# *Identification and comparative analysis of seed storage proteins (SSPs) in sequenced cereal genomes*

A dissertation

Submitted to the Department of Biotechnology

Thapar Institute of Engineering and Technology

In partial fulfilment

of the Requirements for the Degree of

Master of Science in Biotechnology

**By**

**Chandni**

**302001004**

**Supervisor – Dr. Atul Kumar Upadhyay**

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Thapar Institute of Engineering and Technology, Patiala, Punjab

June, 2022

# DECLARATION

I hereby declare that the project work titled **"Identification and comparative analysis of seed storage proteins in sequenced cereal genomes"** is an authentic record of my own work completed at Thapar Institute of Engineering and Technology as part of the 6 month dissertation requirement for the award of the degree of Master of Science in Biotechnology, under the supervision of Dr. Atul Kumar Upadhyay, from January to June, 2022.

Chandni

(Signature of student)

Chandni

302001004

Date: 30-06-2022

# CERTIFICATE

This is to certify that the dissertation entitled **“Identification and comparative analysis of seed storage proteins in sequenced cereal genomes”** submitted by Ms. Chandni (Roll No. 302001004) in partial fulfilment for the requirement for the award of degree of Master of Science in Biotechnology, Department of Biotechnology, Thapar institute of Engineering and Technology, Patiala, Punjab, is an authentic record of students work during the period of six months from January 2022 to June 2022. This report has not submitted for the reward of any other degree or certificate in this or any other university or institute.

Atul Upadhyay

Dr. Atul Kumar Upadhyay

Supervisor

Assistant Professor

Thapar institute of Engineering and Technology

Patiala, Punjab

# ACKNOWLEDGEMENT

I am grateful to Dr. Prakash Gopalan, Director of Thapar Institution of Engineering and Technology (TIET), for giving the opportunity and institute facilities to allow me to complete dissertation work as part of the MSc curriculum requirement. With heartfelt thanks, I offer my warmest feelings to Dr. Atul Kumar Upadhyay, Scientist, TIET, who agreed to take on and guide this dissertation and training. I have no words to express my profound gratitude to him for his illuminating guidance, consistent support, supervision, and keen interest during the dissertation process.

I would like to express my deepest gratitude to HOD, Dr. M S Reddy, Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, for his thoughtful recommendations and foresight. Finally, I'd like to thank my family and friends for providing me the strength to keep going.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# OBJECTIVES

1. Identification of SSPs in different sequenced cereal genomes
2. Annotation of identified SSPs

# ABSTRACT

Seeds are not only propagation and dispersal organs, but they also synthesise and store proteins known as seed storage proteins (SSPs). The high protein content of the plant can be mobilised and utilised for plant care, defence, and growth. Because of their high quality and quantity, SSPs are utilised in food all around the world. SSPs perform a key metabolic and structural role in seeds, storing nitrogen, carbon, and energy during germination. These storage proteins share several characteristics such as they are synthesised in high concentrations in specific tissues during germination and at certain stages of plant development and gets deposited in the form of protein bodies and can be mobilised for seed defence. Various computational and experimental methods are used to identify seed storage proteins and understand its function by in depth annotation studies. In this study, SSPs were identified and analysed in sequenced cereal genomes using various bioinformatics tools.

# 1. INTRODUCTION

As we all know that plants are the principal source of human diet and animal feed. A major part of our human diet contains cereals and legumes. Around 70% of human food comprises cereals and legumes and the remaining 30% comes from animals. Protein content in seeds ranges from -10% (in cereals) to -40% (in certain legumes and oil seeds) of dry weight, making them a significant source of dietary protein (Shewry *et al*., 1995). Cereals and legumes are the world's two most important crops. They not only supply energy, but also proteins for human, animal, and livestock nutrition. Because animal proteins are expensive, individuals in developing nations rely almost entirely on seed protein to meet their protein needs (Mandal and Mandal 2002). Cereal grains have a lower protein content than legume seeds, with an average of 10–12 percent dry weight. They supply about 200 mt (Million tons) of protein for human and cattle nutrition, which is around three times the amount generated from the more protein-rich (20–40 percent) legume seeds (Shewry and Halford, 2002). During specific stages of development, plants acquire store compounds such as starch, lipids, and proteins. Storage proteins accumulate in both vegetative and reproductive tissues, serving as a reservoir for plant growth in later phases. The formation of storage protein is thus helpful to plant survival. Storage proteins are an important source of plant proteins in the diet. (Fujiwara *et al*., 2002).

Seed storage protein is a class of protein that is only present in seeds, where it accumulates in enormous amounts, and can be digested to release its constituent amino acids. These acids are utilized by seedlings as a source of reduced nitrogen, which is required for seedling germination and early growth. Seed storage proteins account for a significant portion of the seed's protein composition and play a key role in seed quality. These storage proteins are important because they determine total protein concentration and have an impact on nutritional quality and food processing functional characteristics (Radhika and Rao, 2015).

Seed storage proteins (SSPs) are specially expressed high quantities in the developing seed, with temporal and spatial regulation of synthesis during endosperm development. SSPs are found mainly in special storage organelles known as protein bodies, where they accumulate in large levels during the mid-maturation stage of seed development. Storage proteins are sequestered into protein bodies, which keeps them apart from the cell's metabolic compartments. Nutrition influences the expression of storage proteins as well. For example, the synthesis of sulphur-rich proteins may be reduced in plants growing in low-sulphur soils (Gaur *et al*., 2018).

Seed storage proteins are produced on the endoplasmic reticulum (ER) and cotranslationally transferred into the lumen. Protein folding, assembly, and posttranslational changes take place in endomembranous compartments such as the ER, Golgi apparatus, and vacuole. Proteins can be found in protein bodies produced from the ER (many cereal prolamins) or in the vacuole (albumins, globulins, and some cereal prolamins) (Kermode and Bewley, 1999).

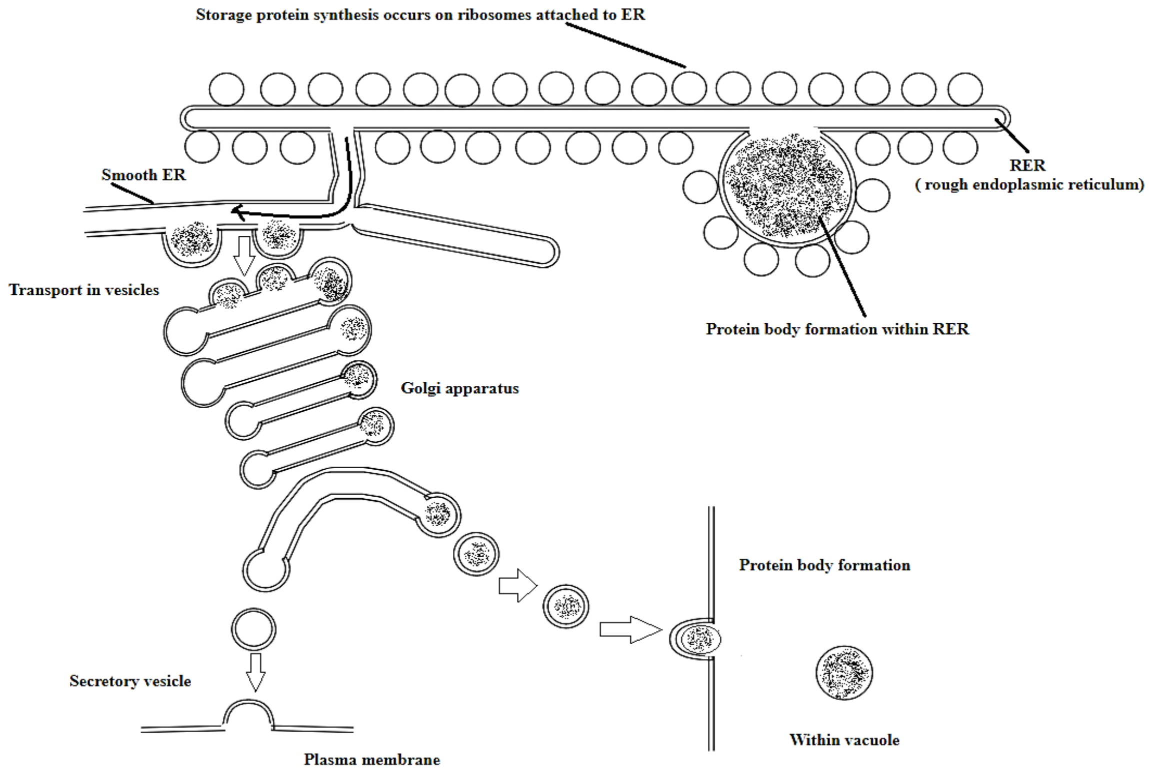


Figure 1. Diagrammatic representation of storage protein synthesis, processing, trafficking, and deposition in the cell's endomembrane system

Its main function is to serve as a reserve for nitrogen, carbon, and sulphur. The great majority of individual proteins found in mature seeds play structural or metabolic tasks. The seed storage proteins reflect the quality and total protein content of a specific species and are used to display low-profile protein data. For example, certain cereal seeds have low levels of Lysine, Threonine, and Tryptophan, while legume seeds have high levels of Cysteine and Methionine. As a result, the nutritious quality of the seeds is reduced for some reason (Wakasa and Takaiwa, 2013).

## 1.1 CLASSIFICATION OF SEED STORAGE PROTEINS (SSPs)

Thomas Burr Osborne (1924) classified them into four groups based on their extraction and solubility:

| | |
|---|---|
| **Albumins** | • Water soluble |
| **Globulins** | • Dilute salt soluble |
| **Prolamins** | • Alcohol/water mixtures |
| **Glutelins** | • Dilute acid or alkali |

### 2s albumin storage protein

The sedimentation coefficients of 2s albumins define them (S20.w). They are broadly dispersed in dicot seeds and include approximately 25 albumin proteins. 2S albumins in ungerminated seeds have been found in protein bodies, representing 40% of total seed protein. They begin to break down upon seed germination, supplying amino acids for seed growth. 2S albumins are a good source of nitrogen because they are high in arginine and glutamine. They are composed of a single precursor protein that is proteolytically cleaved, resulting in the loss of a linker peptide as well as short peptides from both the N and C termini. They have been most widely studied in the Cruciferae, specifically oilseed rape (where they are known as Napins) and Arabidopsis. (Pedro F.N. Souza, 2020).

### Prolamins

It is regarded as one of the most important endosperm storage proteins in all cereal grains (except rice and oat). Examples are wheat, barley, rye, maize, sorghum, millets, and other grains. Proline and amide nitrogen produced from glutamine are generally abundant. Alcohol/water combinations (e.g., 60-70 percent (v/v) ethanol) are easily soluble. The Structure is much more varied than in 7S and 11S globulins. The presence of repeating amino-acid sequences on one or more short peptide motifs (Methionine, Glycine, Histidine, etc). As an example, consider Zein. The α-Zein account for 75 to 80 percent of total proteins in maize and are divided into two groups: M 19,000 and M 22,000. Both have a similar structure, with distinct N and C-terminal domains bordering the sequences. Cereals are a major source of Zein proteins (Shewry and Tatham, 1990).

## Globulins

Globulin storage proteins can be found in both dicots and monocots. These are found in the endosperm's embryo and outer aleurone layer (Kriz, 1989). It dissolves quickly in dilute salt solutions and has a sedimentation value of about 7. Based on their sedimentation coefficient, they are classified into two groups: 7S Vicilins and 11S Legumins. 11S storage globulins are found in the starchy endosperm and are also found in various cereal grains. 11S globulins are composed of six subunit pairs, each of which consists of an acidic subunit (MW 40,000) and a basic unit (MW 20,000) joined by a single disulfide bond, whereas 7S globulins are trimeric proteins ranging in size from 150,000 to 190,000 that lack cysteine residues (Singh and Shepherd, 1985). Before human consumption, it is normally eliminated by milling (wheat), polishing (rice), and pearling (barley).

## Glutelins

Glutelins dissolve in dilute acids and bases. They are often prolamin-like proteins found in grass seeds such as wheat. The majority of wheat ingested by humans is processed from white flour, which is made by grinding the germ (embryo) and bran (pericarp, testa, nucellar layer, and aleurone layer). As a result, it belongs to starchy endosperm cells and includes high levels of starch and gluten. As previously discussed, gluten proteins form a continuous matrix in mature dry endosperm cells. When flour and water are combined to make a dough, the protein matrices in individual cells come together to form a continuous network. This imparts viscoelastic qualities to the dough, allowing it to be fermented and baked into leavened bread or processed into pasta, noodles, and a variety of other cuisines (Field *et al.*, 1983). Irrespective of solubility, it is a key storage protein in rice and derives from the same ancestral gene as leguminous 11S globulin.

# 2. REVIEW OF LITERATURE

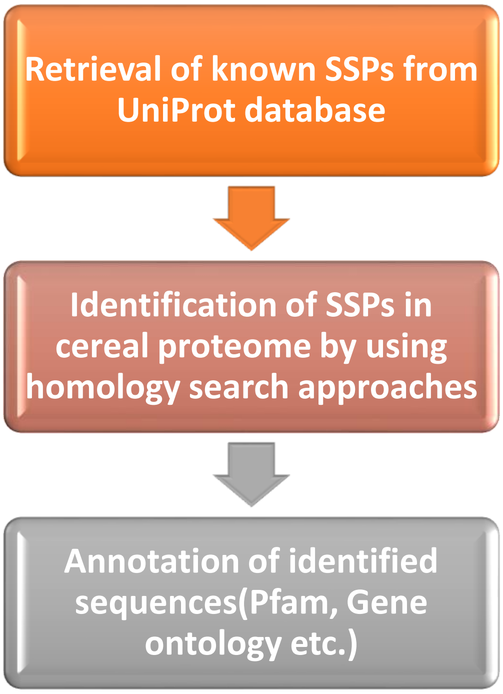
- Ge *et al*., 2021 concluded that high-molecular-weight glutenin subunits (HMW-GSs) derived from the *Thinopyrum elongatum* genome exhibited great potential for future wheat dough quality enhancement. Furthermore, the *T. elongatum* genome, which contains few short CD peptides, will be a good genomic choice for wheat production with reduced celiac disease.
- According to Gaur *et al*., 2018 foxtail millet (*Setaria italica* (L.) P. Beauv.) has been recognized as a nutritious grain crop for generations. The current study demonstrates the presence of distinct seed storage proteins with high quantities of essential amino acids that have not before been discovered in any other cereals or legumes. Furthermore, in addition to the strategy of overproducing free amino acids, the genes encoding these proteins might be used as transgenes to create efficient sinks for the manufacture of nutritious grains rich in essential fatty acids.
- Sharma *et al*., 2019 discovered a NAC-like TF with 90% sequence similarity to the OsNac5 gene cDNA. Furthermore, multiple sequence alignment of NAC-like TF-derived protein sequences with other NAC protein sequences from different crop plants revealed that sequences are more conservative at the N-terminal region, forming a more conserved sub-domain of the NAC family, whereas the C-terminal region was discovered to be highly diverse among different NAC genes. The NAC family of genes has been demonstrated to affect a variety of developmental processes, including seed development, embryo development, shoot apical meristems, fiber growth, leaf senescence, and cell division. NAC-like TF producing a protein for the rice gene was found to be elevated during the S2 stage in the leaves and panicles of the parent HPR14 as well as the five RILs, whereas the wheat NAC gene, NAM-B1, was found to be involved in nutrient remobilization from leaves to developing grains.
- According to Foley *et al*., 2011 *Lupinus angustifolius* has at least 16 seed storage protein genes divided into four families, only five of which have previously been described. An examination of each gene's expression patterns throughout seed development indicated that all of these genes have similar expression profiles and are most highly expressed after flowering. Comparative research with other legumes has revealed details about the evolution of these genes, including indications of gene

duplication following speciation in some cases. Lupin seeds, like other grain legume seeds, contain allergenic proteins, and our research has indicated that all seven members of the β conglutin family are potential allergens for lupin-sensitive people. These results indicate that lupins can be studied further at multiple levels, including molecular biology, physiological, biochemical, and nutritional levels.

