous sequences in protein databases and to recognize a query amino acid sequence (Altschul *et al.,* 1990).

PPI prediction

The Database of *Interacting Proteins (DIP) (http://dip.doe-mbi.ucla.edu)* was applied to predict the proteins' interactions. It is a biological database that gathered experimentally defined interactions between proteins. It combines data from a range of sources to make a single, constant set of protein-protein interactions (Salwinski *et al.,* 2004). If the protein of interest is not present in the database, it is also practical to done sequence similarity *(BLAST)* and motif searches to find closely related proteins. The interaction pattern of them might give insights into the potential but not yet determined query proteins' interactions (Salwinski *et al.,* 2004).

HHPred model generation

Formal sequence quest procedures examine sequence databases such as non-redundant databases or *UniProt*; however, *Hhpred* searches a vast diversity of databases such as *CDD, COGs, PDB, Pfam, SCOP* and *SMART* (Varma *et al.,* 2015). Also, one of the fastest servers for structure prognostication and remote protein homology prognostication is *HHpred*. *HHpred* acts based on the pairwise analogy profile of *HMMs* or *hidden Markov* models (Varma *et al.,* 2015).

Three-dimensional structure

The computational prediction of a protein structure from its amino acid sequences considerably simplifies its function prediction, subsequently (Gazi *et al.,* 2018). An online server *PS*^{2}*-v2* (PS Square version 2), a template-based method was utilized for the structure of the *HP* prediction. The protein modeling by this online server proved the function of it latter. Also, *PFP-FunDSeqE* has been applied to evaluate the protein fold patterns based on functional domain information and evolutionary data combination.

The quality of the model was evaluated by *VERIFY3D* by assigning a structural class according to the location and environment of each residue position and by comparing the results to good structures. Environments of residues correlated three parameters: the local secondary structure, the residue area that is buried and the side-chain area fraction covered by polar atoms (Eisenberg *et al.,* 1997).

Functional prediction

Different tools were applied for accurate functional assignments of the desired *HPs* such as *CDD-Blast, Pfam, HmmScan, SMART, Scanprosite, MOTIF, INTERPROSCAN, CATH, SUPERFAMILY*, and *Protonet*.

PlantPReS website

PlantPReS (http://www.proteome.ir/) has a proteomic database of over 35086 stress-responding proteins of 577 virtual articles. It contains over 10,600 proteins of the unique stress response.

This provides information about the plant type, registration number, protein name, stress type, tissue, and growth stage of all stress-response proteins. *PlantPReS* also provides customized *BLAST* tool searches for sequences with common ancestors. In addition, a filtration mode is provided in *PlantPReS* for several analyzes. The text or graphic format of the results can be displayed (Kumar and Shanker 2018).

Results

The sequence of hypothetical protein

HPs were subjected to *NCBI Blast* search to gain the primary data. The sequences of our *HPs* were chosen to specify their function by different tools. The *FASTA* format of them are as follows:

LOC105974023:

MISQRGIEANPAKIEAITSMAPPTSIKKVQQLNGCL
AALNRFISRSADKGLPFFKILRGGKKFEWNEDCQRA
FTELKAYLTSPPLLTKPQPGDTLFLYLAISADAISA
VLIRDGEKGHQPIYYISRALQGPEHRYTNMEKLALA
LINAARKLRPYFQSHQVIVLTNYPLKQILRSPETSG
RLAKWAIELSEYGVEFKPRPAIKAQILADFVVEMTT
SEESTSIPTWAINVDGSSTATGGGAGIASNNEAEYE
ALIAGIRLALAAGARKLIVHSDSQLVVNQVLGNYEA
KEESMAKYLALALTLLSKLDSYEIKQVPRANNIDAD
KLARLGSSMASIGSRKITLLTASQPEIVSTDGVNCA
EESEPCWITPITNYLKSGELPTDIAQAKKIKVRAAR
FLMIGEDLYKRGFSSPYLKCLNPSAADYVLREVHEG
ICGNHLSGRNLALKILRQGYYWPTMHEDAKKLVQRC
KPCQEHANILHLPAALMQPIDSPIPFAQWGVDLVGP
FPPATGGRKFLIVAVDYFTKWVEAEPLARIR

LOC109219975:

MPRPVFYEEEVDNTNRLIRDELRYNRRSLSKEHDEL
LIKLTLEQKSVYDRIITAVHEDKGGLFFLYGHGGTG
KTFIWRTLSSAIRSKGDIVLTVASSGIASLLLPGGR
TAHSRFAIPLNATEDSTCNIKQRSPLAKLIVETKLI
IWDEAPMMHRYCFEALDRTLRDILRFKDASNLDRPF
GGKIVVLGGDFRQILPVITKGTRQDIVNAALNSSYL
WNHCHVLKLTKNMRLEGNQVESHLNDLRQFSDWVLA
IGDGMIENFVDGIEKVHIPDVLLINNCDDPISVIVE

Physicochemical characterization

In this study, we analyzed the physicochemical characteristics of two *HPs* from UCB-1 *pistachio* rootstock for the first time. The physicochemical features of the intended proteins are summarized in table 1. Isoelectric point (pI) of the *HPs* were 9.01 and 6.51. pI is the pH at which the amino acid of protein tolerates no net charge and therefore does not move in a direct current electrical field.

