ISSN: 0974-8571 Vol. 13 No. 1 June, 2021

International Journal of Computational Intelligence in Control

IN SILLICO CHARACTERIZATION OF *ZINC METALLOPROTEASE; FTSH6* AND ITS ORTHOLOGUES BY COMPUTATIONAL APPROACH

Rahmat Ullah[1], Mamoona Rauf[1]*, Falak Naz[1], Muhammad Arif[2], Aziz Ud Din[3], Sajidul Ghafoor[3], Naseeb Ullah[3], Mohammad Islam[3], Khalida Zafar[4]

[1]Department of Botany, Garden Campus, Abdul Wali Khan University Mardan, Pakistan
[2]Department of Biotechnology, Garden Campus, Abdul Wali Khan University Mardan, Pakistan
[3]Department of Biotechnology and Genetic Engineering, Hazara University, Mansehra, Pakistan
[4]Department of Botany, Islamia college university, Peshawar, Pakistan

*Correspondence:
Email: mamoona@awkum.edu.pk

ABSTRACT

In this study, the bioinformatics characterization of *ZINC METALLOPROTEASE; FTSH6* and its orthologous in different plants is carried out in order to clarify the molecular mechanism for heat-stress tolerance.

Twenty orthologous are chosen to develop a gene regulatory network around *FTSH6* for a detailed analysis with phytozome. Evolutionary studies were carried out using a single procedure, upstream 1000 bp promoter sequence to the initial gene codon and the coding sequence of orthologous. 1000 *FTSH*6 promoters and their orthologous genes sequences have been analyzed using the EARS search tool for non-coding areas for analysis of relative genomes. Significant peaks of *FTSH6* selected from *Arabidopsis thaliana*. The MEME tool was utilized and motif sequences in the developers were maintained with their exact location in order to identify possible *cis* regulation elements in the sequences corresponding to those peaks preserved in the *FTSH*6 promoters' sequences. A sequence N….T/CCCAA/CG/ATGTA/GTGG…N of the new supposed core element *cis*-regulatory is found. The result of protein-protein interactions for the *Arabidopsis thaliana* network is the functional partners AT1G56180, EGY3, LHCB3 and PTAC5. The protein-protein interactions of *Arabidopsis thaliana* with selected orthologous gene *Cucumis sativus* is LHCB3 which is the functional partner in each other. From these results, the *ZINC METALLOPROTEASE* enzyme has been concluded to be a complex promoter of various plant species and to develop new regulatory elements that control its expression of thermal stress and other types of biotic and abiotic stress which affect the nutritional status of plant development. These studies jointly give a better understanding of *Arabidopsis thaliana* and *FTSH6*'s gene-expression mechanisms. This research will help to alter the expression of the gene to improve thermal stress-tolerance against over-accumulation in plants.

Keywords: *Arabidopsis thaliana*, selenium, orthologous, conserved regulatory sequence, promoter, *FtSH6*, thermomemory

INTRODUCTION

Peptide bonding enzymes are proteases that are removed from other proteins. In many biological reactions they are complex, from infinite proteolysis in which nutritional or stock protein breaks into amino acids or proteins with impairments, through inappropriate proteolysis that separates a specific peptide bond from different regulatory mechanics

(Fellers, 2004). As part of the flow protease representation that controls converters, receptors, kinases and transcriptional factors, inadequate proteology may initiate or disable all types of enzymes (Auf dem Keller *et al.*, 2007). Proteolysis is infinite for flora, the current plant proteolysis information is inadequate and imperfect. In addition to the standard of action of broad-specific proteases, the specific characteristics of exact proteases should be well-educated for many leftovers, especially those that should be important during the shooting period (Wagner *et al.*, 2012). In flora, it is also known that by proteolysis procedures, rising cellular path species are accurate. The MERPOS database (http://merops.sanger.ac.uk, Rawlings *et al.*, 2004) lists 686 proteases and 184 presumed silent counterparts for the *Arabidopsis thaliana* plant organism. The pathway of ubiquitin / proteasome also has 1,300 inheritable aspects (Vierstra, 2003). In the majority of subcellular areas proteases play a control role in protecting proteome homosexuality. In the *Arabidopsis thaliana*, Clp, *FTSH*, DegP (in chloroplasts) and Lon protease (in chloroplasts and mitochondria), four major protease families were identified as critical elements of chloroplast and mitochondrial protein control systems. A relatively bright categorized family of proteases is the *FTSH* family (Filamentation temperature sensitive). The *FTSH* proteases contain a transmembrane sphere N-terminal, and a protease zone of the M41 peptidease family with a zinc-binding motive. The AAA hydrophilic section (ATPase related to several cellular activities) is traced which carries a zinc-binding motif and is therefore considered in the zinc metalloprotease family (Wagner *et al.*, 2012). Metal-proteases (Ito and Akiyama 2005), which can produce eukaryotes and prokaryotes (excluding Archaea) in organelles of endosymbiosis (chloroplast and mitochondria), have an impact on sheat-related *FTSH* proteases, ATP. The archaea appear to consume the safe role of *FTSH* in a membrane-fixed lon protease (Ruepp *et al.*, 2000).

FTSH proteases consist of a N terminal transmembrane slice with a C terminal area covering the AAA ATPase area (Suno *et al.*, 2006). As the AAA dominance is concerned, the FtH proteases are credited to the AAA family of proteins (ATPase related to cellular unpredictability). The highly preserved AAA proteins are classified as approximately 200 to 250 residues that cover the Walker A and B motifs needed to hydrolyze and bind nucleotides. It is therefore called "the second area of homology" which can contain preserved arginine residues important for the oligomeration of nucleotide and hydrolysis in general. As mentioned, various *FTSH* subunits can be adjusted between open and locking positions in crystallographic education. ATP hydrolysis is the reason for an ambitious shift in AAA. The corresponding signs for the move in the internal proteolysis of a well preserved hydrophobic range (Bieniossek *et al.*, 2009). This region fixes and mixes the substratum in a chamber of proteolysis (Hinnerwisch *et al.*, 2005). 12 *FTSH*-coding genes are included in the genome of *Arabidopsis thaliana* (Garcia *et al.*, 2006). More than some *FTSH* genes form homologous sets pointing towards fresh gene replication. Based in their arrangement uniqueity. *FTSH* proteases were nearly exposed to each other, and were able to balance in complexes, such as the *FTSH* 2/8 pair or the *FTSH* 1/5 pair, situated in the chloroplast plant thylakoid membrane (Zhang *et al.*, 2010). *Arabidopsis thaliana* includes five counterparts among the 12 *FTSH* proteases which do not include the zinc binding subject and are therefore probably inactive (Wagner *et al.*, 2011). The proteins are called *FTSH*i. The chloroplasts are targeted entirely by eight out of twelve *FTSH* (*FTSH* 1, 2, 5–9, and 12) and the five inactive *FTSH*i (Ferro *et al.*, 2010); in mitochondria there are three *FTSH* proteases (*FTSH* 3, 4 and 10) (Janska *et al.*, 2010). *FTSH*11 is targeted twice for either organs (Urantowka *et al.*, 2005). In two sets (m*FTSH* and I-*FTSH*) the mitochondrial *FTSH* proteases are separated. The *FTSH*3 and *FTSH*10

m*FTSH* proteases are placed on the inner surface of mitochondria; the mitochondrial matrix appear in the proteolysis domain (Kwasniak *et al.,* 2011). Under controlled conditions, individual knockouts of either *FTSH*3 or *FTSH*10. Growth in the field could be seen in the reduced production of seed and in *FTSH*10 salmon leaves (Wagner *et al.,* 2011). *FTSH*10 has just been shown to be associated with prohibitins such as m-*FTSH* yeast proteases (Piechota *et al.,* 2010). In the inner mitochondrial membrane, the I-*FTSH* protease *FTSH*4 is located on the active side of the membrane but faces the interconnecting space (Gibala *et al.,* 2009). Knock-out of *FTSH*4 causes leaf morphology to change during short-day photoperiod at the late stage of rosette growth (Kicia *et al.,* 2010). The closest homolog to the *FTSH*4 is in *Arabidopsis thaliana* *FTSH*11, and therefore it is also considered an I-*FTSH*. Kwasniak *et al.,* (2011) reviewed the mitochondrial *FTSH* proteases. The chloroplast envelope is the membrane compartment host to most *FTSH* proteins and has 4 active and 5 inactive *FTSH* homologues (Ferro *et al.,* 2010). A function was only tested for *FTSH*11 of these proteases (Chen *et al.,* 2006). Continuous light-sensitive plants are *FTSH*11 knockout mutants, in comparison to the wild type of continuous light-tolerant plant. This is part of the earlier heat intolerance proposed (Chen *et al.,* 2006). There are a limited number of *FTSH*12, *FTSH*9 an d *FTSH*7. There are therefore likely the same features in *FTSH*9 and *FTSH*7 resulting from curre nt gene replication (García *et al.,* 2006), and form a complex protease. Mutants that remove a si ngle mutual protease are not affected by a phenoty pe. *Arabidopsis thaliana* does not consume *FTSH*12, but *FTSH*12 and *FTSH*11 may use overlying features because of its position. Consequently, the terms vary (Garcia *et al.,* 2006). *FTSH*12 and *FTSH*11 appear unlikely to be complex, but unlike analyses, they will improve *FTSH*12, so individual proteases have one or more joint substrates. *FTSH*12 and *FTSH*11 are considered complex. To prevent a detailed breach of this mutant, *FTSH*12 is inappropriately toxic to the embryo. Five inactive *FTSH*i are more likely that they have no protease action, but they can be a proteolytic *FTSH* complex fragment. It has been demonstrated on behalf of *FTSH*2 that not all complex subunits of proteolysis must stay active in order to continue their function (Zhang *et al.,* 2010). There are comparable figures on the appearance of certain *FTSH*i and *FTSH*12 and similarly on embryos, so that one or more *FTSH*i can form an *FTSH*12 proteolysis complex. *FTSH* proteases from the hetero-ligomeric hexametry complex (Zaltsman *et al.,* 2005) are the main proteases studied by *FTSH* in *Arabidopsis thaliana*. Together, *FTSH*1 and *FTSH*5, *FTSH*2 and *FTSH*8 are two protease types in the complex (Garcia *et al.,* 2006). One type of company is approximately 90 percent individual and can replace each other in a complex way in part. Type A and Type B members are nearly 50 percent individual. Deletions of each type (*FTSH*5 and *FTSH*2) cause several phennotypes (Sakamoto *et al.,* 2003). In the lower specified subunits *FTSH*1 and *FTSH*8 no single mutant phénotype exists. The *FTSH* thylacoid system has a most common function of degrading the D1 (Katos *et al.,* 2009) reaction center of Photosystem II, but other substrates, e.g. for *FTSH*1, 2, 5 and 8 have also been described. Horrible pea and b6 cytochrome RieskeFeS conditions in *Chlamydomonas* (Malnoe *et al.,* 2011). In the Kontakt Review (Wagner *et al.,* 2011), further comprehensive explaining of *FTSH* proteases from chloroplast can be provided. *FTSH*6 is also grouped in *FTSH* thylakoid protease because of its high sequence similarity. This protease, however, is not known for a certain location or an accredited function. No significant T-DNA knock-out mutant phenotype exists (Wagner *et al.,* 2011). Sequence analyzes indicated that other plant species are also affected by FtH6 (Gracia *et al.,* 2006). The evolutionary preservation of *FTSH*6 shows that gene duplication is not only absurd, it has been a dissimilar task. Four *FTSH* (selected as *FTSH*i) are

