Plant Bulletin Vol 2, (1) 2023: 11-17 ISSN: 2959-3379 (Print), 2959-3387 (Online)

Research Article

Identification of novel targets environmental stress tolerance genes in plants by using bioinformatics

Hameeda Nasar[1], Atia Yousaf[1], Afroz Rais[1], Hina Ali[2], Shagufta Saddozai[2], Asia saeed[1], Abdul Qadir[3*]
[1]*Department of Botany, Sardar bahdur khan Women ,s University, Quetta. Pakistan*
[2]*Department of Zoology, Sardar Bahdur Khan Women,s University, Quetta, Pakistan*
[3]*Productivity Enhancement of Sugarcane, Pakistan Agriculture Research Council, Islamabad, Pakistan*
**Correspondence: abdulqadirjan@gmail.com*

ABSTRACT

The plant show retorts to the abiotic stresses which influence the biosynthesis of secondary metabolites in the plants. The current research identified novel genes in reference plants *Prunus avium*, *Prunus persica*, *Mangifera indica*, and *Carica* papaya, while as a model plant *Arabidopsis thaliana* was used. For this purpose, bioinformatics tools such as ORF, BLAST, and CDD tools were used. The Est of *Prunus avium*, *Prunus persica*, *Mangifera indica*, and *Carica papaya* were 30582, 26413, 38501, 20333 respectively. The genes HOS10, LOS4, MPK3, ELIP1, and DREB1A were analyzed through ORF, compared by BLAST, and CDD was used for searching of conserved domain database with nucleotide query coverage. The query coverage of all the genes in all four plants varies from 28% to 96% and the frame preferred are mostly +3 and +2, +1 is also present but in the very least form. The current research will help to comprehend the query coverage and define the identity of the selected genes in the cold and warm region's plants.

Keywords: Abiotic Stress, Cold Genes, Cold Plant, Warm Plant, Bioinformatics.

INTRODUCTION

Climate and agriculture both are correlated with each other not only morph physiological but also genetically. The biotic and abiotic factors can influence agriculture (1). Due to climatic changes the abiotic factors such as air, water, humidity, and the main factor is temperature. The degree of cold and hot temperature is the most important factor that influences the plants especially on fruits containing plants. It impacts their reproductive and vegetative parts (2). Temperature can be affected the growth of plants. But some plants can tolerate different environmental stress. If plants tolerate cold temperature it is said to be cold acclimation. (3). Because these plants possess cold-induced genes that will help in climate change. (4).

The modern period of genes concept arises questions about self-determining properties of genes and genes as a unit of inheritance. (5). Bioinformatics can create databases and some methods which are used for genome Sequencing. The improvement in bioinformatics occurs from 1990 to 2000. When software tools are developed. (6). The bioinformatics tools implementation in the field of modern biology causes some advantages as well as some challenges (7). The bioinformatics tools provide a lot of information to the scientist for Research and understanding of the cause of biotic and abiotic stresses (8).

The National Centre for Biotechnology information is a very useful tool for sequencing biology. It contents some online resources such as PubMed, databases, gene Bank, taxonomy Browser, CDD, BLAST, and other resources. (9). BLAST is an NCBI tool that is used for checking similarity and dissimilarity between sequences and proteins. In this way, a scientist will be able to scrutinize the revolutionary relationship between new members of the family and their Parents. (10). By the blast tool, the Quarry sequences can be compared. The BLAST has the

Article History
Received: May 10, 2023, Accepted: June 25, 2023,
Published: July 02, 2023

application by which scientists can identify the function of known query sequences by the function of similar sequences (7). The applications of blast programs are blast quick start Mini-course (11). The development is continuously occurring in blast databases such as strategy files which can be saved and reused at the time of need (12). CDD is a database by which commentary of sequences with concert protein document footprints location and function site as well as concept concluded from the footprint. The document model of the latest CDD release has 57000 of which 15000 were curated by the staff of CDD. These curations were started about 20 years ago and development is continuously occurring in them. (13).

