In-silico Identification and Characterization of Universal Stress Protein (USP) Gene Family in Wheat

By

Hajira Imran

Registration # 00000330580

Supervisor

Dr. Alvina Gul

ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD, PAKISTAN.

AUGUST 2022

In-silico Identification and Characterization of Universal Stress Protein (USP) Gene Family in Wheat

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

By

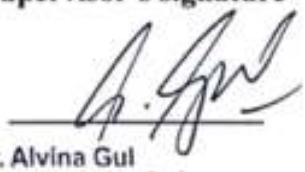
Hajira Imran

Registration # 00000330580

Supervisor

Dr. Alvina Gul

Supervisor's signature



Dr. Alvina Gul
Tenured Associate Professor
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY
ISLAMABAD, PAKISTAN.
AUGUST 2022

Form TH-4

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: (Student Name & Regn No.) **Hajira Imran & 00000330580.**

Titled: "*In-silico* Identification and Characterization of Universal Stress Protein (USP) Gene Family in Wheat" be accepted in partial fulfillment of the requirements for the award of MS PLANT BIOTECHNOLOGY degree with (__A__ grade).

Examination Committee Members

1. Name: **Dr. Faiza Munir** (Internal Examiner) Signature: Dr. Faiza Munir, Assistant Professor, Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

2. Name: **Dr. Rabia Amir** (Internal Examiner) Signature: Assistant Professor, AS & H Department, NUTECH Islamabad

3. Name: **Dr. Roohi Aslam** (External Examiner) Signature:

Supervisor's name: **Dr. Alvina Gul** Signature: Dr. Alvina Gul, Tenured Associate Professor, Deptt of Plant Biotechnology, Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

Date: 25-8-22

Date: 26-08-22

Dr. Muhammad Faraz Bhatti, Head of Department (HoD), Deptt of Plant Biotechnology, Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Head of Department

COUNTERSINGED

Date: 26-08-22

Dr. Hussnain A. Janjua, Principal, Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

Dean/Principal

Thesis Acceptance Certificate

It is certified that the final copy of MS thesis written by **Hajira Imran**, registration No. 00000330580, of ASAB has been vetted by undersigned. Found complete in all aspects as per NUST status/regulation, is free of plagiarism, errors and mistakes and is accepted as partial fulfilment for the award of MS/MPhil degree. It is further certified that necessary amendments are pointed by GEC members of the scholar has also been incorporated in the said thesis.

Signature: _____

Dr. Alvina Gul
Tenured Associate Professor
Dept of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

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

Date: 25-8-22

Signature: _____

Dr. Muhammad Faraz Bhatti
Head of Department (HoD)
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

HOD Plant Biotechnology: **Dr. Muhammad Faraz Bhatti**

Date: 26-08-22

Signature: _____

Dr. Hussnain A. Janjua
Principal
Atta-ur-Rahman School of Applied Biosciences (ASAB)
NUST, Islamabad

Principal ASAB: **Dr. Hussnain A. Janjua**

Date: 26-08-22

Certificate of Plagiarism

It is confirmed that MS thesis entitled "***In-silico*** **Identification and Characterization of Universal Stress Protein (USP) Gene Family in Wheat"** of **Hajira Imran**, registration No. 00000330580 has been me. I undertake that.

1. The thesis has significant new work as compared to already reported one. No figure, table, sentence, or section has been copied from previous work except when placed under quotation marks and duly referenced.
2. The presented work is original, and authors own work as no data fabrication was found in the results i.e., there is no plagiarism.
3. This thesis has been checked for plagiarism and the Turnitin report endorsed by supervisor is attached.

Dr. Alvina Gul
Tenured Associate Professor
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST, Islamabad

Signature of Supervisor

Dr. Alvina Gul

Associate Professor

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

Copyright Statement

Declaration

I certify that this research work titled "*In-silico* Identification and Characterization of Universal Stress Protein (USP) Gene Family in Wheat" is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Hajira Imran

Master of Science in Plant Biotechnology

Registration # 00000330580

Dedication

"My dissertation is dedicated to my Parents and friends"

Acknowledgment

In the name of Allah, the most gracious and most beneficent. I am greatly blessed with the countless blessings of Allah (SWT) that helped me get through this journey and let me complete my work. All the praises for Allah and His Messenger Hazrat Muhammad صلی الله علیه وسلم. My parents, teachers and friends always gave me motivation to do better and better. They have always been a wise counsel and had a sympathetic ear

I am extremely grateful to my supervisor, Dr. Alvina Gul whose constant support throughout my research work helped me complete my research. She has been a there at each step to guide me through the path and always boosted my confidence whenever I felt I couldn't do anymore. She also helped me grow as a person. Secondly, I want to thank my Co-supervisor Dr. Roohi Aslam whose guidance and support throughout the whole thesis made me able to achieve all my goals.

It has been a wonderful experience at NUST, and I am really honored to be its part. NUST provided me with the exposure and opportunities needed for the personal growth as well as acquaintances that will stay with me for the rest of my life. I am thankful to the ASAB faculty. They have helped me throughout my MS. I extend my gratitude to Principal ASAB, Dr. Hussnain A. Janjua, and Head of departments. I also wanted to thank my guidance and examination committee that included Dr. Roohi Aslam, Dr. Faiza and Dr. Rabia who guided me through the research phase.

I am grateful to all the people I met along the way at ASAB. My friends and family have been a great support. I am blessed to have Khadija as my best friend who was with me through every thick and thin. I want to thank my dearest colleague Tuba Sharf Batool who helped me through every step. The love, care and support from my family has been a steppingstone for all the achievements I have so far. I pay my gratitude to all the people who have been with me in this journey and made me what I am today.

Abstract

Climate has changed drastically over the last decade. It is crucial to understand the needs of the plants and their adaptive mechanism that help them survive during adverse environmental conditions. Abiotic stressors mainly salt concentration, osmotic stress, heat stress, drought, flooding, etc. affect plants significantly. In this research work, we have done the identification and characterization of the Universal Stress Protein (USP) gene family in wheat. In-silico approaches such as identification, gene ontologies, chromosomal mapping, circus, and synteny analysis were used to analyze the reported sequences. The study revealed that the domain architecture plays the most significant role in the multi-functional features of this family which is present in all plants. Moreover, the syntenic relationship revealed the conservancy among the monocot genomes. The role of USP in host cells was explored through studies/tools such as subcellular localization and gene ontologies The presence of several regulatory elements also gave insight into stress-specific modulation and regulation. Furthermore, protein modeling of the *TaUSP* genes revealed the presence of binding pockets with functionally important amino acids This work led us to report a total of 107 protein sequences on the ABD genome grouped into 34 *TaUSP* genes. Further instigations such as expression profiling might help in verifying the stress-specific transcriptional modulation of these genes. Hence, this work would be quite useful in developing economically stress-resilient varieties.

Keywords: Abiotic stress, drought, Universal Stress Proteins, regulatory elements

Contents

List of Figures

List of Tables

List of Abbreviations

AANH	Adenine Nucleotide Alpha Hydrolase
ABA	Abscisic Acid
ABRE	Abscisic Acid Response
CRE	Cis Regulatory Elements
DB	Databases
GSDS	Gene Structure and Display Server
JARE	Jasmonic Acid Responsive Elements
LTR	Low Temperature Response
SA	Salicylic Acid
SARE	Salicylic Acid Responsive Elements
ROS	Reactive Oxygen Species
USP	Universal Stress Protein

Chapter One: Introduction

1. Introduction

Any environmental condition that has a detrimental effect on growth, development, grain quality and overall production is referred to as stress. All plants, particularly crops, have a reaction that allows them to release stress or, more accurately, allows them to survive stress during a physically dormant time, such as kernels or stress sensitivity. Biotic and abiotic stresses put a lot of pressure on the overall cellular activity of plants. They are also termed as limiting factors for plant growth and performance because they affect crop yield and production throughout the world, causing global unrest when needs are not fulfilled. These stresses include heat, salinity, drought, and disease attack which results in damage at different molecular levels, including denaturation of proteins, and enzyme inactivation. As a result of these development is halted, and plants may die due to these stresses (Ahmad et al., 2021). Field environments are quite different from the laboratory conditions where plants are tested for different stresses and how they will respond to these stresses. Plant detects these stresses and produces various proteins via signaling pathways, regulating stress tolerance. A wide variety of research has been conducted on how plants responded to these stresses (Yadav et al., 2020). This introductory chapter provides you with an overview of wheat, its taxonomic hierarchy, genome, economic importance, and Universal stress proteins that play a key role in stress regulation and tolerance.

1.1. Overview of wheat

Triticum aestivum commonly called bread wheat is a major cereal crop of the family *Gramineae* or Poaceae. It is cultivated in most parts of the world and is estimated to contribute to a fifth of the total caloric intake by humans. This specie has hexaploidy means it has 3 sets of chromosomes as AABBDD evolved through natural hybridization. Wheat is classified into six major groups:

1) Hard red winter wheat
2) Durum wheat
3) Hard red spring wheat
4) Hard white wheat

5) Soft red winter wheat
6) Soft white wheat

1.1.1. Species of wheat

There are different classifications of the *T. aestivum* genome, the species are characterized as follows:

1) Hexaploid

a) *T. aestivum is* the major cultivated species commonly called bread wheat.

b) *T. spelt* commonly called spelt is also a hexaploid specie.

c) *Triticale is* a hexaploid man-made variety used in livestock by crossing durum with rye.

2) Tetraploid

a) *T. durum* commonly called durum (pasta wheat) is a tetraploid form that is also widely cultivated today.

b) *T. dicoccum (*commonly called Emmer) used to be cultivated in the middle east in ancient times, now its cultivation is limited to certain areas.

c) *T. turgidum is* commonly called the Khorasan specie of wheat. Its name is derived from the area where it used to be cultivated. The nutty flavor is the characteristic of this specie.

3) Diploid

a) T. monococcum is diploid specie commonly termed as Einkorn. It has both wild and cultivated variants. The domestication period of this specie is the same of *emmer*.

According to updated information, there are currently 20 accepted species of wheat named *T. aestivum, T. urartu, T. turgidum, T. dicoccon, T. aethiopicum, T. araraticum, T. boeoticum, T. carthlicum, T. compactum, T. dicoccoides, T. karamyschevii, T. monococcum, T. spelta, T. turanicum, T. polonicum, T. timopheevii* (http://www.itis.gov/)

1.1.2. Morphological Features

Wheat belongs to the family of grasses so is monocotyledonous. The morphological features of all the species of wheat are the same, the genetic makeup, however, differs in all due to which distinct traits are observed. Wheat is an annual cereal crop with a height of 0.7-1.2m. Roots are adventitious or crowned. There is one main cylindrical stem with 3-4 tillers.

Leaves are present on each with auricles, they have a compound spike on top of the plant and usually, the grain is about 30-60mg in weight (Kyozuka, 2014).

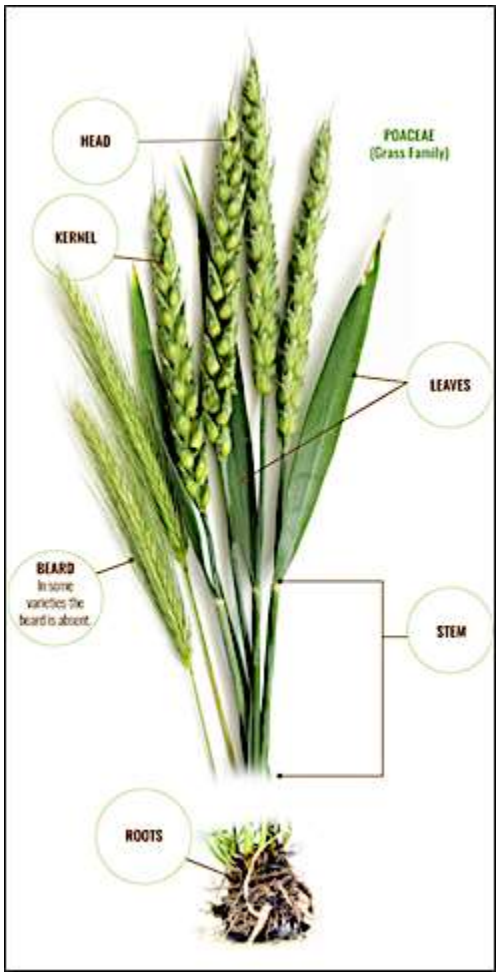


Figure 1: Morphological features of wheat (Desheva, 2013)

1.1.3. Importance of wheat

Wheat is both a nutritively and economically important cereal crop. It is a major part of caloric intake and is fed by millions around the world. The nutritional properties of the wheat are such that carbohydrates constitute 75-80%, total protein content is 10-17%, fats are 2-2.5 %, and minerals constitute about 1.4-2.3% of the total nutritional content (Kumar et al., 2011).

1.1.4. The wheat genome

Histologically, the wheat genome is roughly 100 times bigger than Arabidopsis, 40 times larger than rice, and about six times larger than corn in

terms of the number of base pairs per base pair (bp/bp) (Guan et al., 2020). Due to polyploidy and numerous duplications, bread wheat has a big genome that is made up of repetitive DNA sequences. There are about 810 megabytes in the typical wheat chromosome, 25 times more than the average rice chromosome. A triple structure (ABD genomes) and widespread duplications have resulted in the wheat genome's enormous size (Panchy et al., 2016). More than 85 percent of the genome is made of repetitive, highly methylated regions (Brenchley et al., 2012). The number of genes in higher plants is currently estimated to be between 25,000 and 43,000. According to current research, the wheat plant's phenotype was determined by the impacts of around 30,000 genes (Appels et al., 2018).

1.2. Role of abiotic stressors

Stressor usually represents environmental restrictions that cause disturbances in the cell signaling, aerobic respiration, and photosynthetic pathways (lower root water uptake; dehydration stress), as well as unequal distribution in the cells, which promotes greater ROS generation. Moreover, the stress causes also have many impacts particular to distinct environmental stresses.

1.2.1. Types of stresses

Global climate modifications that increase crop yield are not good for grain quality, exerting further strain on global crop supply. Following are some major abiotic stresses that have less or more effect on plant growth and development. Fig 1 shows some abiotic stressor.

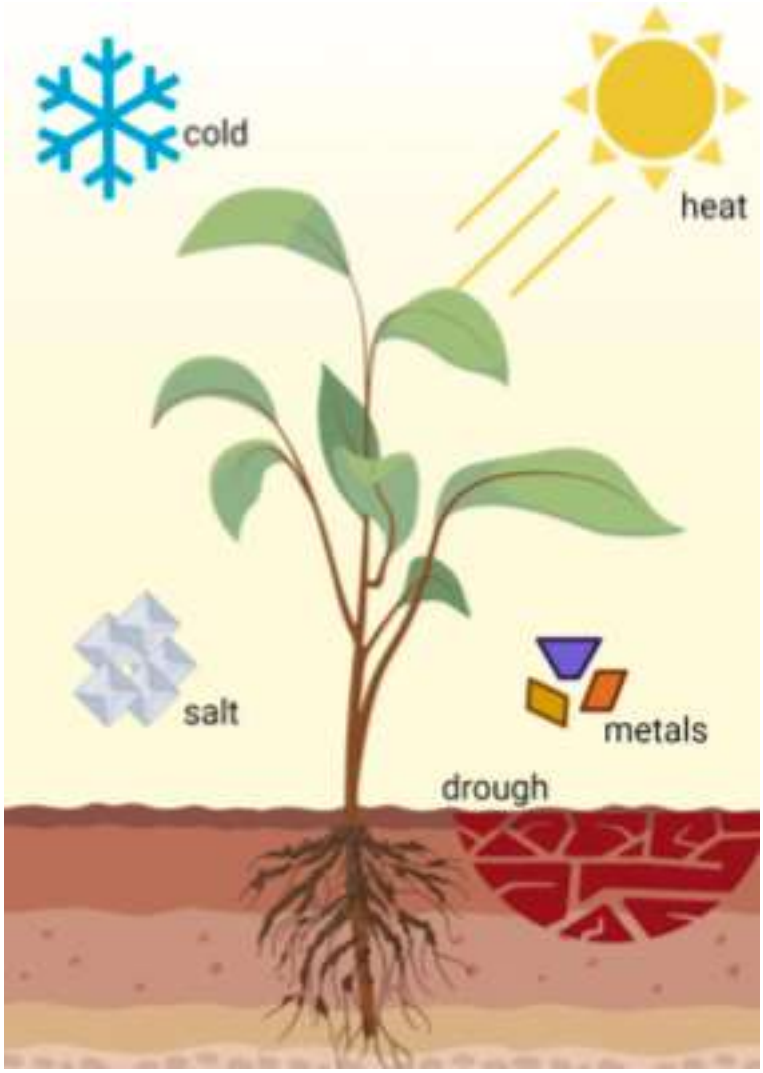


Figure 2: Different abiotic stresses (Sampaio et al., 2022)

1.2.1.1. Drought

Wheat may adjust physiological and biochemical properties to maintain development and growth during independent or coupled drought and heat stress (Sattar et al., 2020). Drought is a polygenic stress and it indicates a lower availability for soil water causing a reduction in root moisture absorption (Senapati et al., 2019). Plant reaction on a microscopic level is owing to an increased accumulation of different osmolytes and hydrophilic proteins. osmotic adjustment, i.e., a reduction of a turgor pressure of the cytosol. Dry spell causes in leaves to stomatal closure due to lower CO_2 absorption, which leads to inequalities between the physiological electron transport and carbon acquisition processes. Internal drying is thus also linked to increased ROS production leading to induction in drought-treated various

ROS enzymes, including as numerous thioredoxin (Trx) variants. Somewhere at the proteomic level, water stress abnormalities in cell function, particularly photosynthesis, lead to changes in many proteins of photosynthesis one of which is Rubisco (Huang et al., 2019).

1.2.1.2. Thermo-tolerance

Wheat yield is extremely vulnerable to temperature changes. Germination and kernel production are key stages in wheat survival and prosperity, and a slight increase has a significant impact on the overall yield (Burke, 2001). Wheat's tolerance process is mostly governed by its genetic makeup, as indicated by the existence of tolerant and sensitive wheat lines. Wheat's genetic make-up prevents it from surviving high-temperature environments by regulating the defense system via upregulation of stress-related genes (Senthil-Kumar et al., 2007). At the microscopic level, radiation and the rising temperatures contribute to increased biomolecular dynamics which results in an enhanced risk of protein malformation. Plant reaction, therefore, involves the induction of multiple downstream processing of heat shock proteins. Furthermore, temperature and evaporation are related to each other, which shows that increased temperature increases the evaporation of water from the soil and improved foliage transpiration, leading typically to water shortages in field settings. Heat therefore also leads to stress and oxidation (Li, B. et al., 2018).

Conversely, colder temperatures contribute to lower biomolecular kinetics, lower cell elasticity, and lower enzyme rate constants. The freezing of ice in the soil leads to a lower intake of water by roots which causes cell hypoxia. As a result, numerous osmolytes and hydrophilic proteins such as dehydrins are accumulated. Thermo-synthetic transportation mechanisms and carbon assimilations processes result in increased photo-inhibition and power loss due to thermal stress (Sattar et al., 2020).

1.2.1.3. Salinity

Salt stress in the soil water solution reflects increased amounts of sodium ions. The reduced soil water potential shows an osmotic impact on plant cells that can lead to the buildup of various compatible solutes. Furthermore, increased

ions in the solution in water deficit create an ionic effect, which advances with stress time and causes harmful ions such as Sodium ions to penetrate plant cells. The reaction to plants is either an active filtration of ions or cellular division into vacuoles which results in cell endurance. These activities are energy-intensive and hence related to increased concentrations of various energy-consuming transport systems. For all stressors of dehydration, the hydrodynamic effect is rapid and frequent while the electrical activity is temporary (Yang & Guo, 2018).

1.2.1.4. Floods

Flooding can contribute to the root system to anoxic conditions, therefore leading to decomposition. Fermentation causes increased organic acid buildup leading to cell cytoplasm acid pH, which has detrimental effects on the functioning of numerous cellular enzymes. Komatsu along with her colleagues have released proteome research on soybean root type flood responses, including many cellular organelle-based investigations.

Metal ions pressure also has harmful ionic effects, entering cells that result in ion exclusion or vacuolar compartmentalization, which is used for soil remediation by plant accumulators.

1.3. Universal Stress Proteins

Proteins have a key role in plant stress, as they directly contribute to the formation of new phenotypes by adjusting the physiological characteristics of the modified environment. In 1994, Marc Wilkins developed the word "proteome," which represents a complete protein in a particular organism at a certain moment. As a result, both plant developmental and health changes, and the environment, unlike the static structure, which is inherited from the parents and defines the genotype of a plant is changing the plant phenotype in terms of epigenome, transcriptome, proteome, and metabolome. Proteins are closely interlinked in the crop stress reaction, as well as in the control of plant expression of genes, transcriptome, and metabolic activity, of structural protein. Protein activity depends on its cellular localization, post-translation changes, and interactive associates and not simply on its molecular construction.

Environmental stress either biotic or abiotic negatively affects crop yield and quality. The adaptive immune response is generated by plants because of stress which leads to the activation of genes that encode stress proteins for its regulation and tolerance (Isokpehi et al., 2011). The universal stress protein family positively regulates the biotic and abiotic stress regulation and tolerance. Extensive research is being done to understand the molecular mechanism related to plant's stress management, but it is still poorly understood. As the genomic sequences are becoming readily available, they can be combined with high through put bioinformatic tools and databases to open new possibilities to investigate the gene families crucial for stress response that are still understudied(Li, W.-T. et al., 2010; Park et al., 2017; Vollmer & Bark, 2018).

1.3.1. Structure of *USPs*

Universal stress proteins were first discovered in *E. coli* where their expression was spiked in response to certain environmental stressors and starvation conditions. Although in plants, the most frequent kind of universal stress proteins has a single domain of *USP*, several other functional motifs are present in other proteins. These may have evolved because of evolutionary pressure against various stressors, resulting in the arrangement of various catalytic motifs with the domain of *USP*. Plants can use a variety of tactics to protect themselves against exogenous pressures because of the process.

Genes that encode *USP*s have 140-160 conserved amino acid residues. They have functional *USP* domains that belong to Pfam accession: PF00582 (Vollmer & Bark, 2018). They have many other functional motifs attached to them. The presence of α/β subdomains in the *USP*s is crucial to many cells' defense signaling pathways and metabolic pathways involved in stress resistance. They may also contain a conserved ATP binding motif G-2X-G9X-G(S/T). *USP*s are upregulated under harsh climatic conditions, malnutrition, bioaccumulation, thermal and oxidative stress, and several other biotic and abiotic stresses.

1.3.2. Functions of *USP*s

In *Arabidopsis thaliana*, there are 44 identified *USP*s that have an evolutionary relationship to bacterial *USPs*. They have high sequence homology with 1 MJH-like sequences. These proteins have several functions such as hydrogen peroxide concentration is regulated by HRU1 under hypoxia (Gonzali et al., 2015). Another *USP* is At*USP*; it is involved in protein and RNA chaperoning and expressed during heat and cold stress (Melencion et al., 2017).

For comparative gene function studies, rice is an excellent crop for its ease of cultivation and propagation as well as its ability to undergo genetic change quickly. Because of the publication of the whole rice genome sequence, a comprehensive analysis of a whole gene family is now possible. Quite recently 44 *USP* genes were identified in rice having functional diversity. While investigating the transcriptional modulation of these genes the researchers identified certain *USP* genes involved in multi-stress regulation, which shows that using the biotechnological tools and engineering techniques the selected Os*USP* genes can be used to produce multi-stress resistant rice as well as other economically important crops (Arabia et al., 2021). *Zea mays* is cultivated worldwide as one of the most important cereal crops. There is a total of 43 *USP* genes found in maize performing a variety of functions and combating a plethora of environmental stress. Mf*USP*1 from *Medicago falcata* plays a significant role in ROS homeostasis, osmotic pressure, salinity, temperature, and some other abiotic factors. Upregulation of this gene provides many agronomic tolerances (Cui et al., 2021).

For the textile sector, cotton is by far the most significant crop in terms of fiber output. Because drought is a key problem for cotton yield, and thus for cotton breeding, isolating the genes, specifically expressed in genotypes with drought tolerance would be beneficial in considering the processes of tolerance against drought. Upland cotton cultivars lack many of the beneficial characteristics of *Gossypium arboreum.* It is well suited to dry land environments due to genes important for drought tolerance and disease resistance. They make *G. arboreum* a useful gene pool for contemporary

cotton cultivar improvement. The two genes G*USP*$_1$ and G*USP*$_2$ have shown structural homology to the Universal stress protein family, they are drought-responsive genes (Hassan et al., 2021; Zahur et al., 2009).

1.4. Significance of *USP*

Plant growth, development, metabolism, and yield are influenced by transcriptional regulation of different plant genes. By interacting with each other through signaling and certain modification during translational and -post-translational events, the intricate regulatory pathways involved in plants' response to stress and control their activities. It has been shown that plants use *USP*s to control defense mechanisms against a variety of external stressors. A significant method for developing stress-tolerant agricultural plants may thus be to regulate the expression of *USP* genes. Determining how the *USP* operates in the biochemical framework behind plant stress tolerance mechanisms is crucial to success. To develop extremely profitable, stress-resistant reaps that may be used in agronomic areas, it may be essential to alter the expression of plant *USP*s. Optimization of *USP* gene expression combined with other approaches, such as genetic engineering and molecular breeding, might generate innovative and highly valuable cultivars because of the physiological relevance of *USP*s in plants. When faced with adverse environmental stressors (such as climate change or excessive temperatures), initiatives that use an understanding of *USP* gene activity are expected to play critical roles in producing future kinds of crops with better stress tolerant abilities. Despite *USP*'s significance, little is known about their molecular characteristics. A broad variety of cell lines and varieties include these proteins, which indicates their importance in tissues, organs, and physiology. Aside from chauffeuring the proteins and RNA, *USP*s are also involved in nucleic acid interaction and hypoxia avoidance in plants.

Plant growth, development, morphology, and production are influenced by the physiological and biochemical processes that are triggered by the regulation of the expression of specific genes. In plants, the signal transduction networks involved in stress reactions govern the activities of one another through cross talking their interaction and their translation and post-translation modification.

It is common knowledge. Plant *USP*s take involved in a range of cell metabolisms to control protection mechanisms against various environmental stressors. Therefore, regulating *USP* gene expression can provide a strong approach to developing agricultural plant stress-tolerant types (Chi et al., 2019a). The effectiveness of this strategy calls on a comprehensive molecular knowledge of the *USP* activities that underlie the stress tolerance reactions of plants. Since plant *USP*s have distinct functions in defensive responses, it may be essential to alter their expression to develop extremely useful, stress-tolerant plants with beneficial applications in farming.

The biological relevance of *USP*s on plants adamantly supports the hypothesis that in association with certain other approaches such as molecular breeding and genetic engineering the regulation of *USP* gene expression in major agricultural plants may generate innovative and highly manufactured crop variants. Therefore, initiatives with knowledge of *USP* gene activity are expected to be important for developing new types of stress-tolerant crops in situations including adverse environmental stressors, such as climate change, excessive heating, and other serious ecological challenges.

1.5. Aims and objectives

This research has the following aims and objectives

1) To identify the universal stress protein gene family in *Triticum aestivum*
2) To conduct phylogenetic analysis of the species based on the presence of *USP*s
3) To identify conserved domains and motifs of *USP*s in wheat
4) To conduct gene duplication analysis
5) To conduct synteny analysis
6) To predict the gene structure and cellular localization of *USP*s in *Triticum aestivum*
7) Cis-regulatory elements analysis of *USP*s of wheat

Chapter Two: Literature review

Overview

This chapter will describe the plant's response to stress and signaling pathway they follow, and all the previous research being done on the Universal stress proteins in different organisms, including bacteria and kingdom plantae. Feeding the ever-increasing world's population is a great challenge. The conflict among the territorial distribution of residential land vs farmland has aggravated the situation more. Food insecurities are increased if combined with decreasing soil fertility, lower agricultural yields, and unpredictable weather patterns. Despite implementation of various food safety strategies, increasing the agricultural yield amidst bad environmental conditions is still a difficult task to achieve. Cereal crops such as wheat, rice, maize, sorghum etc. are on the most affected crops due to changing weather patterns (Dey et al., 2022).

2.1. Plant cell signaling in response to stress

Plants have different adaptive as well as defensive mechanisms to moderate the adverse effects of the stresses inflicted upon them by the changing environmental conditions. To produce response against any stress, plants first need to recognize what type of stress is it. The recognition of these stresses is done by different receptors present on the surface of the plant cells. After the recognition these defensive mechanisms gets stimulated and starts a downstream cascade of signaling molecules that further activates kinases, production of hormone (Fig. 3).

