

**Genome Wide Analysis of Ethylene Responsive Factors
(ERF's) associated with Stress Responses in Wheat
(*Triticum aestivum*)**



Hareem Aamer

00000273464

**DEPARTMENT OF PLANT BIOTECHNOLOGY
ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES
NATIONAL UNIVERSITY OF SCIENCES AND
TECHNOLOGY
ISLAMABAD, PAKISTAN.**

2020

**Genome Wide Analysis of Ethylene Responsive Factors
(ERF's) associated with Stress Responses in Wheat
(*Triticum aestivum*)**



Hareem Aamer

00000273464

Supervisor

Dr. Alvina Gul

Assistant Professor

**DEPARTMENT OF PLANT BIOTECHNOLOGY
ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES
NATIONAL UNIVERSITY OF SCIENCES AND
TECHNOLOGY
ISLAMABAD, PAKISTAN.**

2020

**Genome Wide Analysis of Ethylene Responsive Factors
(ERF's) associated with Stress Responses in Wheat
(*Triticum aestivum*)**

By

Hareem Aamer

00000273464

A thesis submitted in partial fulfillment of the requirements for the
degree of Master of Science in Plant Biotechnology

Thesis Supervisor:

Dr. Alvina Gul

Assistant Professor

**DEPARTMENT OF PLANT BIOTECHNOLOGY
ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES
NATIONAL UNIVERSITY OF SCIENCES AND
TECHNOLOGY
ISLAMABAD, PAKISTAN.**

2020

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by:
(Student Name & Regn No.) **Hareem Aamer (00000273464)**. Titled: **Genome-wide analysis of Ethylene Resonsive Factors associated with stress responses in wheat (*Triticum aestivum*)** be accepted in partial fulfillment of the requirements for the award of **MS in Plant Biotechnology** degree with (_____ Grade).

Examination Committee Members

1. Name: **Dr. Muhammad Qasim Hayat** Signature: _____

2. Name: **Dr. Rabia Amir** Signature: _____

3. Name: **Dr. Marriam Bakhtiar (Co-supervisor)** Signature: _____

Supervisor's name: **Dr. Alvina Gul** Signature: _____

Date: _____

Head of Department

Date

COUNTERSIGNED

Date: _____

Dean/Principal

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Ms. HAREEM AAMER, (Registration No. 00000273464), of Atta-ur-Rahman School of Applied Biosciences, NUST (School/College/Institute) has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: **Dr. Alvina Gul**

Date:

Signature (HOD): _____

Date: _____

Signature (Dean/Principal): _____

Date: _____

Certificate for Plagiarism

It is certified that MS Thesis Titled “**Genome-wide analysis of Ethylene Responsive Factors associated with stress responses in wheat (*Triticum aestivum*)**” of **Ms. Hareem Aamer** Reg No. 00000273464 has been examined by me. I undertake that:

1. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
2. The work presented is original and own work of the author (i.e. there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
3. There is no fabrication of data or results which have been compiled / analyzed.
4. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
5. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor

Signature:_____

Copyright Statement

- Copyright in text of this thesis rests with the student Hareem Aamer. Copies (by any process) both in full or of extracts, may be made only in accordance with instructions given by the Hareem Aamer (author); and lodged in the Library of Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST. Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.
- The ownership of any intellectual property rights which may be described in this thesis is vested in Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the ASAB, which will prescribe the terms and conditions of any such agreement.
- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad.

Dedication

This Thesis is dedicated to my beloved parents for their affection, countless prayers, encouragement and continuous support.

Acknowledgment

I am blessed by Whom, who has created all that exists, who helped me to get through all the way, gave me the potential to do so gave me strength and guidance the one and only Allah Almighty. My parents and friends are my inspiration who never let me give up at any cost and support me to their fullest.