for accumulated chloroplasts. *FTSH*6 in high Current research was aimed at identification and screening of orthologoues of the *FTSH*6 gene among plant species using *Arabidopsis thaliana* plant using *in silico* approach, unravelling *FTSH*6's gene-regulatory network and predict its molecular and biological functions, Predicting the presence of motifs and *cis*-regulatory elements responsible for the activity of the gene expression promoter during the response to thermo-tolerance and evaluation of natural variation in response to thermo-tolerance of different local and / or global ecotypes of plant species containing *FTSH*6 gene orthologoues.

2. MATERIALS AND METHODS

2.1. Characterization of *FtSH6* in bioinformatics

2.1.1. *In silico* expression analysis of *FtSH6*

The *FtSH6* gene has been evaluated at the developmental and structural level by using the affymetrix array data database of the eFp browser to recover expression values (Schmid and Mikolajczyk, 2005). Data from *Arabidopsis thaliana* and other plant species from microarrays have, for this purpose, been obtained (Hruz *et al.*, 2008).

2.1.2. Screening of orthologous *FtSH6* gene and recovery of sequences of promoters

The *FtSH6* genomic sequence was downloaded with multiple searches of phytozome as query sequences using the TAIR 9.0 *Arabidopsis* genome (Rhee and Flanders, 2000) (Goodstein *et al.*, 2013). Known gene sequence of *Arabidopsis* used to confirm BLAST NCBI (Geer, 2009). The closest members of *FtSH* orthologous genes of different species chosen to extract from the initiation codon (ATG) transcription the alleged promoter region of 1000 bp. In genes and 1 kb of the promoter regions for various plant species from NCBI, a phylogenetic relationship was analyzed between the orthologous genes of *FtSH*. A multi-sequence alignment tool for phylogenetic design. The phylogram was viewed by TreeDyn (Edgar 2004).

2.1.3. Search in the promoters for preserved non-coding regions

Analysis by 1000 bp orthologous genetic sequences of the relative genome of the EARS tool (Picot *et al.*, 2010). EARS software reviews all small-sequence arrangements (windows) and aligns all windows pairs all over the world. The preserved arrangements are detected here. A window size of 40bp and a P-value reduction of 0.0001 have been used for this analysis. The EARS result file of the *Arabidopsis* gene promoter and its sequence of orthologous promoters is also individually examined by the software and by where significant peaks are detected (Picot *et al.*, 2010).

2.1.4. Determination of the network of Gene co-expression

The online resource and STRING (Simpson *et al.*, 2012) have been accessed in order to discover useful linkages between Locus proteins and their orthologous genes (http://string-db.org). The system provided unrivaled and broad coverage as well as easy access to predicted and experimental network data, the interactive viewfinder, STRING can approve models, update previews and cluster networks with extensive information on the data and an intensely updated combination of third event features (Simpson *et al.*, 2012).

2.1.5. Analysis of over-representative *cis*-elements

Identification of potential *cis*-regulatory elements in sequences matching the preserved peaks of *FtSH* promoters derived from the *FtSH6* orthologous candidates collected. In representative analyzes of possible cis-elements using MEME two main approaches to this task are employed (Bailey *et al.*, 2009). The preserved non-coding motives derived in the MEME range are characterized by an analysis of biological functions using BLAST proteins and fields investigated by Inter proscan (Bailey *et al.*, 2009).

3: Results and Discussion

3.1.1 Selection of the reference gene and its Orthologues on the basis of Expression Pattern Homolgy by eFP Browser

The selected orthologues *Arabidopsis thaliana FTSH6* with their reference genes *Arabidopsis thaliana FTSH6* were assessed on the basis of tissue based differential Expression analysis at different developmenttttttal stages at 100 threshold By Plant Biology "eFP Browser". The reference gene *FTSH6* shows high expression at mature pollen, cotyledons, sepals, petals, stamen and carpel indicated red coloured parts in **(Figure 3.1)**.

3.1.2 SCREENING OF *FTSH6* ORTHOLOGOUS GENES AND SEQUENCES OF PROMOTERS

Data obtained from NCBI Blast search and Phytozome showing AA Identity, CDS length (bp) and AA length. The reference gene *Arabidopsis thaliana FTSH6* have 100% AA Identity at 2131bp CDS length and 709 AA lengths (the highest amino acid length among the orthologous of *Arabidopsis thaliana)* and selected orthologues *Cucumis sativus (Cucsa.158300.1)* have 73.6% AA Identity at 1954 bp CDS length and 652 (the lowest amino acid length among the orthologous of the *Arabidopsis thaliana* taken as model plant) AA length**.** Simultaneously the other orthologous such as *Boechera stricta* (*Bostr.2902s0260.1)* have 89.1% AA Identity at 2055bp CDS length and 684 AA length, *Capsella grandiflora* (*Cagra.1036s0030.1*) have 85.90% AA Identity at 2061 CDS length and 686 AA length *Capsella rubella* (*Carubv10004704m)* have 83.8% AA Identity at 2058 with 685 AA length, *Eutrema salsugineum (Thhalv10016003m)* have 88.2% AA Identity at 2064bp and 687 AA length, *Brassica rapa (Brara.J01968.1)* have 84.8% AA Identity at 2058bp and 685 AA length, *Glycine max* (*Glyma.18G259700.1*) have 79.5% AA Identity at 2037bp with 678 AA length, *Eucalyptus grandis (Eucgr.J01348.1*) have 81.6% AA Identity at 2052bp and 683 AA length, *Citrus clementina (Ciclev10028305m)* have 78.30% AA Identity at 2031*bp* and 676 AA length, *Citrus sinensis* (*orange1.1g005815m*) have 77.70% AA Identity at 2031*bp* with 676AA length, *Fragaria vesca (mrna04456.1)* have 78.40% AA Identity at 2037*bp* and 678 AA length, *Prunus persica* (*Prupe.3G266900*) have 77.00% AA Identity at 2046bp 681 AA length, *Populus trichocarpa (Potri.017G084000.1) have* 77.90% AA Identity at 2034bp and 677 AA length, *Medicago truncatula (Medtr7g010800.1)* have 76.70% have AA Identity at 2016bp and 671 AA length, *Aquilegia coerulea (Aqcoe3G077900.1)* have 75.50% AA Identity at 2040bp with 679 AA length, *Ananas comosus (Aco002884.1)* have 77.30% AA Identity at 2037bp with 678 AA length, *Panicum virgatum (Pavir.J13145.1)* have 73.30% AA Identity at 2067bp with 688 AA length, *Kalanchoe laxiflora (Kalax.0004s0036.1)* have 70.70% AA Identity at 2031bp with 676 AA length, *Zea Mays (GRMZM2G048836.01)* have 73.90% AA Identity at 2010bp with 691 AA length (the second highest amino acid length among the orthologous of the *Arabidopsis thaliana*), *Gossypium raimondii (Gorai.012G053900)* have 74.20% AA Identity at 2028bp with 675 AA length as shown in the (**Table 3.1**).

3.1.3 Verification of Orthologous genes on the basis of Evolutionary and phylogenetic Trends

onstruction of evolutionary trees based on sequences of amino acid The phylogenetic trees have been built using TreeDyn tools (*http://www.phylogeny.fr*), which explained the orthological genetic verification on the basis of an advance and phylogenetic trend, which shows that branch length of the *Arabidopsis thaliana FtSH6* is 0.92, with comparison of the selected orthologous *Cucumis sativus (Cucsa.158300.1)* has 0.86 branch length as illustrated by **Figure 3.2**.

