Expansion of Protein Domains and Stress Tolerance in Barley: An *In silico* Approach

Villayat Ali[1,*], Priyanka Singh[1] and Malkhey Verma[2,*]
[1]Department of Biochemistry and Microbial Sciences, School of Applied and Basic Sciences, Central University of Punjab, Bathinda, Punjab, 151001, India
[2]School of Biotechnology Institute of Science, Banaras Hindu University Varanasi, Uttar Pradesh, 221005, India
*Corresponding Author email:villayatgoba@gmail.com; malkhey@yahoo.com

ABSTRACT

In eukaryotes, proteins contain tandem repeats of numerous domains which are vital for providing a surface that promotes or mediates in protein-protein interactions as well as ligand binding like DNA and RNA, which are required for biological functions. In this study, we come across that most of the proteins domains in barley are classified as tandem repeats and these domains are assigned to deal with adverse environmental conditions. About 100 to 377 such repeats like pentatricopeptide repeats (PPRs), Leucine-rich repeats (LRRs), zf-RING 2, WD40, Myb_DNA-binding, F-box, and Ankyrin (ANK 2) repeats, which are known for their critical functions in cell physiology, stress tolerance, and development. Based on their functions we can say that barley is a naturally stress tolerant plant. Here, we explore the biological functions of the protein domains of the barley genome through Pfam database, via systematically screening the Pfam domain on HMMER (bio-sequence analysis using profile hidden Markov models), a fast and sensitive homology search that reveals the distribution of protein domain. We recognized 2,872 unique domains in 13,797 proteins out of 24,211 protein sequences in *Hordeum vulgare*. In the 13,797 sequences, 42 percent sequences are multi-domain proteins and 32 percent with multiple occurrences of the same domains. And LRR-8 was identified as one of the largest repeats in the barley genome.

Keywords: *Hordeum vulgare, Protein Domain, Tandem repeats (TRs), PPR, LRR*

INTRODUCTION

Hordeum vulgare is a self-pollinating monocot plant belonging to Poaceae, and commonly known as barley. It has been domesticated and cultivated for at least 17,000 years. It is the fourth important cereal crop besides wheat, maize, and rice in the world (Akar *et al*., 2004). It is also known for its natural tolerance towards salt, drought and fungal disease along with adaptation to diverse environmental conditions (Dai *et al*., 2012).

Plants are subjected to a variety of diverse environmental circumstances during their existence (Han *et al*., 2021). Stress is a term that refers to a set of circumstances that are not favourable to plant growth and development (Lu *et al*., 2020). Abiotic stressors can significantly reduce crop growth and production, and they are a major limiting factor in crop yields around the world (Han *et al*., 2021). Salt and water stress harm 10% of the world's agricultural land, lowering crop yields by more than half, while biotic stresses like pathogens and diseases also cause significant crop losses. Plants have developed many regulatory mechanisms to acquire and react to various stress signals to adapt to changing environmental conditions (Han *et al*., 2021). Whenever a plant is stressed, the relevant regulatory pathway is activated, and the subsequent reactions aid the plant's survival (Lu *et al*., 2020).

Direct structural and functional identification of each of an organism's proteins is excessively costly and complex. Even if no actual

J. Himalayan Ecol. Sustain. Dev. Vol. 16 (2021)

ISSN 0973-7502

experimentation on the organism has been done, genome sequencing is valuable information for studying its molecular biology using bioinformatics tools. When you have an entire genomic sequence, you can start asking broad questions about an organism's metabolic potential and molecular systems. In proteome research, information transfer among associated proteins is crucial. While comparing entire sequence data is a valuable tool for identifying near and direct links, it overlooks the more nuanced correlations among proteins (Yeats *et al.*, 2003). The identification of a protein's domain composition is a more complex way of analysing it. So with the application of *in-silico* analysis and tools, we can identify structural and functional aspects of proteins architecture within a short time (Azlan *et al.,* 2021).

Proteins consist of evolutionarily conserved domains (Lee and Lee, 2009). Protein domains are more typically found in combination with other domains in multi-domain proteins, but they can also be found alone (Bjorklund *et al*., 2006). On the basis of their structure and function, domains are distinguished into two types i.e., functional domain and structural domain (Lee and Lee, 2009). Chothia in 1992 discovered that domains evolve independently and that the mechanism of evolution shows that in large-scale evolution, often unrelated proteins arise with many conserved structural and functional domains. Around 80% of proteins in eukaryotes and two-thirds of proteins in prokaryotes are multi-domain proteins (Lee and Lee, 2009). Predicting protein domains is essential for understanding a protein function at the molecular level, but without a reference genome sequence, we cannot predict a protein domain. Protein sequences are categorized into family and

domain using the Pfam database (Mistry *et al*., 2021). In this study, we aim to explore the biological functions of the protein domains of the barley genome through Pfam database, and for this, we have systematically screened the Pfam domain on HMMER, a fast and sensitive homology search that reveals the distribution of protein domains. In Table 1. we have summarized the primary function of the reported protein domains in the barley genome.

The literature revealed that in eukaryotes most of the domains of the protein contain highly expanded tandem repeats (TRs) in comparison to prokaryotes, which acclimatize the higher organisms to deal with adverse environmental conditions. In the case of plants, it was assumed that the expansions of protein tandem repeats are progressed through internal tandem duplications and recombination of genetic material (Sharma and Pandey, 2016). The basic root of this evolution is the genetic alterations that include mutation, duplication, and recombination. DNA duplication is a random process that may occur in any part of the coding DNA segment, that leads to the simple sequence repeats (Pearson and Sinden, 1998). These repeats later get translated into single amino acid repeats or tandem oligo-peptide repeats, and finally lead to altered protein structure and function (Katti *et al*., 2000). Based on length and functions, tandem repeats are categorized into the following types (Katti *et al*., 2000):

- Shorter oligopeptides repeat (approximately 2 to 20 amino acids), which are structurally or functionally insufficient on their own, and provide supplemental function during repeat interactions.

J. Himalayan Ecol. Sustain. Dev. Vol. 16 (2021)

ISSN 0973-7502

- Moderate oligopeptides are repeats of 20 to 40 amino acids that can form structurally distinct units but not function independently. In addition, several of them develop a three-dimensional structure for communicating protein-protein interactions.

- Oligopeptides repeat more than 100 amino acids that are structurally and functionally competent on their own.

MATERIALS AND METHODS

Protein sequence and analysis tools

Barley protein sequences (*Hordeum vulgare*, UniProt: UP000011116 and Taxonomy ID:112509) were obtained from reference proteome data available at http://www.uniprot.org/downloads. Pfam, a protein family and domain database that is frequently used to explore emerging genomes, as well as to drive experimental job on specific protein and the HHMER software, which automatically creates a Hidden Markov Models (HMM) from the seed alignment and searches it against the pfamseq database (Mistry *et al*., 2021) were used to identify proteins domains.