I am thankful to my supervisor, Dr. Alvina Gul who has supported me throughout the thesis and gave me motivation to do this work. She has enlightened me at every step and built confidence in me. I am grateful to my Co-supervisor Dr. Marriam Bakhtiar who have helped me grow as a person and has given me immense love and support she has been my inspiration I am thankful to her. It is an honor that I am a part of NUST. It has provided me with the opportunities to groom my personality. ASAB has been my home since 2018. I can never forget qualitative time that I spent here. I am obliged to faculty of ASAB for their precious mentoring throughout my journey in ASAB. I would like to extend my gratitude for Principal ASAB, Dr. Hussnain A. Janjua, and Head of department for their endless back end efforts. My guidance and examination committee including Dr. Muhammad Qasim Hayat, Dr. Rabia Amir and Dr. Marriam Bakhtiar who have guided me at each step.

I have started my journey of MS Plant Biotechnology with amazing bunch of people that can never be ignored. Lab is a second home for a researcher, and you become lucky if you find incredible people around you. I am blessed to have my friends Bareera chaudry and Hadiqa Nadeem Qureshi both have been with me in all ups and downs. I am grateful to them for their love and support. I am thankful to my lab colleagues Iqra Rauf, Ayesha fazal, Maria Gillani, Abu bakar Rana, Zeeshan khan, Affan Ahmed, Nitasha Fatima, Nida Fatima Ali, and Ammar Ashar who supported me and made this thesis less hectic by their supportive nature. They are my friends who made my time precious and unforgettable at NUST. I would like to extend my appreciation to Dr. Ijaz Rehman who helped me in tough times as my therapist. In the End I would like to thank my family especially my parents who encouraged and supported me to peruse my MS degree. I cannot forget the prayers of my mother and father.

HAREEM AAMER

Table of Contents

CHAPTER 1: Introduction	1
1.1 Wheat (<i>Triticum aestivum</i>): An overview	1
1.2 Harvesting and growing conditions	1
1.3 Types of Wheat and other related cereals	2
1.4 Wheat classification and evolution	3
1.5 Wheat uses and products	4
1.6 World production of Wheat	5
1.7 Wheat genome	6
1.8 Biotic and Abiotic factors affecting wheat	6
1.9 Wheat production in Pakistan	6
1.10 Wheat productivity constraints and low productivity in Pakistan	7
1.11 Wheat production in future	8
1.12 Wheat production in future	8
1.13 Regulatory functions of plant TF's in response to biotic and abiotic factors	9
1.14 Ethylene Responsive Factors (ERF) and wheat	11
1.15 Network analysis	13
1.16 Protein-Protein interaction networks	14
1.17 Hub proteins	14
1.18 Significance of studies	15
1.19 Aims and Objectives	15
CHAPTER 2: Literature Review	17
2.1 Studies and literature on Wheat	17
2.2 Studies on wheat Transcription Factors (ERF's)	17
2.3 Network analysis	18
2.4 Topological analysis	19
2.5 Databases and information resources	20
2.6 Future studies and research	21
CHAPTER 3: Materials and Methods	22
3.1 Identification of TaERF proteins by Bioinformatic analysis	22
3.2 Phylogenetic-domain-MSA analysis ERFs	22
3.3 ERF proteins motifs analysis and GO analysis	22

3.4 Gene structure and sub-cellular localization of TaERF genes	23
3.5 Scanning protein-protein interactions.....	23
3.6 Construction of PPIs network & withdraw component from network.....	23
3.7 Topological graph, theory analysis of protein network and linkages.....	23
3.8 Backbone network formation through high BC values	23
3.9 Sub network consisting of Hub proteins or Center points	24
3.10 Tools and databases	24
3.11 Methodology	28
CHAPTER 4: Results and Discussion	29
4.1 Identification of TaERF proteins by Bioinformatic analysis	29
4.2 MSA.....	30
4.3 Phylogenetic relationship of ERF's of Triticum aestivum	31
4.4 ERF protein motif analysis.....	42
4.5 Gene structure analysis of Triticum aestivum ERF	43
4.6 Subcellular localization of Triticum aestivum ERF	44
4.7 Gene ontology analysis	45
4.8 Scanning protein-protein interactions	45
4.9 PPIs network construction and withdraw component from the network.....	48
4.10 Topological graph theory analysis of protein network and linkages.....	49
4.11 Backbone Network Construction	51
4.12 Sub Network Construction containing the Hub network	51
4.13 Discussion and Conclusion.....	53
CHAPTER 5: References	139
Appendices	55