Table 1. Physicochemical properties of the *hypothetical proteins*.

Properties	**Values**	
Hypothetical proteins	LOC105974023	LOC109219975
Number of amino acids	535	288
Molecular Weight	58962.16 Da	32652.53 Da
Theoretical pI	9.01	6.51
Total number of negatively charged residues	53	38
Total number of positively charged residues	63	36
Ext. coefficient	67310 $M^{-1}cm^{-1}$	31190 $M^{-1}cm^{-1}$
Instability index	35.83	40.66
Aliphatic index	96.39	103.89
Grand average of hydropathicity (GRAVY)	-0.134	-0.178

At 280 nm, the extinction coefficient of *HPs* ranges 67310 and 31190 $M^{-1}cm^{-1}$. The presence of a high concentration of Cys, Trp, and Tyr shows a higher extinction coefficient of *HPs*. The instability index value of the *HPs* was found to be 35.83 and 40.66. It is suggested that a protein will be stable whose instability index is smaller than 40, a value above 40 predicts that the protein will be unstable. The AI is the relative volume of a protein occupied by aliphatic side chains (A, V, I, and L) and is considered as a positive factor for the raise of thermal stability of globular proteins. The Aliphatic index for the *HPs* was 96.39 and 103.89. The GRAVY of *HPs* is -0.134 and -0.178. The better interaction of protein and water is occurring in low GRAVY.

Sequence analysis

The similarities of the protein sequences were searched against the *PDB* protein structure database and non-redundant *UniProtKB/SwissProt* database. The results of preliminary data revealed the possible recognition of resembling proteins from various organisms. The identities, similarities, and *E*-

values are given in table 2. The results of *PDB* and *UniProtKB/SwissProt* databases were recognized *RNase_HI_like* and *ATP-dependent DNA helicase PIF1-like* as domains hit of these *HPs*. Sequences of the *HPs* were analyzed for functional domain identification using five bioinformatics tools namely *CDD-BLAST, Pfam, HmmScan, SMART,* and *SCANPROSITE*. If the given five tools showed similar domains for proteins, we considered it a 100% confidence level (Gazi *et al.,* 2018). Various confidence levels were determined on the basis of obtained results of these web-tools. One hundred percentage confidence level was considered upon obtaining the same domains *(RNase H-like domain* found in reverse transcriptase and *N-terminal domain* of the *DEAD-box helicase* superfamily) from the five different tools.

PPI prediction

Based on the results of the *PPI* prediction through sequence similarity search, *LOC105974023* and *LOC109219975 HPs* showed a link with *Reverse transcriptase and PIF1* protein *(helicase)*, respectively (E- values, 6e-17 and 1e-12).

HHPred model generation

As shown in the above table, our *HPs* show sequence similarities with few proteins of recognized function. Therefore, *HHpred* was applied to create distant homology models.

The homology derived model was gained by *HHPred* using *4QDY* as the template at the Modeler server. According to the *HHPred* results (Fig. 2), it can be concluded that similar and new sequence similarities were obtained with the unknown proteins, including, reverse transcriptase, integrase, and *DNA* binding protein for *LOC105974023 HP* and proteins like *Exodeoxyribonuclease V, ATP-dependent DNA helicase PIF1 and DNA helicase I/DNA* complex for *LOC109219975 HP*.