Protein domain identification and analysis

HMM of known domains were downloaded from the Pfam database. Reference proteome contained 24,211 protein sequences, which were searched against Pfam HMM database using *hmmscan* program from HMMER package (Finn *et al*., 2011). To consider the relevance of the output sequence and the resulting hit, inclusion thresholds of e-value <= 0.01 were utilized. Using in-house Perl scripts, the output of the *hmmscan* search was examined to extract the domain architecture of each protein sequence. In 13,797 protein sequences, at least one domain was

found. Gene ontology mapping to Pfam domains was downloaded from

http://geneontology.org/external2go/pfam2go. The data obtained from the above analysis were used in Microsoft Excel to generate the histograms.

RESULTS AND DISCUSSION

To identify proteins domains in *Hordeum vulgare* a reference proteome of the barley genome were searched against the Pfam HMM database using hmmscan program from the HMMER package, with e-value <= 0.01. Using our approach, we identified 2,872 unique domains in 13,797 out of 24,211 protein sequences in *Hordeum vulgare*. The remaining 57 % sequences could not be assigned to any known domain at our filtering criteria. Out of 13,797 sequences, 42 percent of sequences are multi-domain proteins (with two or more domains), and 32 percent when proteins with multiple occurrences of the same domains are not counted. It indicates that approximately 10 percent of proteins harbor the same domains multiple times. After the identification of a myriad of protein domains in the barley genome, we choose the most repeated and expanded protein domains which play the key biological functions in the barley genome. Based on biological function and its distribution we categorized our results into different subheadings.

Expansion of Tandem Repeat Domain Families in Barley Genome

Our analysis found that most of the protein domains fall in the category of tandem repeats which are variants in oligopeptides repeats and functions. Researchers previously reported that multicellular organisms have a lot of protein

domain repeats (Ekman *et al.*, 2005). However, it is not the case that only multicellularity determines the presence of multiple repeats, small organisms like yeast and worms have the same proportion of protein repeats. Furthermore, some prokaryotes like *Pseudomonas aeruginosa* and *Escherichia coli* also contains proteins with repeats (Bjorklund *et al.*, 2006).

Our study demonstrates that the barley genome contains well-featured tandem repeats (TRs) families such as pentatricopeptide repeats (PPRs), Leucine-rich repeats (LRRs), zf-RING_2, WD40 and Ankyrin (ANK_2) repeats those play a vital role in development and defence mechanisms, protein domain frequency and its distribution are shown in Fig. 1. The TRs belong to pentatricopeptide repeats (PPRs) families having the length of 35 amino acids (Sharma and Pandey, 2016), and during the study, we found that all the families of PPR comprise of the diverse length of tandem sequence repeats which vary in amino acid sequences (308 protein domains of PPR_2, 156 PPR, and 84 PPR_1), families are found and they are exclusively expanded in terrestrial plants, and their expression is anticipated in mitochondrial and chloroplast genes (Schmitz-Linneweber and Small, 2008). Moreover, the least number of PPR domains are noticed in some prokaryotes like rhodobacter, simkania, ralstonia, and legionella and it has been suggested that the genes get transferred between eukaryotes and prokaryotes through horizontal gene transfer (Schmitz-Linneweber and Small, 2008).

Leucine-rich repeats (LRRs) proteins are commonly well-explored signalling molecules in plants. LRR-motifs are extracellular proteins that are the key components of the plant receptor-like kinase (LRR-RLK) include a major subfamily in plants (Sharma and Pandey, 2016). And it has a length of 20–30 amino acids, and are universally expanded protein domains. Three families (LRR_4, LRR_8, and LRRNT_2), are well established in barley, out of which LRRNT_2 was restricted in plants only. The LRR_8 domain was spotted in 712 proteins, LRR_4 in 118 and LRRNT_2 in 208 proteins. In our analysis LRR-8 was identified as one of the largest repeats in the barley genome likewise, in arabidopsis LRR was recognized as the biggest gene family (Sharma and Pandey, 2016).

The Zn-RING_2 domains are the utmost rich proteins domains in eukaryotic genomes. Zinc finger domains are known for their diverse functions like RNA packaging, DNA recognition, transcriptional stimulation, folding of protein and assemblage, and binding of lipid (Han *et al.*, 2021). Likewise in plants, zinc finger proteins are active in scores of functions ranging from plant growth development to numerous biotic and abiotic stresses (Han *et al.*, 2021). As their functions Zinc fingers have diverse structures. Literature reveals that C2H2 and CCCH (classified based on the position of Cys and His residues), Zinc finger proteins families are counted in plant development activities, while C3HC4 families are related to biotic and abiotic stresses (Han *et al.*, 2021). We identified 183 proteins of the Zn-RING_2 domain in the barley genome (Fig. 1).

WD40 proteins are mostly involved in various cellular processes like cell divisions, vesicle formation, signal transduction, RNA processing and remodelling of chromatin via modification of histone proteins (Tan *et al.*, 2021). Usually, WD40 repeats proteins are known for protein-protein interactions via providing a scaffold, which facilitates the activity of protein-protein

interaction (Tan *et al.*, 2021). The interaction between WD40 repeat proteins, MYB and bHLH transcription factors have been extensively studied in recent years, and flavonoid biosynthesis may be an excellent model for their interactions (Nakatsuka *et al.*, 2008). WD40 consists of around 44 to 60 amino acids (Tan *et al.*, 2021) in this investigation, we found 256 instances of WD40 in 121 proteins, as well as 237 and 200 WD40 repeat family members in arabidopsis and *Oryza sativa* respectively were found by Ouyang *et al*. in 2012).

Ankyrin (ANK_2) which contains 33 amino acids residues are evolutionarily conserved protein domains that occur in a wide diversity of prokaryote and eukaryotes and act as a mediator for protein-protein interactions (Li *et al.*, 2006). ANK repeats, a light-dependent protein in arabidopsis, are involved in cell differentiation and development (Zhang *et al.*, 1992). BOP1, an ANK protein is required for leaf morphogenesis (Ha *et al.*, 2004). Regulation of ethylene signalling and biosynthesis are related to XBAT32 and XBAT35 are ANK containing proteins (Lyzenga *et al*., 2012). Recent studies have revealed that numerous AKN repeats express their reactions in response to biotic and abiotic stresses in plants. Expression of rice OsBIANK1 gene, having coding

ANK protein is changed in pathogen infested rice seedling as related to normal or control, which proposes its contribution in disease resistance (Zhang *et al.*, 2010).

Results discussed above demonstrate that the repeat domain families are remarkable in numbers than non-repeat gene families in the barley genome. Large numbers of TRs, in some cases, assign their large-scale tandem gene duplication or internal genetic duplication within genes families as revealed by Ponting et al. in 1999. Despite of this uncertainty, still continues the selection pressure fast the rate of TRs gain/loss this rapid process create genetic diversity, adjust the plant to adverse conditions along with the changing environment (Richard *et al.*, 2008). So it was demonstrated that conservation of protein sequences is a favourable mechanism from an evolutionary point rather than gaining/loss of TRs, because continuous changes in sequence lengths cause instability in protein structure and its activities (Schaper and Anisimova, 2015), during the course of evolution excessive distribution of TRs families in plants, indicate its typical roles in development and growth of the plant. Despite that its mechanism is yet to be understood (Yang *et al.*, 2010).
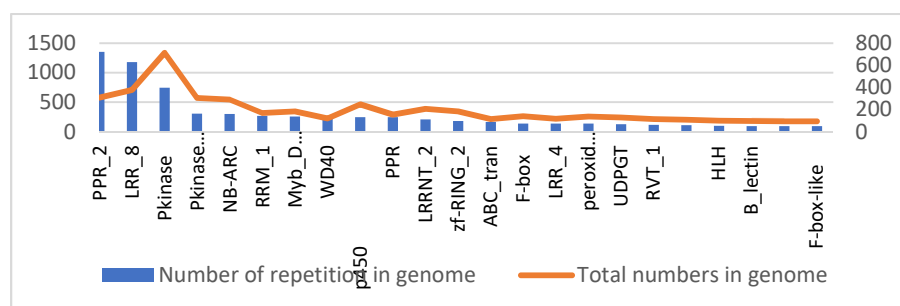

Fig. 1 Distribution of protein domain and its expansion in the barley genome. Blue bar indicates the repetition of single protein domain unit in the genome and orange line indicates the overall number of the protein domain in genome.