2.2. Universal Stress Proteins in bacteria

USPs were first discovered 25 years ago in *E. coli.* Different homologs have been identified so far in many other bacterial species as well. There different types of USPs present in *E. coli.* These types include:

a. USPA
b. USPC
c. USPD

d. USPE
e. USPF
f. USPG

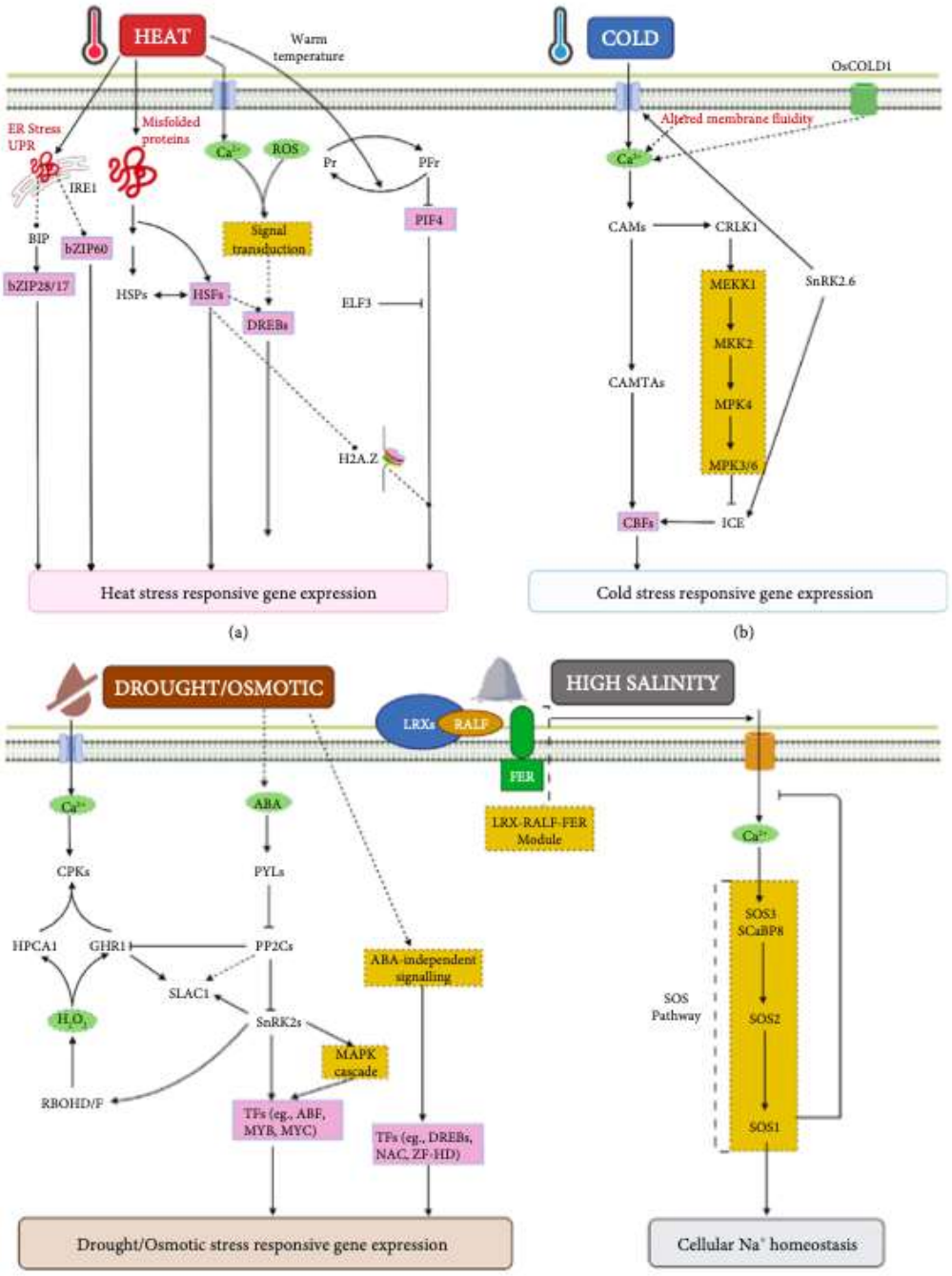


Figure 3: Signaling Pathways followed in response to abiotic stresses (Lohani et al., 2022)

In literature, another USPB was also identified but was removed from the category as it didn't satisfy the criteria of the USP(Vollmer & Bark, 2018). All

these proteins belong to the different classes based on amnio acid composition and structural homology. There are four classes:

I. Class 1: No ATP binding motif (USPA, C and D)
II. Class 2: With ATP binding motif (USPF and G)
III. Class 3 and Class 4: with two USP domains termed E1 and E2 present in tandem (USPE)

The USPs identified from other bacterial species reveled that there are two main classes only first that contains the ATP binding motif (G-2X-G-9X-G-S/T). This class also contains alpha beta domains 4 and 5 respectively. The second class that does not contain this conserved motif and thus unable to bind the ATP molecules.

2.2.1. Structure of bacterial USPs

Bacterial USP have a structural diversity due to presence of different domains in addition to the USP domain (Nachin et al., 2008). They play distinct roles in the cell due to fusion of other domains as well as the presence of functional motifs. The domain architecture of bacterial USP revealed that Na^+/H^+ ion exchanger domain, protein kinase domain, U-box domain, and voltage gated Cl^- channels domain etc. (Fig. 4) shows all the identified domain patterns in the bacterial USPs.

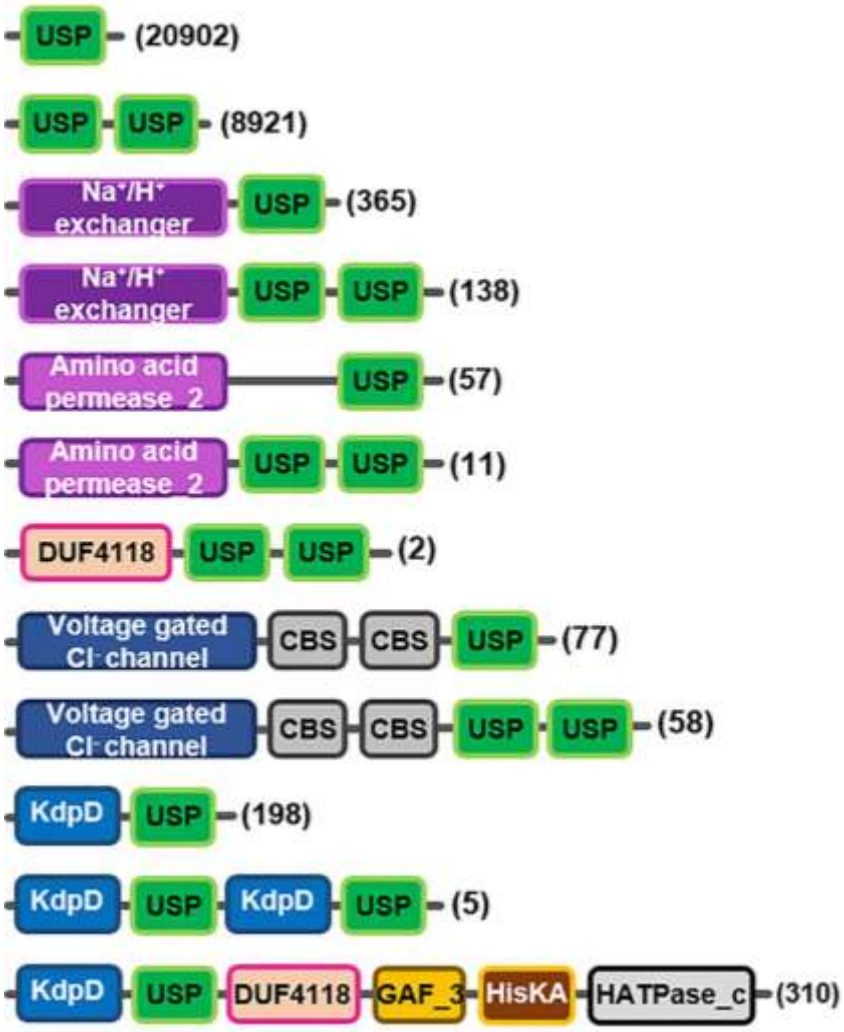


Figure 4: Domain architecture of bacterial USPs (Nachin et al., 2008)

2.2.2. Functions of bacterial USPs

E. coli USPs have their own specific functions. The members of class 1 and class 2 play their role in response to oxidative stress as well as sifting iron ions in the cellular compartments. Other than these functions these proteins also play role in mobility and fixation of several cellular pathways. (Fig. 5) indicates the identified functions performed by the *E. coli* USPs. other functioning of these USPs include proper protein folding, chaperone function, enhanced tolerance to changes in temperature, ubiquitination, binding, remodeling, and proper functioning of the nucleic acids.

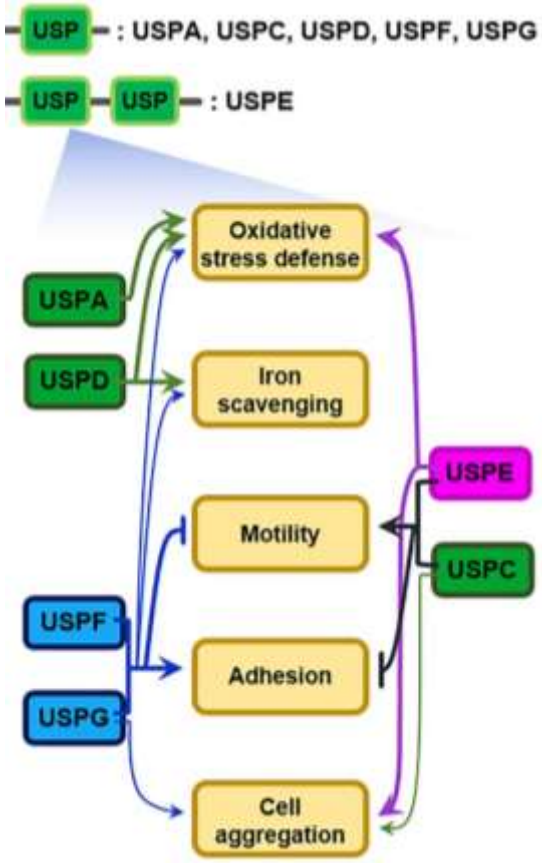


Figure 5: Flow diagram of the functions performed by the bacterial USPs (Nachin et al., 2008)

2.2.3. USPs in other microorganisms

USPA from *Acinetobacter baumanni*i is a gram negative bacteria, and a leading cause of sepsis and pneumonia, helps protect plant against oxidative stress (Elhosseiny et al., 2015). A USP from *Salmonella typhimurium* also plays its role in protection against hypoxia or anoxia (Liu et al., 2007). USPA from *Listeria innocua* showed upregulation upon treatment with acids (Tremonte et al., 2016).

2.3. USPs in Plants

The way USP plays role in defensive mechanisms of archaea, bacteria, and fungi, same is the case with the plants. Over the years USPs have been characterized in various plants and their physiological functions as well as their expression analysis has also been done. Several drought responsive USPs

genes have been identified in viridiplantae and their respective domain architectures are shown in (Fig. 6)(Isokpehi et al., 2011).

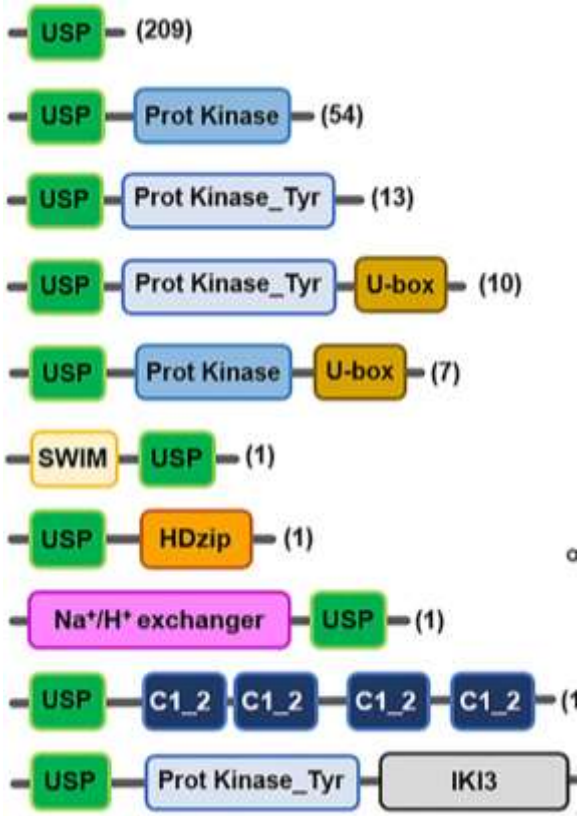


Figure 6: Domain Architecture of Plant USPs (Isokpehi et al., 2011)

2.3.1. USPs in Arabidopsis

There are 41 identified genes in *A. Thaliana* that encode Fifty-three USP genes. The USP were named according to the function they performed and were classified into 4 classes:

1. At*USP*: class that contains single USP domain; 36 out of 53.
2. At*USPUSP*: class with tandem USP domains; 4 out of 53.
3. AtUtyK: class with single USP and an additional protein tyrosine kinase domain; 6 out of 53.
4. AtUK: class with single USP domain with additional protein kinase domain; 7 out of 53 (Bhuria et al., 2019).

New name	Old name	Function
At*USP1*	ATPUB54	Ubiquitination (Wiborg et al., 2008)
At*USP8*	MRH6 (morphogenesis of root hair 6)	Root hair development (Jones et al., 2006)
At*USP9*	RD2 (responsive to desiccation)	Induced in response to desiccation (KlYOSUE & SHINOZAKI)
At*USP12*	HRU1	Regulates ROS production during hypoxia (Gonzali et al., 2015)
At*USP21* **At*USP27***	AtPHOS34 AtPHOS32	Because they were phosphorylated in response to flg22 (Merkouropoulos et al., 2008)

Table 1: Distinct functions performed USPs of *A. Thaliana*

Some of the important functions performed by AtUSP include RNA chaperone functioning under low temperature stress (Fig. 7) (Melencion et al., 2017). Whereas it also acts as a chaperone due to oxidative stress (Jung et al., 2015). At*USP* also interact with other target proteins to give specific functions, such as GTPase ROP2, Thioredoxin h1 and RbohD as ROS producing, protein chaperones, disulphide reductase and inhibiting ABA in response to stomata closure (Gonzali et al., 2015; Kim et al., 2012).

2.3.2. USPs in Cotton

Cotton belongs to the genus *Gossypium* and has at least 52 species and only 4 are cultivated to get the fiber. Each year cotton also gets affected by the abiotic as wells as biotic stresses that causes low quality of the fiber as well as losses to the economy (Edde, 2022). Identified USP in cotton are Gh*USP1* is induce under salinity stress (Li, W. et al., 2015). GUSP1 and GUSP2 are like USPs. Modified USP from *Gossipium arboreum* GUSP1 has increased drought tolerance(Hassan et al., 2021).

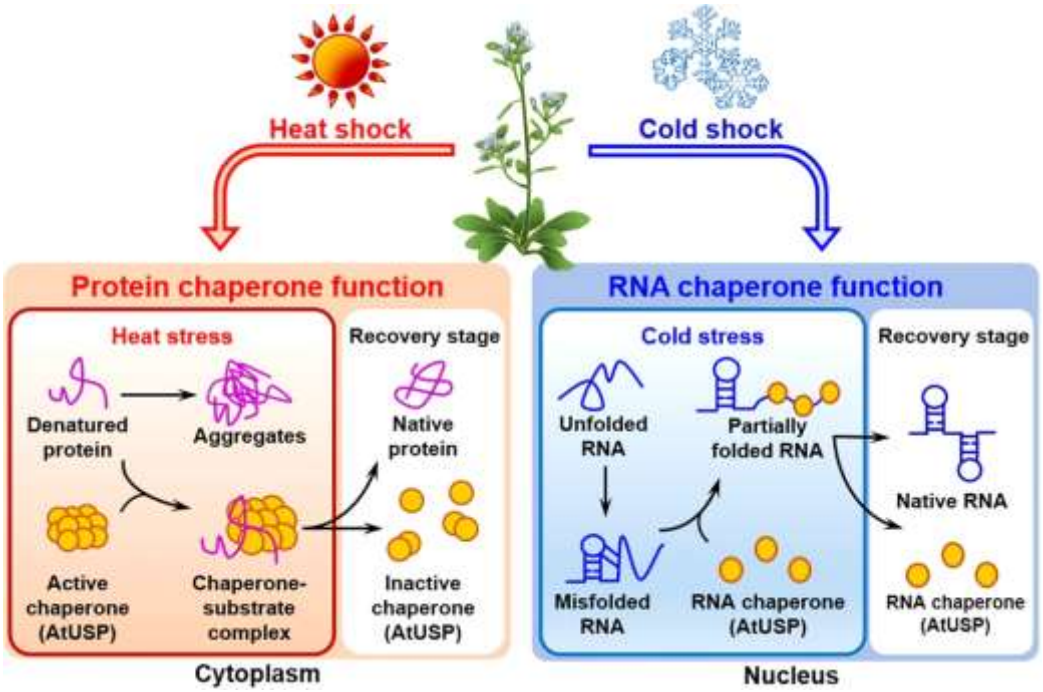


Figure 7:Diagrammatic representation of the functions performed by AtUSPs in response to heat and cold stress (Melencion et al., 2017)

2.3.3. USPs in *Hordeum vulgare*

Twenty-five putative USP sequences were identified in barley. Expression analysis revealed that most of them were sensitive to salt stress rather than being tolerant to osmotic stress (Li, W.-T. et al., 2010).

2.3.4. USPs in *Oryza sativa*

Rice is an important cereal crop. The ease to cultivate and high genetic transforming efficiency it is used as model organism for comparative studies (Basso et al., 2020). Whole genome sequencing has made it easier to study the presence and effects of these genes in cereal crops (Song et al., 2018). *OsUSP1* was the first identified USP in kingdom plantae that has was induced due to ethylene under hypoxic condition(Sauter et al., 2002). Previously 38 USP genes were identified in the rice genome naming *OsUSPs* (Chi et al., 2019b). A recent study has identified 44 *OsUSP* genes and their multi-stress responsive nature has been verified through expression analysis (Arabia et al., 2021).

2.3.5. USPs in *Medicago falcata*

Medicago falcata is more commonly known as alfalfa due to its close resembles. A new USP, *MfUSP1* has been identified. It is a cytosolic protein and is induced in response to ABA, hydrogen peroxide and several other abiotic stress. Overexpression results in enhanced tolerance and also has role in antioxidant defensive mechanism, due to its role in ROS scavenging (Gou et al., 2020).

2.3.6. USPs in *Salicornia brachiata*

Salicornia brachiata is a halophyte and due to their presence in extreme salty environment they are of interest to scientist to study the underlying mechanisms of salinity tolerance (Flowers & Colmer, 2008). Sb*USP*, belongs to USP family, is an uncharacterized protein with increased tolerance to heat, cold, salt and drought stress. Expression in *E. coli* and tobacco showed their role in enhanced tolerance against osmotic and salt stress (Udawat et al., 2016; Udawat et al., 2014)

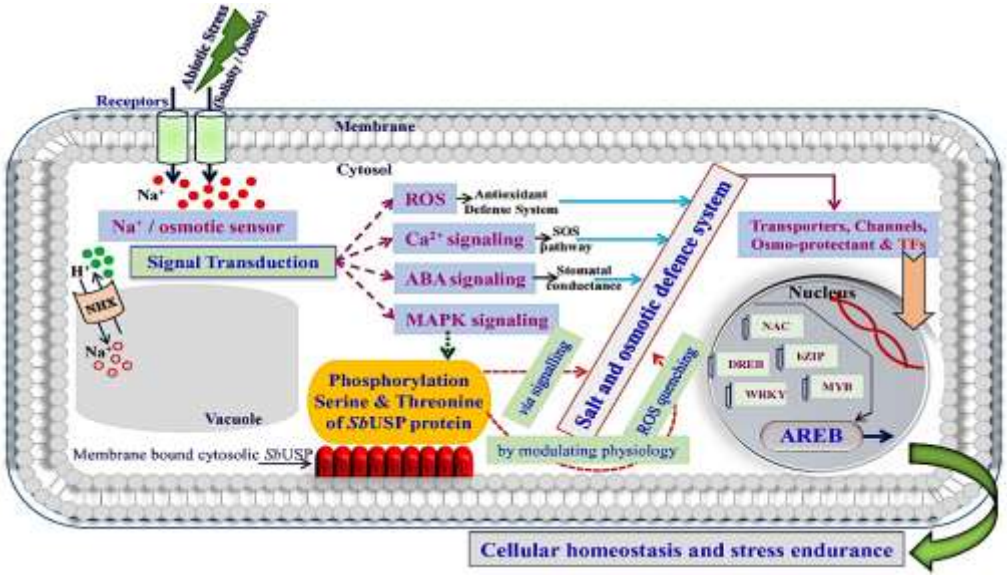


Figure 8:Schematic representation of the ways and activation of SbUSP gene (Udawat et al., 2016).

2.3.7. USPs in Tomato

Tomato is fruit used in making juices, purees, ketchup, and pastes. Tomato has two different varieties. Wild tomato is called *Solanum pennillii* and the

commonly used variety is *Solanum lycopersicum.* Tomato is prone to several abiotic stresses and many studies have been conducted to check the effects of theses stresses on the plants. USPs have also been identified in tomato. *SpUSP* from wild tomato is induced in response to oxidative stress, ABA induced stomatal closure and enhanced photosynthetic ability (Loukehaich et al., 2012). SlRd2 of *S. lycopersicum* interacts with SlCipk6 generates reactive oxygen species in response to oxidative stress (Gutiérrez-Beltrán et al., 2017).

2.3.8. USPs in *Cajunas cajan*

Cajunas cajan commonly called pigeon pea belongs to legume family of plants. 53 genes were identified as USPs and were checked in expression analysis. They also confirm the presence of different domain architecture that might play role in multi-stress responses. The genes had either single USP domain or fused with Na^+/H^+ ion exchanger domain. *C.cajan_29830* and *C.cajan_33874* belonged to USPA class are present in all cultivars (Sinha et al., 2016).

2.3.9. USP in *Salvia miltiorrhiza*

S. miltiorrhiza is an herb and extensively used in Chinese medicine for the treatment of cardiovascular disorders. *SmUSP1,* Sm*USP8* and Sm*USP27* were cloned in *E. coli*. The expression analysis revealed their functioning in response to salinity and heat stress. These results also indicate that these proteins do not work well if multiple stresses are given at once. Rather they show enhanced tolerance to one stress at a time (Wang et al., 2017).

2.3.10. USPs in *Astragalus sinicus*

A. sinicus commonly called Chinese milkvetch and is mostly used in the Chinese medicine. *AsD243,* involved in nodulation, encodes a protein with evolutionary relationship to *USP* family of bacteria having a size of 20-kD. This gene has also sequence similarity with MJ0577 type *USP* protein and acts as molecular switch. *AsD243* expression analysis revealed that under stress condition this gene is expressed in all the plant organs (Chou et al., 2007).

2.3.11. USPs in *Vitis vinifera*

Vitis vinifera commonly called grapevine also have great economic importance in certain countries. Twenty-one USP genes were identified named *VvUSPA.* These genes are induced in response to drought. And other hormonal changes (Cui et al., 2021).

Chapter Three: Materials and Methods

3.1. Overview Of Methodology

The framework opted for the identification and characterization of the gene family is as follows (Fig. 9).

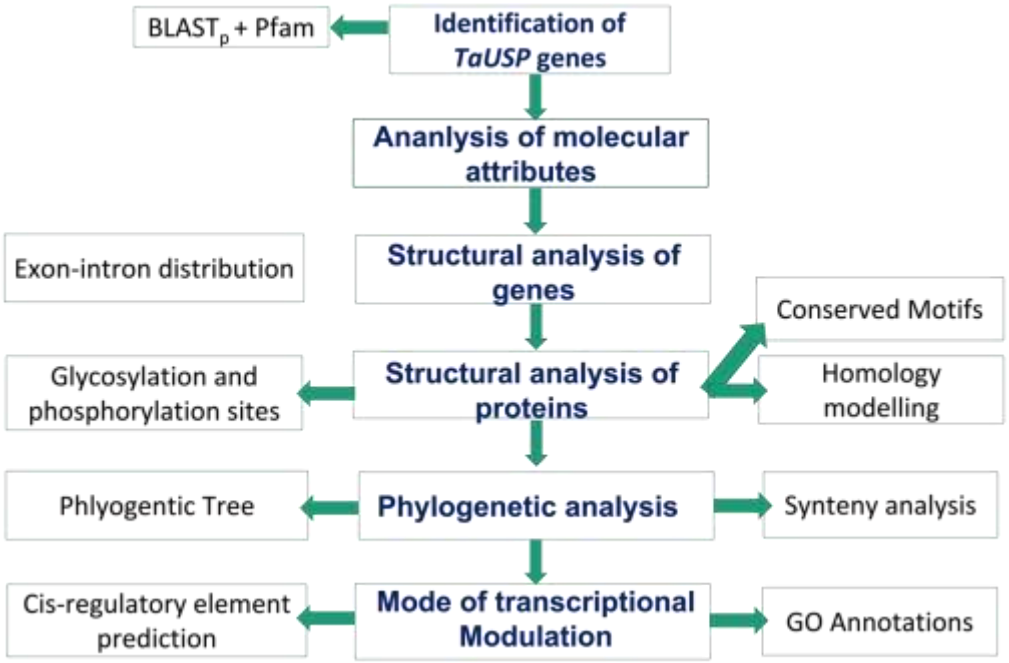


Figure 9: Framework of the study

3.2. Sequence retrieval and characterization of USP genes in Triticum *aestivum*

To characterize the *USP* gene family in *Triticum aestivum*, the amino acid sequence of *Oryza sativa* (Japonica) IRGSp-1.0 *USP* were retrieved from Rice Annotation Project Database(RAP-DB) (https://rapdb.dna.affrc.go.jp/) (Sakai et al., 2013) and were used as query sequences to perform BLASTp against the *Triticum aestivum* IWGSC RefSeq v1.1 genome using 1×10^{-5} as the e-value in ensemble plants database (http://plants.ensembl.org/)(Yates et al., 2022). *USP* genes belongs to adenine nucleotide alpha hydrolase (AANH) superfamily and contains a conserved *USP* domain (PFAM000582), confirmation of USP domains in the retrieved sequences was done using PFAM(http://pfam.xfam.org/) (Mistry et al., 2020). Iso-electric points and

molecular weights were calculated using the ExPasy online server (https://www.expasy.org/resources/comp ute-pi-mw) (Gasteiger et al., 2003).

3.3. Multiple Sequence Alignment, phylogenetic evolutionary analysis, and nomenclature of *USP* genes of *Triticum aestivum*.

Complete protein sequences of *Brachipodium distachyon, Oryza sativa and T. aestivum were* aligned to draw the phylogenetic tree. Multiple sequence alignment of the identified genes was done using Clustalx2.1 software (Larkin et al., 2007). The alignment file was edited using the GeneDoc software (Nicholas, 1997). Phylogenetic analysis was performed using IQ-tree through maximum likelihood using bootstrap value as 1000(Minh et al., 2020). Resultant tree pictures were viewed and edited in iTOL (Letunic & Bork, 2021). Nomenclature of the genes was given according to the *OsUSP* genes identified earlier in 2021 (Arabia et al., 2021)

3.4. Analysis of Gene structure, motif, and amino acid composition

For the analysis of the gene structure and motifs present in the *USP*, complete coding sequences and genomic sequences were taken from ensemble plants database (http://plants.ensembl.org/)(Yates et al., 2022). The exon-intron distribution was graphically visualized using the Gene Structure Display Server (GSDS 2.0)(http://gsds.gao-lab.org/) by comparing the genomic sequence of *TaUSP* genes with the CDS sequences (Hu et al., 2015). For the analysis of the motifs, MEME Suite (version 5.4.1) was used to check the conserved motifs in these genes with the following parameters (1). Total motifs to be found were set to 10. (2). The width of the motif was set to be 6 and 50 as minimum and maximum, respectively. The rest of the parameters were used as default (Bailey et al., 2015).

3.5. Chromosomal mapping and gene duplication analysis

The information related to the positions of USP genes on wheat chromosomes was taken from Ensemble database. The location of the genes was represented using the gene visualization advance tool in TBtools (Chen et al., 2020). Gene

duplication analysis was done keeping in mind two modes of duplication: 1) tandem and 2) segmental duplications respectively. Pair of genes on the same chromosome and separated by <=5 gene positions are termed tandem duplications. Analysis of divergence time and pressure effect of *USP* genes, the Ka (non-synonymous) and Ks (synonymous) values were calculated using the Ka/Ks calculation tool in TB tools. The approximate divergence time was assessed using the formula $\boldsymbol{T = \frac{Ks}{2xMYA}, x = 6.5 \times 10 - 9}$, $\mathbf{MYA = 10 - 6}$ (Panchy et al., 2016).

3.6. Gene Ontology (GO) annotation and Subcellular localization

The GO enrichment analysis was done using the online tool PANTHER version14 (Mi et al., 2019). Protein sequences of wheat *USP* were uploaded to the online server to check the molecular, biological, and cellular functions. Subcellular localization was predicted using the WoLF PSORT webserver(Horton et al., 2007). The heat map was drawn to show the localization of these genes using Heat Map illustrator software on TBtools(Chen et al., 2020).

3.7. Cis-regulatory elements (CREs) analysis of TaUSP promoters

Analysis of CREs present in *USP* of wheat involved extraction of promoter sequences 2.5kbp upstream was done from the ensemble database. Extracted sequences were uploaded to the PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) database to predict different regulatory motifs present in *USP* family of Wheat (Lescot et al., 2002). Graphical representation of these motifs was done using the Toolkit Biologist Tools software along with stacked bar plot, by using the start and ending positions of the respective motif (Chen et al., 2020).