- According to considerable biological, morphological, and genomic evidence, Prior *et al*., 2016 concluded that the *Ralstonia solanacearum* species complex (RSSC) is too complex to be classified as a single species. The taxonomy of this organism needs to be changed to distinguish three phylogenetically distinct groups with distinct biological characteristics and evolutionary ties. Newly isolated bacterial wilt strains can be simply allocated to the proposed system using existing molecular technologies. These modifications will help with a variety of applications, such as creating plant resistance to bacterial wilt, finding new pathogenic variants, quarantine containment management, and the development of diagnostic tests.
- According to Chen *et al*., 2021 the new entire genome of the peanut bacterial wilt pathogen has been sequenced in various habitats. Comparative genomic evaluation of various strain phylotypes indicates genetic diversity and host specificity. The events of horizontal gene transfer and the varied strains with the same host of peanut further supported *R. solanacearum's* wide host range and great adaptability. The evolutionary link between *R. solanacearum* strains was shown to be more closely related to geographic origins than host variation. In short, the findings contribute significantly to our understanding of the pathogenic mechanism of peanut bacterial wilt and offer the theoretical groundwork for future studies on the interaction of *R. solanacearum* with peanuts.
- Shewry and Halford (2002), discovered that cereal storage proteins play a significant role in determining grain quality and end-use qualities. Understanding the structure of these proteins, their physicochemical and functional properties, as well as the biological mechanisms that govern their synthesis, trafficking, and deposition in grain, is critical for future efforts to improve grain end-use quality through genetic engineering.

- According to Shewry *et al*., 1995 current work on seed storage proteins was done to establish a platform for increasing the nutritional and processing qualities of crops by genetic engineering. This is now achievable because of the recent discovery of dependable transformation processes for maize, small grain cereals (wheat and barley), and grain legumes. A detailed overview of storage protein structure and variety is required before attempting to change quality since it reveals the extent to which the proteins' structures can be manipulated without compromising their biological properties.
- According to Paris *et al*., 2021 the current work constitutes a detailed characterization of the -gliadin gene family in durum wheat (*Triticum durum* Desf.) cv. Svevo. Our efforts to sequence and manually annotate these and nearby genes resulted in the correct identification of the γ-gliadin gene family and its location in the durum wheat genome. The findings of this study will provide a solid platform for future proteomic investigations on the role of durum wheat-gliadins in semolina functioning and human health, as well as future breeding initiatives targeted at selecting new durum wheat genotypes with lower immunogenic potential for human consumption.
- Kawakatsu and Takaiwa (2010), discovered that there have been various attempts to use transgenic crops to produce high-value recombinant proteins such as medicines and industrial enzymes. Aside from that, a seed is an excellent mucosal delivery vehicle for oral peptide/protein therapies. Mucosal antigen administration efficiently induces immunity or immunological tolerance at the mucosal locations. Because of the low proteolytic digestibility of PB-I, PB-I-localized 3Crp (three T-cell epitopes generated from the Japanese cedar pollen allergens Cry j 1 and Cry j 2) has greater resistance to enzymatic digestion in the gastrointestinal system, resulting in more efficient suppression of specific IgE suppression when administered orally compared to chemically produced or PB-II-localized versions. This is one of the good examples of how utilizing SSP traits improves research.

## 3. EXPERIMENTAL DESIGN AND METHODS



### 3.1. Identification of SSPs

- The reviewed protein sequences of SSPs were retrieved from the UniProt database (www.uniprot.org) in the Fasta format. The UniProt database contains a large, high-quality, and freely available collection of protein sequences that have been functionally annotated. The UniProt Knowledgebase (UniProtKB) combines reviewed UniProtKB/Swiss-Prot entries to which our expert biocuration team has added data with unreviewed UniProtKB/TrEMBL entries annotated by automated methods (The UniProt Consortium 2021)

### 3.2. Prediction of selected proteome

- Selected sequences of SSPs from UniProt were used as a target sequence.
- The Basic Local Alignment Search Tool (BLAST) (www.blast.ncbi.nlm.nih.gov/Blast.cgi) identifies similar regions between sequences. The programme compares nucleotide or protein sequences and computes the statistical importance of matches. BLAST can be used to infer functional and evolutionary links between sequences, as well as to

discover gene family members. BLASTp (Protein BLAST) matches one or more protein query sequences to a subject protein sequence or a protein sequence database. This is useful when looking for a protein. BLASTp was performed taking the target sequence as a query sequence against database like nr (Non reductant protein sequences), PDB (Protein data bank proteins).

- Selected the template on the basis of percent identity (>60%). The quality of the template structure was assessed on the basis of resolution and R-value of the structure using RCBS protein databank. The template with resolution ranging from 1.5 to 2.0 was taken. R-value tells how well the statistical values of X-ray data are as to the structured model. Downloaded the template sequence in the Fasta format.
- Phylogenetic tree was constructed using Molecular Evolutionary Genetic Analysis (MEGA) (www.megasoftware.net) with the maximum likelihood method. The MEGA software was created with the goal of providing tools for exploring, discovering, and interpreting DNA and protein sequences from an evolutionary perspective. We can create a multiple sequence alignment in MEGA using either the ClustalW or Muscle Algorithms. Using MEGA, we can re-construct a phylogeny using Maximum Likelihood, Minimum Evolution, UPGMA, and Maximum Parsimony methods in addition to Neighbour-Joining (Tamura *et al*., 2021)
- The automated SWISS-MODEL server (www.swissmodel.expasy.org/) was used to simulate the three-dimensional structures of SSPs. To look for templates, the target protein can be specified as amino acid sequence, either in FASTA, Clustal format or as a plain text. SWISS-MODEL accomplishes this task by employing two database search methods: BLAST, which is fast and accurate for closely similar templates, and HHblits, which increases sensitivity in the event of remote homology. When the template search is finished, the templates are ordered based on the expected quality of the generated models, as estimated by the Global Model Quality Estimate (GMQE) and Quaternary Structure Quality Estimate (QSQE). Top-ranked templates and alignments are evaluated to see if they reflect different conformational states or different sections of the target protein. A 3D protein model is automatically built for each specified template by first transferring conserved atom coordinates defined by the target-template alignment (Waterhouse *et al*., 2018).

## 3.3. Annotation of the predicted sequences

- The predicted sequences were annotated with the help of Interpro (www.ebi.ac.uk/interpro) and the Pfam database (www.pfam.xfam.org).
- Pfam database is to provide a complete and accurate classification of protein families and domains. It is used to analyse novel genomes and metagenomes, as well as to drive experimental work on specific proteins and systems. A seed alignment for each Pfam family provides a representative set of sequences for the entry. Using the HMMER software (http://hmmer.org/), a profile hidden Markov model (HMM) is automatically constructed from the seed alignment and searched against the Pfamseq sequence database (Mistry *et al*., 2021).
- The gene ontology terms of the predicted sequences were found using the Gene Ontology database (Interpro). Interpro uses signatures to classify the proteins into various families and provides the functional analysis of these proteins. Gene ontology (GO) is a set of concepts in a domain and their relationships to each other are represented as a hierarchy. It provides three set of GO terms (molecular function, cellular component, and biological process) having relations operating between them.

# 4. RESULTS

## 4.1 Identification of SSPs

### 4.1.1 Retrieval of SSPs from proteomes

The UniProt database yielded a total of 223 reviewed seed storage protein sequences. However, we have to select the reviewed proteomes of rice and wheat since we need to detect SSPs in cereal genomes. Rice had 23 reviewed sequences while wheat had 10 reviewed sequences. We have to choose the reviewed sequence of further classified SSPs i.e., albumin, prolamin, and globulin for structure prediction and homology.

| S.No. | UniProt ID | Protein | Organism |
|---|---|---|---|
| 1 | Q10RX7 | ABC transporter C family member 13 | *Oryza sativa* subsp. japonica (Rice) |
| 2 | Q6Z7B0 | Heat shock 70 kDa protein BIP1 | *Oryza sativa* subsp. japonica (Rice) |
| 3 | Q42456 | Aspartic proteinase oryzasin-1 | *Oryza sativa* subsp. japonica (Rice) |
| 4 | Q10AZ7 | Protein GLUTELIN PRECURSOR ACCUMULATION 3 | *Oryza sativa* subsp. japonica (Rice) |
| 5 | Q84LM2 | Vacuolar-processing enzyme beta-isozyme 1 | *Oryza sativa* subsp. japonica (Rice) |
| 6 | Q10NQ3 | Vacuolar protein sorting-associated protein 9A | *Oryza sativa* subsp. japonica (Rice) |
| 7 | Q6ZJW8 | Polycomb group protein FIE1 | *Oryza sativa* subsp. japonica (Rice) |
| 8 | Q6ZLB0 | bZIP transcription factor RISBZ1 | *Oryza sativa* subsp. japonica (Rice) |
| 9 | Q7X9A8 | bZIP transcription factor RISBZ2 | *Oryza sativa* subsp. japonica (Rice) |
| 10 | B9F1L8 | Dof zinc finger protein 2 (OsDof2) | *Oryza sativa* subsp. japonica (Rice) |
| 11 | Q9SXG8 | Dof zinc finger protein 1 (OsDof1) | *Oryza sativa* subsp. japonica (Rice) |
| 12 | Q6K537 | Dof zinc finger protein 3 | *Oryza sativa* subsp. japonica (Rice) |
| 13 | Q5VNK1 | NAC domain-containing protein 26 | *Oryza sativa* subsp. japonica (Rice) |
| 14 | Q9FTY0 | NAC domain-containing protein 20 | *Oryza sativa* subsp. japonica (Rice) |
| 15 | Q6Z345 | Dof zinc finger protein 4 | *Oryza sativa* subsp. japonica (Rice) |
| 16 | Q5JLR7 | Dof zinc finger protein 5 | *Oryza sativa* subsp. japonica (Rice) |

|  |  |  |  |
|---|---|---|---|
| 17 | Q0ILQ6 | Ras-related protein Rab5A | *Oryza sativa* subsp. japonica (Rice) |
| 18 | P29835 | 19 kDa globulin | *Oryza sativa* subsp. japonica (Rice) |
| 19 | Q01882 | Seed allergenic protein RAG2 | *Oryza sativa* subsp. japonica (Rice) |
| 20 | Q01883 | Seed allergenic protein RAG1 | *Oryza sativa* subsp. japonica (Rice) |
| 21 | Q01881 | Seed allergenic protein RA5 | *Oryza sativa* subsp. japonica (Rice) |
| 22 | P20698 | Prolamin PPROL 17D | *Oryza sativa* subsp. japonica (Rice) |
| 23 | Q0DJ45 | Prolamin PPROL 14E | *Oryza sativa* subsp. japonica (Rice) |

| **S.No.** | **UniProt ID** | **Protein** | **Organism** |
|---|---|---|---|
| 1 | P10388 | Glutenin, high molecular weight subunit DX5 | *Triticum aestivum* (Wheat) |
| 2 | P08489 | Glutenin, high molecular weight subunit PW212 | *Triticum aestivum* (Wheat) |
| 3 | P08488 | Glutenin, high molecular weight subunit 12 | *Triticum aestivum* (Wheat) |
| 4 | P10387 | Glutenin, high molecular weight subunit DY10 | *Triticum aestivum* (Wheat) |
| 5 | P10385 | Glutenin, low molecular weight subunit | *Triticum aestivum* (Wheat) |
| 6 | P10386 | Glutenin, low molecular weight subunit 1D1 | *Triticum aestivum* (Wheat) |
| 7 | P16315 | Glutenin, low molecular weight subunit PTDUCD1 | *Triticum aestivum* (Wheat) |
| 8 | D2KFH1 | Avenin-like a4 | *Triticum aestivum* (Wheat) |
| 9 | P02861 | Glutenin, high molecular weight subunit PC256 | *Triticum aestivum* (Wheat) |
| 10 | P02862 | Glutenin, high molecular weight subunit PC237 | *Triticum aestivum* (Wheat) |

| **S.No.** | **UniProt ID** | **Protein** | **Organism** |
|---|---|---|---|
| 1 | P93198 | 2S seed storage albumin protein | *Juglans regia* (English walnut) |
| 2 | Q0DJ45 | Prolamin PPROL 14E | *Oryza sativa* subsp. japonica (Rice) |
| 3 | Q2TPW5 | 11S globulin seed storage protein Jug r 4 | *Juglans regia* (English walnut) |

## 4.2 Annotation of the identified putative SSPs

After retrieval of proteomes through the UniProt database, the protein sequences of cereal genomes were used as a target sequences then BLASTp was performed taking the target sequence as a query sequence against database nr (Non-reductant protein sequences). Selected the template sequences on the basis of percent identity (>60%). Downloaded the template sequences in the Fasta format. The predicted sequences were annotated with the help of Interpro and Pfam database.

**The Interpro database was used to find the gene ontology words of the predicted sequences. The GO terms of the reviewed seed storage protein sequences in cereal genomes (wheat and rice) are listed in the table below:**

### *Oryza sativa indica*

| S.No. | DESCRIPTION | GENE ONTOLOGY (GO) TERMS | | |
|---|---|---|---|---|
| | | **Biological process** | **Molecular function** | **Cellular component** |
| 1 | EEC81618.1:1-436 hypothetical protein OsI_25134 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 2 | AAL10017.1:55-375 transcription activator REB | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 3 | EAZ03220.1:6-141 hypothetical protein OsI_25368 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 4 | BAA36492.1:1-421 bZIP protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 5 | AAL10017.1:1-420 transcription activator REB | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |

| | | | | |
|---|---|---|---|---|
| 6 | EAZ03217.1:63-167 hypothetical protein OsI_25365 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 7 | ADV78543.1:1-203 small GTP binding protein Rab5 | None | GTP binding (GO:0005525)<br>GTPase activity (GO:0003924) | None |
| 8 | B8ASK4.1:1-497 Vacuolar-processing enzyme beta-isozyme 1 | Proteolysis (GO:0006508)<br>Proteolysis involved in cellular protein catabolic process (GO:0051603) | Peptidase activity (GO:0008233)<br>Cysteine-type endopeptidase activity (GO:0004197) | None |
| 9 | EEC82867.1:1-376 hypothetical protein OsI_27737 | None | Protein binding (GO:0005515) | None |
| 10 | EAZ03220.1:7-131 hypothetical protein OsI_25368 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 11 | A2WJP3.1:1-320 NAC domain-containing protein 20 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |
| 12 | APH07720.1:1-318 NAC protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |
| 13 | APH07721.1:1-326 NAC protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |
| 14 | APH07722.1:1-330 NAC protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |
| 15 | APH07723.1:1-322 NAC protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |

| | | | | |
|---|---|---|---|---|
| 16 | EEC72846.1:1-241 hypothetical protein OsI_06594 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 17 | EEC72845.1:8-105 hypothetical protein OsI_06593 | None | None | None |
| 18 | AAW80680.1:20-150 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 19 | EAY95200.1:28-405 hypothetical protein OsI_17020 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 20 | EEC73768.1:1-363 hypothetical protein OsI_08442 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 21 | AAW80680.1:20-149 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 22 | EEC73921.1:26-147 hypothetical protein OsI_08771 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 23 | EEC73921.1:214-307 hypothetical protein OsI_08771 | None | None | None |

## *Oryza sativa*

| **S.No.** | **DESCRIPTION** | **GENE ONTOLOGY (GO) TERMS** | | |
|---|---|---|---|---|
| | | **Biological process** | **Molecular function** | **Cellular component** |
| 1 | KAB8104546.1:1-431 hypothetical protein EE612_037507 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity GO:0003700) | None |

| | | | | |
|---|---|---|---|---|
| 2 | CAA44001.1:27-165 low molecular weight globulin | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 3 | KAB8093986.1:1-423 hypothetical protein EE612_021007 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity GO:0003700) | None |
| 4 | CAA44001.1:1-165 low molecular weight globulin | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 5 | KAB8104775.1:1-113 hypothetical protein EE612_037871 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 6 | AAK38149.1:1-203 small GTP-binding protein | None | GTP binding (GO:0005525)<br>GTPase activity (GO:0003924) | None |
| 7 | CAA44001.1:1-155 low molecular weight globulin | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 8 | KAB8098941.1:20-150 hypothetical protein EE612_028695 | None | Nutrient reservoir activity (GO:0045735) | None |
| 9 | KAB8098944.1:11-113 hypothetical protein EE612_028695 | None | Nutrient reservoir activity (GO:0045735) | None |
| 10 | KAB8098943.1:16-120 hypothetical protein EE612_028695 | None | Nutrient reservoir activity (GO:0045735) | None |
| 11 | pir\|JC4599\|:21-154 13K prolamin precursor - rice | None | Nutrient reservoir activity (GO:0045735) | None |
| 12 | BAA06875.1:1-509 aspartic protease | Proteolysis (GO:0006508)<br>Lipid metabolic process (GO:0006629) | Aspartic-type endopeptidase activity (GO:0004190) | None |
| 13 | KAB8082805.1:29-522 hypothetical protein EE612_004818 | Proteolysis (GO:0006508)<br>Lipid metabolic process | Aspartic-type endopeptidase activity (GO:0004190) | None |

|  |  |  |  |  |
|---|---|---|---|---|
|  |  | (GO:0006629) |  |  |
| 14 | KAB8104787.1:20-151 hypothetical protein EE612_037915 | None | Nutrient reservoir activity (GO:0045735) | None |
| 15 | KAB8098942.1:20-149 hypothetical protein EE612_028695 | None | Nutrient reservoir activity (GO:0045735) | None |
| 16 | KAB8098943.1:13-119 hypothetical protein EE612_028695 | None | Nutrient reservoir activity (GO:0045735) | None |