3.3 SEARCH IN PROMOTERS FOR PRESERVED NON-CODING REGIONS

A comparative genome-analysis using the 1 kB promotional sequences of the reference genes and its orthologous genes showed that the reference gene promotors of *Arabidopsis thaliana* and their orthologous promoter sequences have significant

peaks in preserved regulatory sequences. The dotted red line shows the meaning threshold of P= 0.0001. The consensus sequences of the reference gene *FTSH6* are found between positions of -72bp to -168bp, which shows that *Arabidopsis thaliana* (*FTSH6*) has evolved with all selected orthologists in non-coding DNA at-168bp upstream in the promoter region, *Arabidopsis thaliana FTSH6* is phylogenically linked to all orthologous genes, as shown in **Figure 3.3**. sequence resembling *FTSH*2 and *FTSH*8. *FTSH*6 has been identified in vitro 36 in light-harvesting PSII degradation for Lhcb1 and Lhcb3, although further studies in highlight acclimation or any other biological procedure still have not shown the in vivo function of *FTSH*6 (*Wagner et al.,* 2012).

*FTSH*6 is particularly manifested at a very low plant level in *Arabidopsis thaliana* and in many other species in the kingdom, such as in dicot plants rapeseed (*Brassica napus*), tomatoes (*Solanum lycopersicum*), monocot wheat (*Triticum aestivum*) and sorghum (*Sorghum bicolor*) (Xue *et al.,* 2015, respectively), at a very low plant rate and in plants at standard development temperatures. Thus, in plants that suggest this metalloprotease is important in response to thermal stress, the heat inducibility of the *FTSH*6 expression seems evolutionary. *FTSH*6 is a major player in the long-term thermotolerance (Sedaghatmehr *et al.,* 2017) recently reported. Current research was aimed at identification and screening of orthologoues of the *FTSH*6 gene among plant species using *Arabidopsis thaliana* plant using *in silico* approach, unravelling *FTSH6*'s gene-regulatory network and predict its molecular and biological functions, Predicting the presence of motifs and *cis*-regulatory elements responsible for the activity of the gene expression promoter during the response to thermo-tolerance and evaluation of natural variation in response to thermo-tolerance of different local and / or global ecotypes of plant species containing *FTSH*6 gene orthologoues.

2. MATERIALS AND METHODS

2.1. Characterization of *FtSH6* in bioinformatics

2.1.1. *In silico* expression analysis of *FtSH6*

The *FtSH6* gene has been evaluated at the developmental and structural level by using the affymetrix array data database of the eFp browser to recover expression values (Schmid and Mikolajczyk, 2005). Data from *Arabidopsis thaliana* and other plant species from microarrays have, for this purpose, been obtained (Hruz *et al.*, 2008).

2.1.2. Screening of orthologous *FtSH6* gene and recovery of sequences of promoters

The *FtSH6* genomic sequence was downloaded with multiple searches of phytozome as query sequences using the TAIR 9.0 *Arabidopsis* genome (Rhee and Flanders, 2000) (Goodstein *et al.,* 2013). Known gene sequence of *Arabidopsis* used to confirm BLAST NCBI (Geer, 2009). The closest members of *FtSH* orthologous genes of different species chosen to extract from the initiation codon (ATG) transcription the alleged promoter region of 1000 bp. In genes and 1 kb of the promoter regions for various plant species from NCBI, a phylogenetic relationship was analyzed between the orthologous genes of *FtSH*. A multi-sequence alignment tool for phylogenetic design. The phylogram was viewed by TreeDyn (Edgar 2004).

2.1.3. Search in the promoters for preserved non-coding regions

Analysis by 1000 bp orthologous genetic sequences of the relative genome of the EARS tool (Picot *et al.,* 2010). EARS software reviews all small-sequence arrangements (windows) and aligns all windows pairs all over the world. The preserved arrangements are detected here. A window size of 40bp and a P-value reduction of 0.0001 have been used for this analysis. The EARS result file of the *Arabidopsis* gene promoter and its sequence of orthologous promoters is also individually examined by the software and by where significant peaks are detected (Picot *et al.*, 2010).

2.1.4. Determination of the network of Gene co-expression

The online resource and STRING (Simpson *et al.*, 2012) have been accessed in order to discover useful linkages between Locus proteins and their orthologous genes (http://string-db.org). The system provided unrivaled and broad coverage as well as easy access to predicted and experimental network data, the interactive viewfinder, STRING can approve models, update previews and cluster networks with extensive information on the data and an intensely updated combination of third event features (Simpson *et al.*, 2012).

2.1.5. Analysis of over-representative *cis*-elements

Identification of potential *cis*-regulatory elements in sequences matching the preserved peaks of *FtSH* promoters derived from the *FtSH6* orthologous candidates collected. In representative analyzes of possible cis-elements using MEME two main approaches to this task are employed (Bailey *et al.*, 2009). The preserved non-coding motives derived in the MEME range are characterized by an analysis of biological functions using BLAST proteins and fields investigated by Inter proscan (Bailey *et al.*, 2009).

3: Results and Discussion

3.1.1 Selection of the reference gene and its Orthologues on the basis of Expression Pattern Homolgy by eFP Browser The selected orthologues *Arabidopsis thaliana FTSH6* with their reference genes *Arabidopsis thaliana FTSH6* were assessed on the basis of tissue based differential Expression analysis at different developmenttttttal stages at 100 threshold By Plant Biology "eFP Browser". The reference gene *FTSH6* shows high expression at mature pollen, cotyledons, sepals, petals, stamen and carpel indicated red coloured parts in **(Figure 3.1)**.

3.1.2 SCREENING OF *FTSH6* ORTHOLOGOUS GENES AND SEQUENCES OF PROMOTERS

Data obtained from NCBI Blast search and Phytozome showing AA Identity, CDS length (bp) and AA length. The reference gene *Arabidopsis thaliana FTSH6* have 100% AA Identity at 2131bp CDS length and 709 AA lengths (the highest amino acid length among the orthologous of *Arabidopsis thaliana)* and selected orthologues *Cucumis sativus (Cucsa.158300.1)* have 73.6% AA Identity at 1954 bp CDS length and 652 (the lowest amino acid length among the orthologous of the *Arabidopsis thaliana* taken as model plant) AA length**.** Simultaneously the other orthologous such as *Boechera stricta* (*Bostr.2902s0260.1)* have 89.1% AA Identity at 2055bp CDS length and 684 AA length, *Capsella grandiflora* (*Cagra.1036s0030.1*) have 85.90% AA Identity at 2061 CDS length and 686 AA length *Capsella rubella* (*Carubv10004704m)* have 83.8% AA Identity at 2058 with 685 AA length, *Eutrema salsugineum (Thhalv10016003m)* have 88.2% AA Identity at

2064bp and 687 AA length, *Brassica rapa (Brara.J01968.1)* have 84.8% AA Identity at 2058bp and 685 AA length, *Glycine max (Glyma.18G259700.1*) have 79.5% AA Identity at 2037bp with 678 AA length, *Eucalyptus grandis (Eucgr.J01348.1*) have 81.6% AA Identity at 2052bp and 683 AA length, *Citrus clementina (Ciclev10028305m)* have 78.30% AA Identity at 2031*bp* and 676 AA length, *Citrus sinensis* (*orange1.1g005815m*) have 77.70% AA Identity at 2031*bp* with 676AA length, *Fragaria vesca (mrna04456.1)* have 78.40% AA Identity at 2037*bp* and 678 AA length, *Prunus persica* (*Prupe.3G266900*) have 77.00% AA Identity at 2046bp 681 AA length, *Populus trichocarpa (Potri.017G084000.1) have* 77.90% AA Identity at 2034bp and 677 AA length, *Medicago truncatula (Medtr7g010800.1)* have 76.70% have AA Identity at 2016bp and 671 AA length, *Aquilegia coerulea (Aqcoe3G077900.1)* have 75.50% AA Identity at 2040bp with 679 AA length, *Ananas comosus (Aco002884.1)* have 77.30% AA Identity at 2037bp with 678 AA length, *Panicum virgatum (Pavir.J13145.1)* have 73.30% AA Identity at 2067bp with 688 AA length, *Kalanchoe laxiflora (Kalax.0004s0036.1)* have 70.70% AA Identity at

2031bp with 676 AA length, *Zea Mays (GRMZM2G048836.01)* have 73.90% AA Identity at 2010bp with 691 AA length (the second highest amino acid length among the orthologous of the *Arabidopsis thaliana*), *Gossypium raimondii (Gorai.012G053900)* have 74.20% AA Identity at 2028bp with 675 AA length as shown in the (**Table 3.1**).

3.1.3 Verification of Orthologous genes on the basis of Evolutionary and phylogenetic Trends

onstruction of evolutionary trees based on sequences of amino acid The phylogenetic trees have been built using TreeDyn tools (*http://www.phylogeny.fr*), which explained the orthological genetic verification on the basis of an advance and phylogenetic trend, which shows that branch length of the *Arabidopsis thaliana FtSH6* is 0.92, with comparison of the selected orthologous *Cucumis sativus (Cucsa.158300.1)* has 0.86 branch length as illustrated by **Figure 3.2**.