3.8. Identification of Glycosylation and phosphorylation sites in *TaUSP* genes

For the prediction of the glycosylation sites in *TaUSP* genes, NetNGlyc 1.0 online server was used using the preset 0.5 threshold value (Gupta et al., 2004). The phosphorylation sites were predicted using NetPhos 3.1 with a prediction on all three sites i.e., serine, threonine, and tyrosine both generic as well as kinase-specific phosphorylation, respectively (Blom et al., 2004).

3.9. Protein modelling, disordered region analysis and prediction of binding sites

Protein modeling for the four *TaUSP* genes was done through homology modeling using the SWISS-MODEL workspace (https://swissmodel.expasy.org/) (Waterhouse et al., 2018). Secondary structure of the model build were identified using the SOPMA webserver (Geourjon & Deleage, 1995). Disordered regions were analyzed using the Mobi DB server (https://mobidb.bio.unipd.it) (Piovesan et al., 2021) and the binding pockets were discovered using the DoGSiteScorer webserver (Volkamer et al., 2010).

Chapter Four: Results

4.1. Identification of *USP* Genes in Wheat

$BLAST_p$ search revealed, a total of 107 *USP* genes in the *T. ae*stivum genome. This search was made using the *OsUSP* genes. The nomenclature of the genes is given based on the *OsUSPs*. Moreover, due to their presence on the ABD genome, they are sub-grouped as *TaUSP* (A, B, D). The coding DNA sequence's length ranged from 186bp to 6498bp; (*TaUSP13(A), and TaUSP16(B) respectively*). The length of protein ranged from 121aa to 844aa (*TaUSP7(D), TaUSP27b(D)* respectively). Molecular weight of these proteins varied greatly from 11,495.23 g/mol to 91,760.08 g/mol (*TaUSP42(D), TaUSP8(A) respectively*). The average pI of these genes was 7.1098. **Table 2**

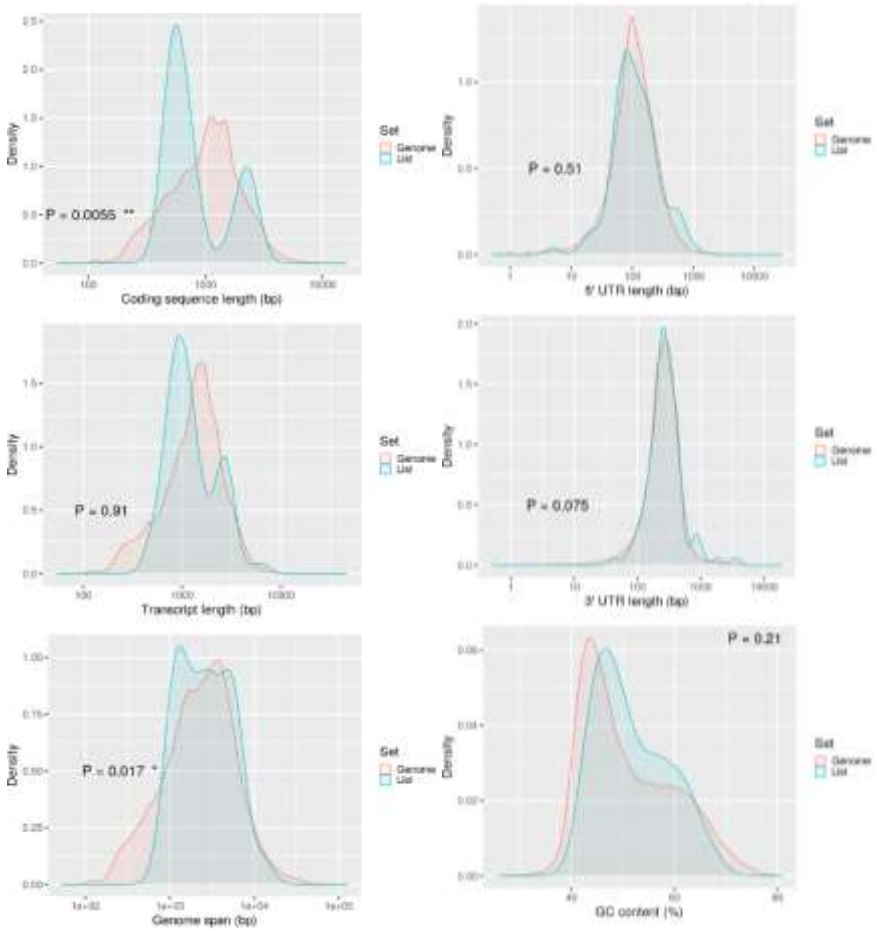


Figure 10: General characteristic features of the identified *TaUSP* genes, Chi square test and student t-Test is applied to check the significant results

4.2. Multiple Sequence Alignment, phylogenetic evolutionary analysis and nomenclature of USP genes of *T. aestivum*.

Multiple sequence alignment of the 107 sequences gave insight into the conserved amino acids present in these sequences. The ligand binding site had a conserved AV/LD sequence. A total of 34 *USP*s were identified comprising of the 107 sequences identified by the $BLAST_p$ search. The phylogenetic analysis was done among 3 species rice, Brachypodium and wheat. It revealed that the tree was divided into two distinct clades. This classification was done based on the domain architecture. *USP* genes contain either a single *USP* domain or *USP* domain with an additional kinase domain belonging to the major classes of protein kinase. *TaUSP10, TaUSP27(a, b), TaUSP34* and *TaUSP35* had an extra U-box domain also attached in addition to *USP* and Protein kinase domain. Fig. 9 shows that not all the genes are present in all three wheat genomes ABD. Many were clustered into the three genomes and the remaining classification was either AB, AD or BD. It also revealed that not all *OsUSP* genes were present in both *B. distachyon* and *T. aestivum* i.e., *OsUSP10, OsUSP20*, *OsUSP32* and *OsUSP42.* No *USP* genes were identified against *OsUSP2, OsUSP6, OsUSP23, OsUSP39* and *OsUSP40.* (Fig. 11)

4.3. Gene structure and motif Analysis and amino acid content of *USP* of wheat

The gene structure of all the identified *TaUSP* in wheat genome was analyzed through webserver. To have better understanding the exon and intron distribution was studied. A moderate variation is seen in the distribution of these intron and exons. For introns, the distribution was from 1 to 10 with *TraesCS1A02G145700.1, TraesCS1D02G144300.1,* belonging to *TaUSP37A, TaUSP37D* have least introns that is one. The sequences that contain *USP* and protein kinase domain had relatively more introns as well as exons. The architectural similarity of the exon-intron distribution indicates that these are also similar at protein level.

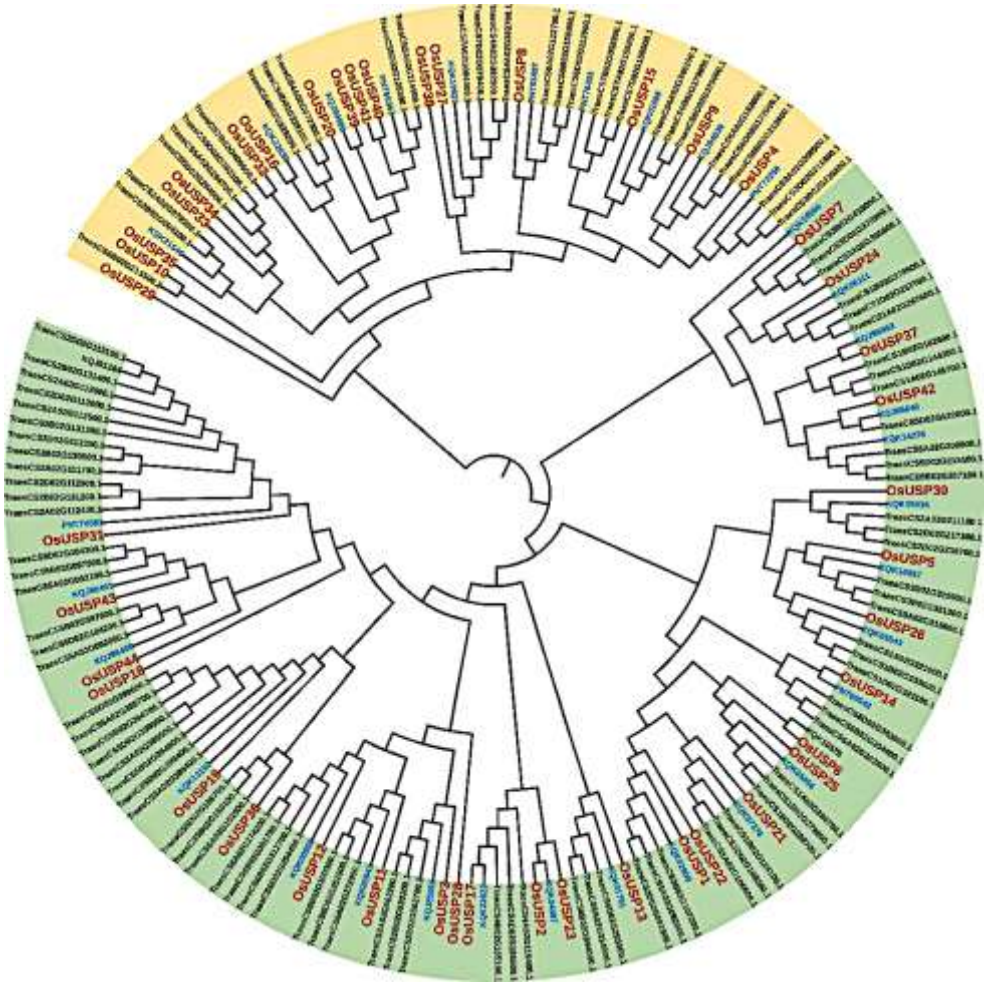


Figure 11: Phylogenetic tree of *B. distachyon*, *O. sativa* and *T. aestivum.*

The exons were also distributed as low as one to maximum as 10. *TaCS4B02G228200.1* and *TaCS4D02G229100.1* has the highest number of exons 10 and these two belong to the *TaUSP16B, TaUSP16D*. Only *TaUSP1A* has 1 exon. The presence of 5'UTR and 3'UTR regions suggests that these genes might have gone alternate splicing. **(**Fig. 12)**.**

Identification of the conserved motifs was done using MEME online server. Variation is indicated on the presence of different motifs in the *TaUSP* genes grouped as 34 gene pairs as total. There were 10 motifs identified. The information of these motifs is present in **Table 3**. There are either 3, 9 or 10 motifs collectively present in the genes. This confirms the presence of single *USP* domain, and a protein kinase domain along with U-box domain in these proteins. From these observations this can be deduced that motifs 1 and 2 are characteristic feature of *USP* domain whereas 3- 9 belong to the protein kinase

and U-box domain. All the *USP*s belonging to Class A had 1-2 motifs whereas the Class B had 6- 10 motifs in total (Fig. 13).

Moreover, the amino acid (aa) composition of these *USP*s indicate that *USP*s with same domain pattern have similar amnio acid content whereas the *USP*s with different domain patterns have slight difference in the amino acid composition (Fig 14). This can be checked by looking at the amino acid composition of *TaUSP3* and *TaUSP35* both have different domain pattern hence they also have different composition. The results of gene structure, motifs and amino acid indicates consistency of the domain architecture present**.**

4.4. Chromosomal mapping and gene duplication analysis

The genes are distributed variably on each of the seven chromosomes of the ABD genome of wheat. Chromosome 5A, 5B and 5D has the highest number of genes as 9,7 and 10, followed by 6A, 6B and 6D with 7, 9, and 7 genes, respectively. Whereas chromosome 7A, 7B and 7D has only 1 gene. Chromosome 1A, 1B and 1D has 5 gene. Chromosome 2A and 2B has 7 gene pairs whereas chromosome 2D have 6 genes. Chromosome 3A, 3B and 3D has 4 gene pairs located on distinct positions. Chromosome 4A,4B and 4B has 4, 2 and 2 gene pairs respectively (Fig. 15(a, b).) Linked genes are indicated through circle gene view in (Fig.16)**.** These results indicates that all genes might not be present on all three genomes of the wheat, and they might have lost in the process of evolution. The existence of more than 1 gene on certain chromosomes indicates that duplication events have occurred over the time that gave rise to these genes on the chromosomes on separate location.

The duplication analysis revealed that 15 gene pairs had duplicated in which only 1 is tandem duplication and rest of the 14 gene pairs were segmental duplication (Fig.17). Divergence time is from 1.51 MYA to 5.01 MYA calculated by the Ka to Ks ratio. As well as the selection pressure was found to be negative as the ka/ks ratio was >1. **Table 4**. Synteny analysis among rice

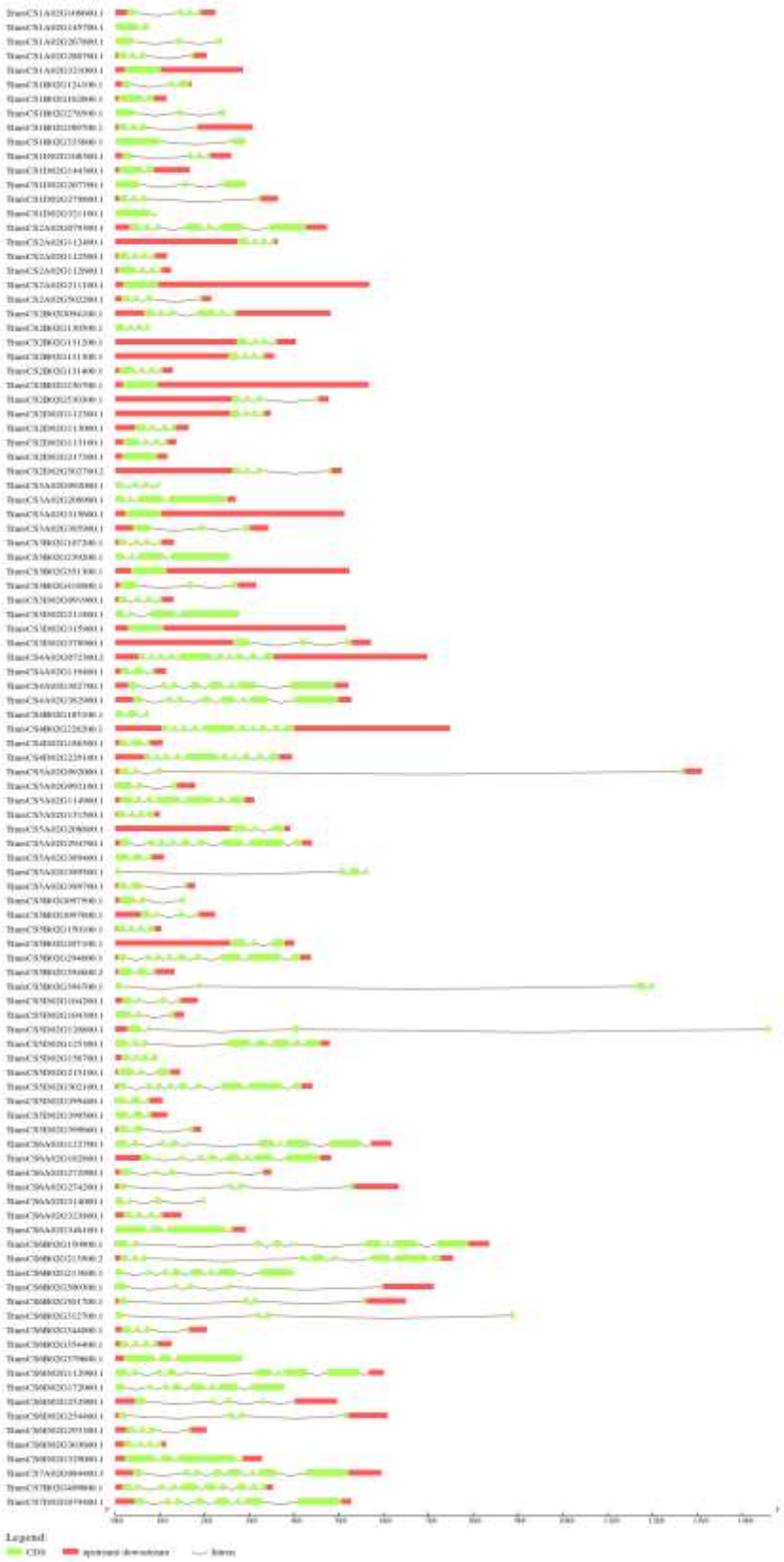


Figure 12: Identified gene structure of *TaUSP* genes

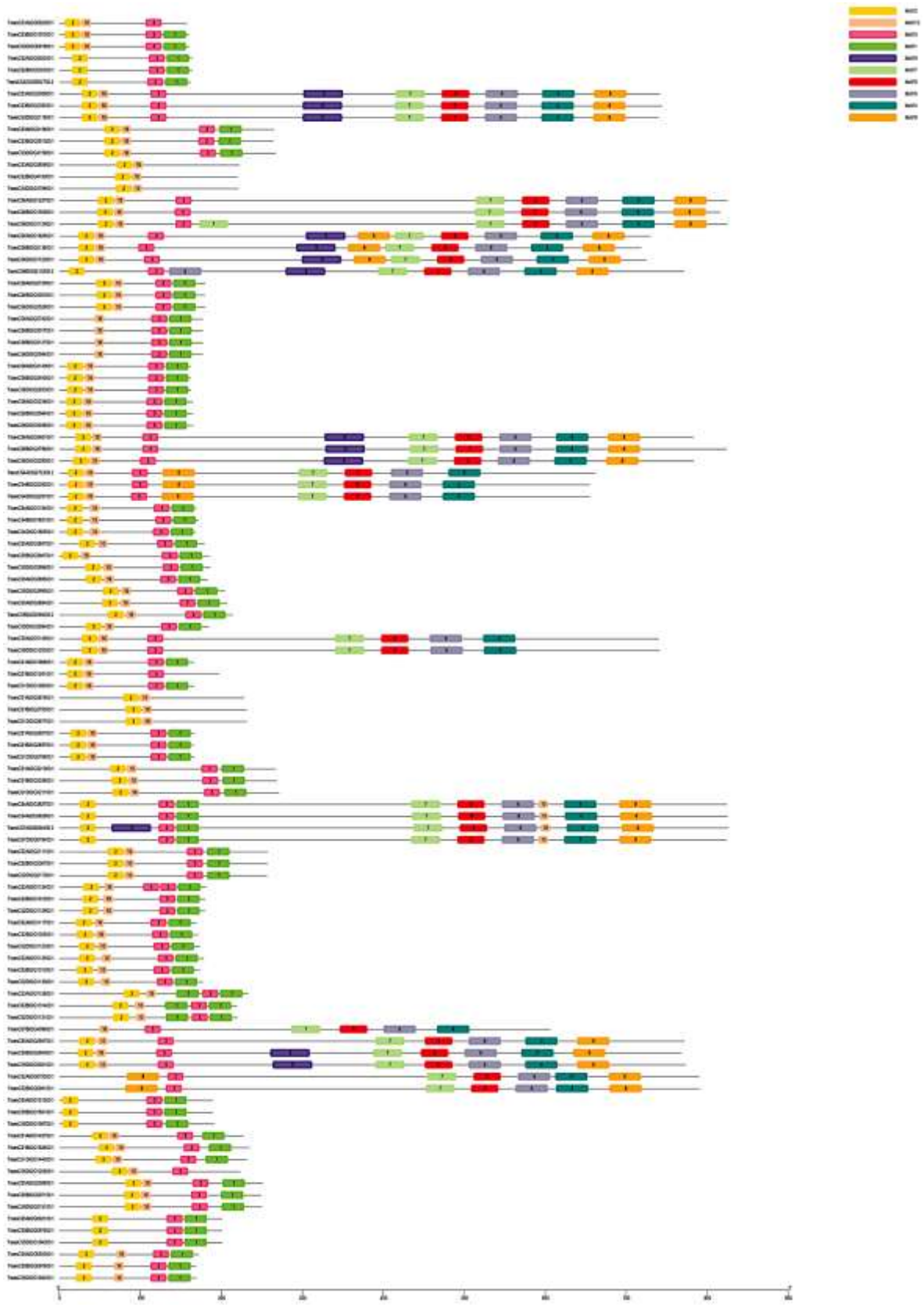


Figure 13: Identified motifs in *TaUSPs*

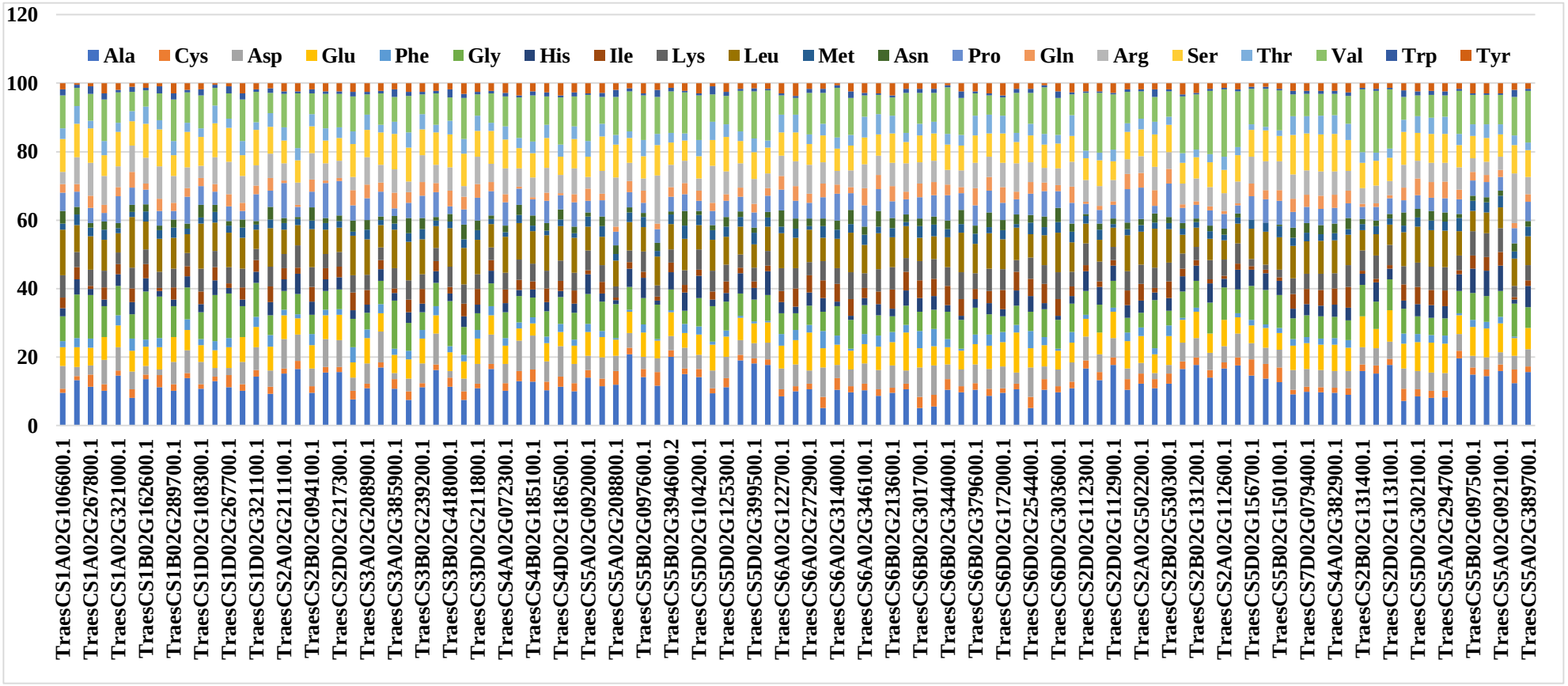


Figure 14: Amino acid composition of *TaUSPs*

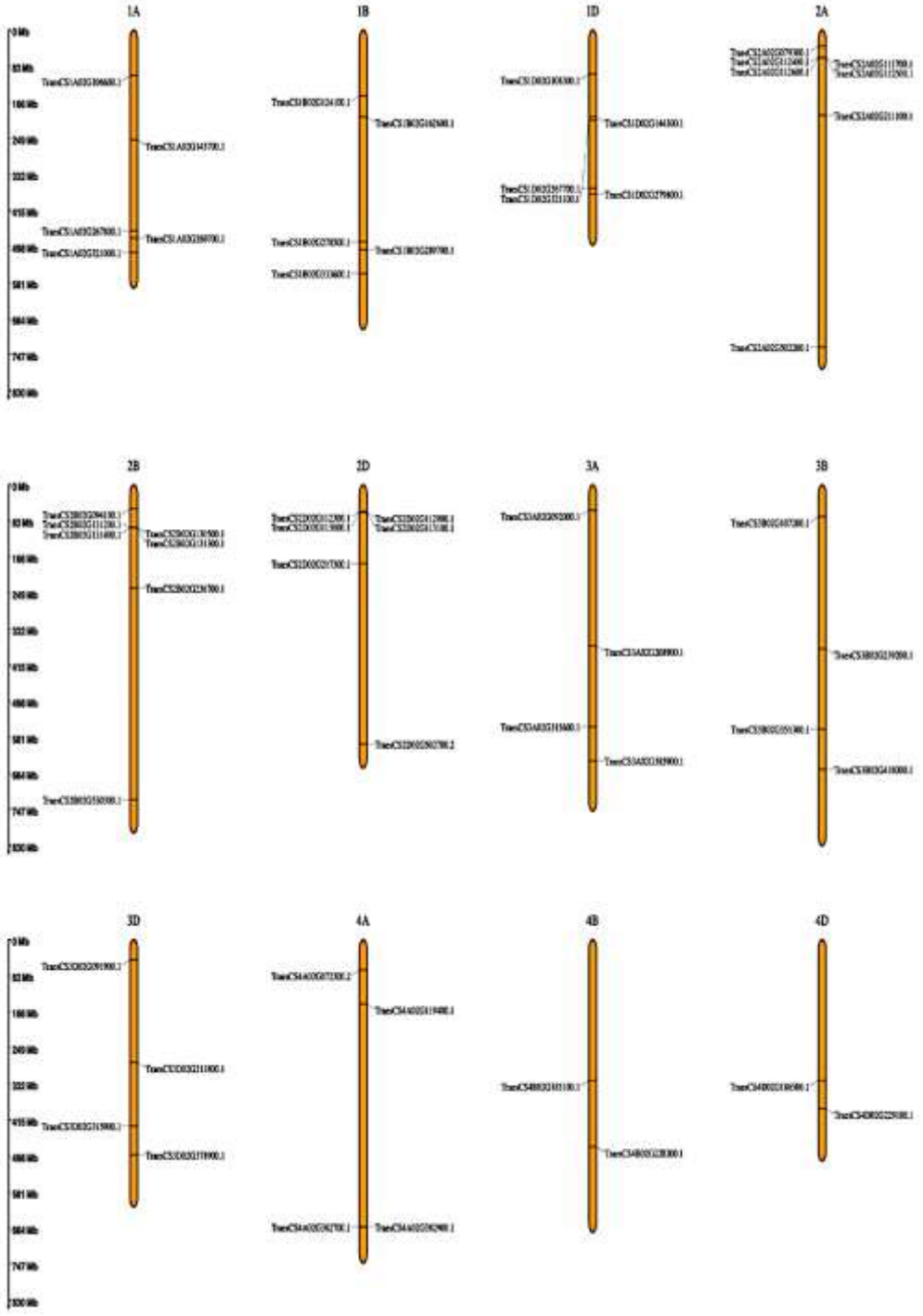


***Figure 15*a: Chromosomal location of *TaUSP* genes (chromosome 1A-4D)**

and wheat showed high conservancy and identified several paralogs. (Fig. 18)