## *Oryza sativa japonica*

| **S.No.** | **DESCRIPTION** | **GENE ONTOLOGY (GO) TERMS** |  |  |
|---|---|---|---|---|
|  |  | **Biological process** | **Molecular function** | **Cellular component** |
| 1 | XP_015645416.1:1-436 bZIP transcription factor RISBZ1-like isoform X1 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 2 | XP_025882428.1:1-310 bZIP transcription factor RISBZ1-like isoform X2 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity GO:0003700) | None |
| 3 | ACT31354.1:20-152 RISBZ1 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 4 | ABF99348.1:3-328 bZIP transcription factor family protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity GO:0003700) | None |
| 5 | XP_015631619.1:55-380 bZIP transcription factor RISBZ2 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 6 | XP_015646665.1:1-163 seed allergenic protein RAG1 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 7 | BAA01997.1:1-162 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 8 | KAF2921952.1:23-188 hypothetical protein DAI22_07g073500 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 9 | BAA01998.1:27-165 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 10 | ACA50505.1:27-166 seed allergenic protein RAG2 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 11 | BAA07711.1:20-160 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 12 | NP_001389964.1:1-155 seed allergenic protein RA5 precursor | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 13 | XP_015645223.1:27-152 alpha-amylase/trypsin inhibitor RA16 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 14 | BAA07773.1:2-109 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 15 | BAA01996.1:1-152 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 16 | KAF2921951.1:1-113 hypothetical protein DAI22_07g073550 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 17 | BAA07774.1:7-113 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 18 | BAA07772.1:1-111 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 19 | BAA01999.1:1-95 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 20 | XP_015631619.1:1-425 bZIP transcription factor RISBZ2 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 21 | ABF99348.1:1-373 bZIP transcription factor family protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 22 | EAZ28911.1:84-369 hypothetical protein OsJ_12953 | Regulation of transcription, DNA-templated (GO:0006355) | DNA-binding transcription factor activity (GO:0003700) | None |
| 23 | EAZ28911.1:1-67 hypothetical protein OsJ_12953 | None | None | None |
| 24 | ACA50505.1:1-166 seed allergenic protein RAG2 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 25 | KAF2921953.1:23-188 hypothetical protein DAI22_07g073500 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 26 | BAA07711.1:18-160 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 27 | AA01998.1:1-165 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 28 | XP_015646665.1:25-163 seed allergenic protein RAG1 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 29 | BAA01997.1:25-162 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 30 | Q01881.2:1-156 Full=Seed allergenic protein RA5 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 31 | KAF2921955.1:1-157 hypothetical protein DAI22_07g073600 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 32 | BAA01996.1:1-153 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 33 | XP_015645223.1:1-152 alpha-amylase/trypsin inhibitor RA16 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 34 | XP_015646664.1:24-160 seed allergenic protein RAG2 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 35 | BAA07773.1:2-109 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 36 | KAF2921954.1:23-143 hypothetical protein DAI22_07g073500 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 37 | KAF2921951.1:25-113 hypothetical protein DAI22_07g073550 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 38 | BAT00615.1:43-147 Os07g0213700 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 39 | BAA01999.1:25-95 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 40 | NP_001388678.1:1-644 heat shock 70 kDa protein BIP1 precursor | None | ATP binding (GO:0005524)<br>ATP-dependent protein folding chaperone (GO:0140662) | None |
| 41 | XP_015619916.1:1-203 Ras-related protein RABF2a | None | GTP binding (GO:0005525)<br>GTPase activity (GO:0003924) | None |
| 42 | EAZ21313.1:1-196 hypothetical protein OsJ_36965 | None | GTP binding (GO:0005525)<br>GTPase activity (GO:0003924) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 43 | ABA99930.2:1-249 Ras-related protein RHN1, putative | None | GTP binding (GO:0005525)<br>GTPase activity (GO:0003924) | None |
| 44 | KAF2908969.1:1-170 hypothetical protein DAI22_12g221000 | None | GTP binding (GO:0005525)<br>GTPase activity (GO:0003924) | None |
| 45 | XP_015636414.1:1-497 Vacuolar-processing enzyme beta-isozyme 1 | Proteolysis (GO:0006508)<br>Proteolysis involved in cellular protein catabolic process (GO:0051603) | Peptidase activity (GO:0008233)<br>Cysteine-type endopeptidase activity (GO:0004197) | None |
| 46 | CAE03020.3:1-517 OSJNBa0091D06.13 | Proteolysis (GO:0006508)<br>Proteolysis involved in cellular protein catabolic process (GO:0051603) | Peptidase activity (GO:0008233)<br>Cysteine-type endopeptidase activity (GO:0004197) | None |
| 47 | XP_015649120.1:1-466 Polycomb group protein FIE1 | None | Protein binding (GO:0005515) | None |
| 48 | XP_015649883.1:62-438 Polycomb group protein FIE1-like isoform X1 | None | Protein binding (GO:0005515) | None |
| 49 | AAS13489.1:1-376 fertilization independent endosperm | None | Protein binding (GO:0005515) | None |
| 50 | XP_015632668.1:9-480 Vacuolar protein sorting-associated protein 9A isoform X1 | Vesicle-mediated transport (GO:0016192) | Guanyl-nucleotide exchange factor activity (GO:0005085) | None |
| 51 | AAP06853.1:124-559 unknown protein | Vesicle-mediated transport (GO:0016192) | Guanyl-nucleotide exchange factor activity (GO:0005085) | None |
| 52 | AAP06853.1:9-41 unknown protein | None | None | None |
| 53 | XP_025879954.1:1-415 Vacuolar protein sorting-associated protein 9A isoform X2 | Vesicle-mediated transport (GO:0016192) | Guanyl-nucleotide exchange factor activity (GO:0005085) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 54 | BAF11545.1:1-371 Os03g0262900 | Vesicle-mediated transport (GO:0016192) | Guanyl-nucleotide exchange factor activity (GO:0005085) | None |
| 55 | ABF95104.1:1-276 Vacuolar sorting protein 9 domain containing protein | Vesicle-mediated transport (GO:0016192) | Guanyl-nucleotide exchange factor activity (GO:0005085) | None |
| 56 | NP_001389964.1:1-160 seed allergenic protein RA5 precursor | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 57 | BAA01996.1:1-157 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 58 | ACA50505.1:1-156 seed allergenic protein RAG2 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 59 | BAA01998.1:1-155 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 60 | KAF2921953.1:23-178 hypothetical protein DAI22_07g073500 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 61 | BAA07711.1:21-149 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 62 | BAA01997.1:1-152 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 63 | XP_015646665.1:1-153 seed allergenic protein RAG1 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 64 | XP_015646664.1:1-151 seed allergenic protein RAG2 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 65 | AAB99797.1:1-157 allergenic protein | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 66 | BAA07773.1:2-100 allergenic protein, partial | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 67 | BAA07774.1:7-104 allergenic protein, partial | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 68 | BAA07772.1:1-101 allergenic protein, partial | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 69 | KAF2921954.1:23-130 hypothetical protein DAI22_07g073500 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 70 | BAT00615.1:43-147 Os07g0213700 | None | Serine-type endopeptidase inhibitor activity (GO:0004867) | None |
| 71 | XP_015632668.1:9-480 Vacuolar protein sorting-associated protein 9A isoform X1 | Vesicle-mediated transport (GO:0016192) | Guanyl-nucleotide exchange factor activity (GO:0005085) | None |
| 72 | XP_015629757.1:1-501 serine/threonine-protein phosphatase BSL1 | None | Protein binding (GO:0005515) | None |
| 73 | AAR88573.1:1-467 putative transcription factor | None | Protein binding (GO:0005515) | None |
| 74 | XP_015621005.1:1-320 NAC domain-containing protein 79 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |
| 75 | XP_015629856.1:1-333 NAC domain-containing protein 79 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |
| 76 | ACJ54899.1:1-295 NAC protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |
| 77 | KAF2950309.1:1-277 hypothetical protein DAI22_01g187600 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |

| | | | | |
|---|---|---|---|---|
| 78 | KAF2950309.1:1-289 hypothetical protein DAI22_01g187600 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) | None |
| 79 | XP_015630971.1:1-1505 ABC transporter C family member 13 | Transmembrane transport (GO:0055085) | ABC-type transporter activity (GO:0140359)<br>ATP binding (GO:0005524) | integral component of membrane (GO:0016021) |
| 80 | EEE58303.1:1-1132 hypothetical protein OsJ_09363 | Transmembrane transport (GO:0055085) | ABC-type transporter activity (GO:0140359)<br>ATP binding (GO:0005524) | integral component of membrane (GO:0016021) |
| 81 | XP_015623741.1:1-373 Dof zinc finger protein 3-like | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 82 | EEE56670.1:1-348 hypothetical protein OsJ_06106 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 83 | ACT31340.1:4-124 Dof-type zinc finger protein 03, partial | None | None | None |
| 84 | AAV43900.1:20-150 putative prolamin 7 | None | Nutrient reservoir activity (GO:0045735) | None |
| 85 | CAA37850.1:19-149 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 86 | AAA50421.1:20-148 prolamine | None | Nutrient reservoir activity (GO:0045735) | None |
| 87 | KAF2930236.1:20-139 hypothetical protein DAI22_05g117900 | None | Nutrient reservoir activity (GO:0045735) | None |
| 88 | BAS93419.1:8-110 Os05g0329200, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 89 | AAA50424.1:24-149 prolamine | None | Nutrient reservoir activity (GO:0045735) | None |
| 90 | KAF2930230.1:6-124 hypothetical protein DAI22_05g118100 | None | Nutrient reservoir activity (GO:0045735) | None |

| | | | | |
|---|---|---|---|---|
| 91 | XP_015638344.1:20-152 prolamin PPROL 14E-like | None | Nutrient reservoir activity (GO:0045735) | None |
| 92 | KAF2930232.1:20-135 hypothetical protein DAI22_05g118200 | None | Nutrient reservoir activity (GO:0045735) | None |
| 93 | AAV44000.1:20-143 putative prolamin 7 | None | Nutrient reservoir activity (GO:0045735) | None |
| 94 | BAS93417.1:20-147 Os05g0329400 | None | Nutrient reservoir activity (GO:0045735) | None |
| 95 | AAB99798.1:22-148 seed prolamine | None | Nutrient reservoir activity (GO:0045735) | None |
| 96 | KAF2930231.1:14-130 hypothetical protein DAI22_05g118101 | None | None | None |
| 97 | AAV43828.1:38-142 putative prolamin 7 | None | Nutrient reservoir activity (GO:0045735) | None |
| 98 | BAS93426.1:16-120 Os05g0331366, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 99 | AAV43993.1:20-125 putative prolamin 7 | None | Nutrient reservoir activity (GO:0045735) | None |
| 100 | AAV43990.1:32-126 putative prolamin 7 | None | None | None |
| 101 | KAF2930237.1:20-131 hypothetical protein DAI22_05g117901 | None | Nutrient reservoir activity (GO:0045735) | None |
| 102 | BAH93086.1:17-109 Os05g0329200 | None | None | None |
| 103 | XP_015645890.1:21-154 13 kDa prolamin C | None | Nutrient reservoir activity (GO:0045735) | None |
| 104 | XP_015645890.1:181-314 13 kDa prolamin C | None | Nutrient reservoir activity (GO:0045735) | None |
| 105 | ADR66990.1:20-94 prolamin protein | None | None | None |
| 106 | KAF2930241.1:24-113 hypothetical protein DAI22_05g119100 | None | None | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 107 | BAS93408.1:1-74 Os05g0328333, partial | None | None | None |
| 108 | XP_025878040.1:20-154 prolamin PPROL 17D-like | None | Nutrient reservoir activity (GO:0045735) | None |
| 109 | KAF2907351.1:12-117 hypothetical protein DAI22_12g090900 | None | Nutrient reservoir activity (GO:0045735) | None |
| 110 | BAT16640.1:24-158 Os12g0269600, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 111 | XP_015619196.1:81-215 prolamin PPROL 17D | None | Nutrient reservoir activity (GO:0045735) | None |
| 112 | CAA43295.1:20-152 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 113 | BAS93410.1:20-73 Os05g0328632 | None | None | None |
| 114 | ABA97368.1:48-153 Prolamin PPROL 17 precursor, putative, expressed | None | Nutrient reservoir activity (GO:0045735) | None |
| 115 | KAB8098940.1:1-53 hypothetical protein EE612_028694, partial | None | None | None |
| 116 | BAS93409.1:17-70 Os05g0328466, partial | None | None | None |
| 117 | NP_001389564.1:1-509 aspartic proteinase oryzasin-1 precursor | Proteolysis (GO:0006508)<br>Lipid metabolic process (GO:0006629) | Aspartic-type endopeptidase activity (GO:0004190) | None |
| 118 | XP_015621906.1:29-522 aspartic proteinase oryzasin-1-like | Proteolysis (GO:0006508)<br>Lipid metabolic process (GO:0006629) | Aspartic-type endopeptidase activity (GO:0004190) | None |
| 119 | XP_015637008.1:1-378 dof zinc finger protein 1 isoform X1 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |

| | | | | |
|---|---|---|---|---|
| 120 | XP_015637009.1:1-363 dof zinc finger protein 1 isoform X2 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 121 | NP_001388848.1:1-363 dof zinc finger protein 2 isoform 1 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 122 | XP_025878040.1:20-156 prolamin PPROL 17D-like | None | Nutrient reservoir activity (GO:0045735) | None |
| 123 | ADB84622.1:20-156 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 124 | BAT16640.1:24-160 Os12g0269600, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 125 | XP_015619196.1:81-217 prolamin PPROL 17D | None | Nutrient reservoir activity (GO:0045735) | None |
| 126 | KAF2907351.1:8-119 hypothetical protein DAI22_12g090900 | None | Nutrient reservoir activity (GO:0045735) | None |
| 127 | AAA50423.1:20-149 prolamine | None | Nutrient reservoir activity (GO:0045735) | None |
| 128 | ABA97368.1:44-155 Prolamin PPROL 17 precursor, putative, expressed | None | Nutrient reservoir activity (GO:0045735) | None |
| 129 | XP_015645890.1:20-156 13 kDa prolamin C | None | Nutrient reservoir activity (GO:0045735) | None |
| 130 | XP_015645890.1:180-316 13 kDa prolamin C | None | Nutrient reservoir activity (GO:0045735) | None |
| 131 | NP_001389960.1:20-156 13 kDa prolamin C precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 132 | BAH95620.1:13-78 Os12g0269600, partial | None | None | None |
| 133 | AAV43900.1:20-149 putative prolamin 7 | None | Nutrient reservoir activity (GO:0045735) | None |
| 134 | CAA37849.1:19-150 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 135 | ABL74532.1:20-151 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 136 | CAA43295.1:20-152 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 137 | Q42465.1:20-151 Prolamin PPROL 14P; Precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 138 | AAA50424.1:45-148 prolamine | None | None | None |
| 139 | ABL74528.1:20-149 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 140 | CAA37850.1:19-148 prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 141 | BAS93419.1:5-109 Os05g0329200, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 142 | AAA50421.1:20-147 prolamine | None | Nutrient reservoir activity (GO:0045735) | None |
| 143 | XP_015638344.1:20-153 prolamin PPROL 14E-like | None | Nutrient reservoir activity (GO:0045735) | None |
| 144 | KAF2930236.1:35-138 hypothetical protein DAI22_05g117900 | None | Nutrient reservoir activity (GO:0045735) | None |
| 145 | KAB8098944.1:8-112 hypothetical protein EE612_028695 | None | Nutrient reservoir activity (GO:0045735) | None |
| 146 | AAV44000.1:20-144 putative prolamin 7 | None | Nutrient reservoir activity (GO:0045735) | None |
| 147 | KAF2930230.1:6-123 hypothetical protein DAI22_05g118100 | None | Nutrient reservoir activity (GO:0045735) | None |
| 148 | XP_015638444.1:20-153 prolamin PPROL 14E-like | None | Nutrient reservoir activity (GO:0045735) | None |
| 149 | XP_015643961.1:40-143 prolamin PPROL 17D-like | None | Nutrient reservoir activity (GO:0045735) | None |
| 150 | BAS93420.1:20-147 Os05g0330150 | None | Nutrient reservoir activity (GO:0045735) | None |
| 151 | KAF2930231.1:14-131 hypothetical protein DAI22_05g118101 | None | None | None |