A

No	Hit	Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	4OL8_B Reverse transcriptase/r	100.0	1.3E-32	2.7E-37	273.1	20.8	212	1-215	263-477 (478)
2	4MH8_A Reverse transcriptase/r	99.9	4.4E-30	9.6E-35	265.4	23.8	338	1-341	253-651 (652)
3	5OVN_B POL protein; FIV, Rever	99.9	1.3E-25	2.9E-30	219.6	12.1	190	1-212	231-428 (428)
4	5OVN_A POL protein; FIV, Rever	99.8	4.5E-22	9.8E-27	201.1	15.5	298	1-327	213-532 (532)
5	1MU2_B POL polyprotein(E.C.2.7	99.7	9.3E-21	2E-25	185.2	12.3	170	1-190	228-409 (426)
6	4G1Q_A Reverse transcriptase/r	99.6	5.5E-18	1.2E-22	172.1	15.8	298	1-330	235-557 (557)
7	4G1Q_B Reverse transcriptase/r	99.6	2E-19	4.3E-24	175.7	4.2	158	1-164	233-401 (428)
8	1MU2_A POL polyprotein(E.C.2.7	99.5	2.3E-15	5E-20	153.2	20.1	297	1-330	233-554 (555)
9	3OYH_B PFV integrase; PROTEIN-	99.2	1.8E-12	3.9E-17	126.2	15.5	133	398-535	33-165 (395)
10	5U1C_D DNA-binding protein 7d,	99.0	1.5E-11	3.2E-16	119.2	10.8	158	363-535	16-188 (383)
11	3F9K_o Integrase (E.C.3.4.23.4	98.3	1.9E-08	4.2E-13	87.7	9.6	94	425-535	1-94 (210)
12	5HOT_B Integrase; integrase, i	98.2	4.1E-08	8.9E-13	91.0	8.6	85	437-535	1-93 (291)
13	3JCA_B Integrase; integration,	98.1	3.4E-08	7.4E-13	90.0	6.2	98	425-535	1-98 (265)
14	5CZ2_E Pr160; integrase, POL,	97.8	4.5E-07	9.8E-12	78.8	6.9	98	425-535	1-98 (210)
15	4FW2_B Integrase; DNA BINDING	97.8	1.5E-06	3.3E-11	79.4	9.5	96	426-535	2-97 (270)
16	3NNQ_B N-terminal domain of Mo	97.7	6.4E-06	1.4E-10	64.0	10.2	104	374-486	7-113 (114)
17	1K6Y_B pol polyprotein (E.C.2.	97.5	4.9E-06	1.1E-10	72.4	8.6	85	437-535	2-93 (212)
18	5T82_A Reverse transcriptase;	97.5	3.3E-06	7.1E-11	59.5	5.6	66	21-88	13-82 (90)
19	5M0Q_H integrase/DNA Complex;	97.4	8.3E-06	1.8E-10	75.0	8.3	93	424-535	3-95 (281)
20	4NZG_D Integrase p46; Structur	97.4	3.4E-05	7.5E-10	58.0	10.3	93	378-479	5-100 (100)
21	3HPH_A Integrase, PC4 and SFRS	97.3	4.2E-05	9.2E-10	67.0	10.5	90	420-535	6-95 (219)
22	2KH4_A Putative uncharacterize	97.0	0.00075	1.6E-08	54.8	13.0	119	89-213	1-132 (147)
23	2EHG_A ribonuclease HI (E.C.3.	96.5	0.0043	9.3E-08	50.1	11.9	113	95-213	2-129 (149)
24	4IBN_A Ribonuclease H (E.C.3.1	96.3	0.0037	8.2E-08	55.0	10.4	124	89-213	11-163 (220)

B

No	Hit	Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	3E1S_A Exodeoxyribonuclease V,	99.7	9E-20	1.9E-24	164.8	15.3	175	39-263	189-368 (574)
2	6HPH_A ATP-dependent DNA helic	99.7	4.3E-19	9.2E-24	154.0	15.8	190	38-244	3-193 (418)
3	5O6B_B ATP-dependent DNA helic	99.7	1.5E-18	3.1E-23	156.1	19.4	191	39-259	1-202 (545)
4	5FTB_A TPR DOMAIN PROTEIN (E.C	99.7	4.4E-18	9.4E-23	148.1	19.4	192	38-244	4-202 (433)
5	5H8O_A DNA helicase I/DNA Comp	99.6	7.6E-18	1.6E-22	168.0	13.4	201	7-245	937-1145(1756)
6	1W36_G EXODEOXYRIBONUCLEASE V	99.6	5E-17	1.1E-21	147.7	15.5	181	41-259	151-380 (608)
7	3UPU_C ATP-dependent DNA helic	99.6	1.6E-16	3.3E-21	140.0	15.3	183	37-244	23-211 (459)
8	4B3F_X DNA-BINDING PROTEIN SMU	99.4	1.2E-14	2.5E-19	133.4	13.1	147	39-234	189-445 (646)
9	5EAN_A DNA replication ATP-dep	99.4	4.5E-15	9.6E-20	143.3	10.6	182	30-234	620-836 (1059)
10	2XZO_A REGULATOR OF NONSENSE T	99.4	4.9E-15	1E-19	135.4	10.0	173	39-234	179-416 (623)
11	5MZN_A PROTEIN; Yeast Helicase	99.4	8.3E-15	1.8E-19	136.6	11.3	182	30-234	234-505 (749)
12	3B85_A Phosphate starvation-in	99.4	1.5E-14	3.1E-19	112.9	10.7	181	39-243	7-207 (208)
13	2WJY_A REGULATOR OF NONSENSE T	99.3	5.1E-15	1.1E-19	138.8	8.2	173	39-234	356-593 (800)
14	2XZL_A ATP-DEPENDENT HELICASE	99.3	8.1E-15	1.7E-19	137.6	8.0	171	39-232	360-593 (802)
15	5WWP_A ORFlab; Middle East res	99.3	7.3E-14	1.6E-18	127.0	13.6	168	39-232	259-447 (600)
16	2ZPA_B Uncharacterized protein	99.2	1.9E-12	4.1E-17	118.2	14.6	196	39-287	175-370 (671)
17	4PJ3_A Aquarius; RNA helicase,	99.1	4.3E-12	9.1E-17	125.7	12.3	185	39-252	783-1134(1475)
18	3VKW_A Replicase large subunit	99.0	1.1E-11	2.4E-16	108.5	11.8	173	41-269	140-331 (446)
19	4NQN_A Replicase polyprotein 1	99.0	3.6E-12	7.5E-17	110.9	7.9	167	38-234	150-321 (423)
20	2L8B_A Protein traI (E.C.3.6.4	99.0	2.3E-11	4.9E-16	94.1	11.4	139	48-228	39-183 (189)
21	3JB9_X Pre-mRNA-splicing facto	98.9	1.7E-11	3.6E-16	120.5	10.6	172	39-234	726-1051(1284)
22	2PJR_F HELICASE PCRA (E.C.3.6.	98.8	9E-11	1.9E-15	105.4	10.8	175	29-232	1-289 (548)
23	3LFU_A DNA helicase II (E.C.3.	98.8	1.3E-10	2.8E-15	105.9	10.9	192	33-259	3-305 (647)
24	4C2U_A DNA HELICASE II (E.C.3.	98.7	9.3E-10	2E-14	100.7	12.8	168	33-232	6-293 (665)