Proteins Domains active in Plant Development and Stress Response

Based on the findings of our analysis we described the most repeated protein domains, which are known for their roles in plant development and stress responses. Repeated sequence containing proteins are actively involved in numerous biological functions like cellular biosynthesis, signal transduction besides that repeated proteins are involved in biotic and abiotic stress stimuli (Schaper and Anisimova, 2015). In this manuscript, we have depicted the functions of the most repeated and expanded proteins domains in the barley genome.

Pentatricopeptide Repeat

In the barley genome, we identified many families of pentatricopeptide repeat (PPR) proteins which constitute one of the greatest gene families in barley after LRR families. Generally, PPR families are forecasted to be located in the mitochondria and chloroplast organelles. Most of the PPR families are distinguished for their role in plant development, EMB175 protein which consists of 14 PPR motifs plays a vital role in embryo development in arabidopsis, cell division was affected by mutation in emb175 which out turn in morphological arrest. EMB175 is also projected to be localized in the chloroplast. Many of the PPR families are projected to be a restorer of cytoplasmic male sterility (CMS) in plants. Rf (Restorer of fertility) locus protein Rf-PPR592 which consists of 14 PPRs motif is distinguished as mitochondrial housed protein, has the ability to re-establish fertility to rf/rf CMS lines (Sharma and Pandey, 2016). While in radish, Rfo is an additional fertility restorer gene accompanied by

many PPR, was cloned by using microsynteny analysis, which suggested to have mitochondrial targeting peptide signal. In *B. napus* (ogu) a CMS mutant was recovered to male fertile plant by transforming Rfo into Ogura (Desloire *et al*., 2003). Much molecular identification forecast PPRs roles in biotic and abiotic stresses. Recent studies reveal that pentatricopeptide repeat (PPR) protein for germination on NaCl (PGN) was recognized as a regulator of biotic and abiotic stress response in arabidopsis. Inactivation of PGN in arabidopsis results in an encounter towards necrotrophic fungal pathogen as well as to hypersensitivity to abscisic acid (ABA), glucose and salinity (Laluk *et al*., 2011). SLG1 (slow growth 1) a mitochondrial PPR protein in arabidopsis, mutant *slg1* cause slow growth and delayed development along with increased response towards ABA and tolerance to drought stress (Sharma and Pandey, 2016). The role of (PPR40) a mitochondrial PPR protein was assigned as a possible connection between mitochondrial electron transport, stress and hormone regulatory control in arabidopsis. PPR40 deficiency leads to increased ROS generation, enhanced the activity of superoxide dismutase and lipid peroxidation, and alteration of multiple stress-sensitive genes in arabidopsis, implying a link between cellular respiration and stress adaptation. In mitochondria, ABA excessively responsive 5 (ABO5) expresses a pentatricopeptide repeat protein necessary for splicing of nad2 (NADH dehydrogenase subunit 2) intron 3. Plants with abo5 mutations had decreased transcript levels of stress-inducible genes like RD29A, COR47, and ABF2 (Sharma and Pandey, 2016). PPR proteins are active in post-

J. Himalayan Ecol. Sustain. Dev. Vol. 16 (2021)

ISSN 0973-7502

transcriptional control of chloroplast and mitochondrial genes, such as RNA maturation, intron splicing, editing and translation initiation, according to many findings (Wang *et al.*, 2021). While many mitochondrial/chloroplast PPR proteins are distinguished to conduct various roles in plant biotic and abiotic mechanisms, but recently Jiang *et al.*, 2015 characterized a nucleo-cytoplasmic localized PPR protein SOAR1, which negatively regulate ABA signalling during seed germination and post-germination (Jiang *et al.*, 2015). As a result of its involvement in a variety of critical biological activities, PPR has a large representation in the barley genome, as stated above.

LRR Repeats

LRR- containing proteins are recognized in all kingdoms. In the barley genome, we identified several families of LRR domains in which LRR_8 was observed 1181 times in 377 proteins. While LRR-RLK (plant receptor-like kinases) having extracellular leucine-rich repeat motif includes one of the largest subfamilies in plants, that LRR-RLK was the most studied signalling molecule in plants. The number of studies of this subfamily justifies their role in numerous developmental and defence-related activities such as cell proliferation, stem cell development and host-specific and non-specific defence action (Sharma and Pandey, 2016). LRX1 and LRX2 are LRR containing proteins, which has been distinguished in cell wall formation. LRX is a leucine-rich repeat having an N-terminal LRR domain and C-terminal extensin domain, while the extension domain is involved in crosslinking with cell wall constituents (Draeger *et al.*, 2015). ERECTA-family (ERL1 and ERL2) leucine-rich-repeat receptor-like kinases (LRR-RLKs) controls stomatal cell differentiation in

arabidopsis (Sharma and Pandey, 2016). The expression pattern of OsGIRL1 (*Oryza sativa* gamma-ray induced LRR-RLK1) in rice was studied under various abiotic stress and phytohormone situations. Under salt, osmotic and heat stress conditions, OsGIRL1 transcript levels were considerably increased. Treatment with salicylic acid (SA) and abscisic acid (ABA) resulted in similar inductions; however, treatment with jasmonic acid resulted in a substantial decrease in OsGIRL1 expression. These results point to OsGIRL1 as a possible stress signal receptor in plants (Sharma and Pandey, 2016).

NB-ARC another vastly studied LRR sub-family, which is expanded in bacteria and plant genomes particularly in the Poaceae family and in barley 304 times NB-ARC domains were identified in 291 proteins. NB-ARC acts as the ATPase domain and its nucleotide-binding state is suggested to modulate the R protein action. In plants, R proteins are key components in pathogen recognition and are usually granted as the plant innate immune system (Van Ooijen *et al.*, 2008). These R proteins demonstrate the general domain network with central NBS at N-terminal and LRR repeats at C-terminal (DeYoung *et al.*, 2006). So the massive presence and expansion of LRR repeats in the barley genome is obvious from its functional point of view, therefore, our analysis reveals 1181 times of LRR_8 protein domains in 113 proteins.