Arabidopsis thaliana can be surviving in every kind of habitat. There are 1001 genome projects which are sequence. The first face of this project have 80 stains which were based on Population scale sequences. Due to various mutations, the present plant of *Arabidopsis thaliana* has diverged from closest 10 million years ago when it was seen in the laboratory. (14) To analyses the cold stress response in *Arabidopsis thaliana* (15). The purpose of this recent bioinformatics study is to search those genes which are involved in environmental conditions such as cold and hot temperatures. For this purpose, choose two plants from cold climate *Prunus avium L, Prunus persica L,* and two from warm *Mangifera indica L, Carica papaya L* These selected plants have their self-sustainability (16).

MATERIALS AND METHODS

The cold stress genes were collected and investigated by various bioinformatics tools such as ORF, BLAST, and CDD. For the investigation, only five genes were selected from the model plant. Which were HOS10, LOS4, MPK3, ELIP1, and DREB1A. These genes were investigated in four plants, out of the two were from cold regions (*Prunus avium L., Prunus persica L)* and two were from warm regions *(Mangifera indica L, Carica papay L).*

Identification of Genes.

Various cold stress genes were studied and selected by different websites and research papers and selected only genes high query covers. 1. http://affymetrix.arabidopsis.info/.2.http://www.arabidopsis.org/info/expression/atgenexpress.jsp. 3.https://www.ncbi.nlm.nih.gov/. The above links provided a lot of cold-related genes out of them selected only five genes (HOS10, LOS4, MPK3, ELIP1, DREB1A) with highest query covers. (17).

Feching the sequences of the selected model genes and their (Open reading frame) ORF.

The sequences of model plants were fetched by NCBI (National Center for Biotechnology and Information) and saved in nucleotide form. The ORF finder and inter the FASTA sequence of identified genes of reference plants (*Arabidopsis thaliana)* and submitted it. After obtaining the results protein sequences were changed into nucleotide sequences. And saved the nucleotide sequences in a file. (18)

Homology search

BLAST is a highly used tool for homology search. Through NCBI EST of every reference, plant was checked to drive the plant name in NCBI. The ESTs of P. *avium L., P. persica, M. indica, C. papay* are 30582, respectively. The FASTA of ORF is subjected to blast then nominated one possibility out of EST, GSS, and NR and in the option of organism tapped the name of reference. The same procedure was applied for all four plants P. *avium, P. persica, M. indica, C. papay*. Nominated only those options which having maximum query covers and identity. (17)

Conserve domain database.

The elucidation of protein function is ended through the conserve domain database. Through these local multiple sequences, the domains of a model plant (*A. thaliana*) and reference plant were compared. Through the conserved domain database, the fasta of the model plant were investigated to check the family and after that, the reference plant's fasta were added to check whether both plants have the same family or not. (17)

RESULTS

Plants in a Cold Climate

To check cold stress-responsive genes in those plants which having cold habitats. Out of millions of cold plants only selected two plants P. *persica and P. avium L* because these are the nutritious and economically beneficial fruits of Pakistan. These plants can tolerate cold conditions which showed that there will be some special types of genes that are helpful for their resistances. To sort out the resistance genes by using NCBI and other websites which were ORF, BLAST, and CDD (Table. 1)

Table 1. homology search for cold stress genes of *P. avium.* and *P. persica* .with *A thaliana.*

Blast of *Prunus persica*				
S.NO	**Gene ID**	**Query cover**	**Accession NO**	**Identity**
1	HOS10	61%	XM_007222892.2	76.02%
2	LOS4	80%	XM020568628.1	68.59%
3	MPK3	96%	XM007205325.2	76.54%
4	ELIP1	63%	XM_007225802.2	71.96%
5	DREB1A	58%	XM007218250.2	71.01%
Blast of *Prunus avium*				
1	HOS10	61%	HE643809.1	73.77%
2	LOS4	80%	XM021955263.1	68.23%
3	MPK3	96%	XM021953490.1	76.72%
4	ELIP1	63%	XM021953490.1	72.44%
5	DREB1A	58%	XM021971179.1	72.92%