3.3 SEARCH IN PROMOTERS FOR PRESERVED NON-CODING REGIONS

A comparative genome-analysis using the 1 kB promotional sequences of the reference genes and its orthologous genes showed that the reference gene promotors of *Arabidopsis thaliana* and their orthologous promoter sequences have significant peaks in preserved regulatory sequences.

The dotted red line shows the meaning threshold of P= 0.0001.

The consensus sequences of the reference gene *FTSH6* are found between positions of -72bp to -168bp, which shows that *Arabidopsis thaliana* (*FTSH6*) has evolved with all selected orthologists in non-coding DNA at-168bp upstream in the promoter region, *Arabidopsis thaliana FTSH6* is phylogenically linked to all orthologous genes, as shown in **Figure 3.3**.

Table 3.1 Representing ***FTSH6*** orthologous genes with GeneInfo Identifiers from different plant species Data obtained from NCBI Blast search and Phytozome

No.	*Organism*	Locus	MRSF	Similarity	AA length	CDS
	Arabidopsis thaliana	*(AT5G15250)*	BRA	100%	709	2130
1	*Boechera stricta*	Bostr.2902s0260.1	BRA	89.10%	684	2055
2	*Capsella grandiflora*	Cagra.1036s0030.1	BRA	85.90%	686	2061
3	*Capsella rubella*	Carubv10000375m	BRA	83.80%	685	2058
4	*Eutrema salsugineum*	Thhalv10016003m	BRA	88.20%	687	2064
5	*Brassica rapa*	Brara.J01968.1	BRA	84.80%	685	2058
6	*Glycine max*	Glyma.18G259700.1	ROS	79.50%	678	2037
7	*Eucalyptus grandis*	Eucgr.J01348.1	ROS	81.60%	683	2052
8	*Citrus clementine*	Ciclev10011227m	SBM	78.30%	676	2031
9	*Citrus sinensis*	orange1.1g005815m	SBM	77.70%	676	2031
10	*Fragaria vesca*	mrna04456.1	ROS	78.40%	678	2037
11	*Prunus persica*	Prupe.3G266900	ROS	77.00%	681	2046
12	*Populus trichocarpa*	Potri.017G084000.1	MAL	77.90%	677	2034
13	*Medicago truncatula*	Medtr7g010800.1	ROS	76.70%	671	2016
14	*Aquilegia coerulea*	Aqcoe3G077900.1	EUD	75.50%	679	2040
15	*Ananas comosus*	Aco002884.1	BRO	77.30%	678	2037
16	*Cucumis sativus*	Cucsa.158300.1	ROS	73.60%	652	1954
17	*Panicum virgatum*	Pavir.J13145.1	ANG	73.30%	688	2067
18	*Kalanchoe laxiflora*	Kalax.0004s0036.1	PEN	70.70%	676	2031
19	*Zea Mays*	GRMZM2G048836.01	ANG	73.90%	691	2010

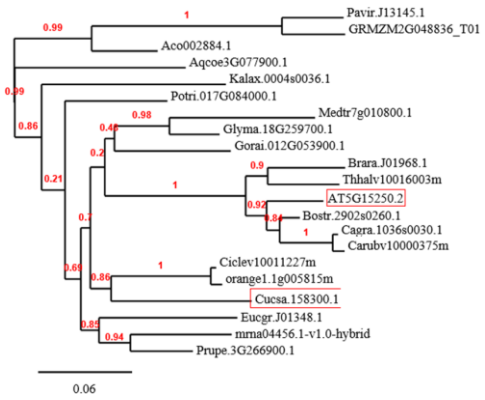


Figure 3.2 Bootstrap Consensus Phylogeny of the *FTSH6* amino acid sequence and its genes in an orthologous. The next adjacent distance tree is given along the branches based on the bootstrap support.

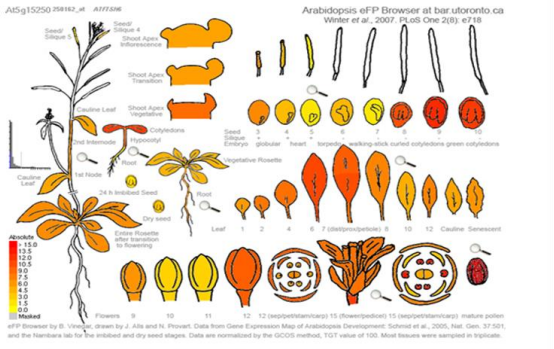


Figure 3.1 Tissue- based differential expression analysis Screening of *FTSH6 (AT5G15250).*

Multispecies plot vs *Arabidopsis thaliana* 40 bp window

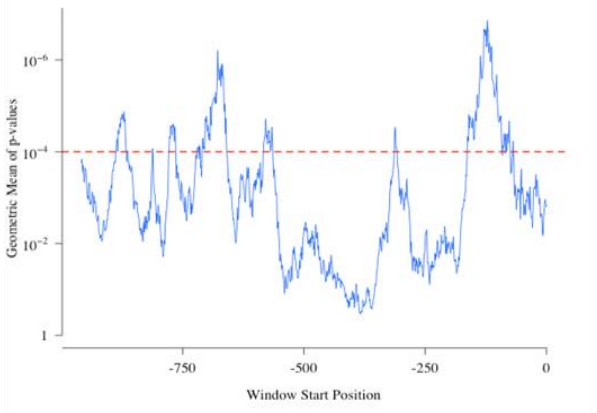
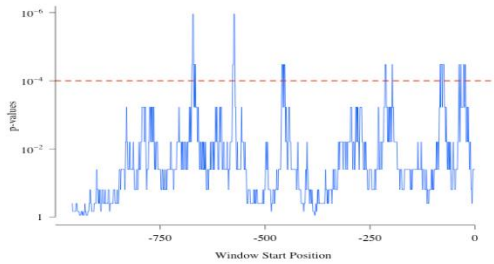


Figure 3.3 Multi-species plot Result of the evolutionarily preserved EAR *FTSH6* sequence identification tool and orthologous genes, The red dotted line specifies the result threshold P= 0.0001. Peaks above the last threshold show also that the window has a highly preserved match for other species. It highlights the consensus sequence only with highest peak crossing the threshold.

A sequence of -561bp to -583bp positions has been maintained in the selected orthologoue gene *Cucumis sativus* (cucsa.158300.1), which means *Cucumis sativus* (cucsa.158300.1) has evolutionarily maintained areas in the promoter region within non- coding DNA at -583bp, and it is clear that Cucumis sativus (cucsa.158300.1) evolutionary has been developed phylogenetically from *Arabidopsis thaliana FtSH6* , and is present in the promoter region as shown in **Figure 3.4**

Cucumis sativus vs *Arabidopsis thaliana* 40 bp window

Figure 3.4 Result of the EAR tool of the



Promoter *FTSH6* and its orthologous *Cucumis sativus (Cucsa.158300.1)* for the identification of the evolutionary conservation regions. The red dotted line indicates the P= 0.0001 threshold. Peaks above this threshold reflect the preservation in the window of other species.

***Zea mays* vs *Arabidopsis thaliana* 40 bp window**

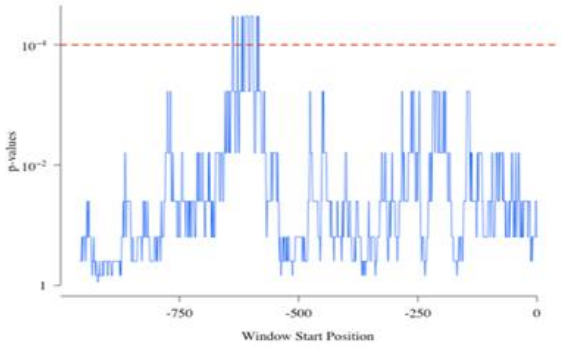


Figure 3.5 Result of the EAR tool of the Promoter *FTSH6* and its orthologous *Zea mays (GRMZM2G082249)* for the identification of the evolutionary conservation regions. The red dotted line indicates the P= 0.0001 threshold. Peaks above this threshold reflect the preservation in the window of other species.

A sequence of -561bp to -583bp positions has been maintained in the selected orthologoue gene *Zea mays* (*GRMZM2G082249*), which means *Zea mays (GRMZM2G082249)* has evolutionarily maintained areas in the promoter region within non- coding DNA at-570bp, and it is clear that *Zea mays* (*GRMZM2G082249*) evolutionary has been developed phylogenetically from *Arabidopsis thaliana FtSH6* , and is present in the promoter region. The marked red line shows the P=0.0001 threshold. The peaks above this threshold reflect the preservation of the window in other species. As shown in **Figure 3.5**, (a consensus) is shown with a highest peak crossing the threshold. A sequence of -275bp and -298bp positions has been maintained in the selected orthologoue gene *Aquilegia coerulea (Aqcoe3G077900.1)*, which means *Aquilegia coerulea (Aqcoe3G077900.1)* has evolutionarily maintained areas in the promoter region within non- coding DNA at -298bp, and it is clear that *Aquilegia coerulea (Aqcoe3G077900.1)* evolutionary has been developed phylogenetically from *Arabidopsis thaliana FtSH6* , and is present in the promoter region. The marked red line shows the P=0.0001 threshold. The peaks above this threshold reflect the preservation of the window in other species. As shown in **Figure 3.6**, (a consensus) is shown with a highest peak crossing the threshold. A sequence of -220bp and -250bp positions has been maintained in the selected orthologoue gene *Ananas comosus (Aco002884.1)*, which means *Ananas comosus (Aco002884.1)* has evolutionarily maintained areas in the promoter region within non- coding DNA at -250bp, and it is clear that *Ananas comosus (Aco002884.1)* evolutionary has been developed phylogenetically from *Arabidopsis thaliana FtSH6* , and is present in the promoter region. The marked red line shows the P=0.0001 threshold. The peaks above this threshold reflect the preservation of the window in other species. As shown in **Figure 3.7**, (a consensus) is shown with a highest peak crossing the threshold.