Ankyrin Repeats

Ankyrin repeats are found often in eukaryotes, and in the barley genome, we found ANK_2 domain repeats. The sequence motif of Ankyrin-repeats is studied via multiple sequence alignment to find out the conserved amino acid residues which are necessary for folding and

J. Himalayan Ecol. Sustain. Dev. Vol. 16 (2021)

ISSN 0973-7502

stability. Only a few amino acids in the ANK repeating units are conserved as a feature of degenerate repeats, while several hydrophobic sites are maintained that are required to maintain the secondary structure. Transmembrane domains (AtANKTM) are known as the most abundant groups which consist of ANK protein repeats (Sharma and Pandey, 2016). In animals, numerous ANK transmembrane proteins have been characterized. Arabidopsis Ankyrin Repeat (AKR) was first characterized by ANK proteins in plants and the gene expression of ANK is light-dependent (Zhang *et al.*, 1992). Consequently, examination demonstrates that AKRP and EMB506 are interacting protein partners and both possess five ANK repeats and both the partners of the AKRP-EMB506 protein complex are vital for morphogenesis and organogenesis during the life cycle (Sharma and Pandey, 2016). TIP1 (TIP growth defective 1) is another ANK protein that is vital for cell differentiation and encode for S-acyltransferase which is requisite for the usual growth of cell and root hair development in plants. Additionally, XBAT32 an ANK containing ubiquitin-mediated protein which controls lateral root growth in the arabidopsis plant (Sharma and Pandey, 2016). Many studies are carried out to explicate the function of ANK repeats protein regarding plant growth and development, response towards biotic and abiotic stresses (Lopez-Ortiz *et al.*, 2020). Few of them are distinguished to play roles under stress conditions. The AKT1, inward-rectifier K^+ channel which consists of five ANK repeats toward its C-terminus and numerous transmembrane domains and the AKT1 channel act as essential in root potassium uptake (Alemán *et al.*, 2011). After further studies, it was revealed that the mutant of akt1 shows ABA hypersensitivity which improves drought-tolerant towards water-deficient situations by enhancing stomatal closure. AKT1 regulatory *cipk23* mutant also shows similar activity and it was proposed that both proteins play the role of adverse controller of plant drought response (Nieves-Cordones *et al.*, 2011). In plants concentrations of salicylic acid are required to respond effectively to plant pathogens for disease resistance. A defence protein containing an ankyrin repeat, NPR1/NIM1 (Non-expressor of PR1/Non-inducible Immunity1) are detect in plants (Sedgwick and Smerdon, 1999). Due to insufficient gene expression of PR-1 (pathogenesis-related) genes, plants lacking NPR1/NIM1 are unable to respond to various SAR-induction (Systemic Acquired Resistance) and become prone to pathogen infection (Sharma and Pandey, 2016).

WD40 Proteins

WD40 proteins domain are present across all eukaryotes and in the barley genome we noticed 256 times in 121 proteins. They are the most clustered interactor as indicated by various interactome studies and their architectural findings indicated that this feature results in their capability to interact with multiple proteins, peptides by serving their rigid scaffold for protein interaction (Tan *et al.*, 2021). They are associated with diverse functions like signal transduction, cell cycle control, cell-cell communication and apoptosis (Tan *et al.*, 2021). In the arabidopsis, they modulate biosynthesis of flavonoids by creating proteins complexes with cooperation with the MYB transcription factor which controls the expression of biosynthetic genes for anthocyanins and proanthocyanidins (Matsui *et al.*, 2008). Extensive analysis of CYP71 (CYCLOPHILIN71), a WD40 repeat protein

J. Himalayan Ecol. Sustain. Dev. Vol. 16 (2021)

ISSN 0973-7502

proposed their involvement in chromatin-based silencing of genes and therefore play a vital role in the control of gene mechanism. Mutant *cyp71* in plants displayed abnormal growth, like defects in lateral organ growth development, abnormal activity in the apical meristem and sessile root growth. Association of CYP71 rightly with histone H3, signifying its part in chromatin-mediated gene silencing (Sharma and Pandey, 2016). In plants, WD40-repeat protein (HOS15) vital for gene repression which is connected with abiotic stress endurance via histone deacetylation. Loss of the hos15 gene cause plants to have cold hypersensitivity. These results suggested that HOS15 play a vital role in gene activation/suppression through histone transition in plant acclimation towards cold stress circumstances (Zhu *et al.*, 2008).

Others exclusively Proteins Domains found in Barley Genomes

Besides repeat domain proteins we found various other domain proteins, some of them are exclusively expanded in plant genomes which are S_locus_glycop domain, an exclusively expanded in plant species and uniquely in the *Poaceae* family. Out of 4857 S_locus_glycop protein domains in the whole plant kingdom, the Poaceae family contains 2014 domains (https://www.ebi.ac.uk/services). Many higher plants evolved a mechanism of self-incompatibility that prevents self-fertilization or inbreeding and thus endorse outcrossing. The execution of self-incompatibility is highly examined in the Brassica plant and possibly finds out the mechanism that, how recognition between pollen and the stigma in self-incompatibility mechanism is controlled by S locus, which consists of three highly polymorphic

genes, S-receptor kinase (SRK), S-locus protein 11 (SP11), also called S-locus cysteine-rich protein; SCR and S-locus glycoprotein (SLG). The specificity of S-haplotype stigma is determined by SRK encoded a membrane-spanning serine/threonine kinase and the specificity of S-haplotype pollen is determined by SP11 encoded a small cysteine-rich protein. During self-pollination, it is hypothesized that SP11 is exuded from the pollen coat which communicates with its related SRK of the stigma papilla cell to trigger the self-incompatibility response (Wang *et al.*, 2019). Both SLG and SRK are highly homologous to each other and also secreted from the stigma but it is not mandatory for S-haplotype specificity of the stigma, however, it intensifies the self-incompatibility reaction, but the mechanism is still contentious (Takayama *et al.*, 2001).

Besides that LRRNT_2 and B3 are entirely restricted to plant genomes and both the domains are tremendously expanded in the Poaceae family, while LRRNT_2 a leucine-rich repeats proteins which develop in the cell wall of plants and defend the plant from microbial pathogen and insects by inhibiting the pectin-depolymerizing action of PGs which are secreted by microbial pathogen and insects. Recent studies in *Phaseolus vulgaris* reveals that four *pgip* genes are linked with a diversified biochemical function as well as the gaining of new recognition specificities. The four pgip genes which differ in the number of amino acids reveal different inhibitory capabilities against the PGs of *Botrytis cinerea, Colletotrichum gloeosporioides, Stenocarpella maydis, Fusarium moniliforme*, and *Aspergillus niger* (Raiola *et al.*, unpublished work (Di Matteo *et al.*, 2003). PGIPs also promotes the development of

oligogalacturonides (OGS) which evoke a variety of defence responses (Kalunke *et al*., 2015).

Likewise, the B3 DNA binding domain (DBD) is highly conserved found only in transcription factors from higher plants including factors involved in auxin-regulated and abscisic acid–regulated transcription. Also, RAV1 and RAV2 are characterized by DNA binding proteins from *Arabidopsis thaliana*. AP2, a DNA-binding domain conserved in arabidopsis APETALA2 and tobacco EREBP exhibit homology with the N-terminal regions of RAV1 and RAV2, while the C-terminal possess homology with C-terminal B3 domain, of VP1/ABI3 transcription factors (Kagaya *et al*., 1999).