Open reading frame topographies

All the frame of cold stress-responsive genes was preferred approximately because all the frame was positive. The model plant *A. thaliana* was showed +3 frame three times while the reference plant showed a two times frame +3. The genes HOS10 and ELIP1 preferred similar frames (+1) and (+3) respectively. For the ORF results the features of reference plant P. *avium* were compared with the features of selected model plant *A. thaliana* the frame of all the cold stress-responsive genes was preferred likewise the frame of P. *persica*. Both model plant and reference plant have one +1, one +2, and three +3 frame but genes first (HOS10) and fourth (ELIP1) have a similar frame (Table. 2)

Table 2. Open reading frame topographies for cold stress-responsive genes in *P. persica* and *P. persica*, compared with *A. thaliana*.

S.no	**Gene ID**	**Gene length**			**Frame preferred**			**Frame length (amino acid)**		
		Prunus persica	***Prunus avium***	***Arabidopsis thaliana***	***Prunus persica***	***Prunus avium l.***	***Arabidopsis thaliana***	***Prunus persica***	***Prunus avium .***	***Arabidopsis thaliana***
1	HOS0	948	888	639	+1	+1	+1	315	295	212
2	LOS4	1542	1491	615	+1	+3	+3	513	496	204
3	MPK3	1113	2133	411	+2	+2	+3	370	710	136
4	ELIP1	588	1113	597	+3	+3	+3	195	370	198
5	DREB1A	651	588	597	+3	+3	+2	216	195	198

Conserve Domain Database

The Conserve domain database CDD search of model plant *A. thaliana* and reference plants P. *avium and P. persica* presented the same family of conserved domains. For CDD both plants were compared with model plants one by one. Approximately all genes showed super families except one gene which was HOS10 (Table. 3)

Table. 3. Conserve domain database identification and association of cold stress-responsive genes in *P. avium* and *P. persica*.

S.no	**Gene ID**	**Families of conserved domain**	***Prunus avium***	***Prunus persica***
1	HOS10	PLNO3091	Present	Present
2	LOS4	DEAD_like helicase super family	Present	Present
3	MPK3	PKc_like super family	Present	Present
4	ELIP1	Chloroa_b_bind super family	Present	Present
5	DREB1A	APz2 super family	Present	Present

Plants in the warm climate

After collecting the results of plants of cold regions two more plants were selected which belonged from warm areas. To identify the warm stress-responsive genes *M. indica* and *P. papaya* which belong from warm areas and also determined the accession number of the selected genes. Here all the features of ORF, quarry covers, and families of CDD will be determined by Bioinformatics tools. (Table. 4)

Table 4. Homology search for cold stress genes of *M. indica* and *C. papaya* with *A. thaliana.*

Blast of *Mangifera indica*				
S.NO	**Gene ID**	**Quary cover**	**Accession NO**	**Identity**
1	HOS10	61	XM_044651845.1	77.72%
2	LOS4	79%	XM_044652740.1	72.06%
3	MPK3	96%	XM_044626422.1	76.99%
4	ELIP1	57%	XM_04464444.1	71.68%
5	DREB1A	58%	XM_044649098.1	76.76%
Blast of *Carica papay*				
1	HOS10	62%	XM_022044626.1	76.57%
2	LOS4	77%	XM_022053051.1	74.40%
3	MPK3	96%	XM_022034066.1	79.55%
4	ELIP1	63%	XM_022043856.1	72.65%
5	DREB1A	28%	XM_022053063.1	73.26%

Open reading frame topographies

The features of ORF for *M. indica* were frame preferred and gene and amino acid length. The frame of all genes was preferred approximately. The selected model plant *A. thaliana* have three +3 frame while the reference plant shows only one +3 frame. The last two genes show frame similarities, which were +3 and +2 respectively**.** The same process of ORF was applied for plant *C. papaya* to collect the ORF features. Here three genes show the same frame preferred. The frame of model plant was mostly +3 (Table. 5).