Fig.2 HHpred output displaying probable similarities of two HPs: A) LOC105974023 and B) LOC109219975

Three-dimensional structure

PS^2 was used to determine the three-dimensional structure of the *HPs* (Fig. 3). The server used a *2zd1A* template to model these proteins (Fig.4).

Functional analysis

The results of the annotated function of *HPs* also showed similarities with the *RNase H-like* domain (in *reverse transcriptase*) and helicase enzyme for *HPs* which confirms the previous findings.

PlantPRes analysis

Based on the results of PlantPReS, a list of 84 and 87 identical proteins was found for *LOC105974023* and *LOC109219975 HPs*, respectively. They were related to different stress responses. The first two proteins have the highest similarity as summarized in Table.3. Totally, 10 proteins were related to salinity stress.

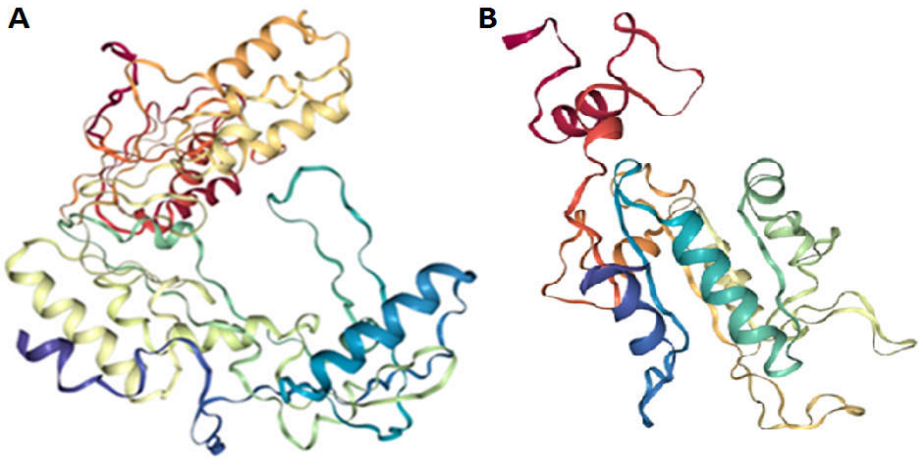

Fig. 3. Three-dimensional structures of *HPs* by PS^2 : A) LOC105974023; B) LOC109219975