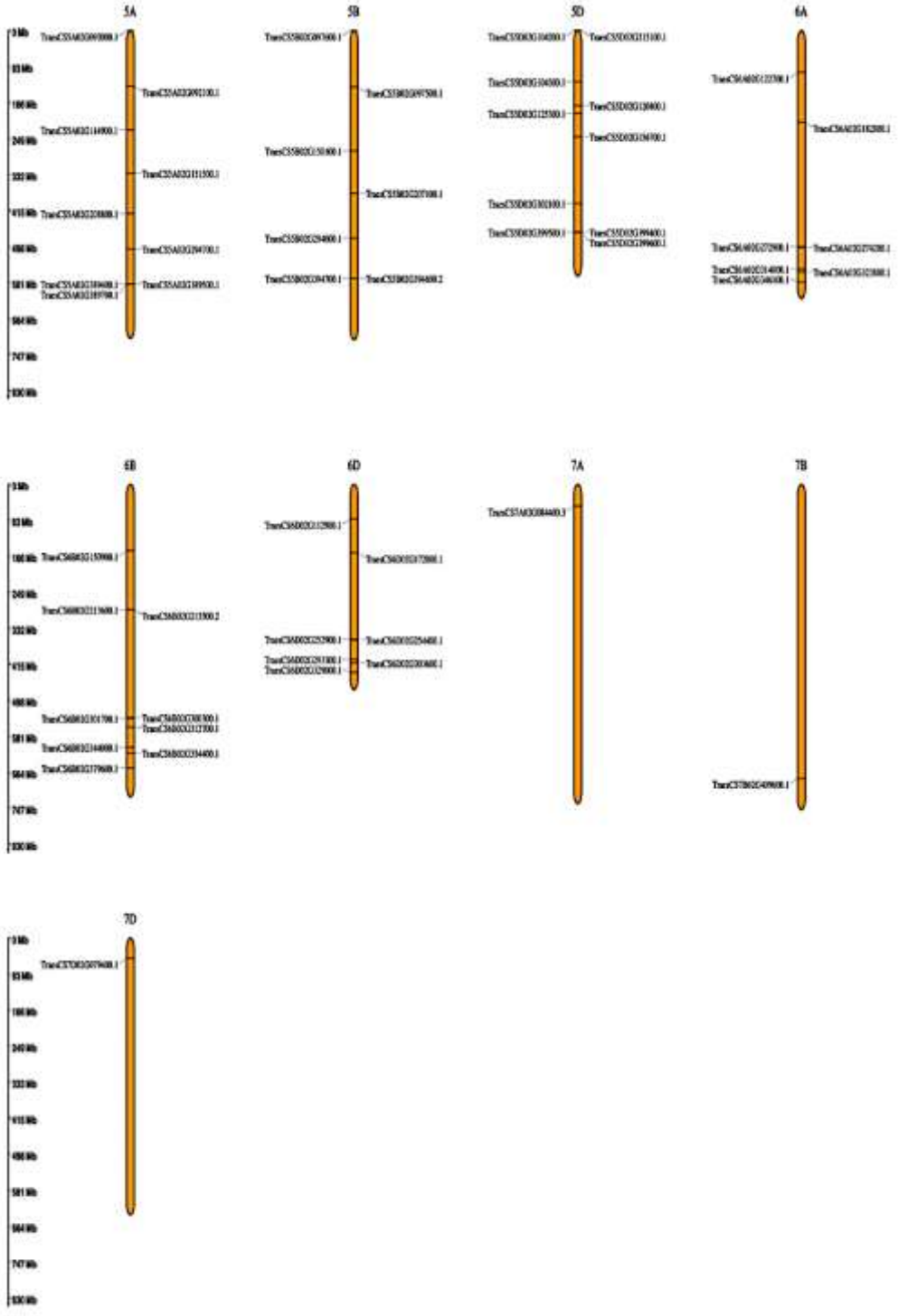


Figure 15b: Chromosomal location of *TaUSP* genes (chromosome 5A-7D)

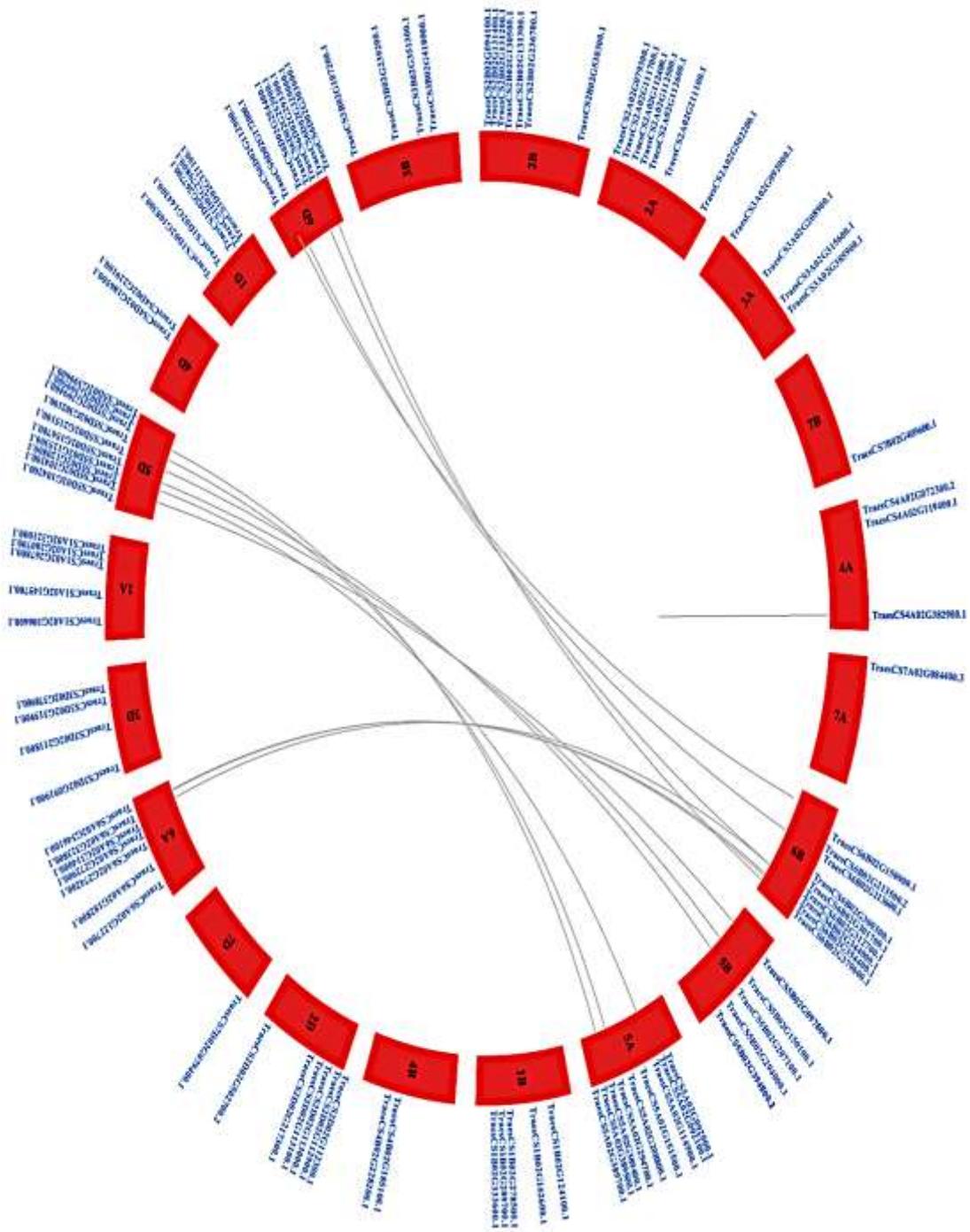


Figure 16: Linked genes

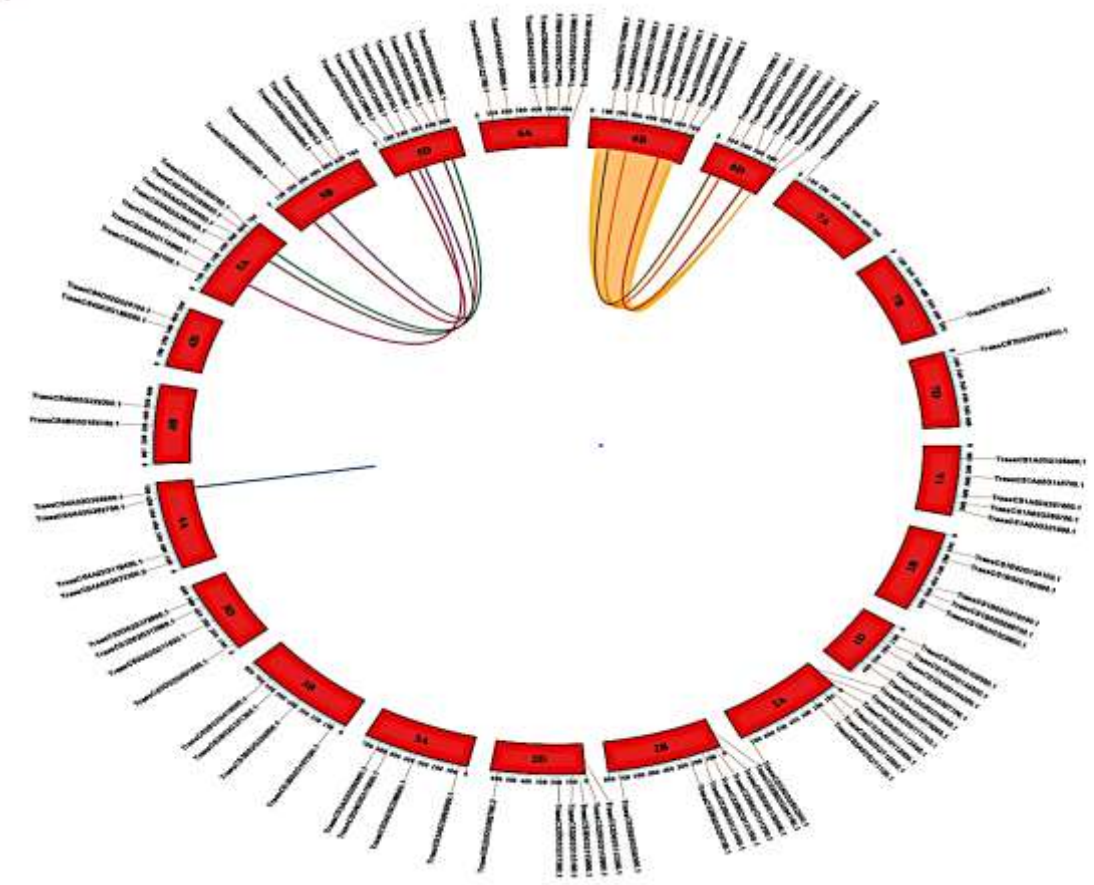


Figure 17:Circos plot for the duplicated gene pairs

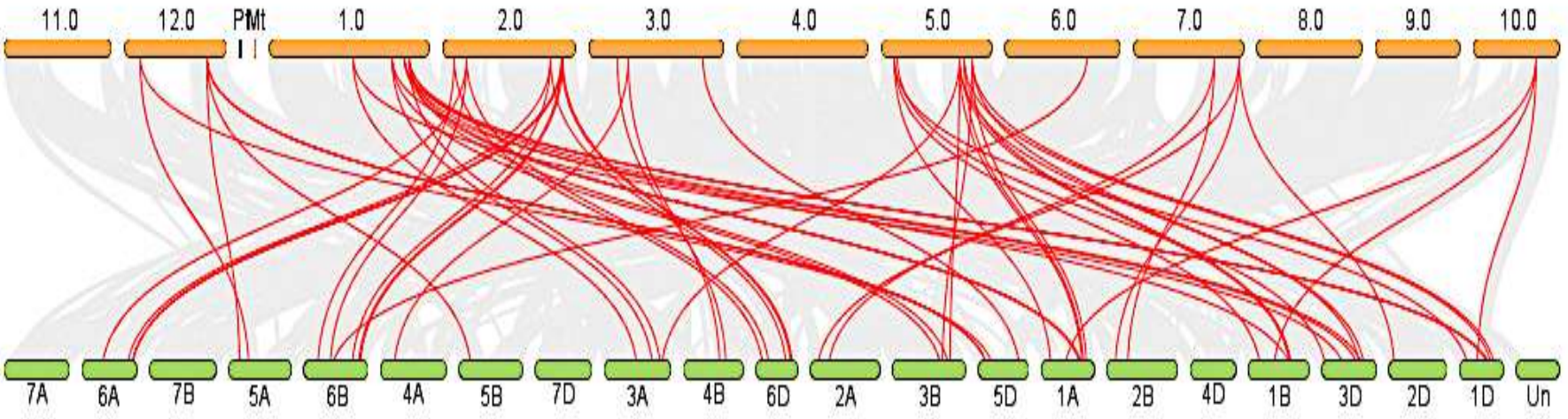


Figure 18:Syntenic relationship between rice and wheat. orange color indicates rice genome, whereas green color indicates wheat genome. the red lines indicate the paralogs duplicated over time.

4.5. Gene Ontology (GO) annotation and Subcellular localization

Predicting the subcellular localization of the identified genes revealed that these genes show expression in many compartments of the cell performing a variety of functions. The results revealed that these proteins are localized within the cytoplasm, mitochondria, endoplasmic reticulum, cytoskeleton, chloroplast, and nucleus. They are also present in extracellular matrix, part of the plasma membrane. some of them are also present in peroxisomes that indicates their function in anoxic conditions and peroxidase activity. Intensity of the heat map indicates that these genes are strongly localized within cytoplasm and chloroplast Their presence in Golgi complex, on the nuclear plasma membrane, vacuoles elucidated their enormous functioning in the various biological, molecular, and cellular processes. (Fig. 19)

Functional annotations were checked through the web server PANTHER. The different Go ontologies based on cellular processes, biological processes, and molecular functions were predicted. A total of 29 out of 107 *TaUSP* genes were mapped to eight biological processes as protein phosphorylation (GO:0006468), phosphorylation (GO:0016310), phosphate-containing compound metabolic process (GO:0006796), phosphorus metabolic process (GO:0006793), cellular protein modification process (GO:0006464), Protein modification process (GO:0036211), macromolecule modification (GO:0043412) and no sequence found a match with biological regulation (GO:0065007). **(**Fig. 20)

Furthermore, talking about the cellular processes, so out of 107 only 6 sequences were mapped with the PANTHER IDs. There were mapped to 9 cellular processes such as cytoplasm (GO:0005737), intracellular anatomical structure (GO:0005622), nucleus (GO:0005634), cellular anatomical entity (GO:0110165), cellular component (GO:0005575), intracellular membrane-bounded organelle (GO:0043231), membrane-bounded organelle (GO:0043227), intracellular organelle (GO:0043229), organelle (GO:0043226), whereas 67 sequences were categorized as unclassified. (Fig. 21)

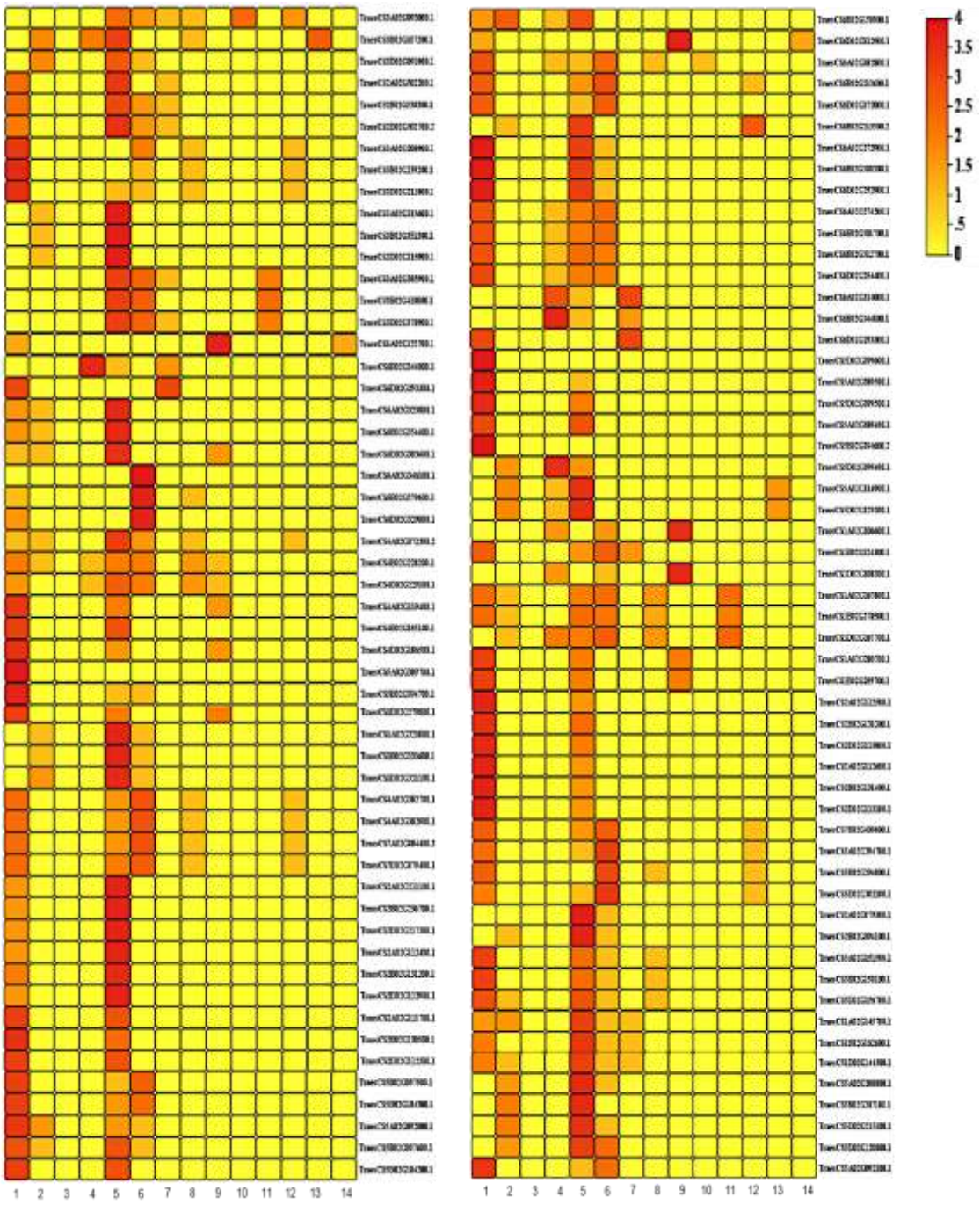


Figure 19: Subcellular Localization. The localization pattern of the genes is indicated. The number from 1-14; 1: Cytoplasm, 2: Mitochondria, 3: Endoplasmic reticulum, 4: Cytoskeleton, 5: Chloroplast, 6: Nucleus, 7: Extracellular, 8: Plasma, 9: Peroxisome, 10: Golgi complex, 11: Nuclear Plasma, 12: Vacuole, 13: Chloro- Mito, 14: Cytoskeleton-nucleus, indicates cellular compartments. Whereas the color gradient from red to yellow indicates the most to least presence using the log base 2 values

Lastly, talking about the molecular functions, sequences were mapped to functions like AMP binding (GO:0016208), protein kinase activity (GO:0004672), phosphotransferase activity, alcohol group as acceptor (GO:0016773), kinase activity (GO:0016301), transferase activity, transferring phosphorus-containing groups (GO:0016772), adenyl ribonucleotide Binding (GO:0032559), adenyl nucleotide binding (GO:0030554), purine ribonucleotide binding (GO:0032555), purine nucleotide binding (GO:0017076), ribonucleotide binding (GO:0032553), carbohydrate derivative binding (GO:0097367), ATP binding (GO:0005524), purine ribonucleoside triphosphate binding (GO:0035639), nucleotide binding (GO:0000166), nucleoside phosphate binding (GO:1901265), anion binding (GO:0043168), small molecule binding (GO:0036094), catalytic activity, acting on a protein (GO:0140096) and the last one being transferase activity (GO:0016740). (Fig. 22).

4.6. Cis-regulatory elements analysis of *USP* promoters of wheat

The promoter sequences downloaded from the database were used to identify different cis regulatory elements (CREs) which might control the activation of genes under certain conditions. So, this analysis revealed that a total 7 types or classes of CREs are present on the sequences such as light responsive elements (LREs), hormone responsive elements (HREs), development related elements, promoter elements, abiotic stress related elements, biotic stress-responsive elements and then was a 7^{th} class of elements with functions still unknown (Fig. 23). The result of each class is discussed below.

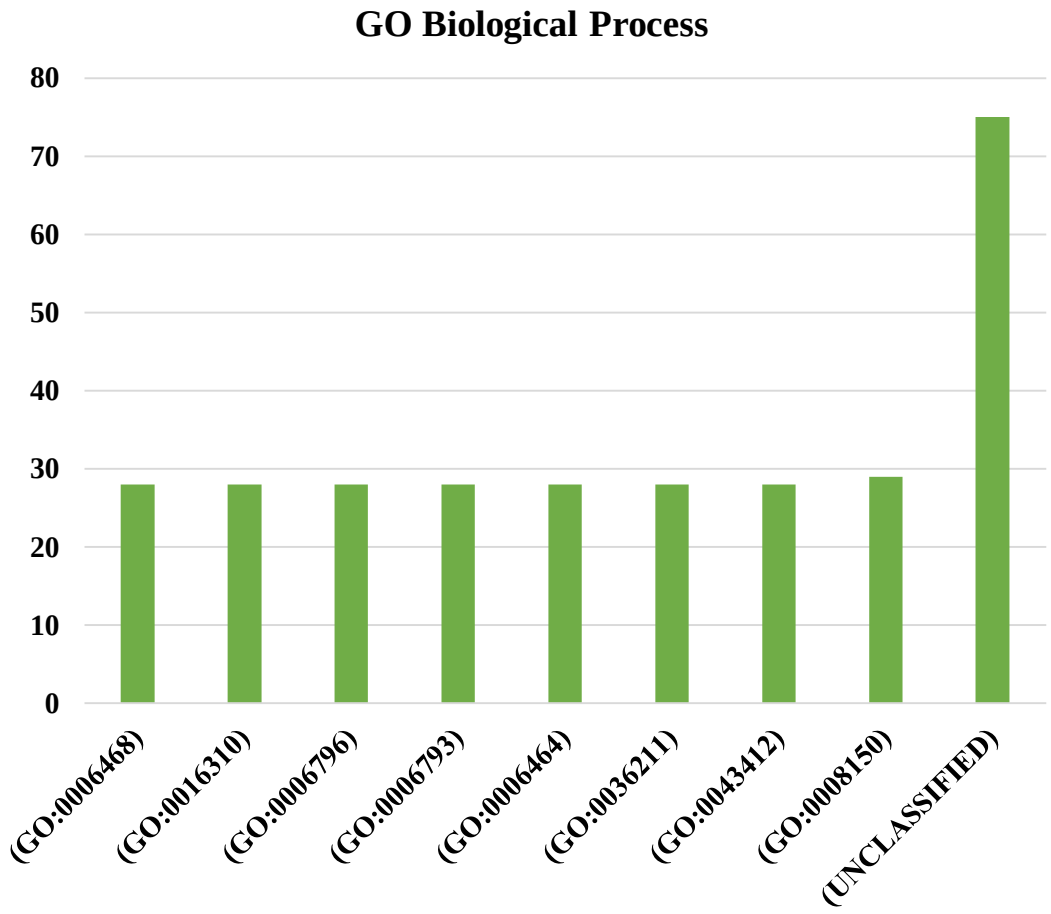


Figure 20: GO Biological Process

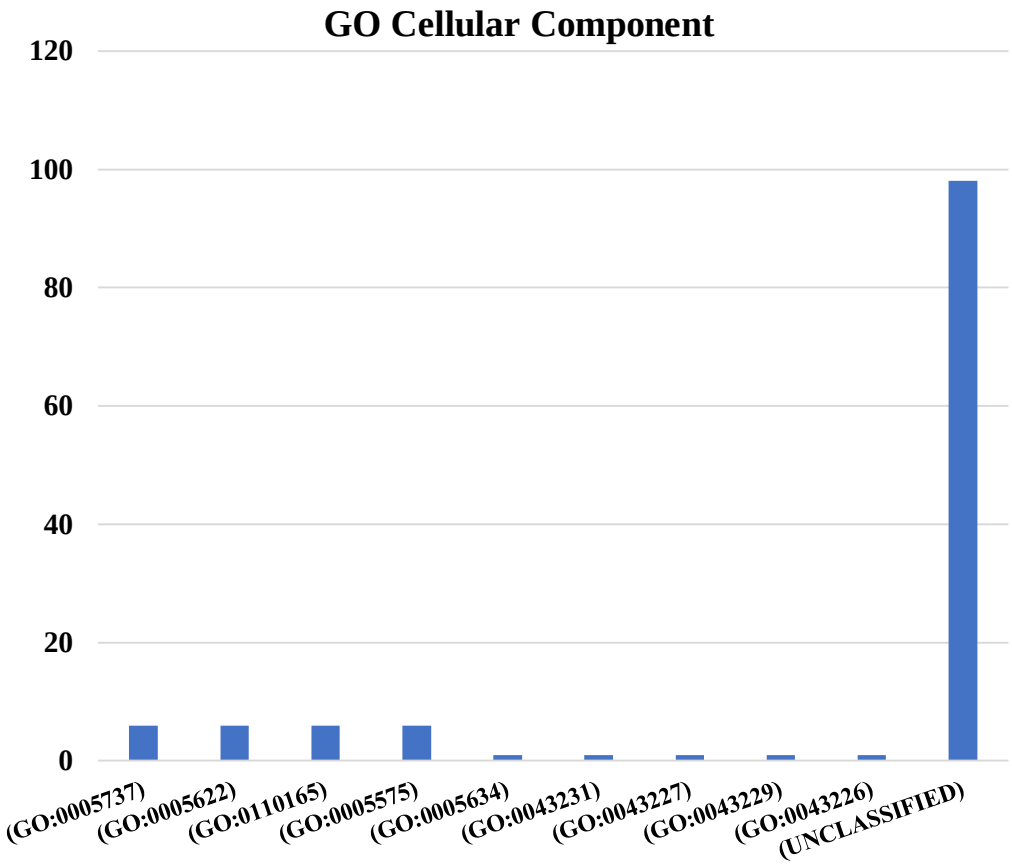


Figure 21: GO Cellular functions

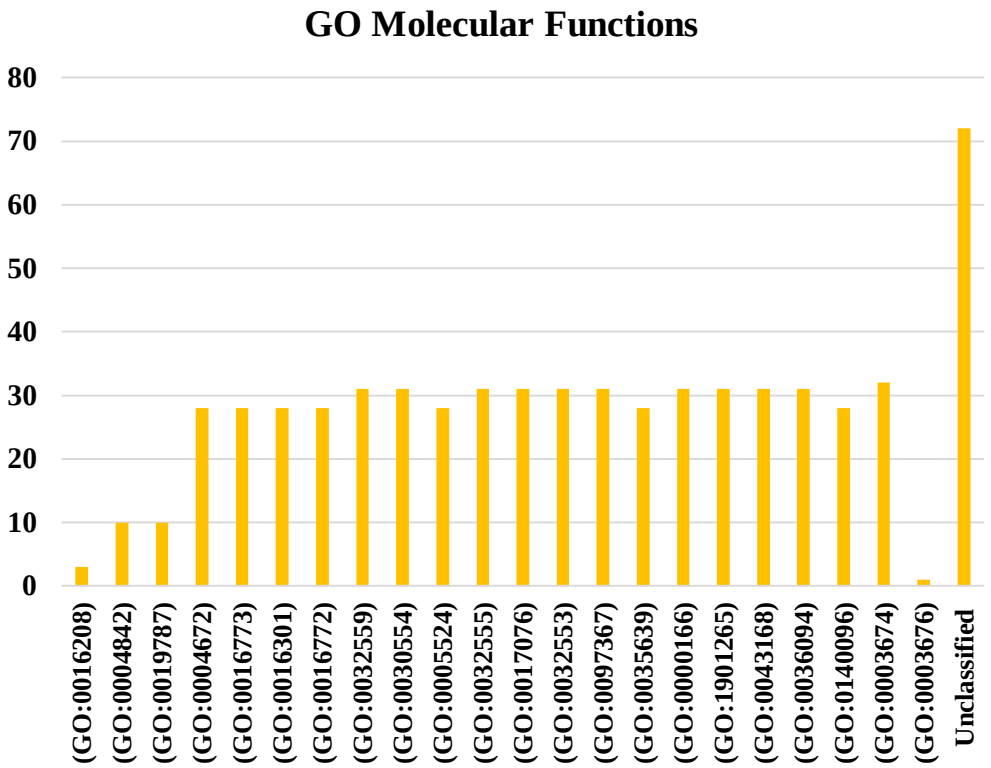


Figure 22: GO Molecular processes

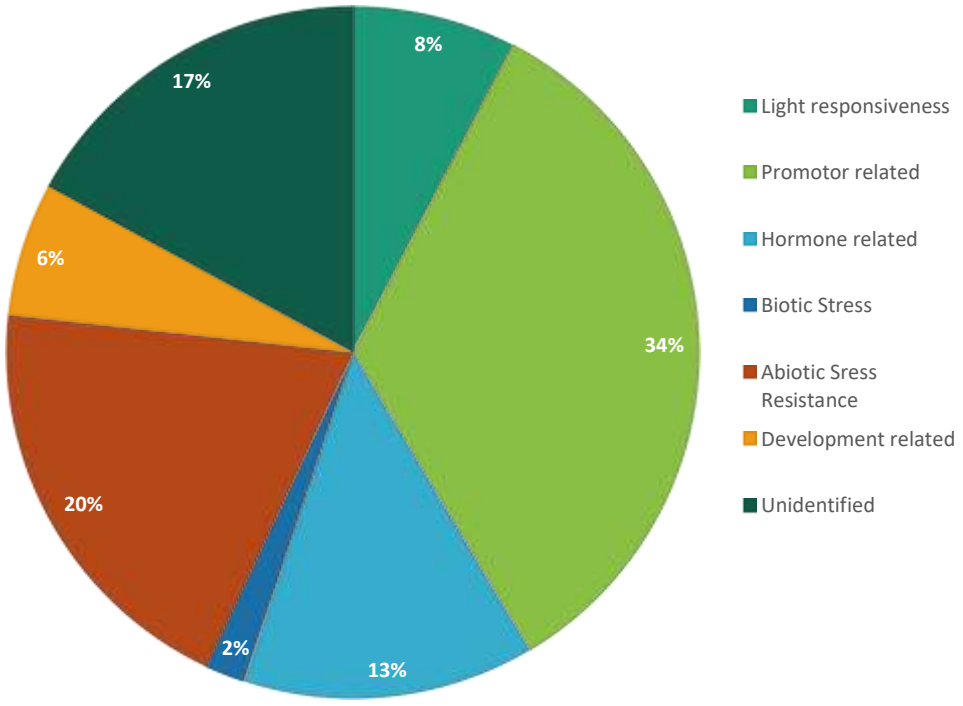


Figure 23: The relative percentages of 7 different cis-regulatory elements (CREs) in the 2.5kbp upstream of the TaUSP genes.