| | | | | |
|---|---|---|---|---|
| 152 | BAS93417.1:20-147 Os05g0329400 | None | Nutrient reservoir activity (GO:0045735) | None |
| 153 | KAF2930232.1:20-135 hypothetical protein DAI22_05g118200 | None | None | None |
| 154 | XP_015638405.1:20-153 prolamin PPROL 14E-like | None | Nutrient reservoir activity (GO:0045735) | None |
| 155 | BAS93426.1:17-121 Os05g0331366, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 156 | ADB84624.1:8-115 prolamin-like protein | None | Nutrient reservoir activity (GO:0045735) | None |
| 157 | AAV43828.1:38-143 putative prolamin 7 | None | Nutrient reservoir activity (GO:0045735) | None |
| 158 | AAV43990.1:29-127 putative prolamin 7 | None | None | None |
| 159 | BAD45844.1:33-140 putative prolamin | None | Nutrient reservoir activity (GO:0045735) | None |
| 160 | BAH93539.1:51-158 Os06g0507001, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 161 | AAB99798.1:44-147 seed prolamine | None | None | None |
| 162 | XP_015637008.1:1-378 dof zinc finger protein 1 isoform X1 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 163 | EAY95200.1:28-405 hypothetical protein OsI_17020 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 164 | XP_015637009.1:1-363 dof zinc finger protein 1 isoform X2 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 165 | CAE02073.2:47-360 OSJNBa0005N02.10 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 166 | NP_001388848.1:1-342 dof zinc finger protein 2 isoform 1 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |

| | | | | |
|---|---|---|---|---|
| 167 | EEC73768.1:1-342 hypothetical protein OsI_08442 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 168 | BAA78573.1:3-336 Dof zinc finger protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 169 | NP_001388849.1:53-327 dof zinc finger protein 2 isoform 2 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 170 | XP_015637008.1:1-344 dof zinc finger protein 1 isoform X1 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 171 | CAE02073.2:48-326 OSJNBa0005N02.10 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 172 | XP_015637009.1:51-329 dof zinc finger protein 1 isoform X2 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 173 | NP_001388891.1:1-123 dof zinc finger protein 4 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 174 | NP_001388891.1:189-282 dof zinc finger protein 4 | None | None | None |
| 175 | EEE57725.1:26-147 hypothetical protein OsJ_08221 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 176 | EEE57725.1:213-306 hypothetical protein OsJ_08221 | None | None | None |
| 177 | ACT31342.1:15-106 Dof-type zinc finger protein 05, partial | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |
| 178 | XP_015630447.1:1-211 dof zinc finger protein 5 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677) DNA-binding transcription factor activity (GO:0003700) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 179 | EAZ13599.1:1-205 hypothetical protein OsJ_03515 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 180 | EAY75883.1:1-204 hypothetical protein OsI_03802 | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 181 | BAA78576.1:4-154 Dof zinc finger protein | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |
| 182 | ACT83312.1:28-114 Dof-type zinc finger protein 20, partial | Regulation of transcription, DNA-templated (GO:0006355) | DNA binding (GO:0003677)<br>DNA-binding transcription factor activity (GO:0003700) | None |

## *Triticum aestivum* (Wheat)

| S.No. | DESCRIPTION | GENE ONTOLOGY (GO) TERMS | | |
|---|---|---|---|---|
| | | **Biological process** | **Molecular function** | **Cellular component** |
| 1 | AAS67318.1:1-125 HMW glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 2 | AEL99901.1:1-118 high molecular weight glutenin subunit Dx5, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 3 | AHW49379.1:1-122 HMW glutenin, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 4 | AIE47878.1:36-117 high molecular weight glutenin subunit 1Dx2 | None | Nutrient reservoir activity (GO:0045735) | None |
| 5 | ASZ79030.1:41-122 high molecular weight glutenin subunit 1Slx2.3 | None | Nutrient reservoir activity (GO:0045735) | None |
| 6 | P08489.1:1-124 high molecular weight glutenin subunit PW212 | None | Nutrient reservoir activity (GO:0045735) | None |
| 7 | AFB35208.1:76-236 LMW-GS | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 8 | AAB48474.1:74-234 low-molecular-weight glutenin storage protein, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 9 | prf\|\|1209306A:84-244 gliadin gamma | None | Nutrient reservoir activity (GO:0045735) | None |
| 10 | AAX98174.1:130-290 low molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 11 | ACX46512.1:145-305 LMW-m glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 12 | QKX46045.1:147-307 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 13 | AMY62702.1:158-318 S-type low molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 14 | ABG45900.1:134-294 low molecular weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 15 | QKX46041.1:146-306 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 16 | ABB17938.1:127-287 gamma-gliadin/LMW-glutenin chimera Ch4 precursor, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 17 | AGK83314.1:148-308 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 18 | P04730.1:84-244 Gliadin B-III | None | Nutrient reservoir activity (GO:0045735) | None |
| 19 | ABB17936.1:138-298 gamma-gliadin/LMW-glutenin chimera Ch2 precursor, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 20 | QKX46052.1:1-33 low-molecular-weight glutenin subunit | None | None | None |

| | | | | |
|---|---|---|---|---|
| 21 | BAA22614.2:124-284 LMM glutenin 2, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 22 | AAV92074.1:144-304 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 23 | AAV92017.1:143-303 low molecular weight glutenin, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 24 | AAV92013.1:133-293 low molecular weight glutenin, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 25 | QZP12101.1:1-458 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 26 | QZP12101.1:605-648 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 27 | AHC72161.1:605-646 high molecular weight glutenin subunit 1Dy protein | None | Nutrient reservoir activity (GO:0045735) | None |
| 28 | ARJ58937.1:1-325 high molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 29 | ARJ58937.1:560-603 high molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 30 | AVE15896.1:1-373 high molecular weight glutenin, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 31 | CAC84118.1:1-194 glutenin high molecular weight subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 32 | CAC84120.1:1-174 glutenin high molecular weight subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 33 | XP_044451036.1:303-346 glutenin, high molecular weight subunit 12 isoform X1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 34 | ACO56367.1:615-658 HMW glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 35 | P08488.1:617-660 high molecular weight glutenin subunit 12 | None | Nutrient reservoir activity (GO:0045735) | None |
| 36 | ABO86195.1:609-652 Dy-type high molecular weight subunit protein | None | Nutrient reservoir activity (GO:0045735) | None |
| 37 | AKP95632.1:630-673 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 38 | ACD44935.1:594-637 HMW glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 39 | AAU04841.1:612-655 high molecular weight glutenin subunit 1Dy10.1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 40 | AKP95633.1:1-140 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 41 | SCW25215.1:674-717 HMW glutenin y-type subunit By8 precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 42 | KAF6990880.1:1-173 hypothetical protein CFC21_008029, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 43 | AEN55440.1:706-749 y-type high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 44 | SCW25212.1:659-702 HMW glutenin y-type subunit 1By7 precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 45 | AYM46701.1:1-201 high molecular weight glutenin subunit 1Dy12 | None | Nutrient reservoir activity (GO:0045735) | None |
| 46 | AYM46701.1:606-649 high molecular weight glutenin subunit 1Dy12 | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 47 | AAF23507.1:670-713 glutenin, high molecular weight subunit type y precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 48 | QKX46002.1:131-302 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 49 | QKX46002.1:1-31 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 50 | AFG73615.1:132-302 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 51 | AFG73619.1:130-300 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 52 | AQZ22280.1:134-304 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 53 | QKX46003.1:131-301 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 54 | QKX46010.1:109-279 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 55 | BAB78756.1:120-290 low-molecular-weight glutenin subunit group 6 type IV | None | Nutrient reservoir activity (GO:0045735) | None |
| 56 | ACT98429.1:136-306 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 57 | BAA22613.2:115-285 LMM glutenin 1, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 58 | ACR10430.1:110-277 low molecular weight protein, partial | None | Nutrient reservoir activity (GO:0045735) | None |

| | | | | |
|---|---|---|---|---|
| 59 | QKX46037.1:249-295 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 60 | QKX46037.1:124-186 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 61 | QKX46037.1:1-35 low-molecular-weight glutenin subunit | None | None | None |
| 62 | BAB78757.1:234-279 low-molecular-weight glutenin subunit group 6 type IV | None | Nutrient reservoir activity (GO:0045735) | None |
| 63 | BAB78757.1:109-171 low-molecular-weight glutenin subunit group 6 type IV | None | Nutrient reservoir activity (GO:0045735) | None |
| 64 | ANU06100.1:255-300 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 65 | ANU06100.1:130-192 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 66 | BAB78754.1:258-303 low-molecular-weight glutenin subunit group 6 type IV | None | Nutrient reservoir activity (GO:0045735) | None |
| 67 | BAB78754.1:133-195 low-molecular-weight glutenin subunit group 6 type IV | None | Nutrient reservoir activity (GO:0045735) | None |
| 68 | AQZ22278.1:257-302 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 69 | AQZ22278.1:132-194 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 70 | CAD58622.1:236-281 low molecular weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 71 | CAD58622.1:111-173 low molecular weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |

| | | | | |
|---|---|---|---|---|
| 72 | ACT98428.1:259-304 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 73 | ACT98428.1:134-196 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 74 | AGK83402.1:233-278 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 75 | AGK83402.1:108-170 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 76 | ACY08809.1:134-196 low molecular weight glutenin subunit A3-1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 77 | AGK83396.1:235-280 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 78 | AGK83396.1:110-172 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 79 | AGK83142.1:239-284 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 80 | AGK83142.1:114-176 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 81 | AGK83307.1:235-280 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 82 | AAN32705.1:101-146 gamma-gliadin, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 83 | AAR29050.1:1-468 HMW glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 84 | AAR29050.1:543-658 HMW glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 85 | P08488.1:545-660 high molecular weight glutenin subunit 12 | None | Nutrient reservoir activity (GO:0045735) | None |
| 86 | ABO86195.1:537-652 Dy-type high molecular weight subunit protein | None | Nutrient reservoir activity (GO:0045735) | None |
| 87 | AAX86829.1:1-384 HMW glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 88 | AAU04841.1:1-465 high molecular weight glutenin subunit 1Dy10.1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 89 | AAU04841.1:612-655 high molecular weight glutenin subunit 1Dy10.1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 90 | AKP95632.1:1-483 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 91 | AKP95632.1:626-673 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 92 | ABX64443.1:558-673 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 93 | CAC84120.1:1-174 glutenin high molecular weight subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 94 | XP_044451036.1:1-214 glutenin, high molecular weight subunit 12 isoform X1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 95 | XP_044451036.1:231-346 glutenin, high molecular weight subunit 12 isoform X1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 96 | AEN55440.1:1-371 y-type high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |

| | | | | |
|---|---|---|---|---|
| 97 | AEN55440.1:702-749 y-type high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 98 | ARJ58937.1:1-173 high molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 99 | ARJ58937.1:560-603 high molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 100 | AHC72161.1:605-646 high molecular weight glutenin subunit 1Dy protein | None | Nutrient reservoir activity (GO:0045735) | None |
| 101 | QZP12101.1:605-648 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 102 | AHW49381.1:1-140 HMW glutenin, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 103 | ACD44935.1:1-214 HMW glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 104 | ACD44935.1:522-637 HMW glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 105 | ACD44935.1:255-447 HMW glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 106 | AKP95633.1:249-468 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 107 | SCW25215.1:602-717 HMW glutenin y-type subunit By8 precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 108 | SCW25212.1:587-702 HMW glutenin y-type subunit 1By7 precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 109 | AAF23507.1:670-713 glutenin, high molecular weight subunit type y precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 110 | AGK83207.1:169-340 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 111 | P10385.1:185-356 low molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 112 | AAS10192.1:185-357 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 113 | AGE13924.1:46-111 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 114 | AGE13924.1:184-222 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 115 | BAB78762.1:37-102 low-molecular-weight glutenin subunit group 11 type VI | None | Nutrient reservoir activity (GO:0045735) | None |
| 116 | BAB78762.1:174-212 low-molecular-weight glutenin subunit group 11 type VI | None | Nutrient reservoir activity (GO:0045735) | None |
| 117 | AYP71749.1:54-119 low molecular weight glutenin subunit A3-3 | None | Nutrient reservoir activity (GO:0045735) | None |
| 118 | AYP71749.1:193-231 low molecular weight glutenin subunit A3-3 | None | Nutrient reservoir activity (GO:0045735) | None |
| 119 | AYP71752.1:62-127 low molecular weight glutenin subunit A3-6 | None | Nutrient reservoir activity (GO:0045735) | None |
| 120 | AYP71752.1:201-239 low molecular weight glutenin subunit A3-6 | None | Nutrient reservoir activity (GO:0045735) | None |
| 121 | AHN55168.1:93-158 low-molecular-weight glutenin subunit Glu-A3, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 122 | AHN55168.1:231-264 low-molecular-weight glutenin subunit Glu-A3, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 123 | ACP27640.1:95-160 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |

| | | | | |
|---|---|---|---|---|
| 124 | ACP27640.1:232-270 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 125 | ACJ66627.1:208-273 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 126 | ACJ66627.1:308-346 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 127 | ACP27638.1:127-192 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 128 | ACP27638.1:265-303 low molecular weight glutenin | None | None | None |
| 129 | ACZ51336.1:184-249 low molecular weight glutenin subunit A3-2, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 130 | ACZ51336.1:322-360 low molecular weight glutenin subunit A3-2, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 131 | AHN55167.1:200-265 low-molecular-weight glutenin subunit Glu-A3, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 132 | AHN55167.1:338-371 low-molecular-weight glutenin subunit Glu-A3, partial | None | None | None |
| 133 | AGK83212.1:186-251 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 134 | AGK83212.1:323-355 low-molecular-weight glutenin subunit, partial | None | None | None |
| 135 | AGK83310.1:324-356 low-molecular-weight glutenin subunit, partial | None | None | None |
| 136 | ABG76006.1:200-265 low molecular weight glutenin subunit GF-1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 137 | ABG76006.1:338-376 low molecular weight glutenin subunit GF-1 | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 138 | XP_044452584.1:200-265 low molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 139 | AGK83260.1:188-253 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 140 | AGK83260.1:326-364 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 141 | AAB48479.1:183-248 low-molecular-weight glutenin storage protein | None | Nutrient reservoir activity (GO:0045735) | None |
| 142 | AAB48479.1:321-359 low-molecular-weight glutenin storage protein | None | Nutrient reservoir activity (GO:0045735) | None |
| 143 | AGO17750.1:200-265 LMW-GS, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 144 | AGO17750.1:338-371 LMW-GS, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 145 | AGK83137.1:165-230 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 146 | AGK83137.1:303-341 low-molecular-weight glutenin subunit, partial | None | None | None |
| 147 | AAS10189.1:338-376 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 148 | AGK83149.1:199-264 low-molecular-weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 149 | AGK83149.1:336-374 low-molecular-weight glutenin subunit, partial | None | None | None |
| 150 | AQZ22274.1:202-267 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 151 | AQZ22274.1:340-378 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 152 | AAS10188.1:208-273 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 153 | AAS10188.1:346-384 low molecular weight glutenin | None | Nutrient reservoir activity (GO:0045735) | None |
| 154 | ACT98424.1:181-246 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 155 | ACT98424.1:318-356 low-molecular-weight glutenin subunit | None | None | None |
| 156 | AQZ22276.1:181-246 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 157 | AQZ22276.1:317-355 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 158 | ACY08811.1:182-247 low molecular weight glutenin subunit A3-2 | None | Nutrient reservoir activity (GO:0045735) | None |
| 159 | ACY08811.1:320-358 low molecular weight glutenin subunit A3-2 | None | None | None |
| 160 | BAB78763.1:216-281 low-molecular-weight glutenin subunit group 12 type VI | None | Nutrient reservoir activity (GO:0045735) | None |
| 161 | BAB78763.1:353-391 low-molecular-weight glutenin subunit group 12 type VI | None | None | None |
| 162 | AGO17731.1:182-247 LMW-GS, partial | None | None | None |
| 163 | AGO17731.1:319-352 LMW-GS, partial | None | Nutrient reservoir activity (GO:0045735) | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 164 | BAB78764.1:214-279 low-molecular-weight glutenin subunit group 12 type VI | None | Nutrient reservoir activity (GO:0045735) | None |
| 165 | BAB78764.1:351-389 low-molecular-weight glutenin subunit group 12 type VI | None | None | None |
| 166 | ACT98430.1:215-280 low-molecular-weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 167 | ACT98430.1:352-390 low-molecular-weight glutenin subunit | None | None | None |
| 168 | P02861.1:1-101 high molecular weight glutenin subunit PC256 | None | Nutrient reservoir activity (GO:0045735) | None |
| 169 | CAC83002.1:131-220 HMW glutenin, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 170 | ABF82252.1:735-824 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 171 | AKW50842.1:741-830 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |
| 172 | AAB02788.1:726-815 HMW glutenin subunit Ax2 | None | Nutrient reservoir activity (GO:0045735) | None |
| 173 | AAF23506.1:671-766 glutenin, high molecular weight subunit type x precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 174 | BAN29068.1:705-794 high molecular weight glutenin subunit, partial | None | Nutrient reservoir activity (GO:0045735) | None |
| 175 | SCW25219.1:725-814 HMW glutenin x-type subunit 1Ax21 precursor | None | Nutrient reservoir activity (GO:0045735) | None |
| 176 | ABK54365.1:726-815 high molecular weight glutenin subunit | None | Nutrient reservoir activity (GO:0045735) | None |