While Apetala 2 (AP2) domain is a huge family of DNA binding protein, which is exclusively specific to many groups of plants transcription factors, (AP2/ERF) family which comprises other four major subfamilies, AP2, RAV, ERF and dehydration-responsive element-binding (DREB) protein. The domain was first seen in the arabidopsis and found to play the roles of transcription regulator to numerous environmental biotic and abiotic stress signals like salt, cold, and drought (Lin *et al*., 2008) and play a vital role in the initiation of floral meristem and floral organ tissue determination (Sharma *et al*., 2020), the establishment of floral homeotic genes for identifying sepals and petals. Also known as the ABC model for flowering development in arabidopsis (Sharma *et al*., 2020). Previous studies reveal that in barley, in addition to AP2/ERF DNA binding domain, *HvRAF* (*Hordeum vulgare* root abundant factor), a novel family of ethylene response factor (ERF) had been identified from the early seedlings. It was found that the expression of *HvRAF* was aroused in

retaliation to diverse phytohormones like salicylic acid, ethylene, and methyl jasmonate which, defend plants during diverse environmental stresses (Jung *et al*., 2007).

Universally expanded Protein Domain Families

Along with the plants restricted domains we found many protein domains that are universally expanded in all organisms and the barley genome was not an exceptional one. Fig. 2, showing the frequency of single-domain protein. From the top 30 core protein domains, most of the domains are universally expanded protein domains in the barley genome. Most of the protein domains are involved in vital and conserved regulatory mechanisms such as protein-protein interaction, metabolism signalling pathways, synthesis of signalling molecules, DNA/RNA binding proteins, post-transcriptional processing, plant defence mechanisms, protein ubiquitination, and degradation mechanisms. We found two families of LRR domains (LRR_8 and LRR_4) while the LRR_8 domain was observed 1181 times in 377 proteins. LRR are leucine-rich repeats which are often involved in the protein-protein interactions (Kobe and Deisenhofer, 1994)*. Several genes also encode for the protein kinase domain which phosphorylates proteins in signalling pathways. Protein kinases are involved in numerous biological processes. Most notably, tyrosine kinases are abundant in barley suggesting that tyrosine phosphorylation plays a crucial role in its signalling network (Lehti-Shiu and Shiu, 2012).

P450 domain was observed 247 times in 246 proteins in barley. Class II P450s are involved in the biosynthesis of all types of hormones in plants, including oxygenation of fatty acids for synthesis of cutins, almost every pathway of secondary metabolisms such as lignification and

the synthesis of flower pigments and defence chemicals (which include aromas, flavours, antioxidants, phytoestrogens, xenobiotics detoxification (Singpho and Sharma, 2021). P450s play a role in the detoxification of hazardous activated oxygen species (Singpho and Sharma, 2021) all these crucial functions justify the numbers of P450 proteins in the barley genome. While DEAD proteins also known as DEAD-box helicase whose intend to unwind nucleic acids and involve in a wide range of RNA metabolism, including splicing of pre mRNA, nucleo-cytoplasmic transport, RNA decay, organellar gene expression and translation (Aubourg *et al.*, 1999). Likewise in barley plant, a recent study found HVD1 (*Hordeum vulgare* DEAD-box protein) which presume to encode an ATP-dependent RNA helicase, a member of which is also known as the DEAD-box family but its C-terminal amino acid sequence contain unique motifs, despite that its presumed amino acids sequence, contain eight consensus motifs which are usually spotted in the membrane of the DEAD-box protein family. In addition, it has five repeats of RGG, recognized as an RNA-binding motif because its hydrophobic C-terminus demonstrates a novel DEAD box protein that is promoted by salt stress in plants. In barley under salt stress conditions, HDV1 mRNA was observed to be induced eight-fold higher than its normal condition also others stress like water stress, cold stress and abscisic acid (ABA) treatment rise the expression of HDV1 mRNA. HDV1 has first reported the DEAD-box gene in plants whose mRNA was assembled during salt stress (Nakamura *et al.*, 2004).

Zf-RING_2 domain is universally expanded proteins that contain 40 to 60 residues of Zn-finger which binds two atoms of zinc and is possibly involved in moderating protein-protein

interactions which play vital roles in various biological courses (Han *et al.*, 2021). The zinc finger is known as an important transcription factor in eukaryotic organisms that regulate genes expression through mainly interacting with target DNA. Zinc finger proteins are classified into several forms based on the position of Cys and His residues, such as C2H2, C2HC, C2C2, C2HCC2C2, C2C2C2C2 etc, while all of these C2H2 types are well characterized and widely exist in animals, humans and plants. In the plant, they are active in growth, development and environmental stress responses (Han *et al.*, 2020).

In barley, we have observed 185 proteins of the Myb_DNA-binding domain and 99 proteins domain of bHLH. And from a recent study, it was revealed that in plants large fraction of coding regions encode transcription factors (TFs) and within them, MYBs and bHLHs (basic helix-loop-helix) have typical intrinsic properties which are specific to plants only. Both of these TFs families proteins can act as hetero or homodimers, which are associated with proteins of others proteins families, or form a complex of MYB/bHLH which modulates distinct cellular processes like responses to biotic and abiotic stresses, cell death, cell wall synthesis, circadian clock, hormone signalling and biosynthesis of specialized metabolites (Vailleau *et al.*, 2002; Seo and Mas, 2014). In many cases it has been described that MYB/bHLH complexes, are paralleled evolved and it has been linked with the diverse flexibility in developmental and metabolic activities in higher plants (Feller *et al.*, 2011).

One of several major protein families found in eukaryotic cells such as yeast, animals and plants is the F-box domain carrying proteins. In our study, we found 140 F-box protein domains in the

barley genome. According to current research, there are at least 694 and 687 F-box genes in the *Arabidopsis thaliana* and *Oryza sativa* genomes, respectively (Gagne *et al.*, 2002; Jain *et al.*, 2007) making the F-box super-family one of its largest in plants. F-box genes are renowned for their functions in protein ubiquitination and degradation, but they have also been identified to influence several important processes in plants, including embryogenesis, hormonal responses, seedling development, floral organogenesis, senescence, and disease resistance (Xu *et al.*, 2009; Jain *et al.*, 2007). Likewise, we observed 138 proteins domains of Peroxidase which act as an antioxidant defence in plants that detoxify ROS species in plants and have tissue-specific functions like hydrogen detoxification system in the chloroplast, ROS scavenging in the cytosol, mitochondria, and peroxisome. During harsh environmental conditions, peroxidase activity is increased. It was found that under salt stress