4.6.1. Promoter-related elements

There are 7 identified promoter-related elements present in these genes. These are CAAT-box, TATA-box, AT˜TATA-box, A-box, AT-rich element, TAAT, unnamed-1. One of the most abundant CREs are located upstream to start codon: TATA-box and CAAT-box at -35 and -10 positions respectively also called core promoters. In the present sequences, the CAAT-box is the most abundant. **(**Fig. 24)

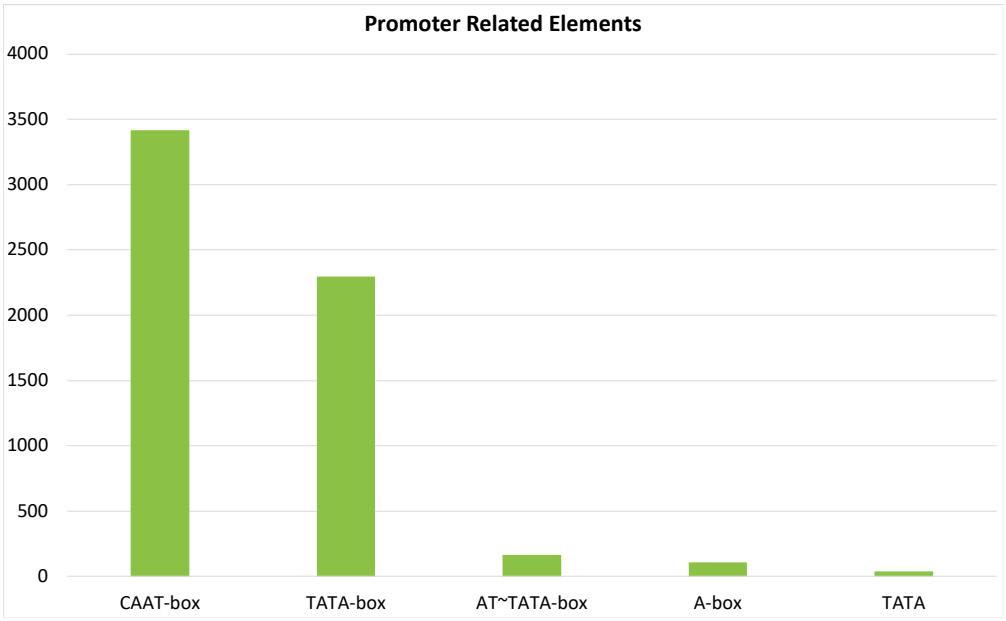


Figure 24: Promoter related elements. Six different elements are identified as CAAT-box, TATA-box, AT~TATA-box, A-box, TATA. CAAT-box and TATA-box are the two most abundant elements within the promoter regions.

4.6.2. Light responsive elements

Light responsive elements (LRE) identified were G-box, TCT-motif, AE-box, GATA-motif, Sp1, GT1-motif, Box 4, ACE, LAMP-element, I-box GA-motif, Gap-box, ATCT-motif, 3-AF1, L- box, chs-CMA1a, chs-CMA2, chs-Unit 1 m 1, AT1-motif, ATC-motif, TCCC-motif, Pc-CMA1c, GTGGC-motif. These are supposed to activate different photosystems and hence give response to light. The results also indicated two or more than two elements of these class are near each other, which indicates that more than one element is required for the activation of these promoters. G-box is the most abundant LRE present in the sequences. The presence of light responsive elements these sequences

strongly tells that these genes might be involved in the activation of pathways regulated by the light. **(**Fig. 25)

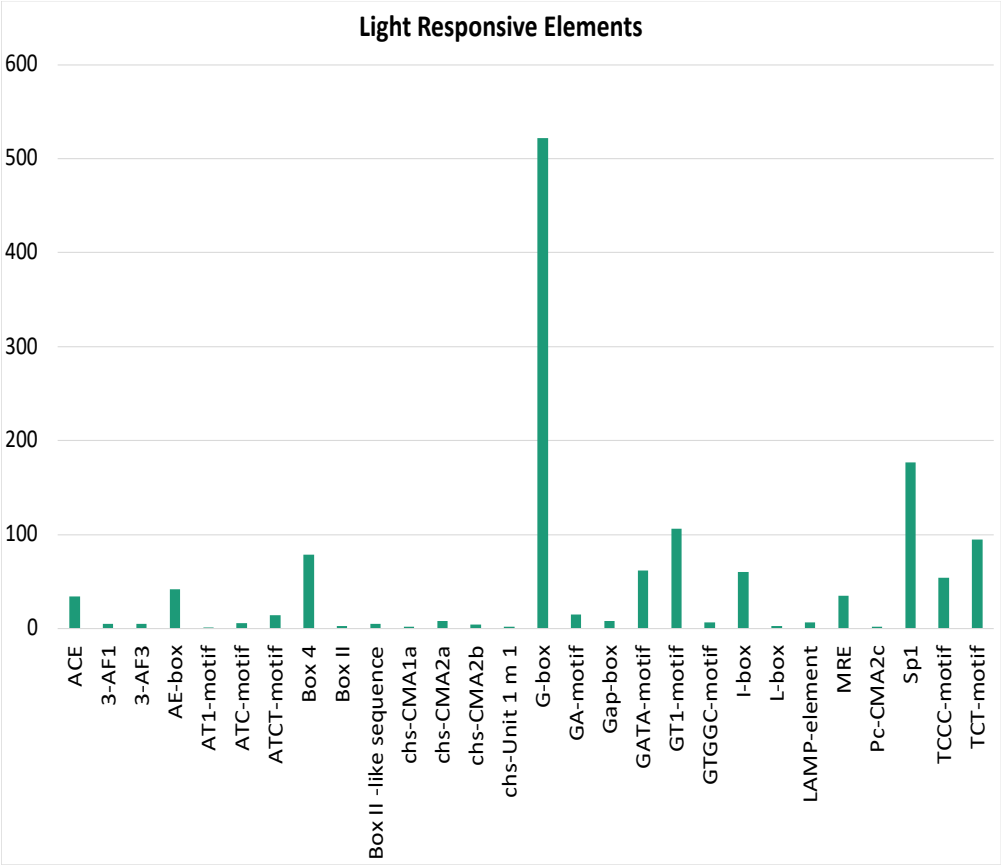


Figure 25: Light responsive elements. G-box, TCT-motif, AE-box, GATA-motif, Sp1, GT1-motif, Box 4, ACE, LAMP-element, I-box GA-motif, Gap-box, ATCT-motif, 3-AF1, L- box, chs-CMA1a, chs-CMA2, chs-Unit 1 m 1, AT1-motif, ATC-motif, TCCC-motif, Pc-CMA1c, GTGGC-motif

4.6.3. Hormone related elements

The upstream regions of the *USP*s also contained the regulatory elements related to the hormones. A total of 18 hormone related elements belonging to 6 different classes have been identified. Class 1 is abscisic acid (ABA) related: ABRE, ABRE3a, ABRE4, AT~ABRE, class 2 is auxin related: AuxRE, AuxRE-Core, TGA-box, TGA-element, CGTCA-motif (jasmonic acid), JERE and TGACG-motif are included. Class 4 belongs to salicylic acid (SA): TCA, TCA-element, and salicylic acid responsive elements (SARE), class 5 is gibberellic acid: P-Box, TATC-box, and gibberellic acid responsive element (GARE-motif) and the last class 6 is the ethylene (ETH) which has ethylene responsive elements (ERE). The results indicate that these proteins also get

activated by the change in hormones and effect these pathways giving responses according to them. **(**Fig. 26).

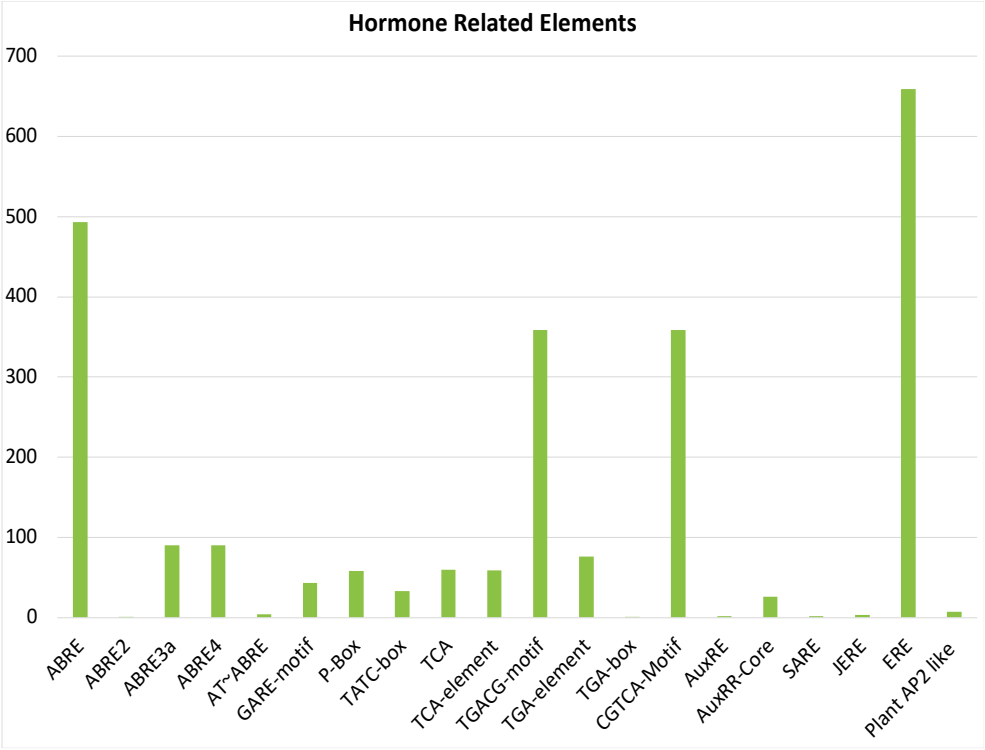


Figure 26: Hormone related elements. A total of 18 hormone related elements have been identified. Abscisic acid (ABA) related: ABRE, ABRE3a, ABRE4, AT~ABRE, auxin related: AuxRE, AuxRE-Core, TGA-box, TGA-element, CGTCA-motif, JERE and TGACG-motif, TCA

4.6.4. Development related responsive elements

A total of 25 development related responsive elements identified are AAGAA-motif, CCGTCC-box, AC-I, AC-II, as-I, CGGTCC-box, circadian, CAT-box, dOCT, E_2Fb, F-box, GCN4-motif, HD-zip 1, HD-zip 3, MSA like, NON, CARE, re2f1, RY element, NON-box, O2 Site, Unnamed__8, Unnamed__10, Unnamed__12 and Unnamed__14. These factors play a distinguished role in in the cellular development process including the cell cycle and the cell proliferation pathways. Some of the genes might also control circadian pathways, and the pathways involved in zein metabolism. These motifs also indicate that these might play role in tissue specific expression of genes related to developmental process. **(**Fig. 27)**.**

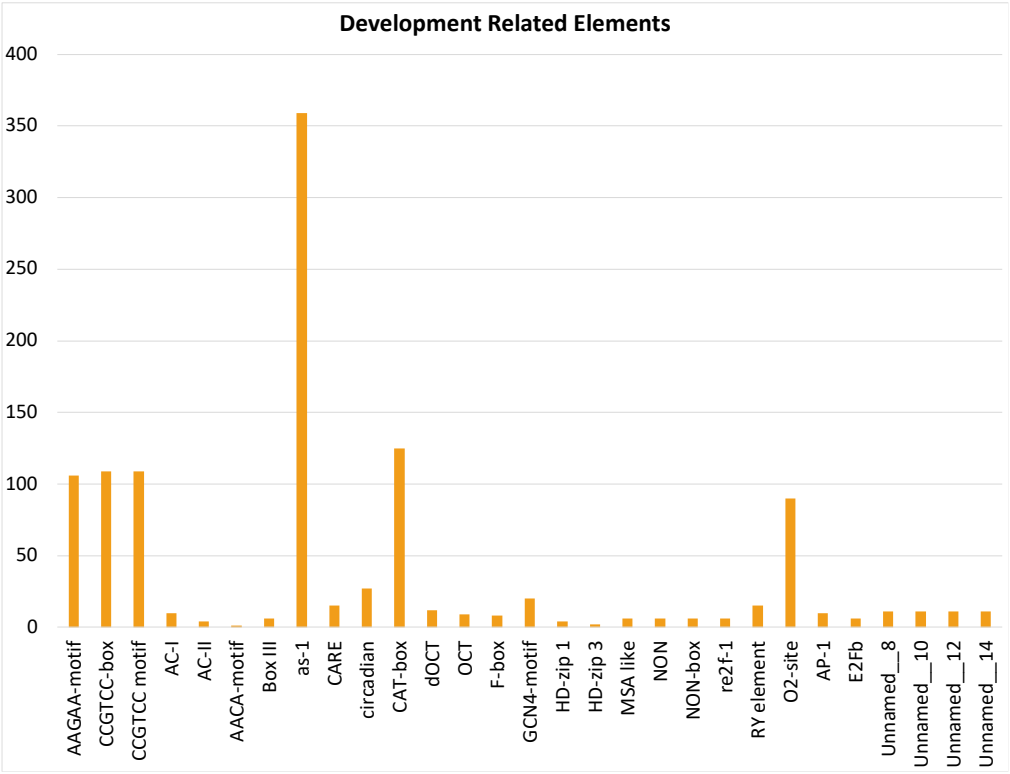


Figure 27: Development related elements. A total of 25 development related responsive elements are identified i.e., AAGAA-motif, CCGTCC-box, AC-I, AC-II, as-I, CGGTCC-box, circadian, CAT-box, dOCT, E2Fb, F-box, GCN4-motif, HD-zip 1, HD-zip 3, MSA like, N

4.6.5. Abiotic stress-responsive elements

A total of 18 out of 113 elements were identified as abiotic stress responsive elements. CCAAT-box, Drought responsive elements DRE core, DRE 1, GC-motif, LTR; low temperature response also called response to cold, MBS, MBS 1, MYB along with its binding and recognition sites, MYC, MYB like site, STRE is for low pH, osmotic pressure, AT-rich sequence and ARE. (Fig. 28)**.**

4.6.6. Biotic stress-related elements

Among 113 sequences 4 biotic stress related elements were identified and participate in wound healing and response to pathogen attack. The identified motifs are Box S, W-box, WRE 3 and WUN-motif. **(**Fig. 29).

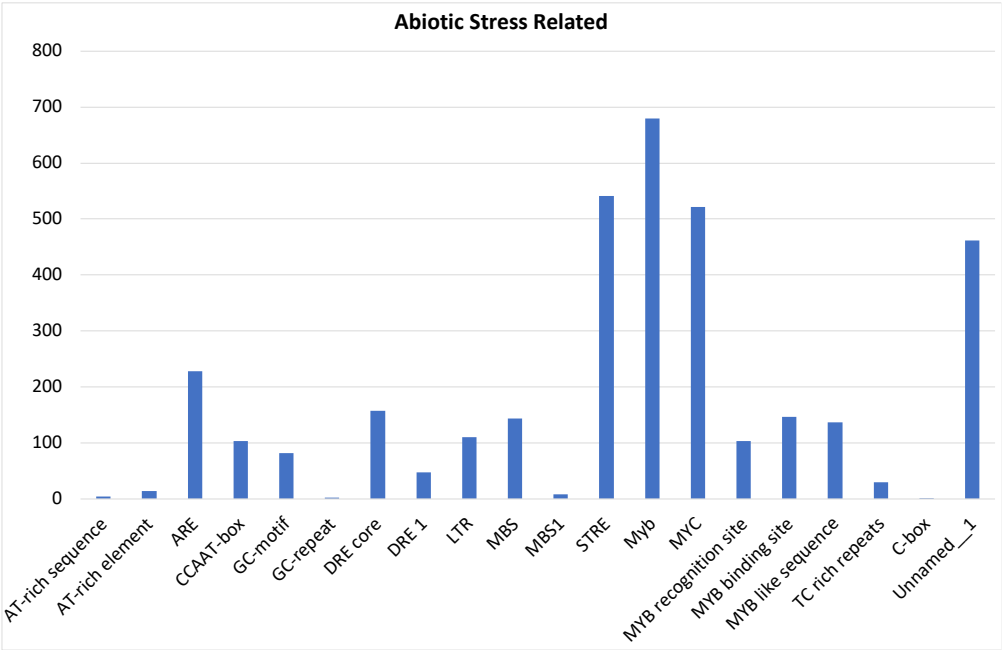


Figure 28: Abiotic stress related elements. A total of 18 out of 113 elements were identified as abiotic stress responsive elements. CCAAT-box, Drought responsive elements DRE core, DRE 1, GC-motif, LTR

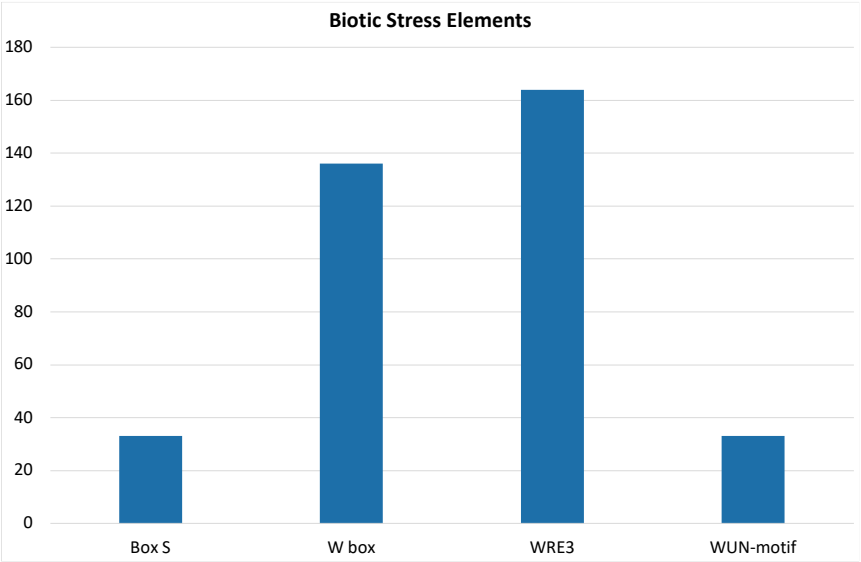


Figure 29: Biotic stress related elements. The identified motifs are Box S, W-box, WRE 3 and WUN-motif.

4.6.7. Unidentified

The last class contains the unnamed unidentified sequences with no known function but are present in quite a number. They are Unnamed__2,

Unnamed__4, Unnamed__6, Unnamed__16. Unnamed__4 with a motif sequence CTCC is the most abundant. **(**Fig. 30)**.**

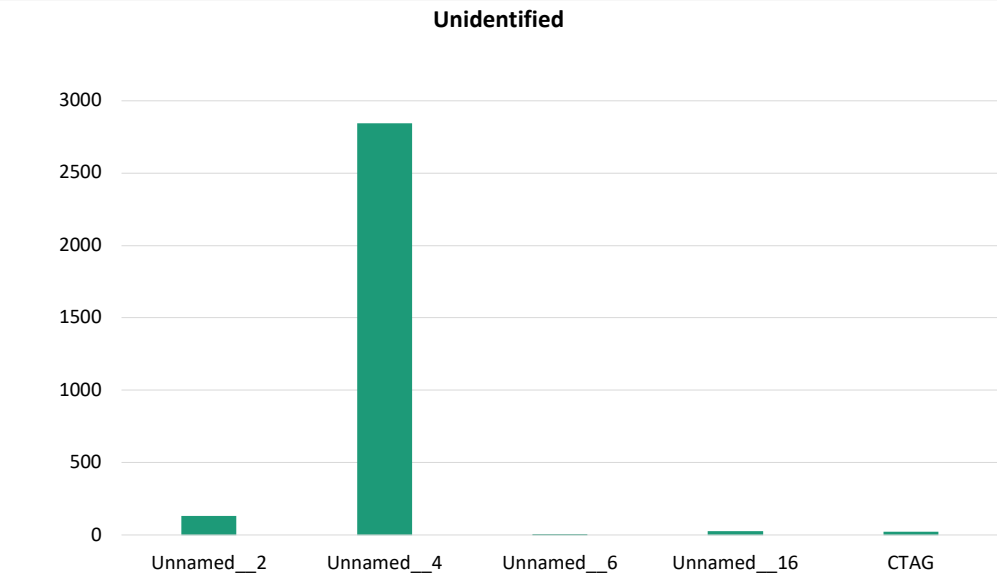


Figure 30: Unidentified. They are Unnamed__2, Unnamed__4, Unnamed__6, Unnamed__16. Unnamed__4 with a motif sequence CTCC is the most abundant.

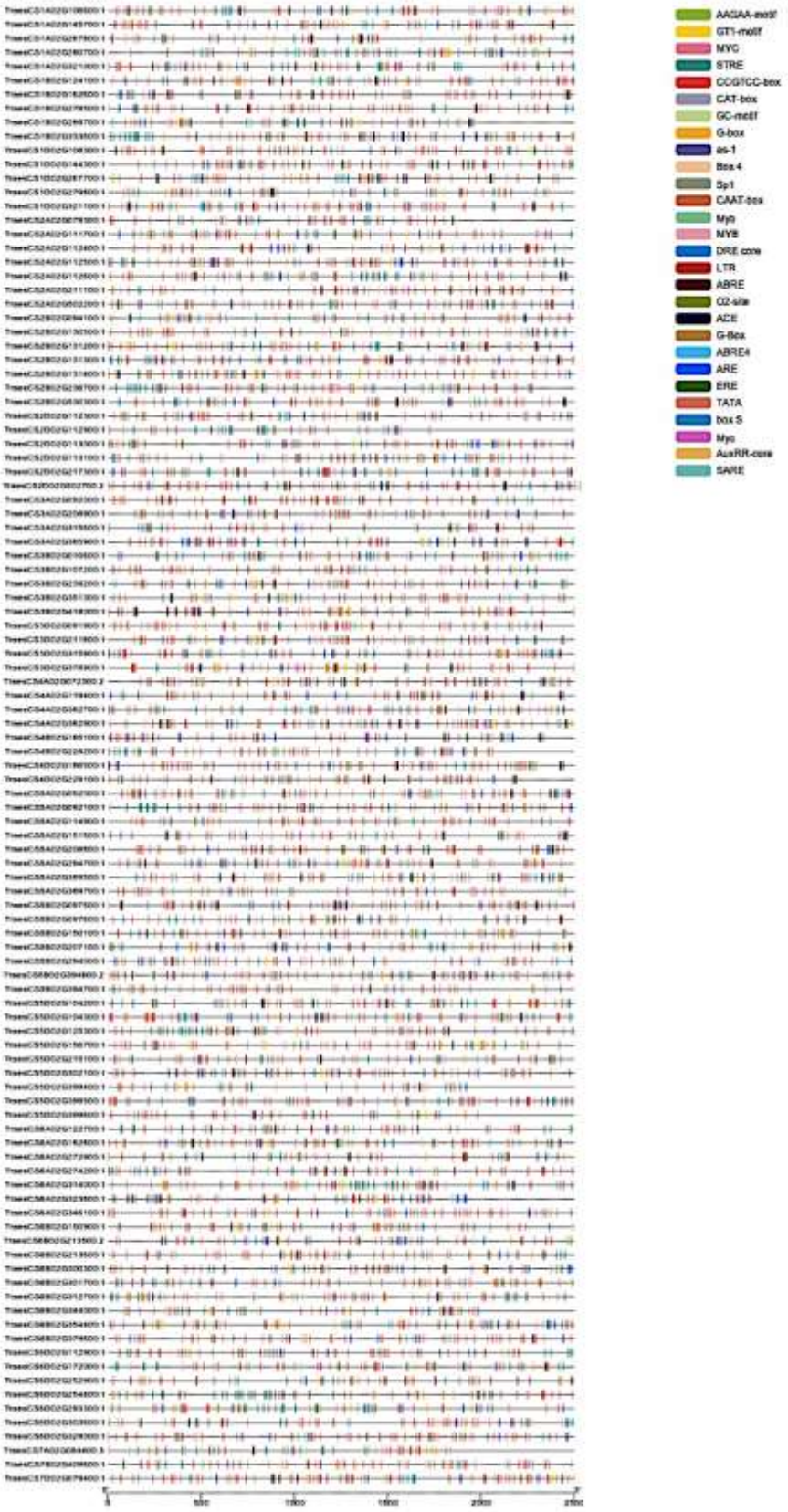


Figure 31: Cis-regulatory elements (CREs). Graphical representations of abundant CREs present in the promoter region. The image was taken using the simple bio sequence illustrator in TBtools. Using the start and end sites of the motifs

4.7. Identification of Glycosylation and phosphorylation sites in *TaUSP* genes

Phosphorylation is one of the significant post translational modifications and plays a key role in activation, deactivation, and regulation of the cellular pathways. In this process Protein kinases phosphorylates amino acids such as serine, threonine, and tyrosine. A substantial number of phosphorylated sites are predicted in 107 sequences. *TraesCS7A02G084400.3(TaUSP27A)* has the highest number of phosphorylation sites and *TraesCS1D02G108300.1* has the lowest sites 122 and 5, respectively. **Table 4.**

Glycosylation is another post translational modifications that help proteins in proper folding and giving them their characteristics functionality and plays role in the stability of these protein structures. This study has revealed 81 glycosylation sites in 19 out of 34 *TaUSP's* have been predicted. Highest glycosylation sites were present in *TaUSP10* i.e., 7; followed by *TaUSP32* with 6 sites, *TaUSP27a* and *TaUSP35* with 4 sites, *TaUSP8* and *TaUSP27b* with 3 sites, *TaUSP4* and *TaUSP14* with 2 sites while the rest 11 *TaUSP* have only 1 glycosylation site. N-glycosylation score more than 0.5 and jury score 9/9 indicates high specificity to glycosylation event and predicts that protein might have stable glycosylated mediated structure **Table 5.**

4.8. Protein modelling, disordered region analysis and prediction of binding sites

Three dimensional structures of four *TaUSP* genes *TaUSP4, TaUSP10, TaUSP21* and *TaUSP30,* chosen from the two groups, were modelled, using crystal structure of human IRAK1 (PDB_6BFN.1. A), crystal structure of USP from Arabidopsis Thaliana At3g01520 (PDB_2GM3.1. A) and hypothetical protein (PDB_1MJH.1. A) respectively. Secondary structure of all these proteins was predicted using the SOPMA webserver and a-helices, b-sheets, extended strands, and random coils were observed in between 31.91%-46.3%, 4.59%-6.63%, 9.85%-20.48% and 37.22%-45.91% respectively. **Table 6**.

The *TaUSP4* and *TaUSP30* has 19% and15.2% disordered region whereas, *TaUSP10 and TaUSP21* has no disordered region and solely contains the

functional domains. The predicted protein models were validated by Ramachandran plot (Fig. 32-35). The favored regions were above 90%. **Table 7.**

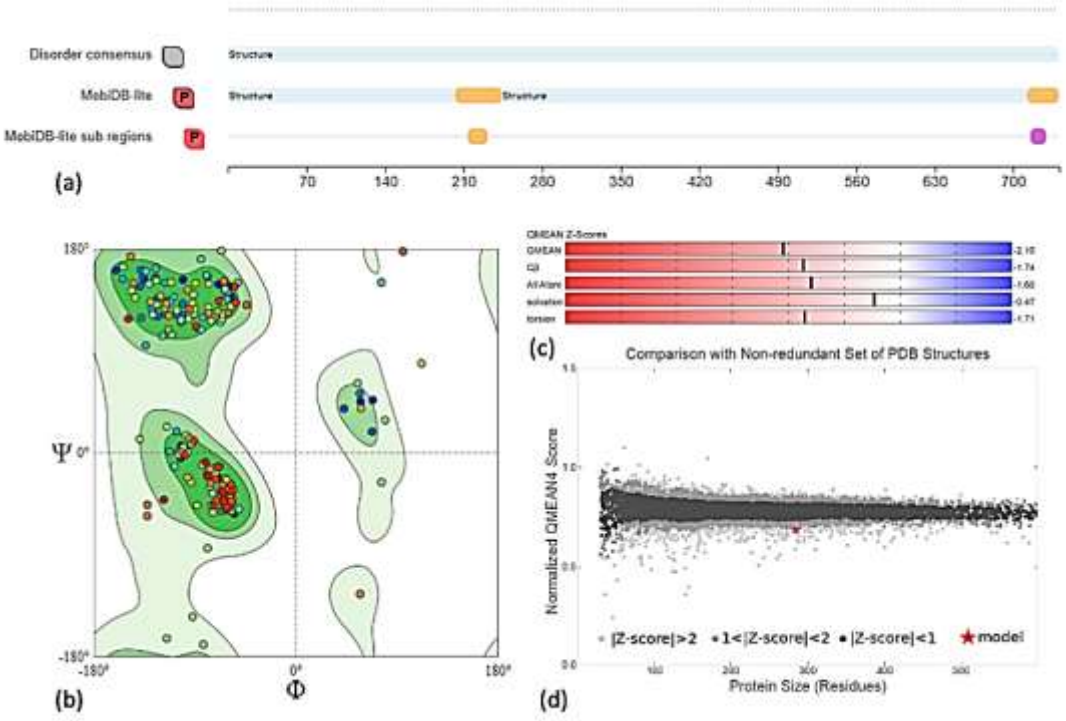


Figure 32:TaUSP4 a). Disordered regions, b). Ramachandran Plot, c) Q-mean Z-scores and d). relative protein size in Non redundant protein databank.