| | | | | |
|---|---|---|---|---|
| 177 | AIU47923.1:865-974 x-type HMW-GS 1Slx2.2 | None | Nutrient reservoir activity (GO:0045735) | None |
| 178 | AOZ35396.1:738-820 HMW glutenin 1Mgx | None | Nutrient reservoir activity (GO:0045735) | None |
| 179 | AOZ35396.1:367-390 HMW glutenin 1Mgx | None | None | None |
| 180 | P02862.1:1-39 high molecular weight glutenin subunit PC237 | None | Nutrient reservoir activity (GO:0045735) | None |
| 181 | D2KFH1.1:1-172 Avenin-like a4 | None | Nutrient reservoir activity (GO:0045735) | None |
| 182 | XP_044366523.1:1-171 Avenin-like a4 | None | Nutrient reservoir activity (GO:0045735) | None |
| 183 | P0CZ08.1:1-175 Avenin-like a3 | None | Nutrient reservoir activity (GO:0045735) | None |
| 184 | XP_044438991.1:1-168 Avenin-like a1 | None | Nutrient reservoir activity (GO:0045735) | None |
| 185 | P0CZ10.1:1-181 Avenin-like a6 | None | Nutrient reservoir activity (GO:0045735) | None |

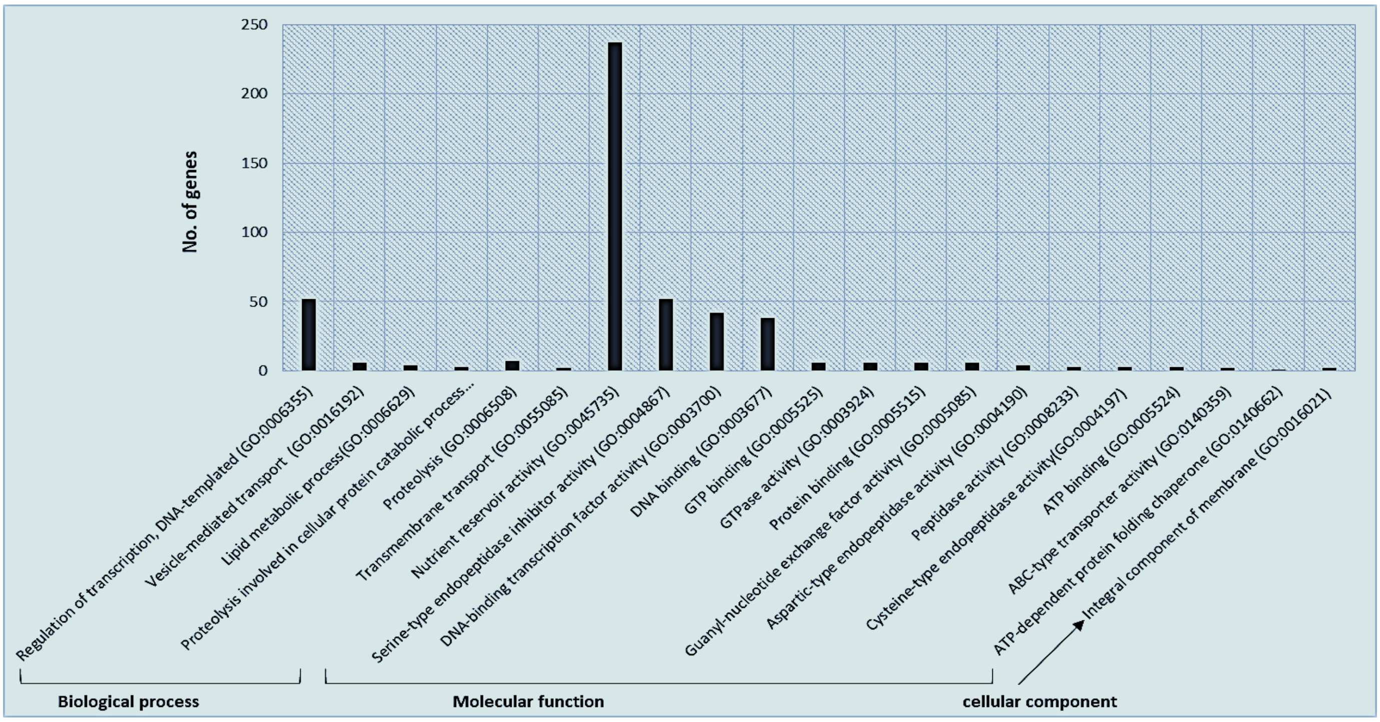


**Figure 2: Gene Ontology analysis of selected SSP sequences**

Out of 406 protein sequences, 368 have GO annotations, which corresponds to several GO terms classified as a cellular component, molecular function, and biological activity. There is a list of 21 GO terms from which we choose the top five. In decreasing order, the top five GO terms are listed: Nutrient reservoir activity (GO: 0045735) > Regulation of transcription, DNA templated (GO: 0006355) > Serine-type endopeptidase inhibitor activity (GO: 0004867) > DNA-binding transcription factor activity (GO: 0003700) > DNA binding (GO: 0003677). In comparison to other GO terms, nutrition reservoir activity has the greatest place in figure 2, indicating that it is the top activity demonstrated by protein sequences as a molecular function. Proteome nutrient reservoir activity states that it aids in the storage of nourishing substrates. The top second position is obtained through transcriptional regulation and DNA templated activity. It refers to the process of modifying the frequency, rate, or range of cellular DNA-templated transcription. Following this, Serine-type endopeptidase inhibitor activity has risen to third place. It inhibits, stops, or lowers the activity of serine-type endopeptidases, which are enzymes that catalyse the hydrolysis of nonterminal peptide bonds in a polypeptide chain, with a serine residue (and a histidine residue) at the active center. DNA-binding transcription factor activity is ranked fourth out of 21 GO terms. Through non-covalent binding to a specific double-stranded genomic DNA sequence within a cis-regulatory region, this transcription regulator activity influences gene set transcription. Figure 2 shows DNA binding activity in the top 5th position. The interaction of a gene with DNA in a non-covalent and selective manner is termed as DNA binding activity.

**GO annotations using the Interpro database, the Pfam database is now being used to classify predicted protein sequences in cereal genomes (wheat and rice), as indicated in the table below:**

## *Oryza sativa*

| S.No. | Query name | Description | Family | Pfam id |
|---|---|---|---|---|
| 1 | KAB8104546.1:1-431 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 2 | CAA44001.1:27-165 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 3 | KAB8093986.1:1-423 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 4 | CAA44001.1:1-165 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 5 | KAB8104775.1:1-113 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 6 | AAK38149.1:1-203 | Ras family | Ras | PF00071.25 |
| 7 | CAA44001.1:1-155 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 8 | KAB8098941.1:20-150 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 9 | KAB8098944.1:11-113 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 10 | KAB8098943.1:16-120 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 11 | pir\|JC4599\|:21-154 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 12 | BAA06875.1:1-509 | Eukaryotic aspartyl protease | Asp | PF00026.26 |
| 13 | KAB8082805.1:29-522 | Eukaryotic aspartyl protease | Asp | PF00026.26 |
| 14 | KAB8104787.1:20-151 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 15 | KAB8098942.1:20-149 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 16 | KAB8098943.1:13-119 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |

## *Oryza sativa indica*

| S.No. | Query name | Description | Family | Pfam id |
|---|---|---|---|---|
| 1 | EEC81618.1:1-436 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 2 | AAL10017.1:55-375 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 3 | EAZ03220.1:6-141 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 4 | BAA36492.1:1-421 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 5 | AAL10017.1:1-420 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 6 | EAZ03217.1:63-167 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 7 | ADV78543.1:1-203 | Ras family | Ras | PF00071.25 |
| 8 | B8ASK4.1:1-497 | Peptidase C13 family | Peptidase_C13 | PF01650.21 |
| 9 | EEC82867.1:1-376 | WD domain, G-beta repeat | WD40 | PF00400.35 |
| 10 | EAZ03220.1:7-131 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 11 | A2WJP3.1:1-320 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 12 | APH07720.1:1-318 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 13 | APH07721.1:1-326 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 14 | APH07722.1:1-330 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 15 | APH07723.1:1-322 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 16 | EEC72846.1:1-241 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 17 | EEC72845.1:8-105 | None | None | None |
| 18 | AAW80680.1:20-150 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 19 | EAY95200.1:28-405 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 20 | EEC73768.1:1-363 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 21 | AAW80680.1:20-149 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |

|  |  |  |  |  |
|---|---|---|---|---|
| 22 | EEC73921.1:26-147 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 23 | EEC73921.1:214-307 | None | None | None |

## *Oryza sativa japonica*

| S.No. | Query name | Description | Family | Pfam id |
|---|---|---|---|---|
| 1 | XP_015645416.1:1-436 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 2 | XP_025882428.1:1-310 | bZIP transcription factor | bZIP_1 | PF00170.24 |
| 3 | ACT31354.1:20-152 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 4 | ABF99348.1:3-328 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 5 | XP_015631619.1:55-380 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 6 | XP_015646665.1:1-163 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 7 | BAA01997.1:1-162 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 8 | KAF2921952.1:23-188 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 9 | BAA01998.1:27-165 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 10 | ACA50505.1:27-166 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 11 | BAA07711.1:20-160 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 12 | NP_001389964.1:1-155 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 13 | XP_015645223.1:27-152 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 14 | BAA07773.1:2-109 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |

|  |  |  |  |  |
|---|---|---|---|---|
| 15 | BAA01996.1:1-152 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 16 | KAF2921951.1:1-113 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 17 | BAA07774.1:7-113 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 18 | BAA07772.1:1-111 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 19 | BAA01999.1:1-95 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 20 | XP_015631619.1:1-425 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 21 | BAA11431.1:1-425 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 22 | ABF99348.1:1-373 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 23 | EAZ28911.1:84-369 | Basic leucine-zipper C terminal | bZIP_C | PF00170.24 |
| 24 | EAZ28911.1:1-67 | None | None | None |
| 25 | ACA50505.1:1-166 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 26 | KAF2921953.1:23-188 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 27 | BAA07711.1:18-160 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 28 | BAA01998.1:1-165 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 29 | XP_015646665.1:25-163 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 30 | EAZ03220.1:3-141 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 31 | BAA01997.1:25-162 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 32 | Q01881.2:1-156 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 33 | KAF2921955.1:1-157 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 34 | BAA01996.1:1-153 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |

| | | | | |
|---|---|---|---|---|
| 35 | XP_015645223.1:1-152 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 36 | XP_015646664.1:24-160 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 37 | BAA07773.1:2-109 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 38 | KAF2921954.1:23-143 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 39 | KAF2921951.1:25-113 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 40 | BAT00615.1:43-147 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 41 | EAZ03217.1:63-167 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 42 | BAA01999.1:25-95 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 43 | NP_001388678.1:1-644 | Hsp70 protein | HSP70 | PF00012.23 |
| 44 | XP_015619916.1:1-203 | Ras family | Ras | PF00071.25 |
| 45 | EAZ21313.1:1-196 | Ras family | Ras | PF00071.25 |
| 46 | ABA99930.2:1-249 | Ras family | Ras | PF00071.25 |
| 47 | KAF2908969.1:1-170 | Ras family | Ras | PF00071.25 |
| 48 | XP_015636414.1:1-497 | Peptidase C13 family | Peptidase_C13 | PF01650.21 |
| 49 | CAE03020.3:1-517 | Peptidase C13 family | Peptidase_C13 | PF01650.21 |
| 50 | XP_015625561.1:26-493 | Peptidase C13 family | Peptidase_C13 | PF01650.21 |
| 51 | XP_015649120.1:1-466 | WD domain, G-beta repeat | WD40 | PF00400.35 |
| 52 | XP_015649883.1:62-438 | WD domain, G-beta repeat | WD40 | PF00400.35 |
| 53 | AAS13489.1:1-376 | WD domain, G-beta repeat | WD40 | PF00400.35 |
| 54 | XP_015632668.1:9-480 | Vacuolar sorting protein 9 (VPS9) domain | VPS9 | PF02204.21 |
| 55 | EEC74907.1:9-458 | Vacuolar sorting protein 9 (VPS9) domain | VPS9 | PF02204.21 |

|  |  |  |  |  |
|---|---|---|---|---|
| 56 | AAP06853.1:124-559 | Vacuolar sorting protein 9 (VPS9) domain | VPS9 | PF02204.21 |
| 57 | AAP06853.1:9-41 | None | None | None |
| 58 | XP_025879954.1:1-415 | Vacuolar sorting protein 9 (VPS9) domain | VPS9 | PF02204.21 |
| 59 | BAF11545.1:1-371 | Vacuolar sorting protein 9 (VPS9) domain | VPS9 | PF02204.21 |
| 60 | ABF95104.1:1-276 | Vacuolar sorting protein 9 (VPS9) domain | VPS9 | PF02204.21 |
| 61 | NP_001389964.1:1-160 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 62 | BAA01996.1:1-157 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 63 | ACA50505.1:1-156 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 64 | BAA01998.1:1-155 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 65 | KAF2921953.1:23-178 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 66 | BAA07711.1:21-149 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 67 | BAA01997.1:1-152 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 68 | XP_015646665.1:1-153 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 69 | XP_015646664.1:1-151 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 70 | AAB99797.1:1-157 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 71 | BAA07773.1:2-100 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 72 | BAA07774.1:7-104 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 73 | BAA07772.1:1-101 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 74 | KAF2921954.1:23-130 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |

| | | | | |
|---|---|---|---|---|
| 75 | BAT00615.1:43-147 | Protease inhibitor/seed storage/LTP family | Tryp_alpha_amyl | PF00234.25 |
| 76 | XP_015629757.1:1-501 | Galactose oxidase, central domain | Kelch_4 | PF13418.9 |
| 77 | AAR88573.1:1-467 | Galactose oxidase, central domain | Kelch_4 | PF13418.9 |
| 78 | XP_015621005.1:1-320 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 79 | XP_015629856.1:1-333 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 80 | ACJ54899.1:1-295 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 81 | KAF2950309.1:1-277 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 82 | KAF2950309.1:1-289 | No apical meristem (NAM) protein | NAM | PF02365.18 |
| 83 | XP_015630971.1:1-1505 | ABC transporter transmembrane region | ABC_membrane | PF00664.26 |
| 84 | EEE58303.1:1-1132 | ABC transporter transmembrane region | ABC_membrane | PF00664.26 |
| 85 | XP_015623741.1:1-373 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 86 | EEE56670.1:1-348 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 87 | ACT31340.1:4-124 | None | None | None |
| 88 | AAV43900.1:20-150 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 89 | CAA37850.1:19-149 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 90 | AAA50421.1:20-148 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 91 | KAF2930236.1:20-139 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 92 | BAS93419.1:8-110 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 93 | AAA50424.1:24-149 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 94 | KAF2930230.1:6-124 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 95 | XP_015638344.1:20-152 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 96 | KAF2930232.1:20-135 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 97 | AAV44000.1:20-143 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 98 | BAS93417.1:20-147 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 99 | AAB99798.1:22-148 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |

|  |  |  |  |  |
|---|---|---|---|---|
| 100 | KAF2930231.1:14-130 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 101 | AAV43828.1:38-142 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 102 | BAS93426.1:16-120 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 103 | AAV43993.1:20-125 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 104 | AAV43990.1:32-126 | None | None | None |
| 105 | KAF2930237.1:20-131 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 106 | BAH93086.1:17-109 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 107 | XP_015645890.1:21-154 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 108 | XP_015645890.1:181-314 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 109 | ADR66990.1:20-94 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 110 | KAF2930241.1:24-113 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 111 | BAS93408.1:1-74 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 112 | XP_025878040.1:20-154 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 113 | KAF2907351.1:12-117 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 114 | BAT16640.1:24-158 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 115 | XP_015619196.1:81-215 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 116 | CAA43295.1:20-152 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 117 | BAS93410.1:20-73 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 118 | ABA97368.1:48-153 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 119 | KAB8098940.1:1-53 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 120 | BAS93409.1:17-70 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 121 | NP_001389564.1:1-509 | Eukaryotic aspartyl protease | Asp | PF00026.26 |
| 122 | XP_015621906.1:29-522 | Eukaryotic aspartyl protease | Asp | PF00026.26 |
| 123 | XP_015637008.1:1-378 | Dof domain, zinc finger | zf-Dof | PF02701.18 |

| | | | | |
|---|---|---|---|---|
| 124 | XP_015637009.1:1-363 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 125 | CAE02073.2:47-360 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 126 | NP_001388848.1:1-363 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 127 | XP_025878040.1:20-156 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 128 | ADB84622.1:20-156 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 129 | BAT16640.1:24-160 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 130 | XP_015619196.1:81-217 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 131 | KAF2907351.1:8-119 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 132 | AAA50423.1:20-149 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 133 | ABA97368.1:44-155 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 134 | XP_015645890.1:20-156 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 135 | XP_015645890.1:180-316 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 136 | NP_001389960.1:20-156 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 137 | BAH95620.1:13-78 | None | None | None |
| 138 | AAV43900.1:20-149 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 139 | CAA37849.1:19-150 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 140 | ABL74532.1:20-151 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 141 | CAA43295.1:20-152 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 142 | Q42465.1:20-151 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 143 | AAA50424.1:45-148 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 144 | ABL74528.1:20-149 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 145 | CAA37850.1:19-148 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 146 | BAS93419.1:5-109 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 147 | AAA50421.1:20-147 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |

| | | | | |
|---|---|---|---|---|
| 148 | XP_015638344.1:20-153 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 149 | KAF2930236.1:35-138 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 150 | KAB8098944.1:8-112 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 151 | AAV44000.1:20-144 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 152 | KAF2930230.1:6-123 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 153 | XP_015638444.1:20-153 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 154 | XP_015643961.1:40-143 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 155 | BAS93420.1:20-147 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 156 | KAF2930231.1:14-131 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 157 | BAS93417.1:20-147 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 158 | KAF2930232.1:20-135 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 159 | XP_015638405.1:20-153 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 160 | BAS93426.1:17-121 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 161 | ADB84624.1:8-115 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 162 | AAV43828.1:38-143 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 163 | AAV43990.1:29-127 | None | None | None |
| 164 | BAD45844.1:33-140 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 165 | BAH93539.1:51-158 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 166 | AAB99798.1:44-147 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 167 | XP_015637008.1:1-378 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 168 | EAY95200.1:28-405 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 169 | XP_015637009.1:1-363 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 170 | CAE02073.2:47-360 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 171 | NP_001388848.1:1-363 | Dof domain, zinc finger | zf-Dof | PF02701.18 |

| | | | | |
|---|---|---|---|---|
| 172 | NP_001388848.1:1-342 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 173 | EEC73768.1:1-342 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 174 | BAA78573.1:3-336 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 175 | NP_001388849.1:53-327 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 176 | XP_015637008.1:1-344 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 177 | CAE02073.2:48-326 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 178 | XP_015637009.1:51-329 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 179 | NP_001388891.1:1-123 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 180 | NP_001388891.1:189-282 | None | None | None |
| 181 | EEE57725.1:26-147 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 182 | EEE57725.1:213-306 | None | None | None |
| 183 | ACT31342.1:15-106 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 184 | XP_015630447.1:1-211 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 185 | EAZ13599.1:1-205 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 186 | EAY75883.1:1-204 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 187 | BAA78576.1:4-154 | Dof domain, zinc finger | zf-Dof | PF02701.18 |
| 188 | ACT83312.1:28-114 | Dof domain, zinc finger | zf-Dof | PF02701.18 |

## *Triticum aestivum*

| S.No. | Query name | Description | Family | Pfam id |
|---|---|---|---|---|
| 1 | AAS67318.1:1-125 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 2 | AEL99901.1:1-118 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 3 | AHW49379.1:1-122 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |

| | | | | |
|---|---|---|---|---|
| 4 | AIE47878.1:36-117 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 5 | ASZ79030.1:41-122 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 6 | P08489.1:1-124 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 7 | AFB35208.1:76-236 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 8 | AAB48474.1:74-234 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 9 | prf\|\|1209306A:84-244 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 10 | AAX98174.1:130-290 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 11 | ACX46512.1:145-305 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 12 | QKX46045.1:147-307 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 13 | AMY62702.1:158-318 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 14 | ABG45900.1:134-294 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 15 | QKX46041.1:146-306 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 16 | ABB17938.1:127-287 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 17 | AGK83314.1:148-308 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 18 | P04730.1:84-244 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 19 | ABB17936.1:138-298 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 20 | QKX46052.1:1-33 | None | None | None |
| 21 | BAA22614.2:124-284 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 22 | AAV92074.1:144-304 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 23 | AAV92017.1:143-303 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 24 | AAV92013.1:133-293 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 25 | QZP12101.1:1-458 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 26 | QZP12101.1:605-648 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 27 | AHC72161.1:605-646 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 28 | ARJ58937.1:1-325 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 29 | ARJ58937.1:560-603 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |

| | | | | |
|---|---|---|---|---|
| 30 | AVE15896.1:1-373 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 31 | CAC84118.1:1-194 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 32 | CAC84120.1:1-174 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 33 | XP_044451036.1:303-346 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 34 | ACO56367.1:615-658 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 35 | P08488.1:617-660 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 36 | ABO86195.1:609-652 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 37 | AKP95632.1:630-673 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 38 | ACD44935.1:594-637 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 39 | AAU04841.1:612-655 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 40 | AKP95633.1:1-140 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 41 | SCW25215.1:674-717 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 42 | KAF6990880.1:1-173 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 43 | AEN55440.1:706-749 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 44 | SCW25212.1:659-702 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 45 | AYM46701.1:1-201 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 46 | AYM46701.1:606-649 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 47 | AAF23507.1:670-713 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 48 | QKX46002.1:131-302 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 49 | QKX46002.1:1-31 | None | None | None |
| 50 | AFG73615.1:132-302 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 51 | AFG73619.1:130-300 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 52 | AQZ22280.1:134-304 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 53 | QKX46003.1:131-301 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 54 | QKX46010.1:109-279 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 55 | BAB78756.1:120-290 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |

|  |  |  |  |  |
|---|---|---|---|---|
| 56 | ACT98429.1:136-306 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 57 | BAA22613.2:115-285 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 58 | ACR10430.1:110-277 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 59 | QKX46037.1:249-295 | None | None | None |
| 60 | QKX46037.1:124-186 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 61 | QKX46037.1:1-35 | None | None | None |
| 62 | BAB78757.1:234-279 | None | None | None |
| 63 | BAB78757.1:109-171 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 64 | ANU06100.1:255-300 | None | None | None |
| 65 | ANU06100.1:130-192 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 66 | BAB78754.1:258-303 | None | None | None |
| 67 | BAB78754.1:133-195 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 68 | AQZ22278.1:257-302 | None | None | None |
| 69 | AQZ22278.1:132-194 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 70 | CAD58622.1:236-281 | None | None | None |
| 71 | CAD58622.1:111-173 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 72 | ACT98428.1:259-304 | None | None | None |
| 73 | ACT98428.1:134-196 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 74 | AGK83402.1:233-278 | None | None | None |
| 75 | AGK83402.1:108-170 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 76 | ACY08809.1:134-196 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 77 | AGK83396.1:235-280 | None | None | None |
| 78 | AGK83396.1:110-172 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 79 | AGK83142.1:239-284 | None | None | None |
| 80 | AGK83142.1:114-176 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 81 | AGK83307.1:235-280 | None | None | None |

| | | | | |
|---|---|---|---|---|
| 82 | AAN32705.1:101-146 | None | None | None |
| 83 | AAR29050.1:1-468 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 84 | AAR29050.1:543-658 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 85 | P08488.1:545-660 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 86 | ABO86195.1:537-652 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 87 | AAX86829.1:1-384 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 88 | AAU04841.1:1-465 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 89 | AAU04841.1:612-655 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 90 | AKP95632.1:1-483 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 91 | AKP95632.1:626-673 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 92 | ABX64443.1:558-673 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 93 | CAC84120.1:1-174 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 94 | XP_044451036.1:1-214 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 95 | XP_044451036.1:231-346 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 96 | AEN55440.1:1-371 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 97 | AEN55440.1:702-749 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 98 | ARJ58937.1:1-173 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 99 | ARJ58937.1:560-603 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 100 | AHC72161.1:605-646 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 101 | QZP12101.1:605-648 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 102 | AHW49381.1:1-140 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 103 | ACD44935.1:1-214 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 104 | ACD44935.1:522-637 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 105 | ACD44935.1:255-447 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 106 | AKP95633.1:249-468 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 107 | SCW25215.1:602-717 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |

|  |  |  |  |  |
|---|---|---|---|---|
| 108 | SCW25212.1:587-702 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 109 | AAF23507.1:670-713 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 110 | AGK83207.1:169-340 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 111 | P10385.1:185-356 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 112 | AAS10192.1:185-357 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 113 | AGE13924.1:46-111 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 114 | AGE13924.1:184-222 | None | None | None |
| 115 | BAB78762.1:37-102 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 116 | BAB78762.1:174-212 | None | None | None |
| 117 | AYP71749.1:54-119 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 118 | AYP71749.1:193-231 | None | None | None |
| 119 | AYP71752.1:62-127 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 120 | AYP71752.1:201-239 | None | None | None |
| 121 | AHN55168.1:93-158 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 122 | AHN55168.1:231-264 | None | None | None |
| 123 | ACP27640.1:95-160 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 124 | ACP27640.1:232-270 | None | None | None |
| 125 | ACJ66627.1:208-273 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 126 | ACJ66627.1:308-346 | None | None | None |
| 127 | ACP27638.1:127-192 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 128 | ACP27638.1:265-303 | None | None | None |
| 129 | ACZ51336.1:184-249 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 130 | ACZ51336.1:322-360 | None | None | None |
| 131 | AHN55167.1:200-265 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 132 | AHN55167.1:338-371 | None | None | None |
| 133 | AGK83212.1:186-251 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |

|  |  |  |  |  |
|---|---|---|---|---|
| 134 | AGK83212.1:323-355 | None | None | None |
| 135 | AGK83310.1:324-356 | None | None | None |
| 136 | ABG76006.1:200-265 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 137 | ABG76006.1:338-376 | None | None | None |
| 138 | XP_044452584.1:200-265 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 139 | AGK83260.1:188-253 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 140 | AGK83260.1:326-364 | None | None | None |
| 141 | AAB48479.1:183-248 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 142 | AAB48479.1:321-359 | None | None | None |
| 143 | AGO17750.1:200-265 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 144 | AGO17750.1:338-371 | None | None | None |
| 145 | AGK83137.1:165-230 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 146 | AGK83137.1:303-341 | None | None | None |
| 147 | AAS10189.1:338-376 | None | None | None |
| 148 | AGK83149.1:199-264 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 149 | AGK83149.1:336-374 | None | None | None |
| 150 | AQZ22274.1:202-267 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 151 | AQZ22274.1:340-378 | None | None | None |
| 152 | AS10188.1:208-273 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 153 | AAS10188.1:346-384 | None | None | None |
| 154 | ACT98424.1:181-246 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 155 | ACT98424.1:318-356 | None | None | None |
| 156 | AQZ22276.1:181-246 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 157 | AQZ22276.1:317-355 | None | None | None |
| 158 | ACY08811.1:182-247 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 159 | ACY08811.1:320-358 | None | None | None |

|  |  |  |  |  |
|---|---|---|---|---|
| 160 | BAB78763.1:216-281 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 161 | BAB78763.1:353-391 | None | None | None |
| 162 | AGO17731.1:182-247 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 163 | AGO17731.1:319-352 | None | None | None |
| 164 | BAB78764.1:214-279 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 165 | BAB78764.1:351-389 | None | None | None |
| 166 | ACT98430.1:215-280 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 167 | ACT98430.1:352-390 | None | None | None |
| 168 | P02861.1:1-101 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 169 | CAC83002.1:131-220 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 170 | ABF82252.1:735-824 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 171 | AKW50842.1:741-830 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 172 | AAB02788.1:726-815 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 173 | AAF23506.1:671-766 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 174 | BAN29068.1:705-794 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 175 | SCW25219.1:725-814 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 176 | ABK54365.1:726-815 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 177 | AIU47923.1:865-974 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 178 | AOZ35396.1:738-820 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 179 | AOZ35396.1:367-390 | None | None | None |
| 180 | P02862.1:1-39 | High molecular weight glutenin subunit | Glutenin_hmw | PF03157.16 |
| 181 | D2KFH1.1:1-172 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 182 | XP_044366523.1:1-171 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 183 | P0CZ08.1:1-175 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 184 | XP_044438991.1:1-168 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |
| 185 | P0CZ10.1:1-181 | Cys-rich Gliadin N-terminal | Gliadin | PF13016.9 |