Table 5. Open reading frame topographies for cold stress-responsive genes in *M. indica* and *C. papay*, compared with *A. thaliana* L**.**

s.no	**Gene ID**	**Gene length**			**Frame preferred**			**Frame length (amino acid)**		
		Mangifera indica	***Carica papay***	***Arabidopsis thaliana***	***Mangifera indica***	***Carica papay***	***Arabidopsis thaliana***	***Mangifera indica***	***Carica papay***	***Arabidopsis thaliana***
1	HOS10	801	861	639	+2	+3	+1	268	286	212
2	LOS4	1455	1491	615	+2	+3	+3	484	496	204
3	MPK3	1128	1113	411	+1	+1	+3	375	370	136
4	ELIP1	519	273	597	+3	+3	+3	172	90	198
5	DREB1A	708	629	597	+2	+3	+2	235	212	198

Conserve Domain Database

The CDD search showed that only four genes have the same families. The second gene LOS4 showed that DEAD_ like helicase superfamily is present in *M. indica* and absent in *C. papaya*. (Table. 6).

Table 6. Conserve domain database identification and association of cold stress-responsive genes in *C. papaya and M. indica*.

S.no	**Gene ID**	**Families of conserved domain**	***Carica papaya***	***Mangifera indica***
1	HOS10	PLNO3091	Present	Present
2	LOS4	DEAD_like helicase super family	Absent	Present
3	MPK3	PKc_like super family	Present	Present
4	ELIP1	Chloroa_b_bindsuper family	Present	Present
5	DREB1A	AP2 super family	Present	Present

DISCUSSION

This study aimed to identify cold stress genes in plants belonging to cold and hot regions. For this purpose, selected four plants were having geographical distributions according to cold acclimation some plants which had no tolerance against cold stress they maybe face some physical injuries so to avoid this type of defect plants having some cold-responsive genes which can tolerate during cold conditions but may be present in hot weather in recessive form (20). Bioinformatics provided information about the genome and some tools for genomic DNA replication such as comparative, genomics operations, sequence alignment, etc. (19,2).

The homology search expressed that both plants from cold stress show the highest and same query coverage (61%, 80%, 96%, 63%, 56%) but the identities were different in some points. the query coverage of hot regions plants was less than cold regions plants. It was surprising that the gene MPK3 have the highest and same query coverage (96%) in all the four plants either from cold or from warm regions in all the four plants genes HOS10, LOS4, MPK3, ELIP1, DR EB1 A showed the same families accept *C. papaya* in which dead like helicase superfamily was absent. ORF was a very helpful tool it translates the nucleotide sequence into protein sequence and gives detail about gene length amino acid length and frame preferred. In the cold-responsive plant *P. avium* MPK3 show the highest gene length and amino acid length with positive frame length. In the warm responsive plant, *C. papay* LOS 4 shows the highest gene length and amino acid length with positive frame preferring.

CONCLUSION

In conclusion, this research highlights the dynamic responses of plants to abiotic stresses, shedding light on the intricate mechanisms influencing the biosynthesis of secondary metabolites. Through the utilization of bioinformatics tools, novel genes have been identified in *Prunus avium*, *Prunus persica*, *Mangifera indica*, and *Carica papaya*, with *Arabidopsis thaliana* serving as a model plant. The analysis of genes such as HOS10, LOS4, MPK3, ELIP1, and DREB1A using ORF, BLAST, and CDD tools revealed varying query coverage and frame preferences across different plant species. These findings contribute to a deeper understanding of gene expression patterns in plants inhabiting diverse environmental conditions, particularly in cold and warm regions, laying the groundwork for further research in plant stress responses and secondary metabolite biosynthesis.