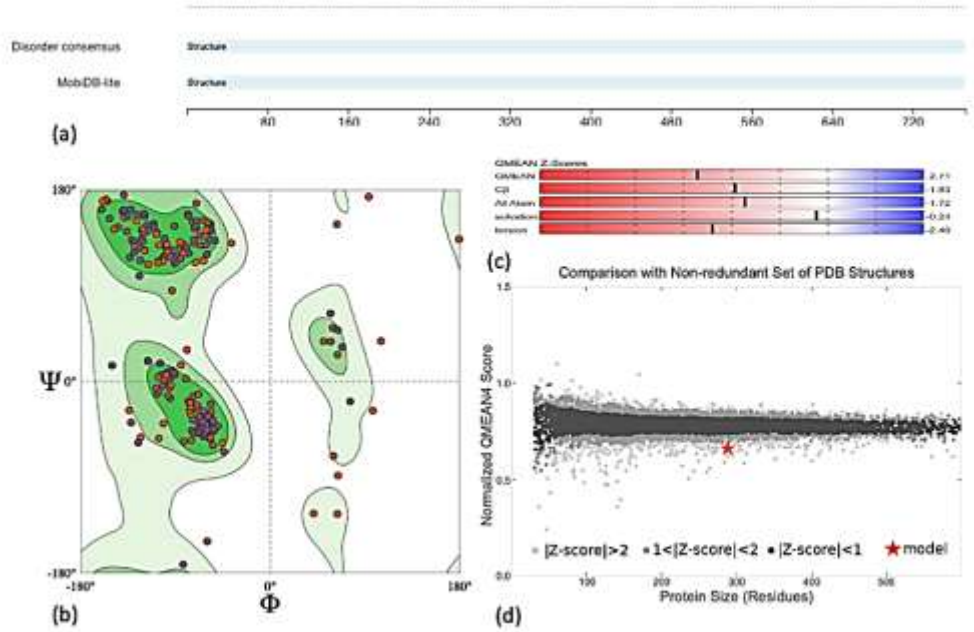


Figure 33: ***TaUSP10*** a). Disordered regions, b). Ramachandran Plot, c) Q-mean Z-scores and d). relative protein size in Non redundant protein databank.

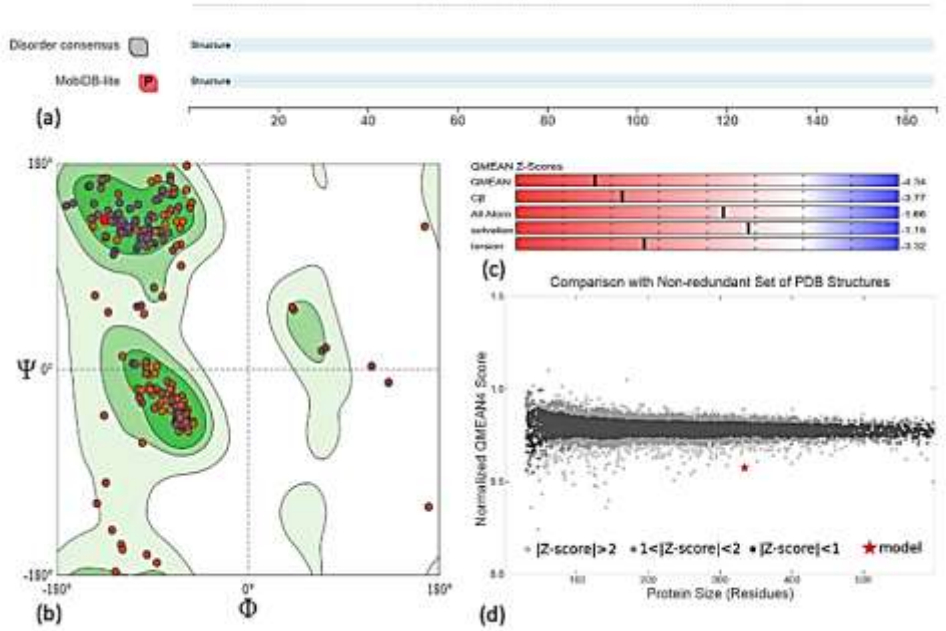


Figure 34: ***TaUSP 21*** a). Disordered regions, b). Ramachandran Plot, c) Q-mean Z-scores and d). relative protein size in Non redundant protein databank.

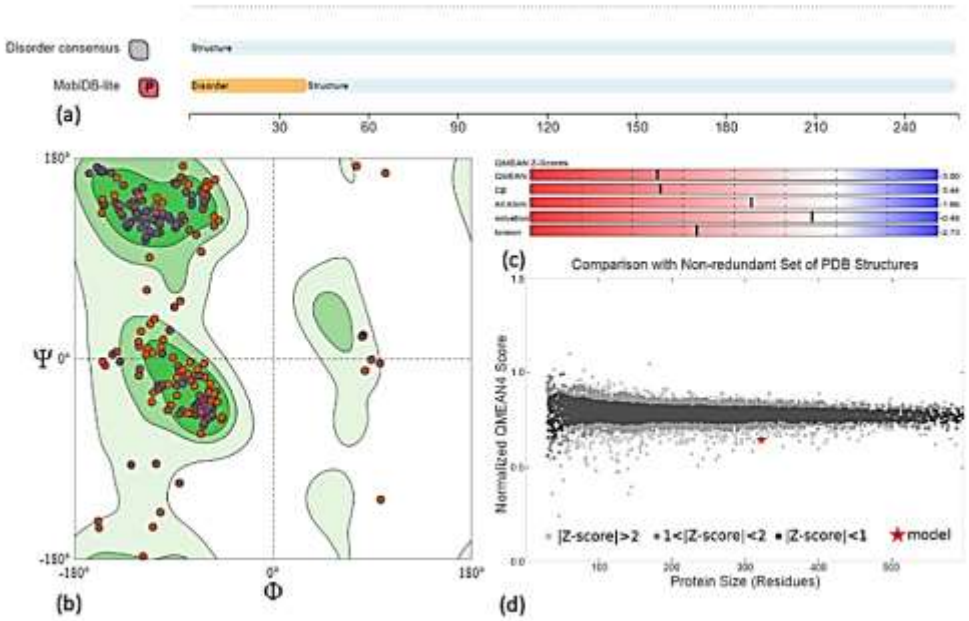


Figure 35:TaUSP30 a). Disordered regions, b). Ramachandran Plot, c) Q-mean Z-scores and d). relative protein size in Non redundant protein databank.

Two binding pockets were predicted in the models plotted (Fig. 36-39). The binding pockets contained several functional amino acids such as alanine (Ala), aspartate (Asp), asparagine (Asn), cysteine (Cys), glutamine (Gln), glycine (Gly), histidine (His), isoleucine (Ile), leucine (Leu), serine (Ser), threonine (Thr), valine (Val), tryptophan (Trp) and tyrosine (Tyr) **Table 8.** The presence of the serine and threonine residues in the binding pockets shows that these might be the sites for phosphorylation and the presence of asparagine give insight into N-glycosylation sites.

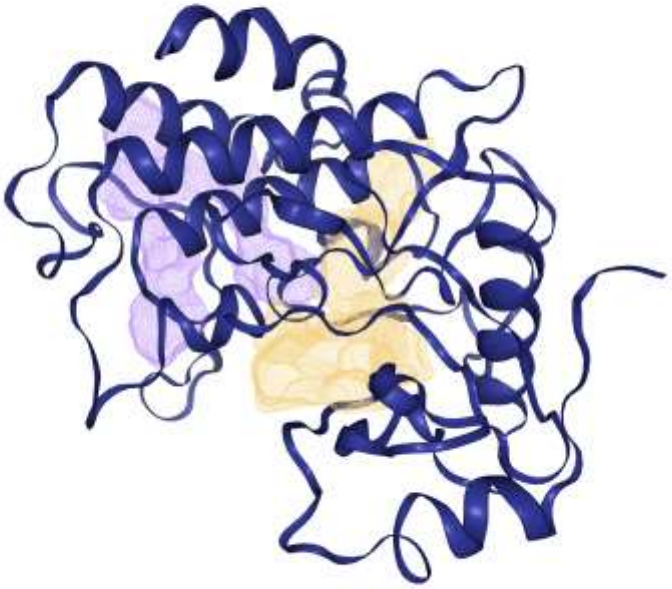


Figure 36:Predicted protein model of TaUSP4 along with pocket binding site in color skin and purple.

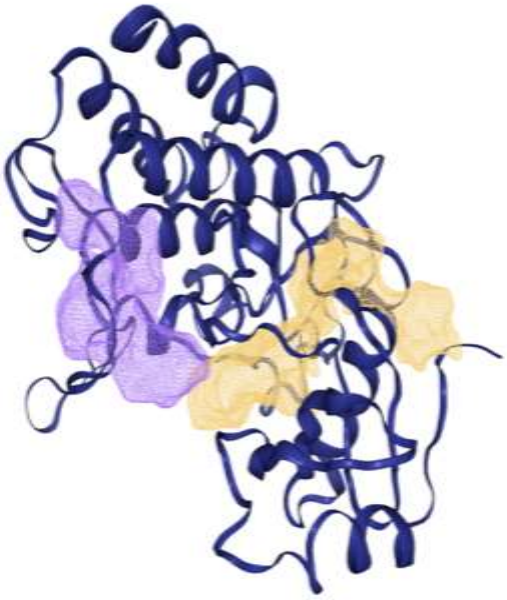


Figure 37: Predicted protein model of TaUSP10 along with pocket binding site in color skin and purple.

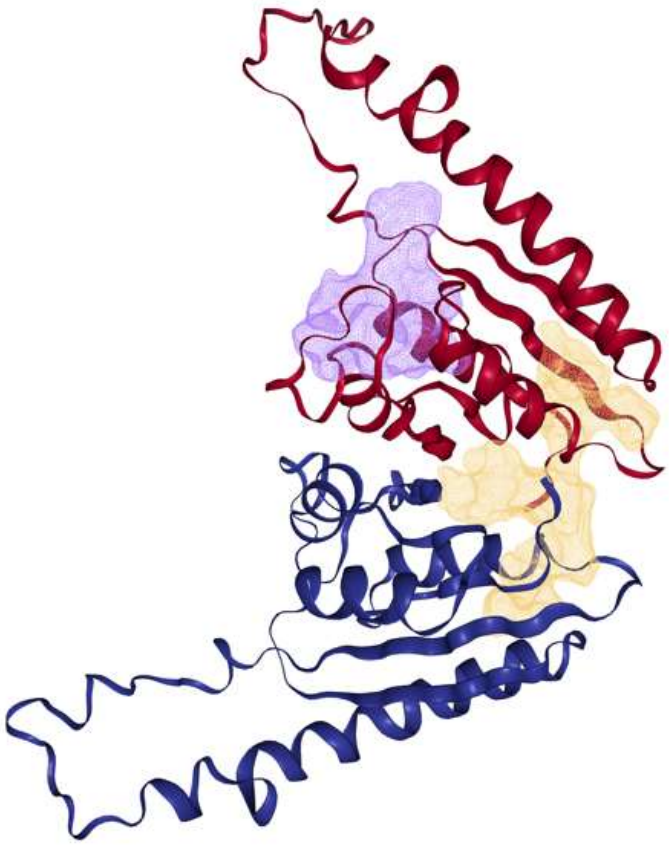


Figure 38:Predicted protein model of TaUSP21 along with pocket binding site in color skin and purple.

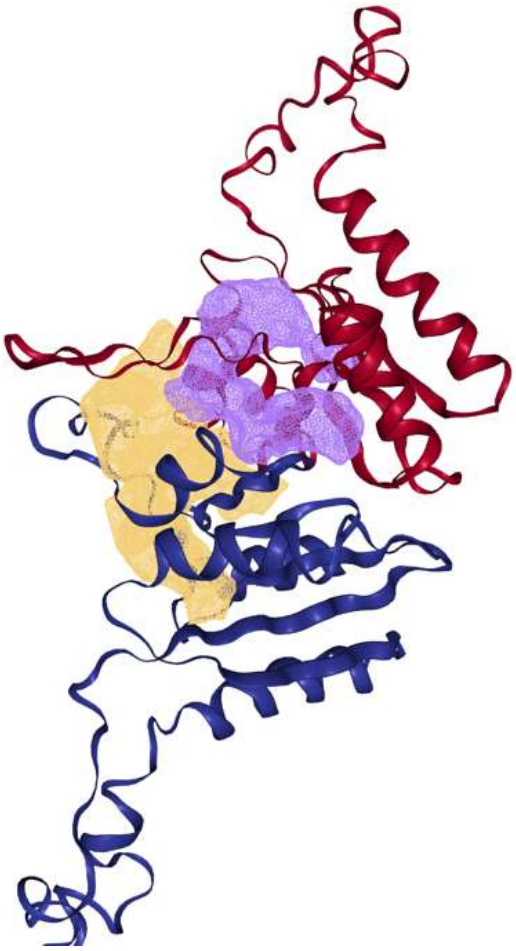


Figure 39:Predicted protein model of TaUSP30 along with pocket binding site in color skin and purple.

Chapter Five: Discussion

Plants being immobile need to withstand extreme conditions such, drought, salinity, extreme temperatures etc. These stresses all lead to stunted growth and development patterns. Advanced molecular studies have allowed researchers to better understand the phenomenon regarding abiotic stress responses in plants (Zhang et al., 2022).

Crop varieties have been improved greatly due to advanced genomics, but similar practices and protocols are difficult to handle for Triticum species due to extremely large and complex genome (Walkowiak et al., 2020). There are several types of stresses that harm the plants throughout their life. (Fig. 40) shows some general signaling pathways involved in response theses abiotic stresses. Different proteins families play role in the0abiotic stress management e.g.

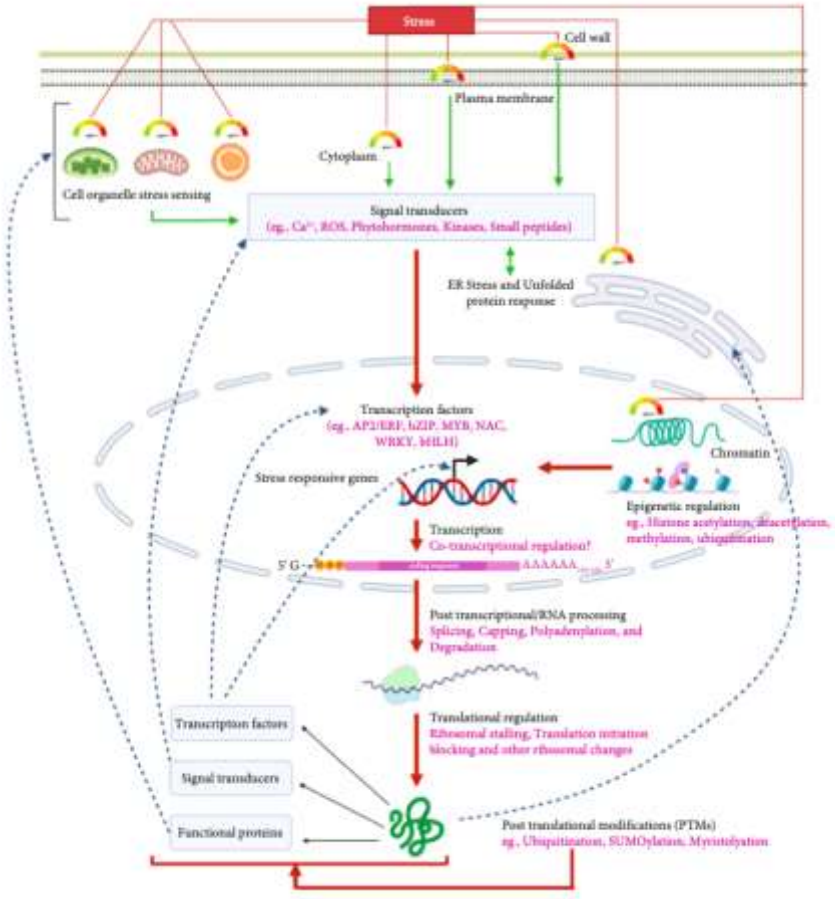


Figure 40: Generalized signaling pathway in response to stress (Lohani et al., 2022)

Universal stress proteins belong to the Adenine Nucleotide Alpha hydrolase superfamily. USPs have been proven to be stimulated in response to several abiotic stresses such as increased temperature, ion imbalances, salinity, drought, waterlogging, etc. Previous studies have revealed the presence of variety of *USP* gene family in different plants, which includes *Oryza sativa japonica*, barley, *Zea mays*, Cotton, Tomato, Potato, Pigeon pea, tobacco, *Medicago falcata* (Isokpehi et al., 2011). In plants there are usually twenty to fifty genes. *Triticum aestivum* and *Brassica napus* have more than 100 genes. The presented study identified a total of 107 *TaUSP* genes in the wheat genome **Table 2**.

Chi- square and t- test revealed significant gene numbers and the way genes are distributed in the genome (Fig 10). Chi-square test analyzed the presence of independent variables (Glen., 2022). Gene structure and identified motifs truly explains the domain architecture of the genes (Fig. 12). Arabia et al. reported the same results in the OsUSP genes. A conserved ATP-binding motif G-2x-G-x-G(S/T) (Freestone et al., 1997) (Fig. 13). Structural analysis of the identified genes along with proteins also showed that domain architecture plays a vital role in this regard (Tkaczuk et al., 2013)**.** The phylogenetic tree showed two main clades that match with domain architecture of *E. coli*. Group A contains a single *USP* domain. On the other hand, group B has peptides with longer lengths as well as an additional functional kinase domain. There are other functional domains other than the protein kinase and gives diversity the functioning of *USP* genes. and protects plants against a variety of abiotic stresses, within these, there are also two classes the ATP binding class and the non-ATP binding class. (Fig.11) (Li, W.-T. et al., 2010; Wang et al., 2017). The identified genes are spanned on all the chromosomes of wheat (Fig. 15(a, b)). Duplication analysis revealed that most of the 15 out of 107 genes were duplicated segmentally and only 2 were tandem duplications. Both types of duplications play a significant in increased number of these genes in wheat (Fig. 17). Further analysis revealed that high number of orthologs are present between rice and wheat. Duplication played

the part in expansion of these genes within monocot lineages (Panchy et al., 2016) Evolutionary studies reveals collinearity in the gene arrangements and helps us infer the ancestral relationships within taxa (Tang et al., 2008). Synteny analysis showed collinearity within monocot genomes (Fig. 18).

Subcellular localization of these genes was predominant in cytoplasm, chloroplast, mitochondria some in nucleus and a few in peroxisomes. This also shows their role in certain cellular pathways and their functioning in response to stress (Fig. 19).(Cui et al., 2021). The role of these genes in redox reactions is proven in many studies(Gonzali et al., 2015). Studies on the expression pattern of these genes indicate regulation and modulation under different stresses. These findings truly describe former studies of stress-specific modulation and regulation of USP genes and their overexpression, results in increased thermotolerance and enhanced tolerance to osmotic stress in *A. thaliana*(Gonzali et al., 2015). Moreover, the *TaUSP* might have similar functionalities in response to stress. To top this argument, we also found several stress-responsive regulatory elements present in the upstream regions of *TaUSP* genes (Fig. 23). These include Sp1, ABRE, LTRs, ARE, MBS, GT motif and TC- rich repeats. Previous studies on tomato USP genes shows that wild variety of these are induced under the influence of ABA, drought, salt, temperature, and ethylene stress(Loukehaich et al., 2012). *OsUSP* genes were upregulated in response to cold temperatures whereas downregulated in response to ABA(Sauter et al., 2002). Promoters of *AtUSP* were highly induced by different stressors and showed multi-stress response. *SbUSP* of *Salicornia brachiata* was expressed in response to temperature, drought, and salt stress(Udawat et al., 2014). There was another study, in which cotton *USP* promoters were activated in response to heavy metals, salts, osmotic stress and gibberellic acid stress(Gorshkova & Pojidaeva, 2021). In general, these studies provide proofs of differential modulation and regulation of *TaUSP* genes under abiotic stresses. And these promoters can induce stress resistance and are excellent to produce variable stress responsive expression of these genes in transgenic plants (Fig. 24-30).

To deeply understand this gene family, proteins of these sequences were also studied. The proper functioning and stability of the protein products depends on the mechanisms of post translational modifications. The two major modifications are glycosylation and phosphorylation. Glycosylation plays role in the proper folding of proteins, strength and signaling pathways(Jayaprakash & Surolia, 2017), on the other hand phosphorylation has its vital role in proteins activation and deactivation through certain alterations in the conformation. It has a major function in the signaling pathways and metabolic processes(Proud, 2019).

Naturally, these modified proteins are reported to be significant in several biological processes. The studies have revealed that modifications in different plants occurs in response to osmotic as well as cold stress. The predicted glycosylation sites give insight into the functional stability of these genes in response to stress(Jayaprakash & Surolia, 2017). There are several studies in which phosphorylation of the proteins resulted in combating against several stressors(Damaris & Yang, 2021). The position specific kinase activity of these proteins can help us further verify their roles using different laboratory techniques. The protein modelling results revealed that active binding pockets of the *TaUSP* contains several functional amino acids including alanine, valine, serine, threonine and glutamine and glutamic acid. Serine and tyrosine residues are essential phosphorylation sites

Chapter Six: Conclusions

This study has investigated the *USP* gene family in wheat. A total of 107 sequences on wheat's ABD genome are reported. MSA revealed the presence of conserved ATP-binding motif in the class 2 *TaUSP*. Phylogenetic analysis revealed the evolutionary relationship of wheat, brachypodium and rice. Four different domain architectures were identified having correlation with their functioning. Evolutionary studies showed that *TaUSP* genes have been conserved over the years and correlate with the monocot and dicot genomes. Duplications over time has led to a variety of these genes in wheat. The presence of the several regulatory elements; such as sp1, TC- rich repeats, ABRE, SARE and JARE etc,. in upstream regions helped further in verifying the multi-stress nature of these proteins. The focus of the study was to conduct a detailed investigation on the functional and structural attributes of *TaUSPs*. It revealed the presence of various functions among the Ta*USP* genes of wheat as well as their characteristic multi-stress nature. Further investigations, such as expression analysis studies will confirm the multi stress nature of these genes. Despite this, additional studies are needed to spot the diverse roles of these genes by producing either knockout lines or overexpressing *TaUSP* genes either solitary or along with different groups. By using advanced tools and techniques, the identified *TaUSP* genes can be used not just to improve wheat cultivars but also will be of great interest to protect other economically important crops to make them more resilient to stress.

References

Ahmad, M., Ali, Q., Hafeez, M., & Malik, A. (2021). Improvement for biotic and abiotic stress tolerance in crop plants. *Biological and Clinical Sciences Research Journal, 2021*(1).

Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., . . . Poland, J. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science, 361*(6403).

Arabia, S., Sami, A. A., Akhter, S., Sarker, R. H., & Islam, T. (2021). Comprehensive in-silico characterization of Universal Stress Proteins in rice (Oryza sativa L.) with insight into their stress-specific transcriptional modulation. *Frontiers in plant science*, 1589.

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research, 43*(W1), W39-W49.

Basso, M. F., Arraes, F. B. M., Grossi-de-Sa, M., Moreira, V. J. V., Alves-Ferreira, M., & Grossi-de-Sa, M. F. (2020). Insights into genetic and molecular elements for transgenic crop development. *Frontiers in plant science*, 509.

Bhuria, M., Goel, P., Kumar, S., & Singh, A. K. (2019). Genome-wide identification and expression profiling of genes encoding universal stress proteins (USP) identify multi-stress responsive USP genes in Arabidopsis thaliana. *Plant Physiology Reports, 24*(3), 434-445.

Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics, 4*(6), 1633-1649.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'Amore, R., Allen, A. M., . . . Bolser, D. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature, 491*(7426), 705-710.

Burke, J. J. (2001). Identification of genetic diversity and mutations in higher plant acquired thermotolerance. In: Wiley Online Library.

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular plant, 13*(8), 1194-1202.

Chi, Y. H., Koo, S. S., Oh, H. T., Lee, E. S., Park, J. H., Phan, K. A. T., . . . Chae, H. B. (2019a). The physiological functions of universal stress proteins and their molecular mechanism to protect plants from environmental stresses. *Frontiers in plant science, 10*, 750.

Chi, Y. H., Koo, S. S., Oh, H. T., Lee, E. S., Park, J. H., Phan, K. A. T., . . . Chae, H. B. (2019b). The physiological functions of universal stress proteins and their molecular mechanism to protect plants from environmental stresses. *Frontiers in plant science*, 750.

Chou, M.-X., Wei, X.-Y., Chen, D.-S., & Zhou, J.-C. (2007). A novel nodule-enhanced gene encoding a putative universal stress protein from Astragalus sinicus. *Journal of plant physiology, 164*(6), 764-772.

Cui, X., Zhang, P., Hu, Y., Chen, C., Liu, Q., Guan, P., & Zhang, J. (2021). Genome-wide analysis of the Universal stress protein A gene family in Vitis and expression in response to abiotic stress. *Plant Physiology and Biochemistry, 165*, 57-70.

Damaris, R. N., & Yang, P. (2021). Protein phosphorylation response to abiotic stress in plants. *Plant Phosphoproteomics*, 17-43.

Desheva, G. (2013). Morphological and agronomical characterization of common wheat landraces (Triticum aestivum L.) from the National Wheat Collection of Bulgaria. *Emirates Journal of Food and Agriculture, 26*, 164-169. doi:10.13140/2.1.4523.5846

Dey, S., Malviya, R., & Gayen, D. (2022). Understanding Abiotic Stress Tolerance in Cereals Through Genomics and Proteomics Approaches. In *Omics Approach to Manage Abiotic Stress in Cereals* (pp. 73-102): Springer.

Edde, P. A. (2022). 4 - Arthropod pests of cotton (Gossypium hirsutum L.). In P. A. Edde (Ed.), *Field Crop Arthropod Pests of Economic Importance* (pp. 208-274): Academic Press.

Elhosseiny, N. M., Amin, M. A., Yassin, A. S., & Attia, A. S. (2015). Acinetobacter baumannii universal stress protein A plays a pivotal role in stress response and is essential for pneumonia and sepsis pathogenesis. *International journal of medical microbiology, 305*(1), 114-123.

Flowers, T. J., & Colmer, T. D. (2008). Salinity tolerance in halophytes. *New Phytologist*, 945-963.

Freestone, P., Nyström, T., Trinei, M., & Norris, V. (1997). The universal stress protein, UspA, of Escherichia coli is phosphorylated in response to stasis11Edited by J. Karn. *Journal of Molecular Biology, 274*(3), 318-324. doi:https://doi.org/10.1006/jmbi.1997.1397

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res, 31*(13), 3784-3788. doi:10.1093/nar/gkg563

Geourjon, C., & Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics, 11*(6), 681-684.

Glen., S. (2022). Chi-Square Statistic: How to Calculate It / Distribution" From StatisticsHowTo.com: Elementary Statistics for the rest of us. Retrieved from https://www.statisticshowto.com/probability-and-statistics/chi-square/

Gonzali, S., Loreti, E., Cardarelli, F., Novi, G., Parlanti, S., Pucciariello, C., . . . Perata, P. (2015). Universal stress protein HRU1 mediates ROS homeostasis under anoxia. *Nature Plants, 1*(11), 1-9.

Gorshkova, D., & Pojidaeva, E. (2021). Members of the Universal Stress Protein Family are Indirectly Involved in Gibberellin-Dependent Regulation of Germination and Post-Germination Growth. *Russian Journal of Plant Physiology, 68*(3), 451-462.

Gou, L., Zhuo, C., Lu, S., & Guo, Z. (2020). A Universal Stress Protein from Medicago falcata (MfUSP1) confers multiple stress tolerance by regulating antioxidant defense and proline accumulation. *Environmental and Experimental Botany, 178*, 104168.

Guan, J., Garcia, D. F., Zhou, Y., Appels, R., Li, A., & Mao, L. (2020). The battle to sequence the bread wheat genome: a tale of the three kingdoms. *Genomics, proteomics & bioinformatics, 18*(3), 221-229.

Gupta, R., Jung, E., & Brunak, S. (2004). NetNGlyc 1.0 Server. *Center for biological sequence analysis, technical university of Denmark available from: http://www. cbs. dtu dk/services/NetNGlyc*.

Gutiérrez-Beltrán, E., Personat, J. M., de la Torre, F., & Del Pozo, O. (2017). A universal stress protein involved in oxidative stress is a phosphorylation target for protein kinase CIPK6. *Plant physiology, 173*(1), 836-852.

Hassan, S., Ahmad, A., Batool, F., Rashid, B., & Husnain, T. (2021). Genetic modification of Gossypium arboreum universal stress protein (GUSP1) improves drought tolerance in transgenic cotton (Gossypium hirsutum). *Physiology and Molecular Biology of Plants, 27*(8), 1779-1794.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research, 35*(suppl_2), W585-W587.

Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., & Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics, 31*(8), 1296-1297.

Huang, H., Ullah, F., Zhou, D.-X., Yi, M., & Zhao, Y. (2019). Mechanisms of ROS regulation of plant development and stress responses. *Frontiers in plant science, 10*, 800.