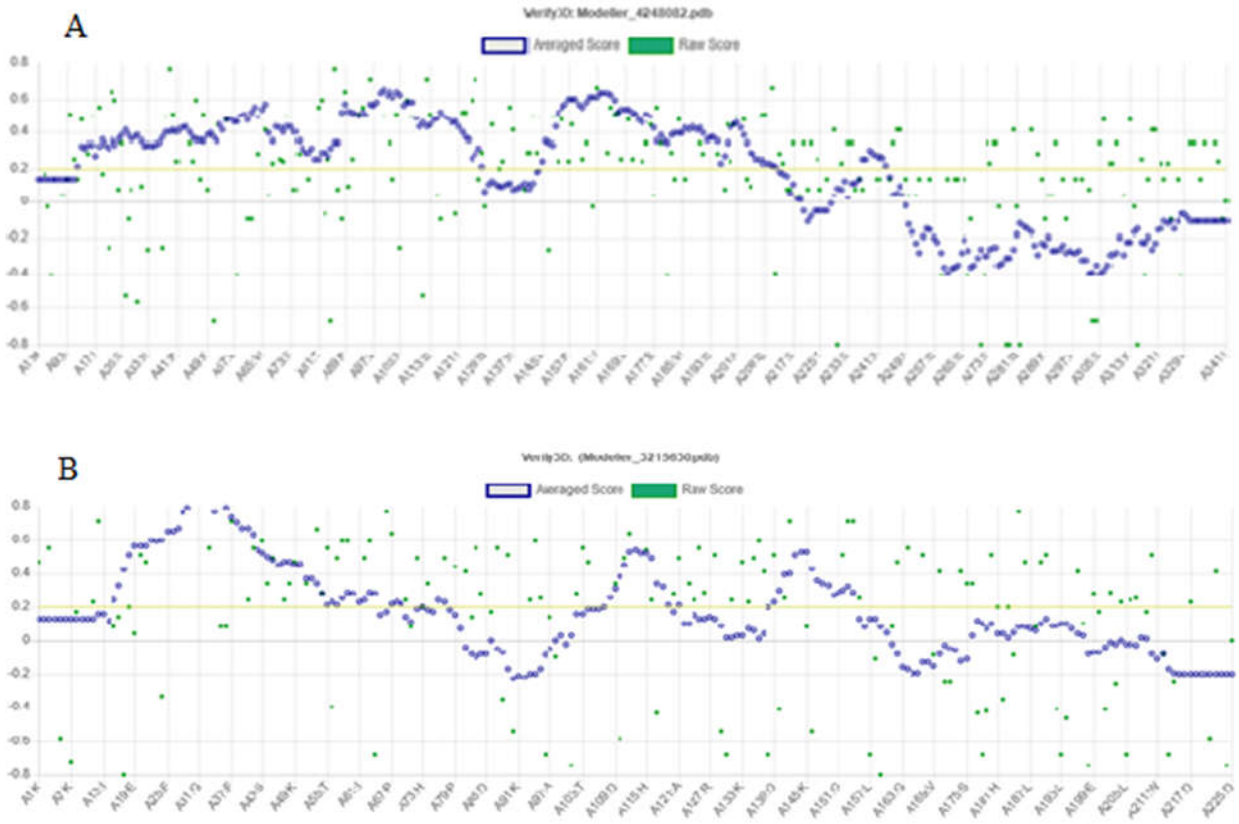

Fig.4. The model quality obtained by *VERIFY3D*: A) LOC105974023; B) LOC109219975

Table 2. The output of the *blastp* searches for similar sequences against non-redundant *UniProtKB/SwissProt* sequences.

Hypothetical proteins	Protein	Organism	Identity	Similarity	Score (bits)	e-value
	Reverse transcriptase	Retrovirus(from transposon opus)	67/222(30%)	109/222(49%)	93.6	1.00E-18
	Reverse transcriptase	Retrovirus (from transposon 297)	59/194(30%)	95/194(48%)	85.9	3.00E-16
	Reverse transcriptase	Retrovirus (from transposon 412)	54/193(28%)	82/193(42%)	77	2.00E-13
LOC105974023	Gypsy retrotransposon integrase-like protein 1	Retrovirus (from transposon 412)	50/174(29%)	79/174(45%)	69.7	2.00E-11
	Retrotransposable element Tf2	Retrovirus	56/229(24%)	106/229(46%)	67	3.00E-10
	ATP-dependent DNA helicase PIF1-like	Nicotiana attenuata	249/288(86%)	268/288(93%)	525	0
	ATP-dependent DNA helicase PIF1-like	Nicotiana sylvestris	247/288(86%)	262/288(90%)	508	4.00E-176
LOC109219975	ATP-dependent DNA helicase PIF1-like	Nicotiana tabacum	218/288(76%)	248/288(86%)	456	1.00E-156
	ATP-dependent DNA helicase pif1-like	Nicotiana tomentosiformis	224/288(78%)	252/288(87%)	441	1.00E-150
	ATP-dependent DNA helicase pif1-like	Capsicum annuum	202/288(70%)	239/288(82%)	420	1.00E-145

Table 3. The most identical proteins to our target proteins in response to abiotic stress.

LOC105974203								
Protein ID	Protein name	Percent identity	Spot Number	Protein source organism	Tissue	Organelle	Stres S	Expression
Q7XTU6	OSJNBb0034I13.10 protein	46.67%	5	Rice (Oryza sativa)	Root	Cell	S	Upregulated
A5CBB4	Putative uncharacterized protein	44.48%	14	Thellungiella (Thellungiella)	Leaf	Cell	S	Upregulated
LOC109219975								
Q10R18	Dehydration-responsive element-binding protein 2E	44.82%	OsNS-121	Rice (Oryza sativa)	shoot	Nucleus	D	Upregulated
Q0WQJ7	ATP binding microtubule motor family protein	42%	H01	Soybean	Hypocotyl	Cell	S	Upregulated

S= Salinity; D=Drought

Discussion

It is the first study that we analyzed the properties of the *UCB-1* rootstock *pistachio HPs*. Based on the achieved results, the *RNase_HI_like* and *N-terminal domain of the DEAD-box helicase* superfamily were recognized as domains hit of these *HPs* by PDB and UniProtKB/SwissProt databases. *RNase H*, which is involved in *DNA* repair, transcription, and replication, is an endonuclease that splits the *RNA* strand of an *RNA/DNA* hybrid in a sequence non-specific manner. The elementary analysis revealed that one of the *HPs* has a high similarity with reverse transcriptase enzyme in different retrotransposons of retroviruses. The upregulation of this protein in *UCB-1* rootstock *pistachio* under salinity stress has been reported in our previous study (Jamshidi Goharrizi *et al.,* 2019).