List of Figures

Figure 1.1: Model of the signaling pathways and transcription factors in stresses.....	10
Figure 1.2: General characteristics of transcription factor families	10
Figure 1.3: Illustration of ERF family Domains	11
Figure 1.4: Model of the ethylene (ET) signaling pathway to ERFs	12
Figure 3.1: Illustrating methodology of research.....	28
Figure 4.1: Multiple sequence alignment view in CLC bio software package 20.0.3.....	30
Figure 4.2: A rooted tree using Clustral W	32
Figure 4.3: Conserved Motif analysis.....	43
Figure 4.4: Representation of localized proteins in different organelles.....	44
Figure 4.5: Representation of GO analysis.....	45
Figure 4.6: Representation of proteins-protein interaction of wheat ERF using string.....	47
Figure 4.7: Protein-protein interaction network.....	49
Figure 4.8: Topological analysis of network.....	51
Figure 4.9: Sub network containing the hub network	52

List of Tables

Table 1.1 Major and minor cereal crops produced around the world.....	2
Table 1.2 Different species of wheat.....	3
Table 1.3 The classification of <i>Triticum aestivum</i>	4
Table 1.4 Wheat production in tonnes.....	5
Table 2.1 Ethylene responsive transcription factors reported in wheat.....	55
Table 2.2 Information resources, sites and databases.....	20
Table 3.1 Tools and databases.....	24
Table 4.1 ERF proteins in different plant species closely related to wheat.....	29
Table 4.2 Ethylene responsive factors & predicted characteristic in wheat.....	62
Table 4.3 ERF proteins with number of introns and exons.....	69
Table 4.4 Predicted sub cellular localizations by Wolfpsort and Plant-mPLOC.....	76
Table 4.5 Predicted Gene ontology by Blast 2 Go software.....	84
Table 4.6 Topological analysis of protein interaction network	113

List of Abbreviations

Ta	<i>Triticum aestivum</i>
ERF	Ethylene Responsive factor
AP2	Apetala 2 domain
B3	DNA binding domain
MAPK	Mitogen activated protein kinases
CDPK	Calcium dependent protein kinases
TaPIE1	Pathogen-Induced ethylene responsive factor 1
PPI	Protein protein interaction
PPIN	Protein protein interaction network
GO	Gene ontology
k	Connectivity degree
BC	Betweenness centrality
et al.	et alia
CC	Closeness centrality
FAO	Food and Agriculture Organization
aa	Amino acid
pI	Isoelectric point
MSA	Multiple Sequence Alignment
NJ	Neighbor joining
P-value	Probability value

Abstract

Wheat is the largest crop that is grown among all major and important crops across globe. Wheat is the oldest edible crop that is domesticated and has been the requisite of civilizations such as Europe, West Asia and many other regions from 8000 years. Of the thousands of types known, the foremost vital is common wheat (*Triticum aestivum*). Bread wheat is a hexaploid specie, with a genome of 17 Giga bites, consisting of linked but separately sustained genomes. Plants have developed remarkable receptive approaches to different stress circumstances, and at the molecular level a cascade of interconnected networks is regulated by signal cascades. Stress responses include different stages such as signal recognition, signal transduction, and gene expression of response related genes. Abiotic and biotic stresses are natural stresses wheat and other plants face which raises a question and challenging environment for survival. The AP2/ERF superfamily consist of conserved AP2 DNA binding domains of about 57–66 amino acids. The AP2/ERF superfamily consists of three families known as Apetala2, ERF, and RAV families, dependent on the number of apetala2 domains and homology based on sequence; the ERF proteins having single AP2 domain; and the RAV family proteins consisting two DNA binding domains that are different, AP2 and B3. The structural Analysis of ERFs in wheat paved way for combating various stress responses in wheat and network analysis showed various linked proteins responsive to various stress factors.

Keywords: Wheat, transcription factors proteins, network analysis