Isokpehi, R. D., Simmons, S. S., Cohly, H. H., Ekunwe, S. I., Begonia, G. B., & Ayensu, W. K. (2011). Identification of drought-responsive universal stress proteins in viridiplantae. *Bioinformatics and biology insights, 5*, BBI. S6061.

Jayaprakash, N. G., & Surolia, A. (2017). Role of glycosylation in nucleating protein folding and stability. *Biochemical Journal, 474*(14), 2333-2347.

Jones, M. A., Raymond, M. J., & Smirnoff, N. (2006). Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in Arabidopsis. *The Plant Journal, 45*(1), 83-100.

Jung, Y. J., Melencion, S. M. B., Lee, E. S., Park, J. H., Alinapon, C. V., Oh, H. T., . . . Lee, S. Y. (2015). Universal stress protein exhibits a redox-dependent chaperone function in Arabidopsis and enhances plant tolerance to heat shock and oxidative stress. *Frontiers in plant science, 6*, 1141.

Kim, H., Goo, E., Kang, Y., Kim, J., & Hwang, I. (2012). Regulation of universal stress protein genes by quorum sensing and RpoS in Burkholderia glumae. *Journal of bacteriology, 194*(5), 982-992.

KlYOSUE, T. r., & SHINOZAKI, K. Function and Regulation of Genes that are Induced by Dehydration Stress in Arabidopsis thaliana. *Structure, 10*, 11.

Kumar, P., Yadava, R., Gollen, B., Kumar, S., Verma, R. K., & Yadav, S. (2011). Nutritional contents and medicinal properties of wheat: a review. *Life Sciences and Medicine Research, 22*(1), 1-10.

Kyozuka, J. (2014). Grass inflorescence: basic structure and diversity. *Advances in botanical research, 72*, 191-219.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948. doi:10.1093/bioinformatics/btm404

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., . . . Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research, 30*(1), 325-327.

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research, 49*(W1), W293-W296.

Li, B., Gao, K., Ren, H., & Tang, W. (2018). Molecular mechanisms governing plant responses to high temperatures. *Journal of Integrative Plant Biology, 60*(9), 757-779.

Li, W., Zhao, F. a., Fang, W., Xie, D., Hou, J., Yang, X., . . . Lv, S. (2015). Identification of early salt stress responsive proteins in seedling roots of upland cotton (Gossypium hirsutum L.) employing iTRAQ-based proteomic technique. *Frontiers in plant science, 6*, 732.

Li, W.-T., Wei, Y.-M., Wang, J.-R., Liu, C.-J., Lan, X.-J., Jiang, Q.-T., . . . Zheng, Y.-L. (2010). Identification, localization, and characterization of putative USP genes in barley. *Theoretical and Applied Genetics, 121*(5), 907-917.

Liu, W.-T., Karavolos, M. H., Bulmer, D. M., Allaoui, A., Hormaeche, R. D. C. E., Lee, J. J., & Khan, C. A. (2007). Role of the universal stress protein UspA of Salmonella in growth arrest, stress and virulence. *Microbial pathogenesis, 42*(1), 2-10.

Lohani, N., Singh, M. B., & Bhalla, P. L. (2022). Biological parts for engineering abiotic stress tolerance in plants. *BioDesign Research, 2022*.

Loukehaich, R., Wang, T., Ouyang, B., Ziaf, K., Li, H., Zhang, J., . . . Ye, Z. (2012). SpUSP, an annexin-interacting universal stress protein, enhances drought tolerance in tomato. *Journal of experimental botany, 63*(15), 5593-5606.

Melencion, S. M. B., Chi, Y. H., Pham, T. T., Paeng, S. K., Wi, S. D., Lee, C., . . . Lee, S. Y. (2017). RNA chaperone function of a universal stress protein in Arabidopsis confers enhanced cold stress tolerance in plants. *International journal of molecular sciences, 18*(12), 2546.

Merkouropoulos, G., Andreasson, E., Hess, D., Boller, T., & Peck, S. (2008). An Arabidopsis Protein Phosphorylated in Response to Microbial Elicitation, AtPHOS32, Is a Substrate of MAP Kinases 3 and 6. *The Journal of biological chemistry, 283*, 10493-10499. doi:10.1074/jbc.M800735200

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research, 47*(D1), D419-D426.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models

and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution, 37*(5), 1530-1534. doi:10.1093/molbev/msaa015

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E. L. L., . . . Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research, 49*(D1), D412-D419. doi:10.1093/nar/gkaa913

Nachin, L., Brive, L., Persson, K.-C., Svensson, P., & Nyström, T. (2008). Heterodimer formation within universal stress protein classes revealed by an in silico and experimental approach. *Journal of molecular biology, 380*(2), 340-350.

Nicholas, K. B. (1997). Genedoc: a tool for editing and annoting multiple sequence alignments. *http://wwwpscedu/biomed/genedoc*.

Panchy, N., Lehti-Shiu, M., & Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant physiology, 171*(4), 2294-2316.

Park, S.-C., Jung, Y. J., Lee, Y., Kim, I. R., Seol, M.-A., Kim, E.-J., . . . Lee, J. R. (2017). Functional characterization of the Arabidopsis universal stress protein AtUSP with an antifungal activity. *Biochemical and biophysical research communications, 486*(4), 923-929.

Piovesan, D., Necci, M., Escobedo, N., Monzon, A. M., Hatos, A., Mičetić, I., . . . Tosatto, S. C. E. (2021). MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Res, 49*(D1), D361-d367. doi:10.1093/nar/gkaa1058

Proud, C. G. (2019). Phosphorylation and signal transduction pathways in translational control. *Cold Spring Harbor perspectives in biology, 11*(7), a033050.

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., . . . Itoh, T. (2013). Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol, 54*(2), e6. doi:10.1093/pcp/pcs183

Sampaio, M., Neves, J., Cardoso, T., Pissarra, J., Pereira, S., & Pereira, C. (2022). Coping with Abiotic Stress in Plants—An Endomembrane Trafficking Perspective. *Plants, 11*(3), 338.

Sattar, A., Sher, A., Ijaz, M., Ul-Allah, S., Rizwan, M. S., Hussain, M., . . . Cheema, M. A. (2020). Terminal drought and heat stress alter physiological and biochemical attributes in flag leaf of bread wheat. *PLOS ONE, 15*(5), e0232974. doi:10.1371/journal.pone.0232974

Sauter, M., Rzewuski, G., Marwedel, T., & Lorbiecke, R. (2002). The novel ethylene-regulated gene OsUsp1 from rice encodes a member of a

plant protein family related to prokaryotic universal stress proteins. *Journal of experimental botany, 53*(379), 2325-2331.

Senapati, N., Stratonovitch, P., Paul, M. J., & Semenov, M. A. (2019). Drought tolerance during reproductive development is important for increasing wheat yield potential under climate change in Europe. *Journal of experimental botany, 70*(9), 2549-2560.

Senthil-Kumar, M., Kumar, G., Srikanthbabu, V., & Udayakumar, M. (2007). Assessment of variability in acquired thermotolerance: potential option to study genotypic response and the relevance of stress genes. *Journal of plant physiology, 164*(2), 111-125.

Sinha, P., Pazhamala, L. T., Singh, V. K., Saxena, R. K., Krishnamurthy, L., Azam, S., . . . Varshney, R. K. (2016). Identification and validation of selected universal stress protein domain containing drought-responsive genes in Pigeonpea (Cajanus cajan L.). *Frontiers in plant science, 6*, 1065.

Song, S., Tian, D., Zhang, Z., Hu, S., & Yu, J. (2018). Rice genomics: over the past two decades and into the future. *Genomics, proteomics & bioinformatics, 16*(6), 397-404.

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., & Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science, 320*(5875), 486-488.

Tkaczuk, K. L., A. Shumilin, I., Chruszcz, M., Evdokimova, E., Savchenko, A., & Minor, W. (2013). Structural and functional insight into the universal stress protein family. *Evolutionary applications, 6*(3), 434-449.

Tremonte, P., Succi, M., Coppola, R., Sorrentino, E., Tipaldi, L., Picariello, G., . . . Fraternali, F. (2016). Homology-based modeling of universal stress protein from Listeria innocua up-regulated under acid stress conditions. *Frontiers in microbiology, 7*, 1998.

Udawat, P., Jha, R. K., Sinha, D., Mishra, A., & Jha, B. (2016). Overexpression of a cytosolic abiotic stress responsive universal stress protein (SbUSP) mitigates salt and osmotic stress in transgenic tobacco plants. *Frontiers in plant science, 7*, 518.

Udawat, P., Mishra, A., & Jha, B. (2014). Heterologous expression of an uncharacterized universal stress protein gene (SbUSP) from the extreme halophyte, Salicornia brachiata, which confers salt and osmotic tolerance to E. coli. *Gene, 536*(1), 163-170.

Volkamer, A., Griewel, A., Grombacher, T., & Rarey, M. (2010). Analyzing the Topology of Active Sites: On the Prediction of Pockets and

Subpockets. *Journal of Chemical Information and Modeling, 50*(11), 2041-2052. doi:10.1021/ci100241y

Vollmer, A. C., & Bark, S. J. (2018). Twenty-five years of investigating the universal stress protein: function, structure, and applications. *Advances in applied microbiology, 102*, 1-36.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., . . . Pozniak, C. J. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature, 588*(7837), 277-283. doi:10.1038/s41586-020-2961-x

Wang, X.-F., Su, J., Yang, N., Zhang, H., Cao, X.-Y., & Kang, J.-F. (2017). Functional characterization of selected universal stress protein from Salvia miltiorrhiza (SmUSP) in Escherichia coli. *Genes, 8*(9), 224.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research, 46*(W1), W296-W303. doi:10.1093/nar/gky427

Wiborg, J., O'Shea, C., & Skriver, K. (2008). Biochemical function of typical and variant Arabidopsis thaliana U-box E3 ubiquitin-protein ligases. *Biochemical Journal, 413*(3), 447-457.

Yadav, S., Modi, P., Dave, A., Vijapura, A., Patel, D., & Patel, M. (2020). Effect of abiotic stress on crops. *Sustainable Crop Production*.

Yang, Y., & Guo, Y. (2018). Unraveling salt stress signaling in plants. *Journal of Integrative Plant Biology, 60*(9), 796-804.

Yates, A. D., Allen, J., Amode, R. M., Azov, A. G., Barba, M., Becerra, A., . . . Chakiachvili, M. (2022). Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Research, 50*(D1), D996-D1003.

Zahur, M., Maqbool, A., Irfan, M., Barozai, M. Y. K., Rashid, B., Riazuddin, S., & Husnain, T. (2009). Isolation and functional analysis of cotton universal stress protein promoter in response to phytohormones and abiotic stresses. *Molecular biology, 43*(4), 578-585.

Zhang, H., Zhu, J., Gong, Z., & Zhu, J.-K. (2022). Abiotic stress responses in plants. *Nature Reviews Genetics, 23*(2), 104-119.

Digital Receipt

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Noor Malik
Assignment title:	Paper check
Submission title:	MS thesis
File name:	MS_thesis.docx
File size:	100.71K
Page count:	53
Word count:	10,027
Character count:	56,979
Submission date:	19-Aug-2022 12:59AM (UTC-0700)
Submission ID:	1882763307


Abstract

Plagiarism Report

Plagiarism Report

MS thesis

ORIGINALITY REPORT

7%	4%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	1%
2	"Rice Research for Quality Improvement: Genomics and Genetic Engineering", Springer Science and Business Media LLC, 2020 Publication	1%
3	www.researchsquare.com Internet Source	<1%
4	www.mdpi.com Internet Source	<1%
5	Submitted to Higher Education Commission Pakistan Student Paper	<1%
6	link.springer.com Internet Source	<1%
7	genominfo.org Internet Source	<1%
8	Chao Bai, Wenwen Yang, Yao Lu, Wei Wei, Zongbao Li, Li Zhang. " Identification of	<1%

Dr. Alvina Gul
Tenured Associate Professor
Deptt of Plant Biotechnology
Allama Iqbal Open University, Islamabad

80

Scanned with CamScanner

signaling, crown root initiation, and root meristem specification and maintenance ", The Plant Journal, 2019
Publication

14	Qinghai Cao, Wuyun Lv, Hong Jiang, Xueling Chen, Xinchao Wang, Yuchun Wang. "Genome-wide identification of glutathione S-transferase gene family members in tea plant (Camellia sinensis) and their response to environmental stress", International Journal of Biological Macromolecules, 2022 Publication	<1 %
15	Submitted to University of Bristol Student Paper	<1 %
16	Wen-Chi Chang. "Regulation of Ku gene promoters in *Arabidopsis* by hormones and stress", Functional Plant Biology, 2008 Publication	<1 %
17	bmcplantbiol.biomedcentral.com Internet Source	<1 %
18	tessera.spandidos-publications.com Internet Source	<1 %
19	Li Gao, Ye Tian, Meng-Ci Chen, Li Wei, Tian-Ge Gao, Hong-Ju Yin, Jin-Lin Zhang, Tanweer Kumar, Lin-Bo Liu, Suo-Min Wang. "Cloning and functional characterization of epidermis-	<1 %

	Circular RNAs Regulating Islet -Cell Autophagy in Type 2 Diabetes Mellitus ", BioMed Research International, 2019 Publication	
9	Submitted to Indian Agricultural Research Institute Student Paper	<1 %
10	Nishat Passricha, Shabnam Saifi, Mohammad W. Ansari, Narendra Tuteja. "Prediction and validation of cis-regulatory elements in 5' upstream regulatory regions of lectin receptor-like kinase gene family in rice", Protoplasma, 2016 Publication	<1 %
11	Sarah Mae Boyles Melencion, Yong Hun Chi, Thuy Thi Pham, Seol Ki Paeng et al. "RNA Chaperone Function of a Universal Stress Protein in Arabidopsis Confers Enhanced Cold Stress Tolerance in Plants", International Journal of Molecular Sciences, 2017 Publication	<1 %
12	Submitted to Columbia Basin College Student Paper	<1 %
13	Jérémy Lavarenne, Mathieu Gonin, Soazig Guyomarc'h, Jacques Rouster et al. " Inference of the gene regulatory network acting downstream of 1 in rice reveals a regulatory cascade linking genes involved in auxin	<1 %

reveals that expansin genes are involved in fibre cell growth in cotton", Research Square Platform LLC, 2019
Publication

25	Miaoze Xu, Peng Jin, Tingting Liu, Shiqi Gao, Tianye Zhang, Fan Zhang, Xiaolei Han, Long He, Jianping Chen, Jian Yang. " Genome-wide identification and characterization of UBP gene family in wheat (L.) ", PeerJ, 2021 Publication	<1 %
26	Christine Ölander, Jesper Edvardsson Rasmussen, Per Olof Eriksson, Göran Laurell, Helge Rask-Andersen, Jonas Bergquist. "The proteome of the human endolymphatic sac endolymph", Scientific Reports, 2021 Publication	<1 %
27	Lingtong Cheng, Lanxin Ma, Lijun Meng, Haihong Shang, Peijian Cao, Jingjing Jin. "Genome-Wide Identification and Analysis of the Class III Peroxidase Gene Family in Tobacco (Nicotiana tabacum)", Frontiers in Genetics, 2022 Publication	<1 %
28	Yang, Yazhou, Mingyang He, Ziguo Zhu, Shuxiu Li, Yan Xu, Chaohong Zhang, Stacy D Singer, and Yuejin Wang. "Identification of the dehydrin gene family from grapevine species and analysis of their responsiveness to	<1 %

specific promoter MtML1 from Medicago truncatula", Journal of Biotechnology, 2019
Publication

20	Xianna Mi, Xiaojing Wang, Han Wu, Lijun Gan, Jing Ding, Yi Li. "Characterization and expression analysis of cytokinin biosynthesis genes in Fragaria vesca", Plant Growth Regulation, 2017 Publication	<1 %
21	Suwen Lu, Yaxian Zhuge, Tianyi Hao, Zhongjie Liu, Mengwei Zhang, Jinggui Fang. "Systematic analysis reveals O-methyltransferase gene family members involved in flavonoid biosynthesis in grape", Plant Physiology and Biochemistry, 2022 Publication	<1 %
22	www.smujo.id Internet Source	<1 %
23	Emilio Gutiérrez-Beltrán, José María Personat, Fernando de la Torre, Olga del Pozo. "A Universal Stress Protein Involved in Oxidative Stress Is a Phosphorylation Target for Protein Kinase CIPK6", Plant Physiology, 2017 Publication	<1 %
24	Limin Lv, Dongyun Zuo, Xingfen Wang, Hailiang Cheng, Youping Zhang, Qiaolian Wang, Guoli Song, Zhiying Ma. "Genome-wide identification of the expansin gene family	<1 %

signaling, crown root initiation, and root meristem specification and maintenance ", The Plant Journal, 2019
Publication

14	Qinghai Cao, Wuyun Lv, Hong Jiang, Xueling Chen, Xinchao Wang, Yuchun Wang. "Genome-wide identification of glutathione S-transferase gene family members in tea plant (Camellia sinensis) and their response to environmental stress", International Journal of Biological Macromolecules, 2022 Publication	<1 %
15	Submitted to University of Bristol Student Paper	<1 %
16	Wen-Chi Chang. "Regulation of Ku gene promoters in *Arabidopsis* by hormones and stress", Functional Plant Biology, 2008 Publication	<1 %
17	bmcplantbiol.biomedcentral.com Internet Source	<1 %
18	tessera.spandidos-publications.com Internet Source	<1 %
19	Li Gao, Ye Tian, Meng-Ci Chen, Li Wei, Tian-Ge Gao, Hong-Ju Yin, Jin-Lin Zhang, Tanweer Kumar, Lin-Bo Liu, Suo-Min Wang. "Cloning and functional characterization of epidermis-	<1 %

various forms of abiotic and biotic stress", BMC Plant Biology, 2012.
Publication

29	Yue Leng, Jian Sun, Jingguo Wang, Hualong Liu, Hongliang Zheng, Minghui Zhang, Hongwei Zhao, Detang Zou. " Genome - wide lncRNAs identification and association analysis for cold - responsive genes at the booting stage in rice (L.) ", The Plant Genome, 2020 Publication	<1 %
30	bioone.org Internet Source	<1 %
31	Assis, Luciana M., João R. Sousa, Nedja Fernanda Dos Santos Pinto, Abrão A. Silva, Antônio Fernando De Melo Vaz, Paulo P. Andrade, Edgar M. Carvalho, and Marcia A. Melo. "B- cell epitopes of antigenic proteins in Leishmania infantum: an in silico analysis", Parasite Immunology, 2014. Publication	<1 %
32	Lanming Gou, Chunliu Zhuo, Shaoyun Lu, Zhenfei Guo. "A Universal Stress Protein from Medicago falcata (MfUSP1) confers multiple stress tolerance by regulating antioxidant defense and proline accumulation", Environmental and Experimental Botany, 2020	<1 %

	Publication	
33	contact14.ics.uci.edu Internet Source	<1 %
34	kupdf.net Internet Source	<1 %
35	mdpi-res.com Internet Source	<1 %
36	peerj.com Internet Source	<1 %
37	uir.unisa.ac.za Internet Source	<1 %
38	Plant Ecophysiology, 2004. Publication	<1 %
39	Yu Chen, Yan Ma, Jing Guo, Xianfeng Guo. " Cloning and expression analysis of the dehydrin gene in peony ( ) ", The Journal of Horticultural Science and Biotechnology, 2018 Publication	<1 %

Exclude quotes	On	Exclude matches	< 3 words
Exclude bibliography	On		

Appendix

Table 2: *TaUSP* genes and their molecular attributes. bp: base pairs; aa: amino acid; pI: Isoelectric point; Mw: Molecular weight, %ID: percentage identity

S.NO	USP	Chr No	Transcript ID	CDS(5'-3')	bp	aa	pI	weight	Uniprot ID	% ID
TaUSP1	TaUSP1(A)	3A	TraesCS3A02G092000.1	57,896,894-57,897,900	474	157	7.7	17,036.69	A0A3B6ED93	91.2
	TaUSP1(B)	3B	TraesCS3B02G107200.1	72,707,382-72,708,700	839	160	6.5	17,257.84	W5D407	91.2
	TaUSP1(D)	3D	TraesCS3D02G091900.1	46,477,314-46,478,621	812	160	6.95	17,270.88	A0A3B6GR47	91.2
TaUSP3	TaUSP3(A)	2A	TraesCS2A02G502200.1	729,983,524-729,985,675	854	164	6.95	17,830.60	A0A3B6B6Y3	77.5
	TaUSP3(B)	2B	TraesCS2B02G530300.1	725,111,649-725,113,924	823	164	6.59	17,888.68	A0A3B6CEG2	76.1
	TaUSP3(D)	2D	TraesCS2D02G502700.2	596,252,069-596,257,135	849	162	6.29	17,655.00	A0A3B6DK70	78.9
TaUSP4	TaUSP4(A)	3A	TraesCS3A02G208900.1	370,238,256-370,240,949	2404	741	6.36	82,254.66	A0A3B6EGE3	93.9
	TaUSP4(B)	3B	TraesCS3B02G239200.1	377,423,224-377,425,803	2253	750	6.7	83,204.87	A0A3B6FML1	89
	TaUSP4(D)	3D	TraesCS3D02G211800.1	283,314,773-283,317,561	2223	740	6.38	82,162.64	A0A3B6GSM4	94.5
TaUSP5	TaUSP5(A)	3A	TraesCS3A02G315600.1	556,547,493-556,552,618	1462	265	4.97	27,797.96	A0A3B6EJK0	97
	TaUSP5(B)	3B	TraesCS3B02G351300.1	561,232,274-561,237,503	1593	264	4.97	27,768.96	A0A3B6FUI5	97.7
	TaUSP5(D)	3D	TraesCS3D02G315900.1	429,303,584-429,308,740	1607	266	4.91	28,005.23	A0A077RV05	97.7

TaUSP7	TaUSP7(A)	3A	TraesCS3A02G385900.1	635,325,005-635,328,435	1246	222	9.72	23,416.75	A0A3B6EM72	80.2
	TaUSP7(B)	3B	TraesCS3B02G418000.1	655,027,561-655,030,722	1194	220	9.72	23,314.70	A0A077RWQ0	88
	TaUSP7(D)	3D	TraesCS3D02G378900.1	496,070,700-496,073,916	1240	121	9.72	23,356.70	A0A3B6GXT1	90.1
TaUSP8	TaUSP8(A)	6A	TraesCS6A02G122700.1	96,440,495-96,446,671	2942	825	6.82	91,760.08	A0A3B6NKQ3	82.6
	TaUSP8(B)	6B	TraesCS6B02G150900.1	152,079,374-152,087,729	2780	816	6.97	90,745.13	A0A3B6PIJ3	81.6
	TaUSP8(D)	6D	TraesCS6D02G112900.1	79,717,057-79,723,059	2641	825	7.14	91,707.08	A0A3B6QEG8	81.5
TaUSP9	TaUSP9(A)	6A	TraesCS6A02G182800.1	211,641,602-211,646,433	3013	730	6.74	81,674.67	A0A3B6NMW2	96.9
	TaUSP9(B)	6B	TraesCS6B02G213600.1	288,636,347-288,640,344	2157	718	6.98	80,457.43	A0A3B6PL27	95.9
	TaUSP9(D)	6D	TraesCS6D02G172000.1	158,443,808-158,447,604	2178	725	6.76	81,301.34	A0A3B6QGL1	96.3
TaUSP10	TaUSP10(B)	6B	TraesCS6B02G213500.2	288,418,522- 288426,073	2,712	771	8.06	85,259.00	A0A3B6PMJ6	94.8
TaUSP11	TaUSP11(A)	6A	TraesCS6A02G272900.1	499,402,075-499,405,572	846	180	5.82	20,009.89	A0A3B6NTC1	89
	TaUSP11(B)	6B	TraesCS6B02G300300.1	538,368,546-538,375,663	1056	180	5.91	20,032.93	A0A3B6PNK0	89
	TaUSP11(D)	6D	TraesCS6D02G252900.1	357,007,445-357,012,404	1189	180	5.77	20,032.89	A0A3B6QJX9	87.8
TaUSP12	TaUSP12(A)	6A	TraesCS6A02G274200.1	500,504,948-500,511,282	1168	177	5.96	20,007.86	A0A341X0W9	95.5
	TaUSP12(B)	6B	TraesCS6B02G301700.1	540,590,481-540,596,976	1485	177	5.96	20,007.86	A0A341X0W9	95.5
	TaUSP12(B)	6B	TraesCS6B02G312700.1	560,188,942-560,197,866	534	177	5.96	19,979.81	A0A3B6PNZ9	94.9

	TaUSP12(D)	6D	TraesCS6D02G254400.1	358,976,424-358,982,515	1493	177	6.04	19,988.86	A0A3B6QH30	95.5
TaUSP13	TaUSP13(A)	6A	TraesCS6A02G314000.1	550,730,106-550,732,129	186	162	5.66	14,871.45	A0A3B6NU99	89.3
	TaUSP13(B)	6B	TraesCS6B02G344000.1	606,283,698-606,285,748	1039	162	7.64	17,039.95	A0A3B6PQQ9	90.5
	TaUSP13(D)	6D	TraesCS6D02G293300.1	404,217,009-404,219,053	1102	162	6.82	17,081.99	A0A3B6QLB9	92.4
TaUSP14	TaUSP14(A)	6A	TraesCS6A02G323800.1	557,093,697-557,095,184	1118	165	8.99	17,769.71	A0A1D6ACE1	93.3
	TaUSP14(B)	6B	TraesCS6B02G354400.1	620,264,708-620,265,976	928	165	8.99	17,769.71	A0A1D6ACE1	93.3
	TaUSP14(D)	6D	TraesCS6D02G303600.1	412,001,031-412,002,174	686	165	8.99	17,799.80	A0A3B6QIP5	93.3
TaUSP15	TaUSP15(A)	6A	TraesCS6A02G346100.1	579,437,114-579,440,032	2639	783	6.99	85,144.17	A0A3B6NTL5	94.6
	TaUSP15(B)	6B	TraesCS6B02G379600.1	654,732,927-654,735,781	2671	824	8.45	89,872.89	A0A3B6PQM3	97.4
	TaUSP15(D)	6D	TraesCS6D02G329000.1	433,475,662-433,478,942	2992	783	7.24	85,101.30	A0A3B6QMJ7	96
TaUSP16	TaUSP16(A)	4A	TraesCS4A02G072300.2	70,308,291-70,315,265	5957	662	8.81	71,920.48	A0A3B6HR34	91.4
	TaUSP16(B)	4B	TraesCS4B02G228200.1	477,222,946-477,230,431	6498	655	8.56	70,926.20	A0A3B6IT29	91.9
	TaUSP16(D)	4D	TraesCS4D02G229100.1	388,924,735-388,928,689	2921	655	8.81	70,924.31	A0A3B6JL32	89.8
TaUSP17	TaUSP17(A)	4A	TraesCS4A02G119400.1	147,444,968-147,446,107	892	169	5.89	18,007.77	A0A3B6HV35	92.1
	TaUSP17(B)	4B	TraesCS4B02G185100.1	324,009,400-324,010,467	894	171	5.59	18,322.14	A0A3B6JHY8	92
	TaUSP17(D)	4D	TraesCS4D02G186500.1	324,009,400-324,010,467	894	171	5.59	18,040.78	A0A3B6JJK8	92.1

TaUSP19a	TaUSP19a(A)	5A	TraesCS5A02G389700.1	585,428,729-585,430,521	769	179	6.4	18,835.58	A0A3B6KLZ5	78.5
	TaUSP19a(B)	5B	TraesCS5B02G394700.1	572,378,743-572,390,785	558	158	7.83	20,255.31	A0A3B6LTB6	78.5
	TaUSP19a(D)	5D	TraesCS5D02G399600.1	465,029,958-465,031,884	795	186	6.33	19,303.96	A0A3B6MYF0	78.5
TaUSP19b	TaUSP19b(A)	5A	TraesCS5A02G389500.1	585,420,932-585,426,583	552	183	7.78	19,150.97	A0A3B6KMR9	66.7
	TaUSP19b(D)	5D	TraesCS5D02G399500.1	465,021,465-465,022,640	985	204	8.5	21,411.57	A0A3B6MZC4	80.8
TaUSP19c	TaUSP19c(A)	5A	TraesCS5A02G389400.1	585,416,279-585,417,372	895	207	6.24	21,570.58	A0A3B6KPM5	80.4
	TaUSP19c(B)	5B	TraesCS5B02G394600.2	572,363,398-572,364,730	1141	214	6.59	22,481.74	A0A3B6LTK0	81.6
	TaUSP19c(D)	5D	TraesCS5D02G399400.1	465,011,742-465,012,857	801	184	7.19	19,413.15	A0A3B6MWL9	91.3
TaUSP20	TaUSP20(A)	5A	TraesCS5A02G114900.1	229,924,846-229,927,957	2551	740	6.67	79,480.08	A0A3B6KF03	87.9
	TaUSP20(D)	5D	TraesCS5D02G125300.1	191,899,666-191,904,463	2444	741	6.6	79,553.25	A0A3B6MNF8	87.9
TaUSP21	TaUSP21(A)	1A	TraesCS1A02G106600.1	104,377,540-104,379,785	1102	166	5.78	17,864.47	Q2TN84	88.2
	TaUSP21(B)	1B	TraesCS1B02G124100.1	150,662,353-150,664,064	752	197	9.04	22,003.50	A0A3B5YUS5	89.8
	TaUSP21(D)	1D	TraesCS1D02G108300.1	101,093,429-101,096,021	1134	166	6.06	17,772.46	A0A3B5ZR06	89
TaUSP24	TaUSP24(A)	1A	TraesCS1A02G267800.1	462,776,185-462,778,584	687	288	10	24,254.90	A0A3B5Y2A8	88.5
	TaUSP24(B)	1B	TraesCS1B02G278500.1	486,846,591-486,849,061	696	231	9.9	24,630.27	A0A3B5YZB3	91.7
	TaUSP24(D)	1D	TraesCS1D02G267700.1	363,618,945-363,621,881	1198	231	9.9	24,552.19	A0A3B5ZXY9	88.5