***Aquilegia coerulea* vs *Arabidopsis thaliana* 40 bp window**

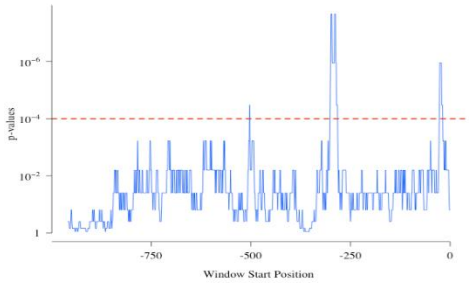


Figure 3.6 Result of the EAR tool of the Promoter *FTSH6* and its orthologous *Aquilegia Coerulea (Aqcoe 3G077900.1)* for the identification of the evolutionary conservation regions. The red dotted line indicates the P= 0.0001 threshold. Peaks above this threshold reflect the preservation in the window of other species.

***Ananas comosus* vs *Arabidopsis thaliana* 40 bp window**

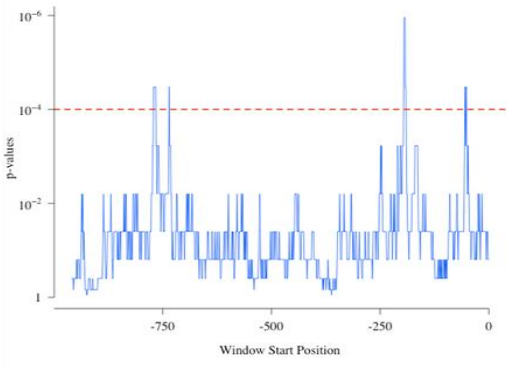


Figure 3.7 Result of the EAR tool of the Promoter *FTSH6* and its orthologous *Ananas comosus (Aco002884.1)* for the identification of the evolutionary conservation regions. The red dotted line indicates the P= 0.0001 threshold. Peaks above this threshold reflect the preservation in the window of other species.

***Prunus Persica* vs *Arabidopsis thaliana* 40 bp window**

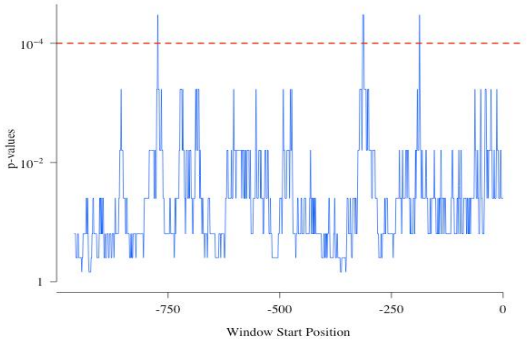


Figure 3.8 Result of the EAR tool of the Promoter *FTSH6* and its orthologous *Prunus Persica (Aco002884.1)* for the identification of the evolutionary conservation regions. The red dotted line indicates the P= 0.0001 threshold. Peaks above this threshold reflect the preservation in the window of other species. It emphasizes the consensus with highest peak crossing the threshold.

A sequence of -145bp to-170bp positions has been maintained in the selected orthologoue gene *Prunus Persica (Prupe.3G266900)*, which means *Prunus Persica (Prupe.3G266900)* has evolutionarily maintained areas in the promoter region within non- coding DNA at -170bp, and it is clear that *Prunus Persica (Prupe.3G266900)* evolutionary has been developed phylogenetically from *Arabidopsis thaliana FtSH6* , and is present in the promoter region. The marked red line shows the P=0.0001 threshold. The peaks above this threshold reflect the preservation of the window in other species. As shown in **Figure 3.8**, (a consensus) is shown with a highest peak crossing the threshold.

A sequence of -552bp to -590bp positions has been maintained in the selected orthologoue gene

Populus trichocarpa (Potri.001G303700.1), which means *Populus trichocarpa (Potri.001G303700.1)* has evolutionarily maintained areas in the promoter region within non- coding DNA at -590bp, and it is clear that *Populus trichocarpa (Potri.001G303700.1)* evolutionary has been developed phylogenetically from *Arabidopsis thaliana FtSH6* , and is present in the promoter region. The marked red line shows the P=0.0001 threshold. The peaks above this threshold reflect the preservation of the window in other species. As shown in **Figure 3.9**, (a consensus) is shown with a curly high peak.

***Populus trichocarpa* vs *Arabidopsis thaliana* 40 bp window**

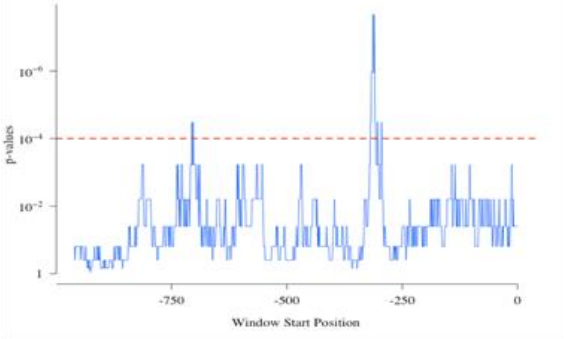


Figure 3.9 Result of the EAR tool of the Promoter *FTSH6* and its orthologous *populus trichocarpa (Potri.001G303700.1)* for the identification of the evolutionary conservation regions. The red dotted line indicates the P= 0.0001 threshold. Peaks above this threshold reflect the preservation in the window of other species. It emphasizes the consensus with highest peak crossing the threshold.

A sequence of -275bp and -320bp positions has been maintained in the selected orthologoue gene *Medicago truncatula (Medtr7g010800.),* which means *Medicago truncatula (Medtr7g010800.)* has evolutionarily maintained areas in the promoter region within non- coding DNA at -320bp, and it is clear that *Medicago truncatula (Medtr7g010800.)* evolutionary has been developed phylogenetically from *Arabidopsis thaliana FtSH6* , and is present in the promoter region. The marked red line shows the P=0.0001 threshold. The peaks above this threshold reflect the preservation of the window in other species. As shown in **Figure 3.10**, (a consensus) is shown with a curly high peak

***Medicago truncatula* vs *Arabidopsis thaliana* 40 bp window**

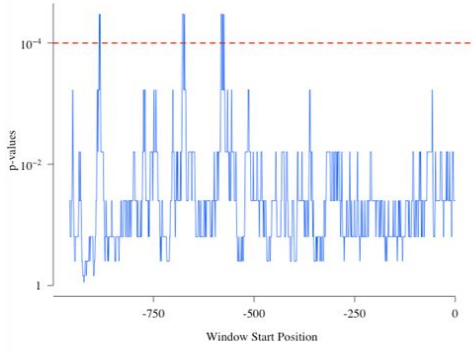


Figure 3.10 Result of the EAR tool of the Promoter *FTSH6* and its orthologous *Medicago truncatula (Medtr7g010800.1)* for the identification of the evolutionary conservation regions. The red dotted line indicates the P= 0.0001 threshold. Peaks above this threshold reflect the preservation in the window of other species.

The orthologous *Citrus clementina (Ciclev10028305 m)* shows a conserved sequence between -280bp and-310bp, which shows that *Citrus clementina (Ciclev10028305 m)* has evolved from *Arabidopsis thaliana FtSH6* is closely linked to *Arabidopsis* because its

promoters have a high sequence identity. The marked red line shows the P=0.0001 threshold. The peaks above this threshold reflect the preservation of the window in other species. as shown in the **Figure 3.11** (a consensus) is shown with a curly high peak.

***Citrus clementina* vs *Arabidopsis thaliana* 40 bp window**

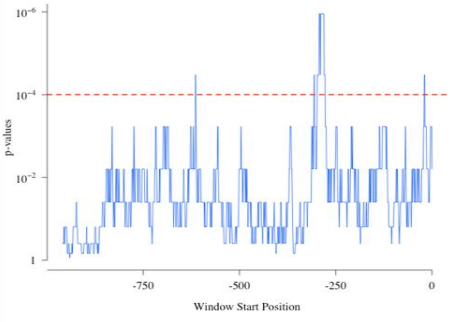


Figure 3.11 Result of EAR *FTSH6* and its orthologous gene *Citrus clementina (Ciclev10028305m)* for the determination of evolutionally preserved areas of the promoter. The red dotted line displays P= 0.0001 in threshold. Peaks over this threshold show that the window has a strongly preserved match in other species.

The orthologous *Citrus sinensis (orange1.1g046605m)* shows a conserved sequence between -276bp to -311bp, which shows that *Citrus sinensis (orange1.1g046605m)* has evolved from *Arabidopsis thaliana FtSH6* is closely linked to *Arabidopsis* because its promoters have a high sequence identity as shown in the (**Figure 3.12**).

***Citrus sinensis* vs *Arabidopsis thaliana* 40 bp window**

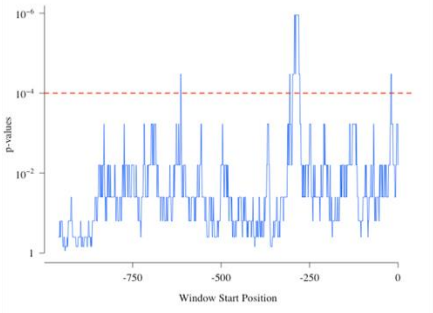


Figure 3.12 Result of EAR *FTSH6* and its orthologous gene *Citrus sinensis (Orange1.1g046605m)* for the determination of evolutionally preserved areas of the promoter. The red dotted line displays P= 0.0001 in threshold. Peaks over this threshold show that the window has a strongly preserved match in other species.