conditions the transcript level of the peroxisomal *APX* gene (*HvAPX1*) increased significantly in barley (Caverzan *et al.*, 2012). Although the Mito_carr domain is universally expanded, in the barley genome we observed 132 proteins sequences. Mitochondrial carrier domains are carrier proteins that facilitate the transport of solutes, metabolism products, cofactors and nucleotides from the inner mitochondrial matrix to the cytoplasm which provides a link between mitochondria and cytosol (Dolce *et al.*, 2014). Along with that, we observed 129 Uridine-5-diphospho glucuronosyltransferase (UDPGT) protein domains which catalyze the glycosylation of the glycosyl group by utilizing UDP-sugar as a donor and transferring different sugars to plant products. It increases their solubility and stability and make it easy for their storage in plant cells and also, determine the nature of their bioactivity (Wang 2009).
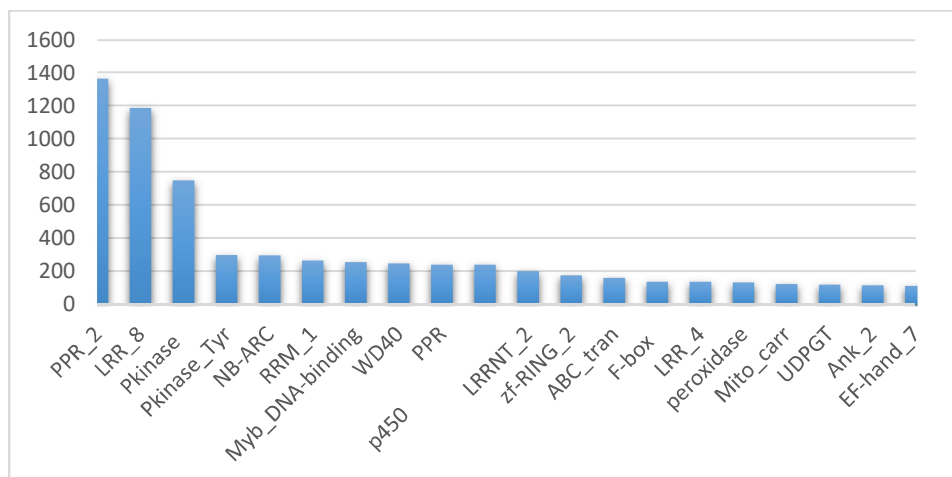

Fig. 2 Distribution of single protein domain in the barley genome.

CONCLUSION

Multi-domain protein repeats are involved in vital biological processes. Our study shows that 42% percent of sequences in the barley genome falls

under multi-domain proteins and numerous of these domains contains around 100 to 712 protein domains units which are associated with plant development and defence mechanism against biotic and abiotic stresses. As from

previous findings, the barley plant was known as a natural tolerant plant in the Poaceae family so going through their functions in this study we can assume that nature endorses barley plants with this trait. Despite the emerging technology of the 21st century, the exact mechanism of tandem gene duplication is yet to be discovered. But without losing our hope new advanced technologies like CRISPR-Cas9 and Next-Generation single-cell sequencing will resolve such issues.

ACKNOWLEDGEMENT

REFERENCES

Akar, T., Avci, M., Dusunceli, F. 2004. Barley: post-harvest operation Ed. Danilo Mejía. *The Central Research Institute for Field Crops*. Ulus, Ankara,Turkey.

Alemán, F., Nieves-Cordones, M., Martínez, V. and Rubio, F. 2011. Root K+ acquisition in plants: the Arabidopsis thaliana model. *Plant and Cell Physiology,* **52** (9):1603-1612.

https://doi.org/10.1093/pcp/pcr096

Aubourg, S., Kreis, M. and Lecharny, A. 1999. The DEAD box RNA helicase family in Arabidopsis thaliana. *Nucleic acids research*, **27** (2):628-636.

https://doi.org/10.1093/nar/27.2.628

Azlan, N. S., Guo, Z. H., Yung, W. S., Wang, Z., Lam, H. M., Lung, S. C., & Chye, M. L. (2021). In silico analysis of acyl-CoA-binding protein expression in soybean. *Frontiers in plant science*, *12*, 553.https://doi.org/10.3389/fpls.2021.646938

Bjurklund, Å.K., Ekman, D. and Elofsson, A. 2006. Expansion of protein domain repeats. *PLoS computational biology*, **2** (8):e114. https://doi.org/10.1371/journal.pcbi.0020114

Caruthers, J.M., Johnson, E.R. and McKay, D B. 2000. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proceedings of the National Academy of Sciences*, **97** (24):13080-13085.https://doi.org/10.1073/pnas.97.24.13080

Caverzan, A., Passaia, G., Rosa, S.B., Ribeiro, C.W., Lazzarotto, F. and Margis-Pinheiro, M. 2012. Plant responses to stresses: role of ascorbate peroxidase in the antioxidant protection. *Genetics and molecular biology*, **35** (4):1011-1019.

https://doi.org/10.1590/S1415-47572012000600016

Chothia, C. 1992. One thousand families for the molecular biologist. *Nature*, **357**:543-544. https://doi.org/10.1038/357543a0

Dai, F., Nevo, E., Wu, D., Comardran, J., Zhou, M., Qiu, L. and Zhang, G. 2012. Tibet is one of the centers of domestication of cultivated barley. *Proceedings of the National Academy of Sciences,* **109**(42): 16969-16973

DeYoung, B.J. and Innes, R.W. 2006. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature immunology*, **7** (12):1243. https://doi.org/10.1038/ni1410

Di Matteo, A., Federici, L., Mattei, B., Salvi, G., Johnson, K.A., Savino, C., De Lorenzo, G., Tsernoglou D. and Cervone, F. 2003. The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proceedings of the National Academy of Sciences*, **100** (17):10124-10128. https://doi.org/10.1073/pnas.1733690100

Dolce, V., Cappello, A.R. and Capobianco, L. 2014. Mitochondrial tricarboxylate and dicarboxylate–Tricarboxylate carriers: from animals to plants. *IUBMB life*, **66** (7):462-471. https://doi.org/10.1002/iub.1290

Draeger, C., Fabrice, T.N., Gineau, E., Mouille, G., Kuhn, B.M., Moller, I., Abdou, M.T., Frey, B., Pauly, M., Bacic A. and Ringli, C. 2015. Arabidopsis leucine-rich repeat extensin (LRX) proteins modify cell wall composition and influence plant growth. *BMC plant biology*, **15** (1):155. https://doi.org/10.1186/s12870-015-0548-8

Ekman, D., Bjurklund, Å. K., Frey-Skött, J. and Elofsson, A. 2005. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *Journal of molecular biology*, **348** (1):231-243. https://doi.org/10.1016/j.jmb.2005.02.007

Feller, A., Machemer, K., Braun, E.L. and Grotewold, E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal*, **66** (1):94-116. https://doi.org/10.1111/j.1365-313X.2010.04459.x

Finn, R.D., Clements, J. and Eddy, S.R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic acids research*, **39** (2):29-37. https://doi.org/10.1093/nar/gkr367

Gagne, J.M., Downes, B.P., Shiu, S.H., Durski, A.M. and Vierstra, R.D. 2002. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis. *Proceedings of the national academy of sciences*, **99** (17):11519-11524. https://doi.org/10.1073/pnas.162339999

Ha, C.M., Jun, J.H., Nam, H.G. and Fletcher, J.C. 2004. BLADE-ON-PETIOLE1 encodes a BTB/POZ domain protein required for leaf morphogenesis in Arabidopsis thaliana. *Plant and Cell Physiology*, 45 (10):1361-1370. https://doi.org/10.1093/pcp/pch201