TaUSP25	TaUSP25(A)	1A	TraesCS1A02G280700.1	478,450,779-478,452,828	731	166	6.31	18,141.73	A0A3B5Y2C9	88.1
	TaUSP25(B)	1B	TraesCS1B02G289700.1	505,110,468-505,113,544	877	166	6.31	18,127.74	A0A1D6RE03	88.1
	TaUSP25(D)	1D	TraesCS1D02G279800.1	377,910,422-377,914,063	985	166	6.31	18,127.74	A0A1D6RE03	89.4
TaUSP26	TaUSP26(A)	1A	TraesCS1A02G321000.1	511,481,306-511,484,164	1300	267	5.35	27,992.20	A0A3B5Y470	94
	TaUSP26(B)	1B	TraesCS1B02G333600.1	560,175,742-560,178,662	1326	268	5.58	28,050.28	A0A3B5Z2J5	93.2
	TaUSP26(D)	1D	TraesCS1D02G321100.1	414,706,779-414,707,723	816	271	5.58	28,050.28	A0A3B5Z2J5	94
TaUSP27a	TaUSP27a(A)	4A	TraesCS4A02G382700.1	660,790,453-660,795,671	3012	824	6.62	90,125.78	A0A3B6I0H7	84.6
	TaUSP27a(A)	4A	TraesCS4A02G382900.1	660,916,337-660,921,619	3039	825	6.26	90,411.95	A0A3B6I0I0	84.6
TaUSP27b	TaUSP27b(A)	7A	TraesCS7A02G084400.3	49,046,290-49,052,241	3515	826	6.32	90,657.06	A0A3B6RCH6	84.6
	TaUSP27b(D)	7D	TraesCS7D02G079400.1	46,838,388-46,843,665	2977	844	6.3	90,220.86	A0A3B6TD12	82.7
TaUSP30	TaUSP30(A)	2A	TraesCS2A02G211100.1	195,309,009-195,314,693	1272	257	4.89	27,277	A0A3B6AWP5	94
	TaUSP30(B)	2B	TraesCS2B02G236700.1	237,343,301-237,348,967	1263	257	4.95	27,231.74	A0A3B6C466	94
	TaUSP30(D)	2D	TraesCS2D02G217300.1	180,876,671-180,877,843	941	256	4.95	27,261.72	A0A3B6DE42	94.4
TaUSP31a	TaUSP31a(A)	2A	TraesCS2A02G112400.1	63,588,179-63,589,310	731	181	6.23	19,006	A0A3B6ASX8	62.7
	TaUSP31a(B)	2B	TraesCS2B02G131200.1	98,255,802-98,257,342	1194	180	5.89	18,988.53	A0A3B6C0C0	62.7
	TaUSP31a(D)	2D	TraesCS2D02G112900.1	62,552,204-62,553,439	961	180	5.73	18,832.28	A0A3B6D7A9	62.7

TaUSP31b	TaUSP31b(A)	2A	TraesCS2A02G111700.1	63,137,318-63,140,788	731	169	7.77	17,987	A0A3B6ARZ3	78.3
	TaUSP31b(B)	2B	TraesCS2B02G130500.1	97,933,152-97,933,921	546	171	7.78	17,924.26	A0A3B6C0H0	75.6
	TaUSP31b(D)	2D	TraesCS2D02G112300.1	62,289,003-62,289,983	738	173	7.77	18,318.00	A0A3B6D7R8	78.3
TaUSP31c	TaUSP31c(A)	2A	TraesCS2A02G112500.1	63,606,214-63,607,374	862	178	7.78	18,518	A0A3B6ATE8	72.9
	TaUSP31c(B)	2B	TraesCS2B02G131300.1	98,263,715-98,264,769	774	173	7.12	18,265.00	A0A3B6C0F4	76.6
	TaUSP31c(D)	2D	TraesCS2D02G113000.1	62,558,657-62,560,305	1283	177	8.36	18,449.22	A0A3B6D7H3	73.6
TaUSP31d	TaUSP31d(A)	2A	TraesCS2A02G112600.1	63,610,223-63,611,477	981	233	5.55	23,600	A0A3B6AS02	64.3
	TaUSP31d(B)	2B	TraesCS2B02G131400.1	98,285,766-98,287,057	983	219	5.4	22,435.48	A0A3B6C307	64.3
	TaUSP31d(D)	2D	TraesCS2D02G113100.1	62,564,109-62,565,483	1046	220	5.28	22,371.39	A0A3B6DAP6	64.3
TaUSP32	TaUSP32(B)	7B	TraesCS7B02G409600.1	678,770,504-678,774,029	2136	606	6.12	67,353.55	A0A3B6STI3	88.9
TaUSP34	TaUSP34(A)	5A	TraesCS5A02G294700.1	503,523,770-503,528,172	2655	772	6.22	86,536.73	A0A3B6KJR2	87.2
	TaUSP34(B)	5B	TraesCS5B02G294000.1	478,482,408-478,486,782	2599	768	6.21	86,207.32	A0A3B6LPW4	85.8
	TaUSP34(D)	5D	TraesCS5D02G302100.1	398,530,847-398,535,258	2617	773	6.17	86,510.83	A0A3B6MV00	86.9
TaUSP35	TaUSP35(A)	2A	TraesCS2A02G079300.1	36,037,402-36,042,137	3010	790	8.8	88,227.26	A0A3B6AQW9	83.1
	TaUSP35(B)	2B	TraesCS2B02G094100.1	54,528,489-54,533,307	3298	791	8.78	88,529.45	A0A3B6BZ91	81.8
TaUSP36	TaUSP36(A)	5A	TraesCS5A02G151500.1	328,839,703-328,840,703	737	189	6.74	19,742.58	A0A3B6KDS2	77

	TaUSP36(B)	5B	TraesCS5B02G150100.1	278,029,271-278,030,305	758	189	7.64	20,014.05	A0A3B6LID3	76
	TaUSP36(D)	5D	TraesCS5D02G156700.1	244,632,387-244,633,356	725	191	7.09	19,948.85	A0A3B6MN76	87.2
TaUSP37	TaUSP37(A)	1A	TraesCS1A02G145700.1	251,936,266-251,937,029	684	227	9.94	23,727.93	A0A3B5XY09	92.9
	TaUSP37(B)	1B	TraesCS1B02G162600.1	199,671,726-199,673,397	1073	235	9.83	24,057.37	A0A3B5YWA3	97.3
	TaUSP37(D)	1D	TraesCS1D02G144300.1	199,671,726-199,673,397	1591	231	10.14	24,057.37	A0A3B5ZRW2	97.3
TaUSP42	TaUSP42(D)	5D	TraesCS5D02G120800.1	172,243,611-172,258,256	793	223	10	11,495.23	A0A3B6MN77	94.7
	TaUSP42(A)	5A	TraesCS5A02G208800.1	422,050,035-422,052,534	963	251	10.42	27,231.74	A0A3B6KIH9	48.4
	TaUSP42(B)	5B	TraesCS5B02G207100.1	375,945,683-375,949,689	1043	249	10.31	26,973	A0A3B6LMD3	63.6
	TaUSP42(D)	5D	TraesCS5D02G215100.1	324,476,246-324,477,708	1046	250	10.42	27,128.16	A0A3B6MRI4	63.6
TaUSP43	TaUSP43(A)	5A	TraesCS5A02G092100.1	127,532,912-127,534,709	1039	201	6.73	21,653.69	A0A3B6KEW3	63.6
	TaUSP43(B)	5B	TraesCS5B02G097500.1	128,852,473-128,854,044	691	201	6.7	21,626.60	A0A3B6LI10	59.3
	TaUSP43(D)	5D	TraesCS5D02G104300.1	118,009,761-118,011,305	828	201	6.73	21,696.66	A0A3B6MMF7	78.4
TaUSP44	TaUSP44(A)	5A	TraesCS5A02G092000.1	127,519,647-127,532,759	1010	172	6.35	18,307.88	A0A3B6KE54	95.1
	TaUSP44(B)	5B	TraesCS5B02G097600.1	128,854,165-128,856,400	1448	169	6.54	18,082.68	A0A3B6LG72	95.1
	TaUSP44(D)	5D	TraesCS5D02G104200.1	118,007,705-118,009,555	1050	169	6.82	18,005.60	A0A3B6MMT4	95.1

Table 3: Identified motifs in *TaUSPs*

MOTIFS	MOTIF SEQUENCE	WIDTH	SITES	E-value
Motifs 1	RGLGAIKRALLGSVSDYCVHHAHCPVVIV	29	64	2.1e-1119
Motifs 2	VLVAVDDSEESKHALEWALDH	21	99	4.0e-837
Motifs 3	DPKEVJCEAVERHHADLLVLG	21	86	4.9e-659
Motifs 4	PAGTFCYIDPEYQQTGKVTTKSDVYALGVVLLQJJTGRPPM	41	26	5.9e-638
Motifs 5	QFQQEVEILSKIRHPNMVLLLGACPEYGCLVYEYM	35	25	2.6e-558
Motifs 6	WQLRFRIAAEVATALLFLHSAKPEPJVHRDLKPANILLDRB	41	26	1.3e-591
Motifs 7	SLRYRRYSIEEIZAATNNFSESLKIGEGGYGPVYKGK	37	28	3.1e-441
Motifs 8	GDWPVEEARRFAELALKCCELRRRDRPDLGTEVLPELNRLR	41	20	4.2e-404
Motifs 9	DVEAEMRRLRLELKQTMDMYNSACKEAJNAKQKAKELHRLKVEEARRYEE	50	9	1.30E-234
Motifs 10	LLRPGDALVLLHV	13	92	4.60E-225

Table 4: Duplicated gene pairs along with their Ka/Ks ratios, divergence, time, selection, and duplication types

Seq 1	Seq 2	Ka	Ks	Ka/Ks	TIME(MYA)	TYPE OF SELECTION	TYPE OF DUPLICATION
TraesCS4A02G382700.1	TraesCS4A02G382900.1	0.017934825	0.154178269	0.116325246	5.01	Negative	Tandem
TraesCS5A02G114900.1	TraesCS5D02G125300.1	0.020656396	0.05649729	0.365617469	1.83	Negative	segmental
TraesCS5A02G294700.1	TraesCS5D02G302100.1	0.013607744	0.056128068	0.242440984	1.82	Negative	segmental
TraesCS5A02G389500.1	TraesCS5D02G399600.1	0.113651145	0.21950221	0.517767654	7.13	Negative	segmental
TraesCS5B02G097600.1	TraesCS5D02G104200.1	0.007860334	0.067842833	0.115860934	2.2	Negative	segmental
TraesCS5B02G150100.1	TraesCS5D02G156700.1	0.026712658	0.124829498	0.213993151	4.05	Negative	segmental
TraesCS5B02G207100.1	TraesCS5D02G215100.1	0.01925771	0.070357353	0.273712823	2.28	Negative	segmental
TraesCS6A02G274200.1	TraesCS6B02G301700.1	0	0.025827516	0	8.39	Negative	segmental
TraesCS6A02G314000.1	TraesCS6B02G344000.1	0.008845311	0.063578137	0.139125049	2.06	Negative	segmental
TraesCS6A02G323800.1	TraesCS6B02G354400.1	0	0.152650077	0	4.96	Negative	segmental
TraesCS6B02G150900.1	TraesCS6D02G112900.1	0.023263314	0.050038168	0.464911382	1.62	Negative	segmental
TraesCS6B02G213600.1	TraesCS6D02G172000.1	0.016021219	0.046478036	0.34470517	1.51	Negative	segmental
TraesCS6B02G300300.1	TraesCS6D02G252900.1	0.004924105	0.079544679	0.061903634	2.58	Negative	segmental
TraesCS6B02G379600.1	TraesCS6D02G329000.1	0.024155579	0.067180242	0.359563745	2.18	Negative	segmental

Table 5. Glycosylation sites in *TaUSPs*

Transcript ID	Position	Region	Score
TraesCS3A02G092000.1	40	NNSY	0.6053
TraesCS3B02G107200.1	40	NNSY	0.6123
TraesCS3D02G091900.1	40	NNSY	0.6062
TraesCS3A02G208900.1	508	NGSL	0.5831
TraesCS3B02G239200.1	507	NGSL	0.5838
	675	NCTE	0.6662
TraesCS3D02G211800.1	507	NGSL	0.5831
TraesCS3A02G315600.1	4	NSSS	0.7264
TraesCS3B02G351300.1	4	NSSS	0.7264
TraesCS3D02G315900.1	4	NSSS	0.7386
TraesCS6A02G122700.1	213	NDTS	0.6361
	564	NVTQ	0.5908
	607	NGSL	0.56
TraesCS6B02G150900.1	212	NDTS	0.6313
	563	NVTQ	0.472
	603	NGSL	0.5591
TraesCS6D02G112900.1	213	NDTS	0.6349
	564	NVTQ	0.5967
	607	NGSL	0.5601
TraesCS6A02G182800.1	59	NQST	0.5403
TraesCS6B02G213600.1	59	NQSR	0.5457
TraesCS6D02G172000.1	59	NQST	0.5381
TraesCS6B02G213500.2	183	NLSI	0.6564
	255	NQSY	0.3817
	273	NSSD	0.7587
	410	NFSE	0.5861
	486	NGSL	0.6153
	743	NLST	0.6423
	749	NHSL	0.4178
TraesCS6A02G272900.1	89	NKSQ	0.6371
TraesCS6B02G300300.1	89	NKSQ	0.6263
TraesCS6D02G252900.1	89	NKSQ	0.6617
TraesCS6A02G274200.1	15	NESS	0.6569
TraesCS6B02G301700.1	15	NESS	0.6569

TraesCS6B02G312700.1	15	NESS	0.6569
TraesCS6D02G254400.1	15	NESS	0.6568
TraesCS6A02G314000.1	5	NLSS	0.7478
TraesCS6B02G344000.1	5	NLSS	0.7472
TraesCS6D02G293300.1	5	NLSS	0.7472
TraesCS6A02G323800.1	149	NATC	0.5428
	161	NGSL	0.3886
TraesCS6B02G354400.1	149	NATC	0.5428
	161	NGSL	0.3886
TraesCS6D02G303600.1	149	NATC	0.5427
	161	NGSL	0.3884
TraesCS5B02G394700.1	66	NYSE	0.6619
TraesCS5A02G114900.1	188	NYSL	0.6949
	707	NMSL	0.6612
TraesCS5D02G125300.1	93	NLSQ	0.7537
	188	NYSL	0.695
	708	NMSL	0.6612
TraesCS1A02G280700.1	154	NASC	0.5519
TraesCS1B02G289700.1	154	NASC	0.5519
TraesCS1D02G279800.1	154	NASC	0.5519
TraesCS4A02G382900.1	528	NGSL	0.6088
	540	NGTR	0.6873
	621	NTTP	0.1359
	810	NHSL	0.4183
TraesCS7A02G084400.3	222	NGTA	0.5537
	246	NRSA	0.56
	530	NGSL	0.5919
TraesCS7D02G079400.1	527	NGSL	0.5924
TraesCS2A02G112400.1	25	NKTV	0.8126
TraesCS2D02G112900.1	24	NKTV	0.8154
TraesCS7B02G409600.1	33	NNSR	0.4764
	53	NDSV	0.596
	149	NCSV	0.7066
	179	NISS	0.7021
	239	NDTM	0.5713
	383	NGTL	0.7058

TraesCS5A02G294700.1	487	NGSL	0.625
TraesCS5B02G294000.1	482	NGSL	0.625
TraesCS5D02G302100.1	487	NGSL	0.625
TraesCS2A02G079300.1	272	NSTL	0.7223
	444	NCSS	0.6943
	493	NRSV	0.6562
	559	NSTP	0.1842
TraesCS2B02G094100.1	270	NSTL	0.7234
	389	NMSL	0.5467
	440	NCSS	0.6772
	491	NRSV	0.6146

Table 6:. Phosphorylation sites in *TaUSPs*

S.NO	USP	Transcript ID	Phosphorylation sites
TaUSP1	TaUSP1(A)	TraesCS3A02G092000.1	16
	TaUSP1(B)	TraesCS3B02G107200.1	16
	TaUSP1(D)	TraesCS3D02G091900.1	15
TaUSP3	TaUSP3(A)	TraesCS2A02G502200.1	13
	TaUSP3(B)	TraesCS2B02G530300.1	12
	TaUSP3(D)	TraesCS2D02G502700.2	13
TaUSP4	TaUSP4(A)	TraesCS3A02G208900.1	73
	TaUSP4(B)	TraesCS3B02G239200.1	65
	TaUSP4(D)	TraesCS3D02G211800.1	70
TaUSP5	TaUSP5(A)	TraesCS3A02G315600.1	26
	TaUSP5(B)	TraesCS3B02G351300.1	26
	TaUSP5(D)	TraesCS3D02G315900.1	24
TaUSP7	TaUSP7(A)	TraesCS3A02G385900.1	23
	TaUSP7(B)	TraesCS3B02G418000.1	20
	TaUSP7(D)	TraesCS3D02G378900.1	23
TaUSP8	TaUSP8(A)	TraesCS6A02G122700.1	65
	TaUSP8(B)	TraesCS6B02G150900.1	120
	TaUSP8(D)	TraesCS6D02G112900.1	72
TaUSP9	TaUSP9(A)	TraesCS6A02G182800.1	85
	TaUSP9(B)	TraesCS6B02G213600.1	87
	TaUSP9(D)	TraesCS6D02G172000.1	79
TaUSP10	TaUSP10(B)	TraesCS6B02G213500.2	77
TaUSP11	TaUSP11(A)	TraesCS6A02G272900.1	16
	TaUSP11(B)	TraesCS6B02G300300.1	16
	TaUSP11(D)	TraesCS6D02G252900.1	16
TaUSP12	TaUSP12(A)	TraesCS6A02G274200.1	14
	TaUSP12(B)	TraesCS6B02G301700.1	14
	TaUSP12(B)	TraesCS6B02G312700.1	14
	TaUSP12(D)	TraesCS6D02G254400.1	15
TaUSP13	TaUSP13(A)	TraesCS6A02G314000.1	11
	TaUSP13(B)	TraesCS6B02G344000.1	12
	TaUSP13(D)	TraesCS6D02G293300.1	10
TaUSP14	TaUSP14(A)	TraesCS6A02G323800.1	9
	TaUSP14(B)	TraesCS6B02G354400.1	9
	TaUSP14(D)	TraesCS6D02G303600.1	9

TaUSP15	TaUSP15(A)	TraesCS6A02G346100.1	86
	TaUSP15(B)	TraesCS6B02G379600.1	86
	TaUSP15(D)	TraesCS6D02G329000.1	90
TaUSP16	TaUSP16(A)	TraesCS4A02G072300.2	70
	TaUSP16(B)	TraesCS4B02G228200.1	68
	TaUSP16(D)	TraesCS4D02G229100.1	68
TaUSP17	TaUSP17(A)	TraesCS4A02G119400.1	9
	TaUSP17(B)	TraesCS4B02G185100.1	10
	TaUSP17(D)	TraesCS4D02G186500.1	10
TaUSP19a	TaUSP19a(A)	TraesCS5A02G389700.1	17
	TaUSP19a(B)	TraesCS5B02G394700.1	12
	TaUSP19a(D)	TraesCS5D02G399600.1	15
TaUSP19b	TaUSP19b(A)	TraesCS5A02G389500.1	16
	TaUSP19b(D)	TraesCS5D02G399500.1	20
TaUSP19c	TaUSP19c(A)	TraesCS5A02G389400.1	13
	TaUSP19c(B)	TraesCS5B02G394600.2	16
	TaUSP19c(D)	TraesCS5D02G399400.1	13
TaUSP20	TaUSP20(A)	TraesCS5A02G114900.1	80
	TaUSP20(D)	TraesCS5D02G125300.1	81
TaUSP21	TaUSP21(A)	TraesCS1A02G106600.1	10
	TaUSP21(B)	TraesCS1B02G124100.1	9
	TaUSP21(D)	TraesCS1D02G108300.1	5
TaUSP24	TaUSP24(A)	TraesCS1A02G267800.1	23
	TaUSP24(B)	TraesCS1B02G278500.1	25
	TaUSP24(D)	TraesCS1D02G267700.1	24
TaUSP25	TaUSP25(A)	TraesCS1A02G280700.1	14
	TaUSP25(B)	TraesCS1B02G289700.1	13
	TaUSP25(D)	TraesCS1D02G279800.1	13
TaUSP26	TaUSP26(A)	TraesCS1A02G321000.1	32
	TaUSP26(B)	TraesCS1B02G333600.1	32
	TaUSP26(D)	TraesCS1D02G321100.1	31
TaUSP27a	TaUSP27a(A)	TraesCS4A02G382700.1	64
	TaUSP27a(A)	TraesCS4A02G382900.1	120
TaUSP27b	TaUSP27b(A)	TraesCS7A02G084400.3	122
	TaUSP27b(D)	TraesCS7D02G079400.1	108
TaUSP30	TaUSP30(A)	TraesCS2A02G211100.1	20

	TaUSP30(B)	TraesCS2B02G236700.1	20
	TaUSP30(D)	TraesCS2D02G217300.1	20
TaUSP31a	TaUSP31a(A)	TraesCS2A02G112400.1	13
	TaUSP31a(B)	TraesCS2B02G131200.1	9
	TaUSP31a(D)	TraesCS2D02G112900.1	10
TaUSP31b	TaUSP31b(A)	TraesCS2A02G111700.1	14
	TaUSP31b(B)	TraesCS2B02G130500.1	14
	TaUSP31b(D)	TraesCS2D02G112300.1	14
TaUSP31c	TaUSP31c(A)	TraesCS2A02G112500.1	10
	TaUSP31c(B)	TraesCS2B02G131300.1	11
	TaUSP31c(D)	TraesCS2D02G113000.1	11
TaUSP31d	TaUSP31d(A)	TraesCS2A02G112600.1	21
	TaUSP31d(B)	TraesCS2B02G131400.1	16
	TaUSP31d(D)	TraesCS2D02G113100.1	15
TaUSP32	TaUSP32(B)	TraesCS7B02G409600.1	63
TaUSP34	TaUSP34(A)	TraesCS5A02G294700.1	82
	TaUSP34(B)	TraesCS5B02G294000.1	81
	TaUSP34(D)	TraesCS5D02G302100.1	77
TaUSP35	TaUSP35(A)	TraesCS2A02G079300.1	63
	TaUSP35(B)	TraesCS2B02G094100.1	59
TaUSP36	TaUSP36(A)	TraesCS5A02G151500.1	17
	TaUSP36(B)	TraesCS5B02G150100.1	15
	TaUSP36(D)	TraesCS5D02G156700.1	16
TaUSP37	TaUSP37(A)	TraesCS1A02G145700.1	28
	TaUSP37(B)	TraesCS1B02G162600.1	29
	TaUSP37(D)	TraesCS1D02G144300.1	30
TaUSP42	TaUSP42(D)	TraesCS5D02G120800.1	21
	TaUSP42(A)	TraesCS5A02G208800.1	27
	TaUSP42(B)	TraesCS5B02G207100.1	24
	TaUSP42(D)	TraesCS5D02G215100.1	24
TaUSP43	TaUSP43(A)	TraesCS5A02G092100.1	16
	TaUSP43(B)	TraesCS5B02G097500.1	16
	TaUSP43(D)	TraesCS5D02G104300.1	15
TaUSP44	TaUSP44(A)	TraesCS5A02G092000.1	13
	TaUSP44(B)	TraesCS5B02G097600.1	14
	TaUSP44(D)	TraesCS5D02G104200.1	16

Table 7:Predicted Secondary structures of *TaUSPs*

Proteins	a-helices	Extended strand (%)	Beta Turn (%)	Random coil
TaUSP4	40.49	9.85	4.59	45.07
TaUSP10	46.3	11.67	4.8	37.22
TaUSP21	35.54	20.48	6.63	37.35
TaUSP30	31.91	15.56	6.61	45.91

Table 8: Properties of the predicted protein models

S.NO.	Model	Ramachandran favored region	Ramachandran outliers	Rotamer outliers	C-beta deviations	Bad Bonds	Bad angles
TaUSP4	**A**	94.35	1.41	0.81	2	1/2,300	18/3122
TaUSP10	**B**	91.99	2.79	1.56	7	01/2,366	38/3206
TaUSP21	**A**	91.46	2.85	0.37	5	0/2476	51/3364
TaUSP30	**A**	90.31	2.5	1.52	9	1/2,520	67/3420

Table 9: Characteristic features of the binding pockets present in *TaUSP*. Å: Angstrom, Ala: Alanine, Asp: Aspartate, Asn: Asparagine, Cys: Cysteine, Gln: Glutamine, Gly: Glycine, His: Histidine, Ile: isoleucine, Leu: leucine, Ser: Serine (Ser), Thr: Threonine

Protein	pockets	Area (SA)	Volume (SA)	AMINO ACIDS	Functional groups			
					Hydrogen bond donors	Hydrogen bond acceptors	Hydrophobic interactions	Hydrophobicity ratio
TaUSP4	p1	1526.76	1152.58	ALA, ARG, ASN, ASP, GLN, GLU, GLY, ILE, LEU, LYS, MET, PRO, SER, THR, TYR, VAL	36	83	56	0.32
	p2	1268.03	716.42	ALA ARG, ASP, CYS, GLN, GLU, GLY, ILE, LEU, LYS, MET, PHE, PRO, SER, THR, TRP, TYR, VAL	18	57	58	0.44
TaUSP10	p1	1309.14	931.26	ARG, ASN, GLN, GLU, GLY, HIS, LEU, LYS, THR, TYR, PRO	15	42	53	0.48
	p2	1062.04	516.8	ALA, ARG, ASN, ASP, CYS, GLU, GLY, HIS,ILE,LEU, LYS, PRO, TYR, VAL	24	70	51	0.35

21A	p1	1615.8	999.55	ALA, ARGG, ASN, ASP, GLU, GLY, ILE, LEU, LYS, PRO, SER, THR, TRP, VAL	28	98	54	0.3
	p2	948.32	847.84	ALA, ARG, ASP, GLY, HIS, ILE, LEU, LYS, PRO, SER, TYR, VAL	26	46	41	0.36
30	p1(skin)	1911.37	1384.29	ALA, ARG, ASP, CYS, GLU, GLY, HIS, ILE, LEU, LYS, MET, PHE,PRO, SER, TRP, TYR, VAL	37	98	93	0.41
	P2(purple)	1073.64	653.03	ALA, ARG, ASP, CYS, GLY, HIS, ILE, LUE,LYS,MET,PHE, PRO,SER,TYR, VAL	25	66	34	0.27
D	p1(skin)	1235.19	916.3	ALA, ARG,ASP, CYS, GLU, GLY, HIS,ILE, LEU, LYS, MET,PHE, SER, THR, TYR, VAL	24	65	80	0.47

	p2(purple)	1083.46	770.63	ALA, ARG, ASN, ASP,HIS, ILE, LEU, PRO, SER, TRP,TYR, VAL	12	39	66	0.56

Track your submissions

In-Silico Identification and Characterization of Universal Stress Protein (USP) Gene Family in Triticum aestivum

Corresponding Author: Alvina Gul

Scientific Reports

2f4a6ff8-cb1d-4963-97e5-8ef920612e6d | v.1.1

Editors Invited *16 Aug 22*

Sent from my iPhone

Begin forwarded message:

From: Scientific Reports <srep@nature.com>
Date: August 9, 2022 at 12:29:19 PM GMT+5
To: "Dr. Alvina Gul" <alvina_gul@asab.nust.edu.pk>
Subject: Scientific Reports - Receipt of Manuscript 'In-Silico Identification and...'

Ref: Submission ID 2f4a6ff8-cb1d-4963-97e5-8ef920612e6d

Dear Dr Gul,

Thank you for submitting your manuscript to Scientific Reports.

Your manuscript is now at our initial Quality Check stage, where we look for adherence to the journal's submission guidelines, including any relevant editorial and publishing policies. If there are any points that need to be addressed prior to progressing we will send you a detailed email. Otherwise, your manuscript will proceed into peer review.

Using the link below, you can check on the status of your submission by creating a new account, or logging in with an existing one.
https://researcher.nature.com/validate?[illegible]

You requested an APC discount or waiver when you submitted your manuscript. If you have not yet completed this request please do it now by following this link: https://article-discounts-and-waivers.springernature.com/request?snapp_id=2f4a6ff8-cb1d-4963-97e5-8ef920612e6d

Kind regards,

Peer Review Advisors
Scientific Reports
--

Springer Nature offers an open access support service to make it easier for our authors to discover and apply for APC funding. For further information please visit http://www.springernature.com/gp/open-research/funding