To investigate the functional analysis, conserved domains were observed because conserved domains are functional units within a protein that play as building blocks in molecular evolution and recombine in different arrangements to create proteins with various functions. The information is then applied for putative functional annotation of protein query sequences according to matches to specific super-family history, identification of proteins with a similar domain. As it was presented in the results, one of our uncharacterized proteins shows a similarity with the *RNase H* domain and integrase protein that is as transposon machinery elements (Majorek *et al.,* 2014).

According to our results, the other *HP* showed a similarity with the helicase enzyme. *DNA* and RNA helicases are enzymes unwinding duplexes and play significant roles in almost transactions of all nucleic acids (Sanan-Mishra *et al.,* 2005). Previous studies reported overexpression of *pea DNA helicase 45 (PDH45)* gene in *tobacco* (Amin *et al.,* 2012) and *rice (Oryza sativa L, CV: IR64)*(Sahoo *et al.,* 2012) which confers tolerance to salinity stress (Sahoo *et al.,* 2012). *DEAD-box helicases* are members of *SF2* superfamily which have a conserved *helicase* core (Ranji *et al.,* 2011). The name of *DEAD-box helicase* is because of their conserved motif containing amino acid sequences of *Asp-Glu-Ala-Asp (D-E-A-D)*(Linder and Jankowsky 2011). The *DEAD-box helicases* have many roles in several aspects of metabolism and in different biological processes in plants like multiple abiotic stress response regulation (Baruah *et al.,* 2017). They are determined to have roles in the activation of abiotic stress tolerance participants in plants (Barak *et al.,* 2014; Liu *et al.,* 2016). Because stress decreases the protein synthesis by the cellular gene expression machinery affecting; therefore, it is clear that molecules involved in the processing of the nucleic acid such as translation factors/helicases are likely to be influenced (Sahoo *et al.,* 2012).

We analyzed the physicochemical properties of these *HPs* of *UCB-1* for the first time. In Table 1 the physicochemical properties of *HPs* are tabulated. The determined Pi will be useful because solubility is minimum and in an electro focusing system mobility is zero at pI. As well as proteins become stable and compact at isoelectric pH, in this regard, computed pI will be beneficial for improving a buffer system for purification by isoelectric focusing method. Moreover, the quantitative study of protein-protein and protein-ligand interactions in the solution can be performed by using computed extinction coefficients. The instability index shows the approximate stability of proteins in a test tube. The proteins with a very high Aliphatic index may indicate stability in a wide temperature range where lower Aliphatic index proteins are not thermally stable and have more flexibility. Besides these indexes, the GRAVY value for a protein is computed by adding the values of hydropathy of all the amino acids and dividing it by the number of residues in the sequence (Ikai 1980).

The *PPIs* are important to perform almost all of the cellular functions. Mostly, proteins mutually interact with another one in a dependent way to carry out a common function (Islam *et al.,* 2015). It is possible to predict the function of a protein according to its interaction with other proteins (Laurie and Jackson 2005). It is very unusual that proteins bring out function with any interactions with other biomolecules. In this regard, in this post-genomic era, *PPI* databases have become the most significant resource for searching biological networks and pathways in cells (Laurie and Jackson 2005). Therefore, a

predicted interaction with *Reverse Transcriptase* enzyme and *helicase* enzymes can provide the *HPs* potential as a part of Transposon elements machinery as well as *PIF1 helicase*. Pif1 family helicases are discovered effectively in all *eukaryotes*, most of the *eubacteria*, and some *archaea bacteria* (Bochman *et al.,* 2010). Whereas these enzymes' functions are different within and between organisms, it is very clear that *Pif1* family *helicases* are vital for the maintenance of both nuclear and mitochondrial genome for plant' survival during stress conditions (Bochman *et al.,* 2010). Moreover, the PS2 server was used to determine the three-dimensional structure of the *HPs*. This server used templates to model those proteins.

In addition, in this study, a similarity with some stress-responsive upregulated genes was determined using *PlantPReS* website. For *LOC105974203 HP*, one of these proteins was a nucleic acid-binding protein in rice *(Oryza sativa),* as a part of the *Retrotransposone* complex, involved in salt stress defense (Feng *et al.,* 2002). Also, a Putative uncharacterized protein up-regulated under salinity stress in *Thellungiella (Thellungiella)* showed sequence similarity with this *HP*. Also, several similarities with some stress-responsive genes were observed for *LOC109219975 HP*. The highest similarity was found with *Dehydration-responsive element-binding protein 2E Rice (Oryza sativa)*, which is induced by abiotic stresses and involved in stress responses of plants (Agarwal *et al.,* 2017). The other protein was *ATP binding microtubule motor family protein*, involved in key processes in plant cells induced by stress conditions (Krtková *et al.,* 2016). Similarly, the results of our previous study revealed the upregulation of these proteins under salinity stress conditions that can be related to the *retrotransposon* and *helicase* activation under salt stress conditions (Jamshidi Goharrizi *et al.,* 2019).