The orthologous *Fragaria vesca (mrna00126.1-v1.0-hybrid)* shows a conserved sequence between -598bp and-642bp, which shows that *Fragaria vesca (mrna00126.1-v1.0-hybrid)* has evolved from *Arabidopsis thaliana FtSH6* is closely linked to *Arabidopsis* because its promoters have a high sequence identity shown in the (**Figure 3.13**).

***Fragaria vesca* vs *Arabidopsis thaliana* 40 bp window**

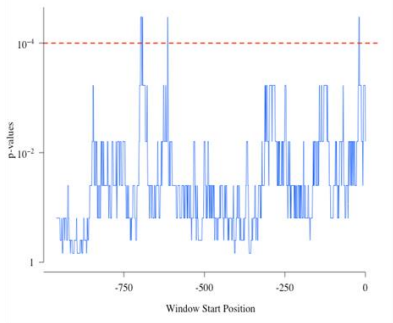


Figure 3.13 Result of EAR *FTSH6* and its orthologous gene *Fragaria vesca (Mrna00126.1-*

V1.0-Hybrid) for the determination of evolutionally preserved areas of the promoter. The red dotted line displays P= 0.0001 in threshold. Peaks over this threshold show that the window has a strongly preserved match in other species. Consensus is highlighted with curly High peak

The orthologous *Gossypium raimondii (Gorai.013G046100.2)* has maintained a sequence between -220bp and -248bp, which means that *Gossypium raimondii (Gorai.013G046100.2)* has evolutionarily conserved regions within non-coding DNA at -248bp in the promoter region¸ which shows that *Gossypium raimondii (Gorai.013G046100.2)* has evolved from *Arabidopsis thaliana FtSH6* is closely linked to *Arabidopsis* because its promoters have a high sequence identity as shown in (**Figure 3.14**).

The orthologuoue *Glycine max (Glyma.18G259700.1*) have conserved sequence(s) between -137bp to -160bp which results that *Glycine max (Glyma.18G259700.1*) have evolutionarily conserved regions within non-coding DNA at -172bp in the promoter region an evident that, *Glycine max (Glyma.18G259700.1*) has been phylogenetically evolved from *Arabidopsis thaliana FtSH6* being taken as model plant and conserved motif is indicated by a red color bracket but the only orthologous which is weakly evolutionary linked with *Arabidopsis thaliana FtSH6* as shown in (**Figure 3.15**).

***Gossypium raimondii* vs *Arabidopsis thaliana* 40 bp window**

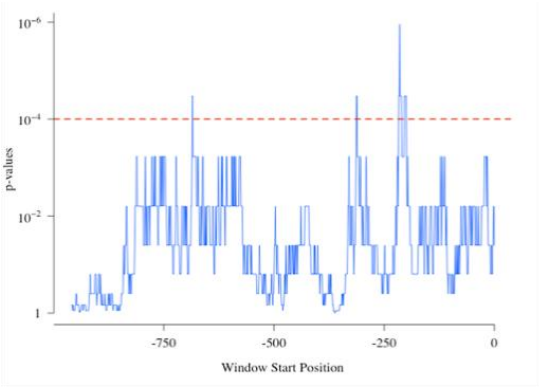


Figure 3.14 Result of EAR Tool for the Identification of Evolutionarily Conserved Regions within the Promoter of *FTSH6* and its Orthologous *Gossypium raimondii (Gorai.013G046100.2.)* The red line shows the meaning threshold of P= 0.0001. Peaks above this threshold indicate that in other species the window has a highly conserved match. Consensus sequence is highlighted with High peak

***Glycine max* vs *Arabidopsis thaliana* 40 bp window**

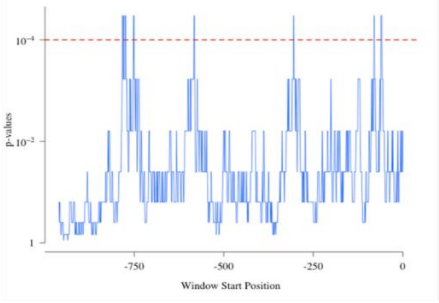


Figure 3.15 Result of EAR Tool For The Identification of Evolutionarily Conserved Regions Within The Promoter of *FTSH6* and its Orthologuous *Glycine max (Glyma.18G259700.1)* The red dotted line shows the threshold of P = 0.001. Peaks above this

threshold indicate that in other species the window has a highly conserved match.

Table 3.2. Position of Conserved Motifs within Promoters of *FtSH6* Meme Tool.

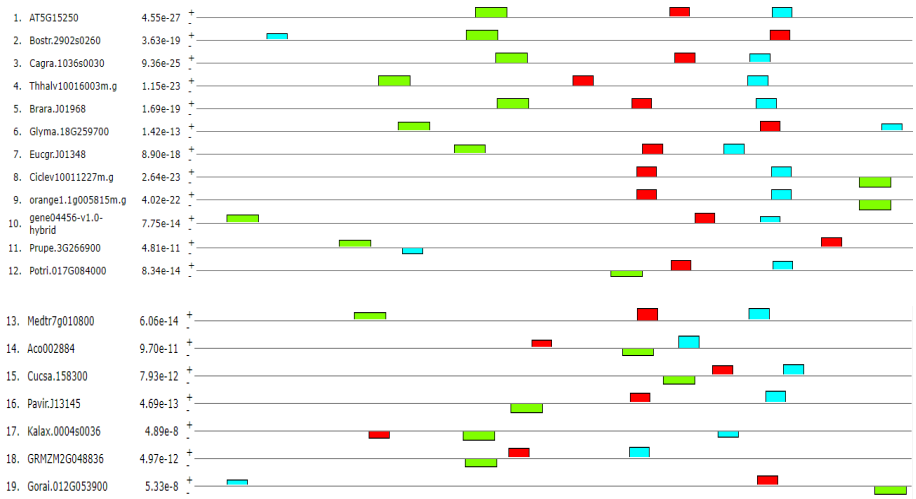
Gene Source	Motif Location
FtSH6 FtSH6	**-661 to -688**
Cucumis sativus (Cucsa.158300.1)	**-723 to -750**

3.4 OVER-REPRESENTATIVE ANALYSIS FOR PUTATIVE *CIS*-ELEMENTS

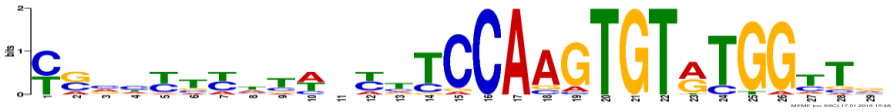
Possible cis-regulatory elements have been identified in the sequences corresponding to the preserved peaks in the *FTSH6* promoter sequences derived from the collected candidates of *FTSH6.* for biological functions and Inter proscan studies which provide for matches based on the highest level of homology (Zhu *et al.,* 1999). We found the peaks above the threshold for P-0.0001, which shows that the other species window matches very well. The motive sequence for "MEME follows "promoter in this region has confirmed that there are consensus sequences between −200 and −1000 positions. The cumulative preservative profile of reference gene *FTSH6* as demonstrated in Figure 3.16, showing Cumulative profiles of the reference gene FTSH6, Distant Species Promoters and Orthologous Promoters. A) 1 kb upstream of the FTSH6 gene was aligned with the EARs tool to identify the evolutionary regulatory regions that have been preserved. B) Display of the positional profile and sequence of consensus 1 with the promoter logo FTSH6 from different plant species. C) Display of the positional profile and sequence of consensus 2 with the promoter logo from various plant species. D) Display the binding sequence at the core.

3.5 EXPLORING PROTEIN-PROTEIN INTERACTION NETWORK. The co-expression network identifies *FTSH6* and its orthologous genes are shown in **(Table 3.3).** The result of STRING for the *Arabidopsis thaliana* network of protein-protein interactions is that, as shown in **Figure 3.17**, the functional partners AT1G56180, EGY3, LHCB3, DEG4 and PTAC5 are encircled by red colour, as shown in **Figure 3.17** and **3.18** Co-expressed proteins are AT1G56180, which have the highest rate of co-expression with the reference gene, including ATP-dependent zinc metalloprotease; unknown protein; unknown protein; unknown biological process; located in chloroplast; expressed in 23 plant structures; expressed in 15 growth stages. The EGY3 Ethylene-dependent gravitropism-deficient and yellow green like 3; S2P-like putative metalloprotease, also contain transmembrane helices near their C-termini and many of them, five of seven, contain a preserved zinc-binding motif HEXXH. EGY1's homologue. The co-expressed protein partner is LHCB3, which may be involved in chlorophyll a-b binding protein 3 chloroplast, the light harvesting complex (LHC) functions as a light receptor, captures photosystems with which it is closely associated and provides excitement energy. Modulates the rate of state transitions of Photosystem II (PSII) and influences the macrostructure of PSII. Involved in the transfer of energy from PSII excitation and separation of charges in thylakoids during photosynthesis. The protein partner PTAC5 was finally found, which is protein disulfide isomerase PTAC5, chloroplastic; exhibits zinc-dependent disulfide isomerase. Required for the development of seedlings and chloroplast under heat stress, probably by maintaining a transcription dependent on plastid-encoded RNA polymerase (PEP), as shown in **Table 3.3 and 3.4**