Han, G., Lu, C., Guo, J., Qiao, Z., Sui, N., Qiu, N. and Wang, B. 2020. C2H2 zinc finger proteins: master regulators of abiotic stress responses in plants. *Frontiers in plant science*, **11**:115. https://doi.org/10.3389/fpls.2020.00115

Han, G. Qiao, Z. Li, Y. Wang, C. and Wang, B. 2021. The Roles of CCCH Zinc-Finger Proteins in Plant Abiotic Stress Tolerance. *International journal of molecular sciences*, **22**(15):8327. https://doi.org/10.3390/ijms22158327

Hollenstein, K., Dawson, R.J. and Locher, K.P. 2007. Structure and mechanism of ABC transporter proteins. *Current opinion in structural biology*, **17** (4):412-418.https://doi.org/10.1016/j.sbi.2007.07.003

Hurwitz, J. and Leis, J.P. 1972. RNA-dependent DNA polymerase activity of RNA tumor viruses I. Directing influence of DNA in the reaction. *Journal of virology*, **9** (1):116-129. https://doi.org/10.1128/jvi.9.1.116-129.1972

Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., Kapoor, S., Tyagi, A.K. and Khurana, J.P. 2007. F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant physiology*, **143** (4):1467-1483. https://doi.org/10.1104/pp.106.091900

Jiang, S.C., Mei, C., Liang, S., Yu, Y.T., Lu, K., Wu, Z., Wang, X.F. and Zhang, D. P. 2015. Crucial roles of the pentatricopeptide repeat protein SOAR1 in Arabidopsis response to drought, salt and cold stresses. *Plant molecular biology*, **88** (4-5):369-385.

https://doi.org/10.1007/s11103-015-0327-9

Jung, J., Won, S.Y., Suh, S.C., Kim, H., Wing, R., Jeong, Y., Hwang, I. and Kim, M. 2007. The barley ERF-type transcription factor HvRAF confers enhanced pathogen resistance and salt tolerance in Arabidopsis. *Planta*, **225** (3):575-588.

https://doi.org/10.1007/s00425-006-0373-2

Kagaya, Y., Ohmiya, K. and Hattori, T. 1999. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic acids research*, **27** (2):470-478.

https://doi.org/10.1093/nar/27.2.470

Kalunke, R.M., Tundo, S., Benedetti, M., Cervone, F., De Lorenzo, G. and D'Ovidio, R. 2015. An update on polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein that protects crop plants against pathogens. *Frontiers in plant science*, **6**:146.

https://doi.org/10.3389/fpls.2015.00146

Katti, M.V., Sami-Subbu, R. and Ranjekar, P.K. Gupta, V.S., 2000. Amino acid repeat patterns in protein sequences: their diversity and structural-functional implications. *Protein Science,* **9** (6):1203-1209.

https://doi.org/10.1110/ps.9.6.1203

Knighton, D.R., Bell, S.M., Zheng, J., Ten Eyck, L.F., Xuong, N.H., Taylor, S.S. and Sowadski, J.M. 1993. 2.0 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with a peptide inhibitor and detergent. Acta Crystallographica Section D: *Biological Crystallography*, **49** (3):357-361.

https://doi.org/10.1107/S0907444993000502

Kobe, B. and Deisenhofer, J. 1994. The leucine-rich repeat: a versatile binding motif. *Trends in biochemical sciences*, **19** (10):415-421.https://doi.org/10.1016/0968-0004(94)90090-6

Laluk, K., AbuQamar, S. and Mengiste, T. 2011. The Arabidopsis mitochondria-localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. *Plant Physiology*, **156** (4):2053-2068.

https://doi.org/10.1104/pp.111.177501

J. Himalayan Ecol. Sustain. Dev. Vol. 16 (2021)

ISSN 0973-7502

Lee, B. and Lee, D. 2009. Protein comparison at the domain architecture level. *BMC bioinformatics*, **10** (15):S5.

https://doi.org/10.1186/1471-2105-10-S15-S5

Lehti-Shiu, M.D. and Shiu, S.H. 2012. Diversity, classification and function of the plant protein kinase superfamily. Philosophical Transactions of the Royal Society B: *Biological Sciences*, **367** (1602):2619-2639.

https://doi.org/10.1098/rstb.2012.0003

Levdansky, E., Romano, J., Shadkchan, Y., Sharon, H., Verstrepen, K.J., Fink, G.R. and Osherov, N. 2007. Coding tandem repeats generate diversity in Aspergillus fumigatus genes. *Eukaryotic Cell*, **6** (8):1380-1391.https://doi.org/10.1128/EC.00229-06

Li, J., Mahajan, A. and Tsai, M.D. 2006. Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry*, **45** (51):15168-15178.

https://doi.org/10.1021/bi062188q

Lin, R.C., Park, H.J. and Wang, H.Y. 2008. Role of Arabidopsis RAP2. 4 in regulating light-and ethylene-mediated developmental processes and drought stress tolerance. *Molecular plant*, **1** (1):42-57.

https://doi.org/10.1093/mp/ssm004

Lopez-Ortiz, C., Peña-Garcia, Y., Natarajan, P., Bhandari, M., Abburi, V., Dutta, S. K., Yadav, L., Stommel, J., Nimmakayala, P. and Reddy, U. K. 2020. The ankyrin repeat gene family in Capsicum spp: Genome-wide survey, characterization and gene expression profile. *Scientific reports*, **10** (1):1-

16.https://doi.org/10.1038/s41598-020-61057-4

Lu, C., Feng, Z., Yuan, F., Han, G. and Wang, B. 2020. The SNARE protein LbSYP61 participates in salt secretion in Limonium bicolor. *Environmental and Experimental Botany*, **176**: 104076.

Lyzenga, W.J., Booth, J.K. and Stone, S.L. 2012. The Arabidopsis RING-type E3 ligase XBAT32 mediates the proteasomal degradation of the ethylene biosynthetic enzyme, 1-aminocyclopropane-1-carboxylate synthase 7. *The Plant Journal*, **71** (1):23-34.https://doi.org/10.1111/j.1365-313X.2012.04965.x

Matsui, K., Umemura, Y. and Ohme-Takagi, M. 2008. AtMYBL2, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in Arabidopsis. *The Plant*, **55** (6):954-967.https://doi.org/10.1111/j.1365-313X.2008.03565.x

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., Tosatto, S. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., and Bateman, A. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Research*, **49**(D1):D412-D419. https://doi.org/10.1093/nar/gkaa913

Nakamura, T., Muramoto, Y., Yokota, S., Ueda, A. and Takabe, T. 2004. Structural and transcriptional characterization of a salt-responsive gene encoding putative ATP-dependent RNA helicase in barley. *Plant science*, **167** (1):63-70.