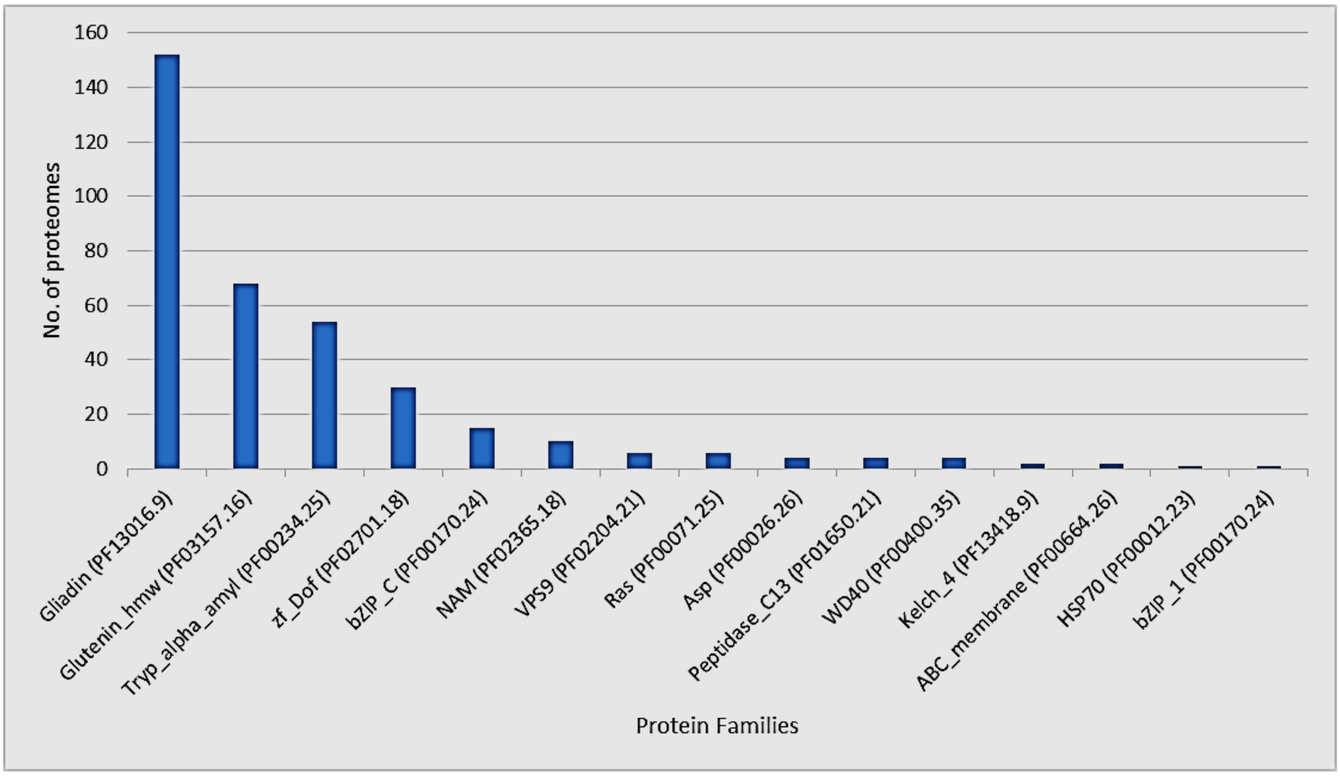


**Figure 3: Classification of SSPs into protein families**

Screened query sequences were subjected to protein family and gene ontology analysis. Pfam database were used for protein family analysis. Out of 412 query proteins maximum number of proteins were found in Gliadin and this amount to 152 and attained a top first position. Gliadins are gluten components that are necessary for bread to rise properly during baking. Gliadins are monomeric molecules in the cell, unlike other gluten proteins such as glutenins, which form extended networks of polymers due to disulfide linkages. Glutenin_hmw contained the second most number of proteins with 68 in number. This group of gluten proteins is supposed to be primarily responsible for the elastic characteristics of gluten and thus doughs. Glutenin high molecular weight subunits are considered as elastomeric proteins because the glutenin network can sustain considerable deformations without breaking and returns to its original shape when the stress is removed. With 54 number of proteins were found in Tryp_alpha_amyl it attains top third position in the list. This is the Plant lipid transfer protein (PLTP) or LTP family. These group of proteins are highly conserved domains ranging in size from 7-9kDa found in higher plant tissues. Lipid transfer proteins, as the name suggests, aid in the transport of phospholipids and other fatty acid groups across cell membranes. zf_ Dof contained the fourth most number of proteins with 30 in number. This family belongs to

Dof domain, zinc finger proteins. The Dof domain is a highly conserved DNA-binding domain found in plant transcription factors known as Dof (DNA-binding with one finger) proteins. The Dof domain contains 52 amino acids and is similar to the Cys2/Cys2 zinc finger DNA-binding domain of GATA1 and steroid hormone receptors, however it has a longer putative loop. At last the top fifth position was attained by bZIP_C family with 15 number of proteins. The bZIP (Basic Leucine zipper domain) domain is found in a wide range of DNA-binding eukaryotic proteins. One section of the domain contains a region that regulates sequence-specific DNA binding features, and also the Leucine zipper, which is essential for the dimerization of two DNA binding areas. The DNA binding region contains several basic amino acids such as arginine and lysine.

## 4.3 Phylogeny and Structure prediction of classified SSPs

As we discussed earlier that SSPs are classified into four classes but the major classes are albumin, prolamin, and globulin. After annotation studies of SSPs in wheat and rice now we have to predict the structure and homology of further classified SSPs. We retrieved the reviewed sequences of classified SSPs from UniProt. Selected sequences was using as a target sequence. Now BLASTp was performed taking the target sequence as a query sequence against databases like nr (Non-reductant protein sequences) and PDB (Protein data bank proteins). Selected the templates on the basis of Percent identity (>60%). Downloaded the template sequence in the Fasta format. The results are discussed below in the form of table:

| **Protein** | **Organism** | **Fasta sequence** | **Database** | **BLASTp** | |
|---|---|---|---|---|---|
| | | | | **Without % identity** | **With % identity (>60%)** |
| **Albumin** | *Juglans regia* (English walnut) | >sp\|P93198\|2SS_JUGRE 2S seed storage albumin protein (Fragment) OS=Juglans regia OX=51240 PE=1 SV=1<br>AALLVALLFVANAAAFRTTITTMEID EDIDNPRRRGEGCREQIQRQQNLN HCQYYLRQQS<br>RSGGYDEDNQRQHFRQCCQQLSQ MDEQCQCEGLRQVVRRQQQQQGL RGEEMEEMVQSARD<br>LPNECGISSQRCEIRRSWF | Non reductant protein sequences(nr) | 100 sequences<br>alb.(nr).txt | 7 sequences<br>alb.(nr.seq.).txt |
| | | | Protein data bank proteins (Pdb) | 4 sequences<br>alb.(pdb).txt | No similarity found. |

<table>
<tr><td rowspan="2"><b>Prolamin</b></td><td rowspan="2"><i>Oryza sativa</i><br>subsp. japonica<br>(Rice)</td><td rowspan="2">>sp|Q0DJ45|PRO7_ORYSJ<br>Prolamin PPROL 14E OS=Oryza sativa subsp. japonica OX=39947 GN=PROLM7 PE=2 SV=1<br>MKIIFVFALLAIAACSASAQFDVLGQ<br>SYRQYQLQSPVLLQQQVLSPYNEFV<br>RQQYGIAAS<br>PFLQSAAFQLRNNQVWQQLALVA<br>QQSHYQDINIVQAIAQQLQLQQFG<br>DLYFDRNLAQAQA<br>LLAFNVPSRYGIYPRYYGAPSTITTLG<br>GVL</td><td>Non reductant protein sequences(nr)</td><td>82 sequences<br>📄<br>prol.(nr).txt</td><td>53 sequences<br>📄<br>prol.(nr.seq.).txt</td></tr>
<tr><td>Protein data bank proteins (Pdb)</td><td>No similarity found.</td><td>No similarity found.</td></tr>
<tr><td rowspan="2"><b>Globulin</b></td><td rowspan="2"><i>Juglans regia</i><br>(English walnut)</td><td rowspan="2">>sp|Q2TPW5|JUGR4_JUGRE 11S globulin seed storage protein Jug r 4 OS=Juglans regia OX=51240 PE=1 SV=1<br>MAKPILLSIYLFLIVALFNGCLAQSGG<br>RQQQQFGQCQLNRLDALEPTNRIE<br>AEAGVIESW<br>DPNNQQFQCAGVAVVRRTIEPNGL<br>LLPQYSNAPQLVYIARGRGITGVLFP<br>GCPETFEESQ<br>RQSQQGQSREFQQDRHQKIRHFRE<br>GDIIAFPAGVAHWSYNDGSNPVVAI<br>SLLDTNNNANQ<br>LDQNPRNFYLAGNPDDEFRPQGQQ<br>EYEQHRRQQQRQQRPGEHGQQQR<br>GLGNNVFSGFDAD<br>FLADAFNVDTETARRLQSENDHRRS<br>IVRVEGRQLQVIRPRWSREEQEREE<br>RKERERERES<br>ESERRQSRRGGRDDNGLEETICTLRL<br>RENIGDPSRADIYTEEAGRISTVNSH<br>TLPVLRWL</td><td>Non reductant protein sequences(nr)</td><td>100 sequences<br>📄<br>glob.(nr).txt</td><td>19 sequences s</td></tr>
<tr><td>Protein data bank proteins (Pdb)</td><td>21 sequences<br>📄<br>glob(pdb).txt</td><td>No similarity found.</td></tr>
</table>

## 4.3.1 Phylogenetic analysis of classified SSPs

In the above table, protein blast results were discussed. Classified SSPs template sequences with protein blast at >60% percent identity was selected for the prediction of homology. The template sequences in the Fasta format were used to analyse phylogeny with maximum likelihood method at MEGA X Software. A phylogenetic tree, also known as a phylogeny, is a diagram that shows the lineages of various species, creatures, or genes from a common ancestor. Phylogenies are important for organising information about biological diversity, structuring classifications, and providing insight into evolutionary events (David Baum, 2008).

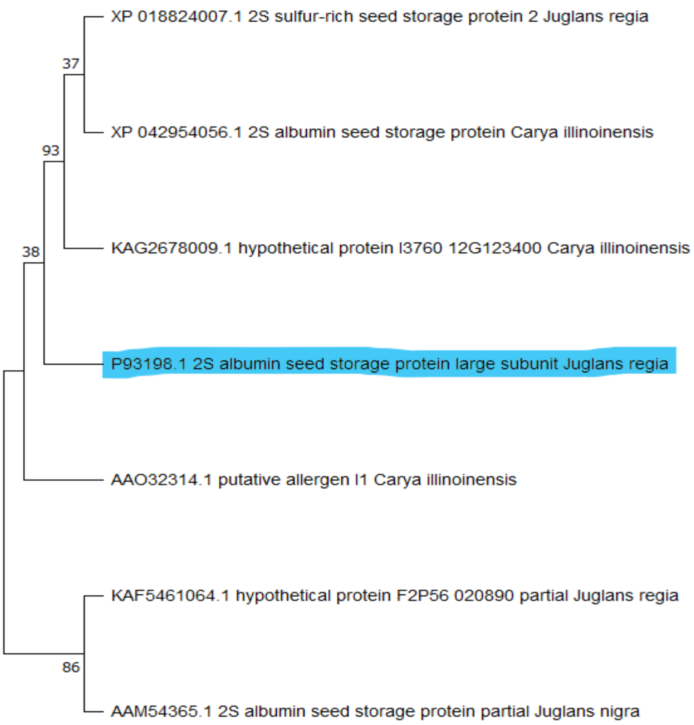


**Figure 4: Phylogenetic tree of albumin protein in *Juglans regia* (English walnut)**

The maximum likelihood technique with a Bootstrap value of 100 was used to infer the evolutionary history. This phylogenetic analysis involved 7 amino acid sequences. There were a total of 162 position in the final dataset. If we choose 100 bootstrap value in a phylogenetic tree it indicates that out of 100, how many times the same branch is observed when repeating the generation of a phylogenetic tree on a resampled set of data. In the figure observed that our query protein shown in blue colour have 93 times repeating in the generation of a phylogenetic tree. So this a good and well accepted bootstrap support.

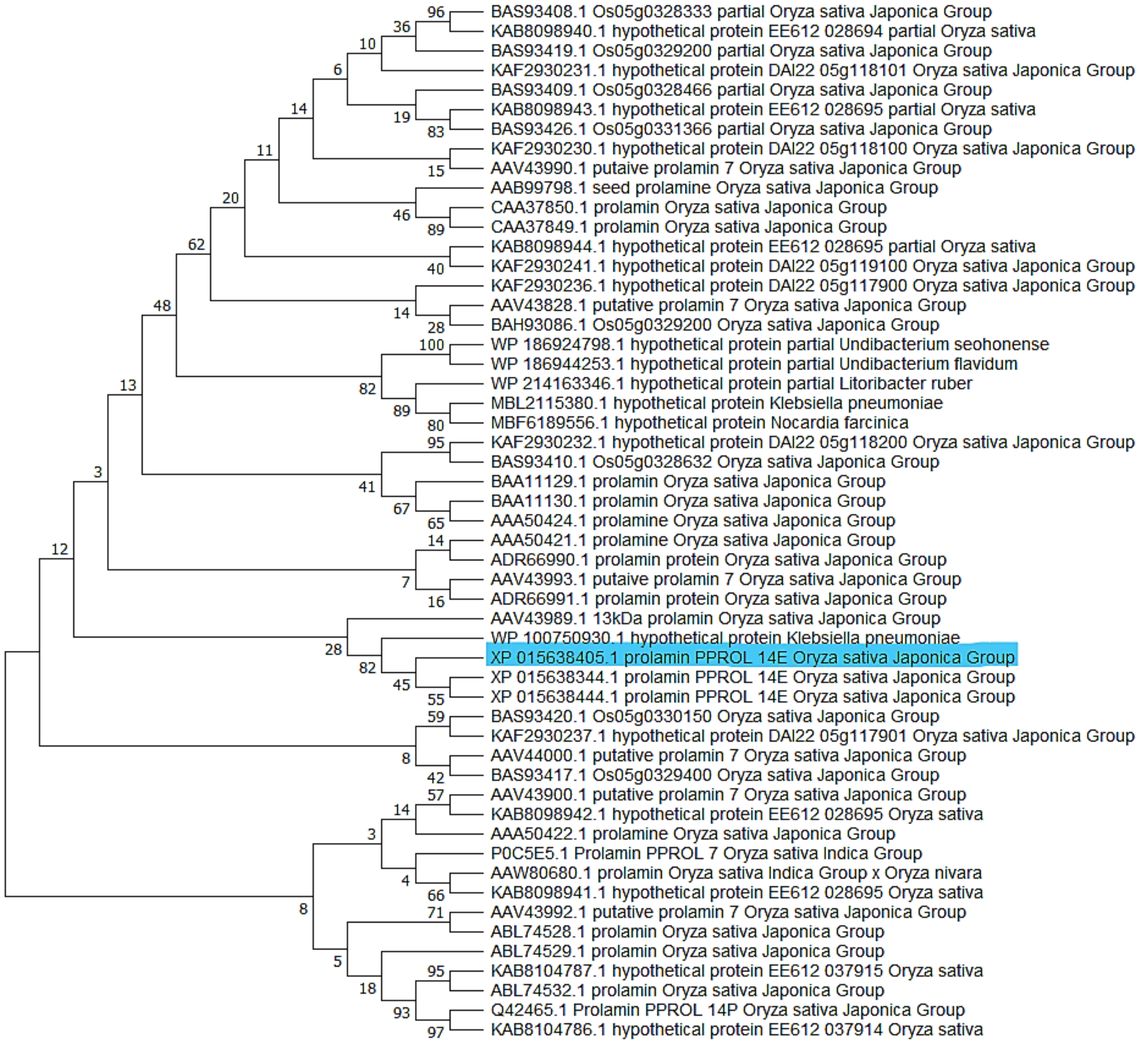


**Figure 5: Phylogenetic tree of prolamin protein in *Oryza sativa* subsp. japonica (Rice)**

This phylogenetic analysis involved 53 amino acid sequences. There were a total of 153 positions in the final dataset. In this figure our query protein shown in blue colour were extracted from UniProt database. After protein blast of this query protein at >60% percent identity 53 template sequences were found. With this template sequences we make phylogenetic tree with 100 bootstrap value. In this our highlighted protein is a common ancestor and all the proteins were linked to this query protein. With bootstap values we were found that how many times the branch is repeating. More the times of branch is repeating more will be support.