(A)



(B)



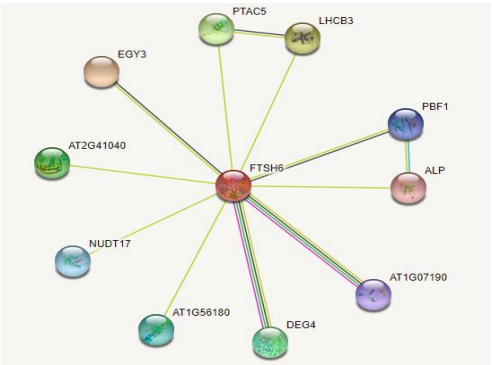
(C)

Log Likelihood Ratio: 380 **Information Content:** 32.1 **Relative Entropy:** 32.3 **Bayes Threshold:** 10.2291

Name	Strand	Start	*p*-value	Sites
4. Carubv10000375m.g	+	650	1.12e-15	AAACTCTCTT TGCCTTTTTACCCTCCAAGTGTATGGTT ATATAAATTT
3. Cagra.1036s0030	+	668	1.12e-15	AAACTCTCTT TGCCTTTTTACCCTCCAAGTGTATGGTT ATATAAATTT
2. Bostr.2902s0260	+	801	1.12e-15	AAACTCTATT TGCCTTTTTACCCTCCAAGTGTATGGTT ACATCAATTT
5. Thhalv10016003m.g	+	526	5.31e-14	AAACTCTGTC TGCCTTTTTAATCTCCAAGTGTATGGTT ATATGAAATT
11. gene04456-v1.0-hybrid	+	696	4.54e-13	ATCCAATGTG CCACTTCATTTCTTCCAAGTGTGTGGCT TAGAGGTAGA
10. orange1.1g005815m.g	+	615	2.68e-12	CAAATTTCCA CCACCTCATAATATCCAAGTGTATGGTT GTATAATTCA
9. Ciclev10011227m.g	+	615	2.68e-12	CAAATTTCCA CCACCTCATAATATCCAAGTGTATGGTT GTAAAATTCA
6. Brara.J01968	+	608	2.68e-12	CATGTGTGTT TGCCCTTTTACCCTCCAAGTGTTTGGAT CATAAGAAAG
14. Medtr7g010800	+	618	1.80e-11	CAATGTGTTC CCCATCAACTTCTTCCAAGTGTGTGGTT ATAACTTGTG
13. Potri.017G084000	+	663	2.89e-11	ATCCAAATGT CCAACTTTTTATTTCCAAGTGTATGGAT GTGGAAGAAA
1. AT5G15250	+	661	6.98e-11	AAACTCTATA TGTCATTTTAGCTTCCAAGTGTATGGTT AAATTTAAGT
8. Eucgr.J01348	+	623	8.24e-11	TTTCAATCCG CCACTTCCGCTTTCCCAAGTGTATGGCT CGAGTCGCAA
7. Glyma.18G259700	+	787	1.24e-10	TCCACACTAC CACTTCAACTTCTTCCAAGTGTGTGGTT GTTACTTACG
12. Prupe.3G266900	+	873	1.45e-9	GACAAACAAA TGGACCACATCATTCCAAGTGTGTGGTT GAGAGGGAGA
20. Gorai.012G053900	+	786	3.91e-9	CATGAATATT TACTGTTTTAATTGCCAAATGTATGGTT GAGACCAAAA
19. GRMZM2G048836	-	568	8.58e-9	GGCGGATAAC CCCCTGTCGTCCCTGCTGGTGAGTGGTG ACTCGTGAGT
16. Cucsa.158300	+	723	1.54e-8	TCTTAAGTCG CCGTTACCTTTTGCACAAGTGTGTGGCG CTTGGAGCAT

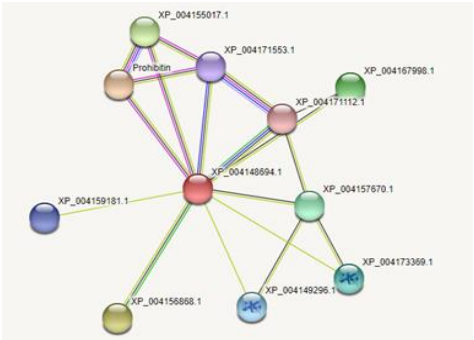
(D) **N..T/CCCAA/CG/ATGTA/GTGG..N**

Figure 3.16 Cumulative profiles of the reference gene *FTSH6 p*romoters and Orthologous Promoters.



		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
FTSH6	ATP-dependent zinc metalloprotease FTSH 6, chloroplastic; Probable ATP-dependent zinc metallopeptidase. Involved in the degradation of the light-harvesting complex of photosystem II (LHC II) during senescence or high light acclimation; In the N-terminal section; belongs to the AAA ATPase family (709 aa)									
Predicted Functional Partners:										
EGY3	Ethylene-dependent gravitropism-deficient and yellow-green-like 3; S2P-like putative metalloprotease, also contain transme...							•		0.831
LHCB3	Chlorophyll a-b binding protein 3, chloroplastic; The light-harvesting complex (LHC) functions as a light receptor, it capture...							•		0.797
PTAC5	Protein disulfide isomerase pTAC5, chloroplastic; Exhibits zinc-dependent disulfide isomerase activity. Required for seedlin...							•		0.759
AT2G41040	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein; Its function is described as methyltransferase...							•		0.734
DEG4	Protease Do-like 4, mitochondrial; Encodes a putative DegP protease (436 aa)					•	•	•		0.662
AT1G56180	ATP-dependent zinc metalloprotease; Unknown protein; Involved in biological_process unknown; Located in chloroplast; Ex...							•		0.642
NUDT17	Nudix hydrolase 17, mitochondrial; Probably mediates the hydrolysis of some nucleoside diphosphate derivatives; Belongs ...							•		0.625
PBF1	N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein; The proteasome is a multicatalytic proteina...					•		•		0.584
AT1G07190	Lon protease; BEST Arabidopsis thaliana protein match is- lon protease 1 (TAIR-AT5G26860.1); Has 106 Blast hits to 106 p...					•	•	•		0.582
ALP	Thiol protease aleurain; May play a role in proteolysis leading to mobilization of nitrogen during senescence and starvation ...							•		0.582

Figure 3.17 Co-Expression Network Identification of the Reference Gene *FTSH6*, the co-expressed proteins encircled with red color,



		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
Predicted Functional Partners:										
Prohibitin	Prohibitin-2-like isoform 1 (290 aa)					•	•	•		0.777
XP_004156868.1	Uncharacterized LOC101220315 (573 aa)							•		0.760
XP_004155017.1	Uncharacterized protein; Prohibitin-1, mitochondrial-like isoform 1 (279 aa)						•	•		0.735
XP_004167998.1	The sequence of the model RefSeq protein was modified relative to its source genomic sequence to represent the inf...							•		0.713
XP_004157670.1	Uncharacterized protein; Derived by automated computational analysis using gene prediction method- Gnomon. Supp...							•		0.706
XP_004173369.1	Chlorophyll a-b binding protein, chloroplastic; The light-harvesting complex (LHC) functions as a light receptor, it capt...							•		0.636
XP_004149296.1	Chlorophyll a-b binding protein 13, chloroplastic-like (264 aa)							•		0.636
XP_004159181.1	Uncharacterized LOC101213430 (403 aa)							•		0.622
XP_004171553.1	Elongation factor Tu, chloroplastic-like (482 aa)					•		•		0.578
XP_004171112.1	Uncharacterized protein; Translation initiation factor IF-2, chloroplastic-like (1023 aa)					•		•		0.573

Figure 3.18 Co-expression relations between *Cucumis sativus* and *Arabidopsis thaliana FtSH6*.

Table 3.3 Co-Expression Network Identification of *FTSH6* and its orthologous genes

Gene Name	Co-expressed gene/protein	Function
FTSH6	EGY3	Ethylene-dependent gravitropism-deficient and yellow-green-like 3; putative metalloprotease S2P-like, also contain transmembrane helices near their C-termini and many of them, five out of seven, contain a preserved zinc-binding motif HEXXH. EGY1's homologue.
FTSH6	LHCB3	Chlorophyll a-b binding protein 3, chloroplastic; the light harvesting complex (LHC) functions as a light receptor, capturing and supplying energy to photosystems closely associated with it. Modulates the rate of state transitions of Photosystem II (PSII) and influences the macrostructure of PSII. Involved in energy transfer of PSII excitation and separation of charges during photosynthesis in thylakoids
FTSH6	PTAC5	Protein isomerase disulfide pTAC5, chloroplastic; zinc-dependent isomerase activity. Required for the development of seedlings and chloroplast under heat stress, probably by maintaining a plastic-encoded transcription of RNA polymerase (PEP)
FTSH6	AT1G56180	Unknown protein; involved in unknown biological process; located in chloroplast; expressed in 23 plant structures; expressed during 15 stages of growth;

Table 3.4 Tissue- based differential expression analysis Screening of *FTSH6*.