Transposon silencing is demonstrated under normal development, while activation of them was found under different biotic and abiotic stresses in plants (Kimura *et al.,* 2001; Wessler 1996). A close relationship between activation of *retrotransposons* and plant defense responses has already been reported (Grandbastien *et al.,* 1997; Grandbastien *et al.,* 1994; Pouteau *et al.,* 1994). Also, *helicases* that are known to express under the effects of different abiotic stresses, playing a significant role in growth stabilizing in plants under stress conditions by some stress-induced pathways regulating (Tuteja *et al.,* 2012). These observations suggest the existence of unknown proteins like retrotransposons and helicase in *UCB-1 pistachio rootstock*. These results may indicate a survival strategy based on the plant ´s biology. By using these elements in hand, researchers could use different strategies to target these elements. Also, they can separate responsible genes toward intended genotypes between plants that are generated from cultured cells of different plant species.

Conclusion

To our knowledge, this is the first *in silico* study of *UCB-1* rootstock *pistachio* uncharacterized proteins' identification. The results of this study showed that most likely, the intended uncharacterized proteins have similarities with *Reverse Transcriptase* and some *Retrotransposon Elements machinery*; as well as, *helicase enzyme* which response to salinity stress. The analysis of this research revealed that these proteins have similarities with proteins that have significant roles in response to salinity stress. We have also assumed that their up-regulation under salinity stress is related to the activation of *retrotransposon* elements and *helicase enzymes*. These observations suggest a novel responsive mechanism to stress conditions. Further studies are required to clarify its promoter and *ORF* construction as well as regulatory responsive elements. These results proposed the *transposon* elements along with the *helicase* enzyme as the producers of stress-induced genetic diversity.

Conflicts of interest

The authors have declared that no competing interests exist.

References

Amirmahani F, Jamshidi Goharrizi K. 2018. Phylogenetic Analysis of Three Long Non-coding RNA Genes: AK082072, AK043754 and AK082467. *J Genet Resour* 4: 56-64.

Barak S, Singh Yadav N, Khan A. 2014. DEAD-box RNA helicases and epigenetic control of

abiotic stress-responsive gene expression. *Plant Signal Behav* 9: e977729.

Barkla Bronwyn J, Vera-Estrella R, Pantoja O. 2013. Progress and challenges for abiotic stress proteomics of crop plants. *Proteom* 13: 1801-1815.

Baruah I, Debbarma J, Boruah H, Keshavaiah C. 2017. The DEAD-box RNA helicases and multiple abiotic stresses in plants: A systematic review of recent advances and challenges. *Plant Omics* 10: 252.

Bharat Siva Varma P, Adimulam YB, Kodukula S. 2015. In silico functional annotation of a hypothetical protein from Staphylococcus aureus. *J Infect Public Heal* 8: 526-532.

Bochman ML, Sabouri N, Zakian VA. 2010. Unwinding the functions of the Pif1 family helicases. *DNA Repair* 9: 237-249.

Eisenberg D, Lüthy R, Bowie J (1997) VERIFY3D: Assessment of protein models with three-dimensional profiles. In: Enzymology BM in, editor. Academic Press.

Epstein L, Beede R, Kaur S, Ferguson L. 2004. Rootstock effects on pistachio trees grown in verticillium dahliae-infested soil. *Phytopath* 94: 388-395.

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y et al. 2002. Sequence and analysis of rice chromosome 4. *Nature* 420: 316.

Ferguson L, Poss J, Grattan S, Grieve C, Wang D, Wilson C et al. 2002. Pistachio rootstocks influence scion growth and ion relations under salinity and boron stress. *J Am Soc Hortic Sci* 127: 194-199.

Galperin MY. 2001. Conserved 'Hypothetical' Proteins: New Hints and New Puzzles. *Comp Funct Genom* 2: 14-18.

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Springer.

Gazi M, Mahmud S, Fahim SM, Kibria MG, Palit P, Islam M et al. 2018. Functional prediction of hypothetical proteins from *Shigella flexneri* and validation of the predicted models by using ROC curve analysis. *Genomics Inform* 16: e26.

Grandbastien M-A, Audeon C, Casacuberta JM, Grappin P, Lucas H, Moreau C et al. 1994. Functional analysis of the tobaccoTnt1 retrotransposon. *Genetica* 93: 181-189.

Grandbastien M-A, Lucas H, Morel J-B, Mhiri C, Vernhettes S, Casacuberta JM. 1997. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. In: Evolution and Impact of Transposable Elements. Springer, 241-252

Ikai A. 1980. Thermostability and aliphatic index of globular proteins. *J Biochem* 88: 1895-1898.

Islam MS, Shahik SM, Sohel M, Patwary NI, Hasan MA. 2015. In silico structural and functional annotation of hypothetical proteins of *Vibrio cholerae* O139. *Genomics Inform* 13: 53.