J. Himalayan Ecol. Sustain. Dev. Vol. 16 (2021)

ISSN 0973-7502

Nakatsuka, T., Haruta, K.S., Pitaksutheepong, C., Abe, Y., Kakizaki, Y., Yamamoto, K. and Nishihara, M. 2008. Identification and characterization of R2R3-MYB and bHLH transcription factors regulating anthocyanin biosynthesis in gentian flowers. *Plant and Cell Physiology*, 49 (12):1818-1829.https://doi.org/10.1093/pcp/pcn163

Nieves-Cordones, M., Caballero, F., Martínez, V. and Rubio, F. 2011. Disruption of the Arabidopsis thaliana inward-rectifier K+ channel AKT1 improves plant responses to water stress. *Plant and Cell Physiology*, 53 (2):423-432.https://doi.org/10.1093/pcp/pcr194

Ouyang, Y., Huang, X., Lu, Z. and Yao, J. 2012. Genomic survey, expression profile and co-expression network analysis of OsWD40 family in rice. *BMC genomics*, **13** (1):100.

https://doi.org/10.1186/1471-2164-13-100

Parret, A. H., Schoofs, G., Proost, P. and De Mot, R. 2003. Plant lectin-like bacteriocin from a rhizosphere-colonizing Pseudomonas isolate. *Journal of bacteriology*, **185** (3):897-908.

https://doi.org/10.1128/JB.185.3.897-908.2003

Pearson, C.E. and Sinden, R.R. 1998. Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. *Current opinion in structural biology*, **8** (3):321-330.https://doi.org/10.1016/S0959-440X(98)80065-1

Ponting, C., Aravind, L., Schultz, J., Bork, P. and Koonin, E.V. 1999. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *Journal of molecular biology*, 289 (4):729-745.https://doi.org/10.1006/jmbi.1999.2827

Richard, G.F., Kerrest, A. and Dujon, B. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiology and Molecular Biology Reviews, **72** (4):686-727. https://doi.org/10.1128/MMBR.00011-08

Schaper, E. and Anisimova, M. 2015. The evolution and function of protein tandem repeats in plants. New Phytologist **206** (1):397-410. https://doi.org/10.1111/nph.13184

Schmitz-Linneweber, C. and Small, I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in plant science*, **13** (12):663-670.

https://doi.org/10.1016/j.tplants.2008.10.001

Sedgwick, S.G. and Smerdon, S.J. 1999. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends in biochemical sciences*, **24** (8):311-316.

https://doi.org/10.1016/S0968-0004(99)01426-7

Seo, P. J. and Mas, P. 2014. Multiple layers of posttranslational regulation refine circadian clock activity in Arabidopsis. *The Plant Cell*, **26** (1):79-87.

https://doi.org/10.1105/tpc.113.119842

Sharma, M. and Pandey, G.K. 2016. Expansion and function of repeat domain proteins during stress and development in plants. *Frontiers in plant science*, **6**:1218.

J. Himalayan Ecol. Sustain. Dev. Vol. 16 (2021)

ISSN 0973-7502

https://doi.org/10.3389/fpls.2015.01218

Sharma, P., Singh, R. and Sehrawat, N. 2020. A critical review on: Significance of floral homeotic APETALA2 gene in plant system. *Journal of Applied Pharmaceutical Science*, **10**(1):124 -130.

Singpho, N. L. and Sharma, J. G. 2021. Importance of Cytochrome P450 gene family from metabolite biosynthesis to stress tolerance: A review. *In IOP Conference Series: Earth and Environmental Science,* **775** (1):012012. IOP Publishing.

Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.S., Watanabe, M., Iwano, M. and Isogai, A. 2001. Direct ligand–receptor complex interaction controls Brassica self-incompatibility. *Nature*, **413** (6855):534.https://doi.org/10.1038/35097104

Tan, L., Salih, H., Htet, N. N. W., Azeem, F. and Zhan., R. 2021. Genomic analysis of WD40 protein family in the mango reveals a TTG1 protein enhances root growth and abiotic tolerance in Arabidopsis. *Scientific reports,* **11**(1):1-10.

https://doi.org/10.1038/s41598-021-81969-z

Toledo-Ortiz, G., Huq, E. and Quail, P.H. 2003. The Arabidopsis basic/helix-loop-helix transcription factor family. *The Plant Cell*, **15** (8):1749-1770.
 https://doi.org/10.1105/tpc.013839

Vailleau, F., Daniel, X., Tronchet, M., Montillet, J.L., Triantaphylides, C. and Roby, D. 2002. A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. *Proceedings of the National Academy of Sciences*, **99** (15):10179-10184.
https://doi.org/10.1073/pnas.152047199

Van Ooijen, G., Mayr, G., Kasiem, M.M., Albrecht, M., Cornelissen, B.J. and Takken, F. L. 2008. Structure–function analysis of the NB-ARC domain of plant disease resistance proteins. *Journal of experimental botany*, **59** (6):1383-1397.https://doi.org/10.1093/jxb/ern045

Wang, C. L., Zhang, Z. P., Oikawa, E., Kitashiba, H. and Nishio, T. 2019. SCR-22 of pollen-dominant S haplotype class is recessive to SCR-44 of pollen-recessive S haplotype class in Brassica rapa. *Horticulture research*, **6**(1):1-11.
https://doi.org/10.1038/s41438-018-0103-5

Wang, X. 2009. Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS letters*, 583 (20):3303-3309.
https://doi.org/10.1016/j.febslet.2009.09.042

Wang, X., An, Y., Xu, P. and Xiao, J. 2021. Functioning of PPR Proteins in organelle RNA metabolism and chloroplast biogenesis. *Frontiers in plant science,* 12:1.
https://doi.org/10.3389/fpls.2021.627501.

Xu, G., Ma, H., Nei, M. and Kong, H. 2009. Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification. *Proceedings of the National Academy of Sciences*, **106** (3):835-840.

https://doi.org/10.1073/pnas.0812043106

Yang, L., Zhu, H., Guo, W. and Zhang, T. 2010. Molecular cloning and characterization of five genes encoding pentatricopeptide repeat proteins from Upland cotton (Gossypium hirsutum L.). *Molecular biology reports*, **37** (2):801.
https://doi.org/10.1007/s11033-009-9610-7

Yeats, C. Bentley, S. and Bateman, A. 2003. New knowledge from old: in silico discovery of novel protein domains in Streptomyces coelicolor. *BMCmicrobiology*, **3**(1):1-20.
https://doi.org/10.1186/1471-2180-3-3

Zhang, H., Scheirer, D.C., Fowle, W.H. and Goodman, H.M. 1992. Expression of antisense or sense RNA of an ankyrin repeat-containing gene blocks chloroplast differentiation in arabidopsis. *The Plant Cell*, **4** (12):1575-1588.
https://doi.org/10.1105/tpc.4.12.1575

Zhang, X., Li, D., Zhang, H., Wang, X., Zheng, Z. and Song, F. 2010. Molecular characterization of rice OsBIANK1, encoding a plasma membrane-anchored ankyrin repeat protein, and its inducible expression in defense responses. *Molecular biology reports* **37** (2):653.

https://doi.org/10.1007/s11033-009-9507-5

Zhu, J., Jeong, J.C., Zhu, Y., Sokolchik, I., Miyazaki, S., Zhu, J. K., Hasegawa, P.M., Bohnert, H.J., Shi, H., Yun, D.J. and Bressan, R.A. 2008. Involvement of Arabidopsis HOS15 in histone deacetylation and cold tolerance. *Proceedings of the National Academy of Sciences,* **105** (12):4945-4950.
https://doi.org/10.1073/pnas.0801029105