FUTURE RECOMMENDATIONS

To propel the field forward, future research endeavors should prioritize functional characterization of the identified genes, particularly those displaying significant sequence conservation across diverse plant species. Experimental validation through gene knockout or overexpression assays would elucidate the specific roles of these genes in stress adaptation and secondary metabolite biosynthesis. Complementary transcriptomic and proteomic analyses could provide a deeper understanding of the dynamic regulatory networks governing gene expression and protein abundance in response to abiotic stresses. Moreover, leveraging this knowledge for targeted genetic engineering of crop plants holds promise for enhancing stress tolerance and improving agricultural productivity. By exploring natural genetic variation in stress-responsive genes and conducting field trials to validate the performance of genetically engineered cultivars under real-world conditions, researchers can contribute to the development of resilient crop varieties tailored to thrive in changing environmental landscapes

REFERENCES

1) Silva, J. D., Rashid, Z., Nhut, D. T., Sivakumar, D., Gera, A., Souza, M. T., & Tennant, P. (2007). *Papaya* (*Carica papaya* L.) biology and biotechnology. Tree and Forestry Science and Biotechnology, 1(1), 47-73.

2) Zhang, C., & Whiting, M. (2012). The occurrence of the protruding pistil in sweet cherry (*Prunus avium* L.) and its consequence on fertilization. Scientia horticulture, 140, (2)149-156.

3) Miura, K., & Furumoto, T. (2013). Cold signaling and cold response in plants. International journal of molecular sciences, 14(3), 5312-5337.

4) Thomashow, M. F. (1999). Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annual review of plant biology, 50(1), 571-599.

5) Portin, P., & Wilkins, A. (2017). The evolving definition of the term “gene”. Genetics, 205(4), 1353-1364.

6) Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. Briefings in bioinformatics, 20(6), 1981-1996.

7) Stover, N. A., & Cavalcanti, A. R. (2017). Using NCBI BLAST. Current Protocols Essential Laboratory Techniques, 14(1), 11-1.

8) Ahmad, T., & Al-Senaidy, A. M. (2013). Software Tools in Bioinformatics: A Survey on the Importance and Issues Faced in Implementation.

9) Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., & Yaschenko, E. (2007). Database resources of the national center for biotechnology information. Nucleic acids research, 36(1), 13-21.

10) Madden, T. (2013). The BLAST sequence analysis tool. The NCBI handbook,2, 425-436

11) Wheeler M, D., & Bhagwat, M. (2007). BLAST quickstart. In Comparative genomics (pp. 149-175). Humana Press.adden, T. (2003). The BLAST sequence analysis tool. The NCBI handbook.

12) Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. BMC bioinformatics, 10(1), 1-9.

13) Yang, M., Derbyshire, M. K., Yamashita, R. A., & Marchler-Bauer, A. (2020). NCBIs conserved domain database and tools for protein domain analysis. Current protocols in bioinformatics, 69 (1).

14) Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., & Weigel, D. (2011). Whole-genome sequencing of multiple

Arabidopsis thaliana populations. Nature genetics, 43(10), 956-963.

15) Amme, S., Matros, A., Schlesier, B., & Mock, H. P. (2006). Proteome analysis of cold stress response in *Arabidopsis thaliana* using DIGE-technology. Journal of experimental botany, 57(7), 1537-1546.

16) Budak, N. H. (2017). Bioactive components of Prunus avium L. black gold (red cherry) and *Prunus avium* L. stark gold (white cherry) juices, wines and vinegars. Journal of food science and technology, 54(1), 62-70.

17) Sadia, B., Baloch, I. A., Bazai, Z. A., Noor, W., & Qambrani, S. S. (2020). 15. Computational identification and annotation of salt stress responsive genes in *Solanum melongena* L. Pure and Applied Biology (PAB), 9(1), 132-139.

18) Karp, P. D., Paley, S., & Romero, P. (2002). The pathway tools software. Bioinformatics, 18(1), 225-232.

19) Karp, P. D., Latendresse, M., Paley, S. M., Krummenacker, M., Ong, Q. D., Billington, R., & Caspi, R. (2016). Pathway Tools version 19.0 update:

20) Knight, M. R., & Knight, H. (2012). Low-temperature perception leading to gene expression and cold tolerance in higher plants. New Phytologist, 195(4), 737-751.