Jamshidi Goharrizi K, Baghizadeh A, Kalantar M, Fatehi F. 2019. Assessment of changes in some biochemical traits and proteomic profile of UCB-1 pistachio rootstock leaf under salinity stress. *J Plant Growth Regul*. https://doi.org/10.1007/s00344-019-10004-3

Jamshidi KG, Amirmahani F, Salehi F. 2019. Assessment of changes in physiological and biochemical traits in four pistachio rootstocks under drought, salinity and drought+ salinity stresses. *Physiol Plant*. doi: 10.1111/ppl.13042

Kiarash JG, Wilde HD, Amirmahani F, Moemeni MM, Zaboli M, Nazari M et al. 2018. Selection and validation of reference genes for normalization of qRT-PCR gene expression in wheat (*Triticum durum* L.) under drought and salt stresses. *J Ggenet* 97: 1433-1444.

Kimura Y, Tosa Y, Shimada S, Sogo R, Kusaba M, Sunaga T et al. 2001. OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses. *Plant Cell Physiol* 42: 1345-1354.

Krtková J, Benáková M, Schwarzerová K. 2016. Multifunctional microtubule-associated proteins in plants. *Front Plant Sci* 7: 474.

Kumar S, Shanker A. 2018. Bioinformatics Resources for the stress biology of plants. In: Biotic and Abiotic Stress Tolerance in Plants. Springer.

Laurie AT, Jackson RM. 2005. Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinform* 21: 1908-1916.

Lesk A, Chothia C. 1986. The response of protein structures to amino-acid sequence changes. *Philos Trans Royal Society of London Series A, Math Phys Sci* 317: 345-356.

Linder P, Jankowsky E. 2011. From unwinding to clamping the DEAD box RNA helicase family. *Mol Cell Biol* 12: 505.

Liu Y, Tabata D, Imai R. 2016. A cold-inducible DEAD-box RNA helicase from Arabidopsis thaliana regulates plant growth and development under low temperature. *PloS One* 11: e0154040.

Majorek KA, Dunin-Horkawicz S, Steczkiewicz K, Muszewska A, Nowotny M, Ginalski K et al. 2014. The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nuc Acids Res* 42: 4160-4179.

Moazzam Jazi M, Ghadirzadeh Khorzoghi E, Botanga C, Seyedi SM. 2016. Identification of Reference Genes for Quantitative Gene Expression Studies in a Non-Model Tree Pistachio (*Pistacia vera* L.). *PLoS One* 11: e0157467.

Moazzzam Jazi M, Seyedi SM, Ebrahimie E, Ebrahimi M, De Moro G, Botanga C. 2017. A genome-wide transcriptome map of pistachio (*Pistacia vera* L.) provides novel insights into salinity-related genes and marker discovery. *BMC Genomics* 18: 627.

Morgan D, Epstein L, Ferguson L. 1992. Verticillium wilt resistance in pistachio rootstock cultivars: assays and an assessment of two interspecific hybrids. *Plant Disease* 76: 310-313.

Mostek A, Börner A, Badowiec A, Weidner S. 2015. Alterations in root proteome of salt-sensitive and tolerant barley lines under salt stress conditions. *J Plant Physiol* 174: 166-176.

Pouteau S, Grandbastien MA, Boccara M. 1994. Microbial elicitors of plant defence responses activate transcription of a retrotransposon. *Plant J* 5: 535-542.

Ranji A, Shkriabai N, Kvaratskhelia M, Musier-Forsyth K, Boris-Lawrie K. 2011. Features of double-stranded RNA-binding domains of RNA helicase A are necessary for selective recognition and translation of complex mRNAs. *J Biol Chem* 286: 5328-5337.

Sahoo RK, Gill SS, Tuteja N. 2012. Pea DNA helicase 45 promotes salinity stress tolerance in IR64 rice with improved yield. *Plant Signal Behav* 7: 1042-1046.

Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. 2004. The database of interacting proteins: 2004 update. *Nuc Acids Res* 32: D449-D451.

Sanan-Mishra N, Pham XH, Sopory SK, Tuteja N. 2005. Pea DNA helicase 45 overexpression in tobacco confers high salinity tolerance without affecting yield. *Proc Natl Acad Sci* 102: 509-514.

Schwarz D, Rouphael Y, Colla G, Venema JH. 2010. Grafting as a tool to improve tolerance of vegetables to abiotic stresses: Thermal stress, water stress and organic pollutants. *Sci Hortic* 127: 162-171.

Srinivasan B, Kempaiah Nagappa L, Shukla A, Balaram H. 2015. Prediction of substrate specificity and preliminary kinetic characterization of the hypothetical protein PVX-123945 from *Plasmodium vivax*. *Exp Parasitol* 151-152: 56-63.

Tuteja N, Singh S, Tuteja R. 2012. Helicases in improving abiotic stress tolerance in crop plants. In: Improving Crop Resistance to Abiotic Stress, Wiley VCH Verlag GmbH & Co. KGaA.

Varma PBS, Adimulam YB, Kodukula S. 2015. In silico functional annotation of a hypothetical protein from *Staphylococcus aureus*. *J Infect Public Healt* 8: 526-532.

Wessler SR. 1996. Plant retrotransposons: turned on by stress. *Curr Biol* 6: 959-961.