Gene	Function
prohibition	Prohibition 2 like isoforms
XP-004156868.1	Uncharacterized LOC1012203155
XP-0041550117.1	Uncharacterized protein, prohibition-1, mitochondrial like isoform1
XP-004149296.1	Chlorophyll a-b binding protein 13, chlropplastic-like
XP-004149269.1	Chlorophyll a-b binding protein. Chlroplastic light harvesting complex function as a light receptor
XP-004171553.1	Elongation factor Tu, chloroplastic-like
XP-0041711122.1	Uncharacterized protein: Translation initiation if 2 chloroplastic like

EFERENCES

kiyama, Y., Kihara, A. and Ito, K., 1996. Subunit of a proton ATPase F0 sector is a substrate of the *FtsH* protease in Escherichia coli. *FEBS letters*, *399*(1-2), pp.26-28.

Auf dem Keller, U., Doucet, A. and Overall, C.M., 2007. Protease research in the era of systems biology. *Biological chemistry*, *388*(11), pp.1159-1162.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, *37*(suppl_2), pp.W202-W208.

Barker, M., de Vries, R., Nield, J., Komenda, J. and Nixon, P.J., 2006. The deg proteases protect Synechocystis sp. PCC 6803 during heat and light stresses but are not essential for

removal of damaged D1 protein during the photosystem two repair cycle. *Journal of Biological Chemistry*.

Barrett, A.J., Rawlings, N.D. and Woessner, J.F., 2003. The Handbook of Proteolytic Enzymes, 2nd ed. *Academic Press, ISB*, *120*(7), pp. 96-104.

Bieniossek, C., Niederhauser, B. and Baumann, U.M., 2009. The crystal structure of apo-*FtsH* reveals domain movements necessary for substrate unfolding and translocation. *Proceedings of the National Academy of Sciences*, pp.pnas-0910708106.

Chen, J., Burke, J.J., Velten, J. and Xin, Z., 2006. *FtsH*11 protease plays a critical role in Arabidopsis thermotolerance. *The Plant Journal*, *48*(1), pp.73-84.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, *5*(1), p.113.

Feller, U., 2004. Proteolysis in Noodén LD, Plant cell death processes. *San Diego Academic Press*, pp. 107-123.

Ferro, M., Brugiere, S., Salvi, D. and Seigneurin-Berny, D., 2010. AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol Cell Proteomics*, *9*, pp.1063-1084.

García L. M., Sjödin, A., Jansson, S. and Funk, C., 2006. Protease gene families in Populus and Arabidopsis. *BMC Plant Biology*, *6*(1), p.30.

Geer, L.Y., Marchler-Bauer, A., Geer, R.C., Han, L., He, J., He, S., Liu, C., Shi, W. and Bryant, S.H., 2009. The NCBI biosystems database. *Nucleic acids research*, *38*(suppl_1), pp.D492-D496.

Gibala, M., Kicia, M., Sakamoto, W., Gola, E.M., Kubrakiewicz, J., Smakowska, E. and Janska, H., 2009. The lack of mitochondrial At*FtsH*4 protease alters Arabidopsis leaf morphology at the late stage of rosette development under short-day photoperiod. *The Plant Journal*, *59*(5), pp.685-699.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D.S., 2011. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*, *40*(D1), pp.D1178-D1186.

Hinnerwisch, J., Fenton, W.A., Furtak, K.J., Farr, G.W. and Horwich, A.L., 2005. Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell*, *121*(7), pp.1029-1041.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W. and Zimmermann, P., 2008. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Advances in bioinformatics*, *2008*.

Ito, K. and Akiyama, Y., 2005. Cellular functions, mechanism of action, and regulation of *FtsH* protease. *Annu. Rev. Microbiol.*, *59*, pp.211-231.

Janska, H., Piechota, J. and Kwasniak, M., 2010. ATP-dependent proteases in biogenesis and maintenance of plant mitochondria. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1797*(6-7), pp.1071-1075.

Kato, Y., Miura, E., Ido, K., Ifuku, K. and Sakamoto, W., 2009. The variegated mutants lacking chloroplastic *FtsH*s are defective in D1 degradation and accumulate reactive oxygen species. *Plant Physiology*, *151*(4), pp.1790-1801.

Kicia, M., Gola, E.M. and Janska, H., 2010. Mitochondrial protease At*FtsH*4 protects ageing Arabidopsis rosettes against oxidative damage under short-day photoperiod. *Plant signaling & behavior*, *5*(2), pp.126-128.

Kihara, A., Akiyama, Y. and Ito, K., 1995. *FtsH* is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proceedings of the National Academy of Sciences*, *92*(10), pp.4532-4536.

Komenda, J., Tichý, M., Prášil, O., Knoppová, J., Kuviková, S., de Vries, R. and Nixon, P.J., 2007. The exposed N-terminal tail of the D1 subunit is required for rapid D1 degradation during photosystem II repair in Synechocystis sp PCC 6803. *The Plant Cell*, *19*(9), pp.2839-2854.

Kwasniak, M., Pogorzelec, L., Migdal, I., Smakowska, E. and Janska, H., 2011. Proteolytic system of plant mitochondria. *Physiologia plantarum*, *145*(1), pp.187-195.

Liu, X., Yu, F. and Rodermel, S., 2010. Arabidopsis chloroplast *FtsH*, var2 and suppressors of var2 leaf variegation: a review. *Journal of integrative plant biology*, *52*(8), pp.750-761.

Malnoë, A., Wollman, F.A., De Vitry, C. and Rappaport, F., 2011. Photosynthetic growth despite a broken Q-cycle. *Nature communications*, *2*, p.301.

Picot, E., Krusche, P., Tiskin, A., Carré, I. and Ott, S., 2010. Evolutionary analysis of regulatory sequences (EARS) in plants. *The Plant Journal*, *64*(1), pp.165-176.

Piechota, J., Kolodziejczak, M., Juszczak, I., Sakamoto, W. and Janska, H., 2010. Identification and characterization of

high molecular weight complexes formed by matrix AAA proteases and prohibitins in mitochondria of *Arabidopsis thaliana*. *Journal of Biological Chemistry*, *285*(17), pp.12512-12521.

Rawlings, N.D., Tolle, D.P. and Barrett, A.J., 2004. MEROPS: the peptidase database. *Nucleic Acids Research*, *32*(suppl_1), pp.D160-D164.

Rhee, S.Y. and Flanders, D.J., 2000. Web-based bioinformatic tools for Arabidopsis researchers. *Arabidopsis: A Practical Approach, Oxford University Press, Oxford*, pp.225-265.

Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N. and Baumeister, W., 2000. The genome sequence of the thermoacidophilic scavenger Thermoplasma acidophilum. *Nature*, *407*(6803), p.508.

Sakamoto, W., Zaltsman, A., Adam, Z. and Takahashi, Y., 2003. Coordinated regulation and complex formation of yellow variegated1 and yellow variegated2, chloroplastic *FtsH* metalloproteases involved in the repair cycle of photosystem II in Arabidopsis thylakoid membranes. *The Plant Cell*, *15*(12), pp.2843-2855.

Schmid, C., and Mikolajczyk, K., 2005. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, *27*(10), pp.1615-1630.

Sedaghatmehr, M., Mueller-Roeber, B. and Balazadeh, S., 2016. The plastid metalloprotease *FtsH*6 and small heat shock protein HSP21 jointly regulate thermomemory in Arabidopsis. *Nature communications*, *7*, p.12439.

Simpson, J.C., Joggerst, B., Laketa, V., Verissimo, F., Cetin, C., Erfle, H., Bexiga, M.G., Singan, V.R., Hériché, J.K., Neumann, B. and Mateos, A., 2012. Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway. *Nature cell biology*, *14*(7), p.764.

Suno, R., Niwa, H., Tsuchiya, D., Zhang, X., Yoshida, M. and Morikawa, K., 2006. Structure of the whole cytosolic region of ATP-dependent protease *FtsH*. *Molecular cell*, *22*(5), pp.575-585.

Urantowka, A., Knorpp, C., Olczak, T., Kolodziejczak, M. and Janska, H., 2005. Plant mitochondria contain at least two i-AAA-like complexes. *Plant molecular biology*, *59*(2), pp.239-252.

Van Wijk, K.J., 2015. Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annual Review of Plant Biology*, *66*, pp.75-111.

Vierstra, R.D., 2003. The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends in plant science*, *8*(3), pp.135-142.

Wagner, R., Aigner, H. and Funk, C., 2012. *FtsH* proteases located in the plant chloroplast. *Physiologia plantarum*, *145*(1), pp.203-214.

Wagner, R., Aigner, H., Pružinská, A., Jänkänpää, H.J., Jansson, S. and Funk, C., 2011. Fitness analyses of *Arabidopsis thaliana* mutants depleted of *FtsH* metalloproteases and characterization of three *FtsH*6 deletion mutants exposed to high light stress, senescence and chilling. *New Phytologist*, *191*(2), pp.449-458.

Xue, G.P., Drenth, J. and McIntyre, C.L., 2015. TaHsfA6f is a transcriptional activator that regulates a suite of heat stress protection genes in wheat (Triticum aestivum L.) including previously unknown Hsf targets. *Journal of Experimental Botany*, *66*(3), pp.1025-1039.

Zaltsman, A., Ori, N. and Adam, Z., 2005. Two types of *FtsH* protease subunits are required for chloroplast biogenesis and photosystem II repair in Arabidopsis. *The Plant Cell*, *17*(10), pp.2782-2790.

Zhang, D., Kato, Y., Zhang, L., Fujimoto, M., Tsutsumi, N. and Sakamoto, W., 2010. The *FtsH* protease heterocomplex in Arabidopsis: dispensability of type-B protease activity for proper chloroplast development. *The Plant Cell*, pp.tpc-110.