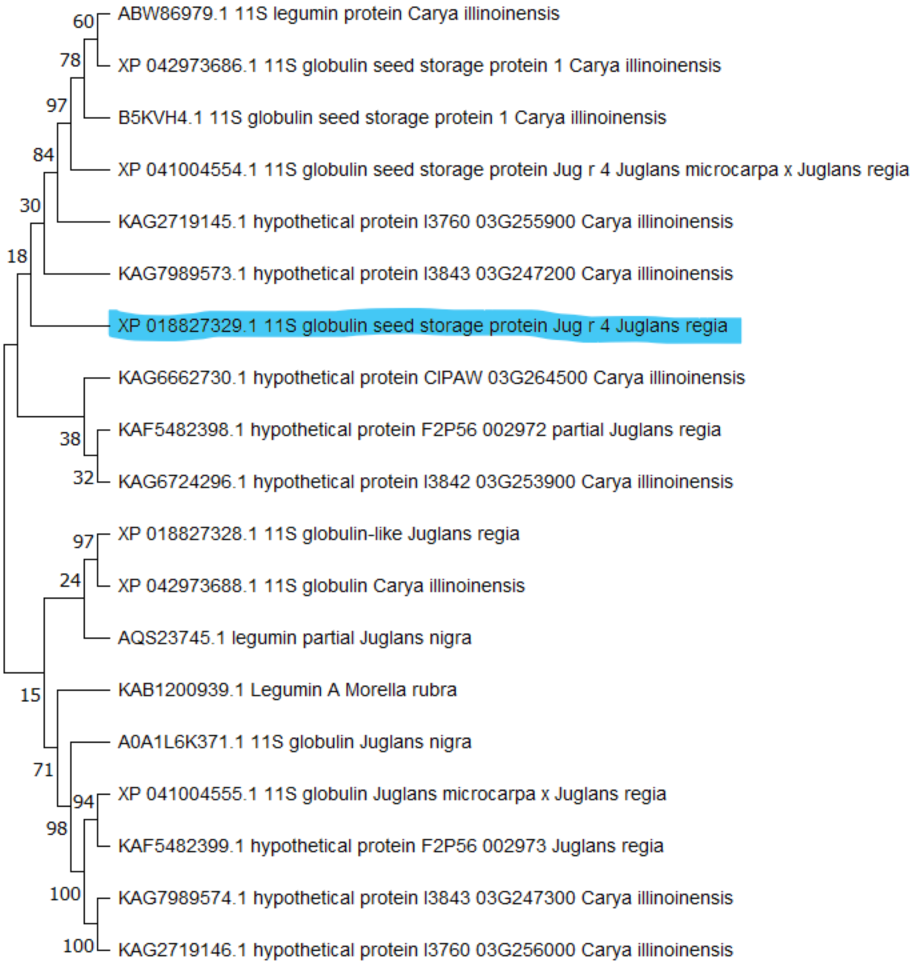


**Figure 6: Phylogenetic tree of globulin protein in *Juglans regia* (English walnut)**

This phylogenetic analysis involved 19 amino acid sequences. There were a total of 551 position in the final dataset. Similarly in this case 11s globuin seed storage protein Jug r 4 (*Juglans regia*) is our ancestor and all other proteins were linked to these protein with some common features, creatures, etc. In this figure we have found that some branches are repeating 100 times so this support very good structure.

### 4.3.2. Structure modelling of classified SSPs

As discussed in the above table only albumin and globulin seed storage protein have template sequences in the Fasta format when BLASTp was performed. The template sequences for structure prediction showed more than 90 % identity to their orthologs were considered structure modelling. These proteins were modelled using Swiss model server. Templates were searched against the PDB (Protein Data Bank).

#### a) <u>6S3F</u>

6s3f.1.B (Moringa seed protein Mo-CBP34) template was used for modelling of 2s albumin protein. This showed coverage of 0.17% and sequence identity of 41.67%. Structure quality parameters with 0.07 GMQE scores, QMEAN of 0.71, Z-score lied close to 0 and Local quality Estimate scores more than 0.6 confirmed the quality of the modelled structure. The reported resolution of this entry is 1.68Å and the Rfree value is 0.231.

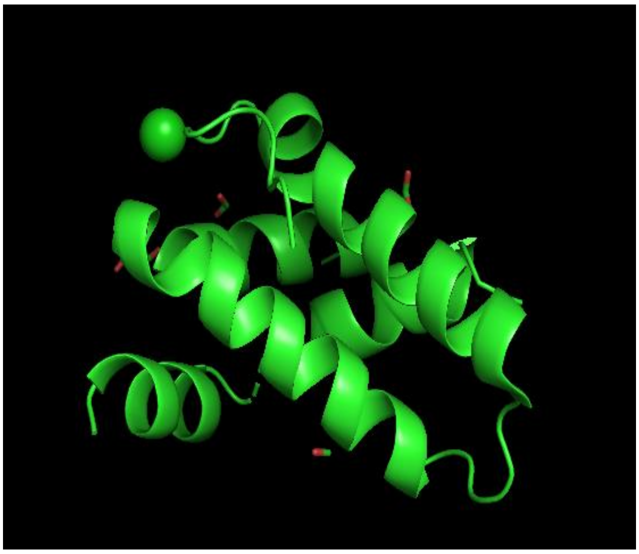


**Figure 7: Structure modelling of 6S3F**

#### b) <u>3QAC</u>

3qac.1.A (Structure of amaranth 11S globulin seed storage protein from *Amaranthus hypochondriacus* L.) template was used for modelling of 11s globulin protein. This shows sequence identity of 55.77% and query coverage 0.82%. Structure quality parameters with 0.62 GMQE scores, QMEAN of 0.69, Z-score lied close to 0 and Local quality Estimate scores more than 0.6 confirmed the quality of the modelled structure. The reported resolution of this entry is 2.27Å and the Rfree value is 0.249.

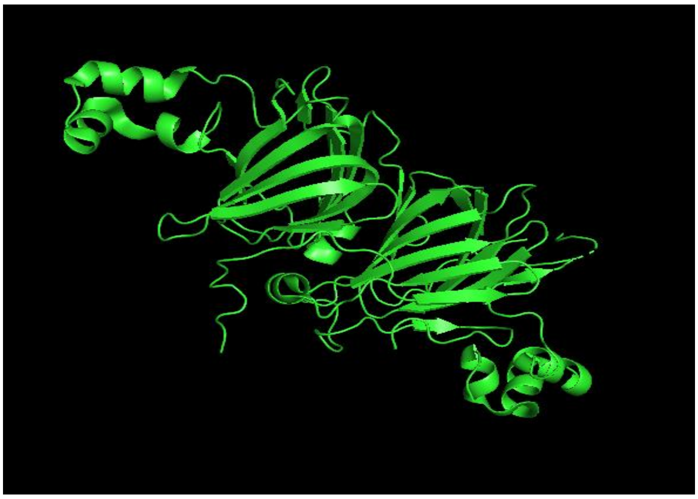


**Figure 8: Structure modelling of 3QAC**

# 5. DISCUSSION

Seed storage proteins (SSPs) are proteins that accumulate significantly in the developing seed. These proteins get translated at faster rate during seed germination and used as a major source of reduced nitrogen for the growing seedlings. The main function of SSPs is to act as a nitrogen, carbon, and sulfur reserve source. There is no mention of any enzymatic activity linked with the seed storage proteins. These proteins are generally not found in non-seed organs. It can accumulate within membrane-bound organelles called protein bodies. They are synthesized at high levels in specific tissues and various developmental stages. Based on Osborne solubility and extraction SSPs classified into four types: Albumin (Mostly distributed in dicot seeds and having sedimentation coefficient S.20.w. examples oil seed rape, Arabidopsis), Prolamin (Major endosperm storage proteins of all cereal grains (except rice and oat) and readily soluble in alcohol/water mixtures), Globulin (These are present in the endosperm's embryo and outer aleurone layer. It is easily soluble in a dilute salt solution), Glutelins (Prolamin-like proteins in certain grass seeds example Wheat. These are soluble in dilute acids or bases. It imparts baking quality to wheat). During this study, various computational tools were used to identify and analyze SSPs in sequenced cereal genomes. The protein sequences of cereal grains were downloaded from the UniProt database. Reviewed protein sequences were selected as a query sequence then BLASTp were performed against non-redundant protein database of NCBI database nr. The protein sequences that matched the entries with higher identity and lower e-value were chosen as they were more likely to be cross-reactive. These predicted protein sequences were annotated using Pfam and Gene Ontology (GO) database. Pfam database were used for protein family analysis and GO database providing their molecular function, biological process and cellular component. By annotation with Pfam maximum number of proteins were found in Gliadin family as comparison with other families in both the cereal genomes i.e., wheat and rice. Gliadins are gluten components that are necessary for bread to rise properly during baking. Gliadins are monomeric molecules in the cell, unlike other gluten proteins such as glutelins, which form extended networks of polymers due to disulfide linkages. Similarly, annotation with Gene ontology database nutrient reservoir activity mostly seen in all the protein sequences of rice and wheat. But if we compare the cereal genomes of rice and wheat, wheat (*Triticum aestivum*) protein sequences were shown most of this activity while rice shown least. Nutrient reservoir activity functions in the storage of nourishing substrates. For homology and structure prediction further classified SSPs were used. As discussed in above similarly reviewed protein sequences of SSPs were downloaded by using UniProt database. Selected sequences were used as a query sequences. The protein blast of the target protein was performed to select a template protein based on query coverage and the percent identity. An appropriate template protein was selected by looking at its structure resolution and R-factor. Selected template protein sequences were used against query sequence to analyse phylogeny. The phylogenetic analysis of protein sequences were run at 100 bootstrap value. Bootstrap value tell us support of the branches of phylogenetic tree. If the bootstrap value is 95% means the branch support is good and acceptable or if the bootstrap value is <50% then they are not acceptable. The models of the selected sequences were built using the Modeller.

# 6. CONCLUSION

In the present study genome wide identification and comparative analysis of SSPs in sequenced cereal genomes like rice and wheat were done by using computational tools. Identification of SSPs were done by using UniProt database. Annotation of SSPs in rice and wheat by using Pfam and Gene Ontology database after downloading the template sequence by protein blast at >60% percent identity. Gliadin family and Nutrient reservoir activity was taken top most position in the list of the protein sequences. The protein sequences with the activities mentioned above, have mainly function in the storage of nourishing substrates. After this we have checked the evolutionary relationship among the identified SSPs by phylogeny analysis by MEGA X software. To get structural insights, 3d-molecular model of albumin and globulin were generated by SWISS-MODEL Program.

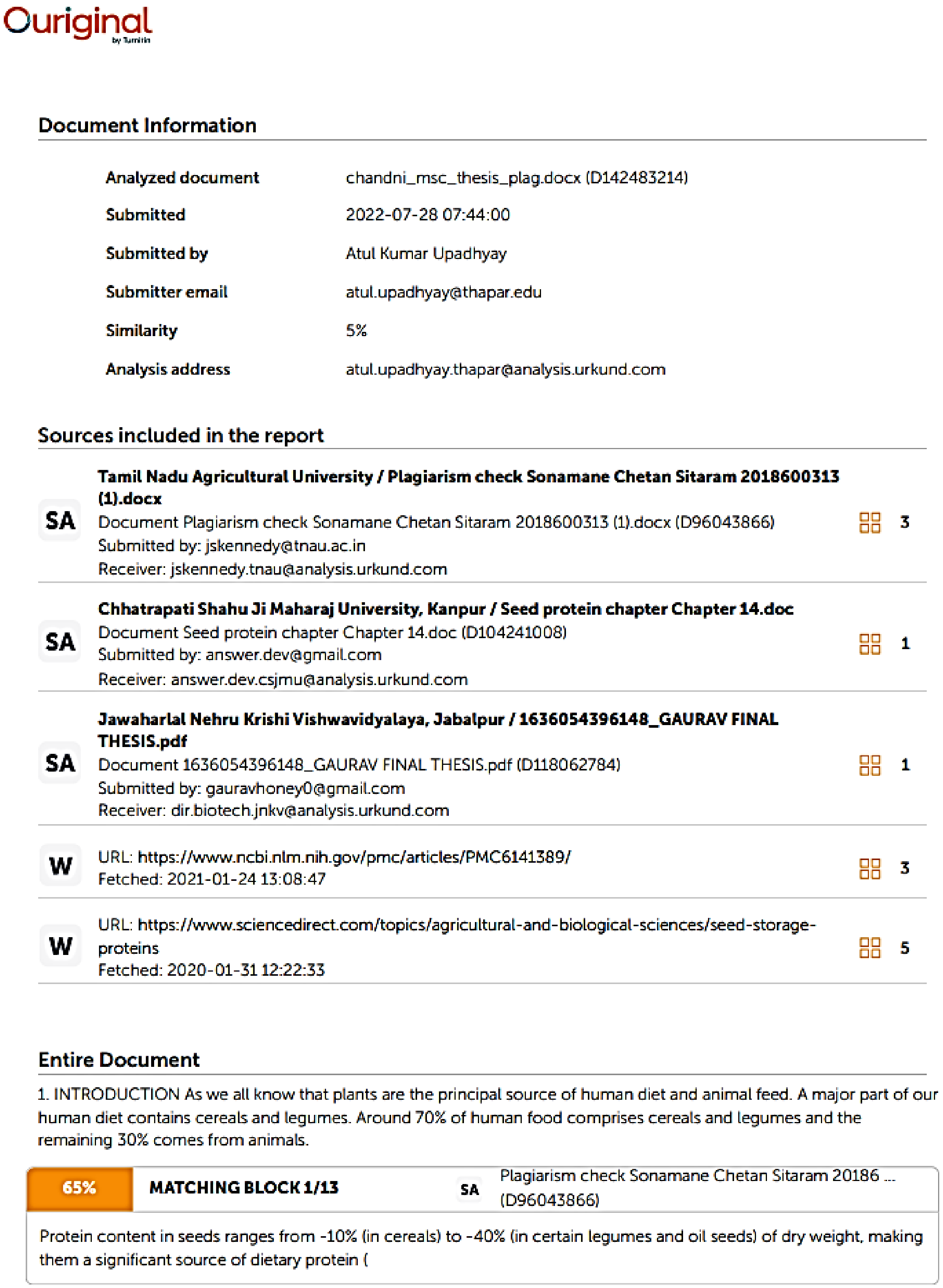
# 7. References

1. Baum, D. (2008). Reading a phylogenetic tree: the meaning of monophyletic groups. Nature Education, 1(1), 190.
2. Bocsanczy, A. M., Espindola, A. S., & Norman, D. J. (2019). Whole-genome sequences of Ralstonia solanacearum strains P816, P822, and P824, emerging pathogens of blueberry in Florida. Microbiology Resource Announcements, 8(3), e01316-18.
3. Casey, R. (1999). Distribution and some properties of seed globulins. Seed proteins, 159-169.
4. Chen, K., Wang, L., Chen, H., Zhang, C., Wang, S., Chu, P., Li, S., Fu, H., Sun, T., Liu, M., Yang, Q., Zou, H., & Zhuang, W. (2021). Complete genome sequence analysis of the peanut pathogen Ralstonia solanacearum strain Rs-P. 362200. BMC microbiology, 21(1), 1-15.
5. Field, J. M., Shewry, P. R., & Miflin, B. J. (1983). Solubilisation and characterisation of wheat gluten proteins: correlations between the amount of aggregated proteins and baking quality. Journal of the Science of Food and Agriculture, 34(4), 370-377.
6. Foley, R. C., Gao, L. L., Spriggs, A., Soo, L. Y., Goggin, D. E., Smith, P., Atkins, C. A., & Singh, K. B. (2011). Identification and characterisation of seed storage protein transcripts from Lupinus angustifolius. BMC Plant Biology, 11(1), 1-15.
7. Fujiwara, T., Nambara, E., Yamagishi, K., Goto, D. B., & Naito, S. (2002). Storage proteins. The Arabidopsis Book/American Society of Plant Biologists, 1.
8. Gaur, V. S., Sood, S., Tiwari, S., & Kumar, A. (2018). Genome-wide identification and characterization of seed storage proteins (SSPs) of foxtail millet (Setaria italica (L.) P. Beauv.). 3 Biotech, 8(10), 1-16.
9. Ge, W., Gao, Y., Xu, S., Ma, X., Wang, H., Kong, L., & Sun, S. (2021). Genome-wide identification, characteristics and expression of the prolamin genes in Thinopyrum elongatum. BMC genomics, 22(1), 1-12.
10. Kawakatsu, T., & Takaiwa, F. (2010). Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains. Plant biotechnology journal, 8(9), 939-953.

11. Kermode, A. R., & Bewley, J. D. (1999). Synthesis, processing and deposition of seed proteins: the pathway of protein synthesis and deposition in the cell. In Seed proteins (pp. 807-841). Springer, Dordrecht.
12. Kriz, A. L. (1989). Characterization of embryo globulins encoded by the maizeGlb genes. Biochemical genetics, 27(3), 239-251.
13. Kriz, A. L. (1999). 7S globulins of cereals. In Seed proteins (pp. 477-498). Springer, Dordrecht.
14. Luo, G., Shen, L., Zhao, S., Li, R., Song, Y., Song, S., Yu, K., Yang, W., Li, X., Sun, J., Wang, Y., Gao, C., Liu, D., & Zhang, A. (2021). Genome-wide identification of seed storage protein gene regulators in wheat through coexpression analysis. The Plant Journal, 108(6), 1704-1720.
15. Mandal, S., & Mandal, R. K. (2000). Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. Current Science, 576-589.
16. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., Tosatto, S. C., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. Nucleic acids research, 49(D1), D412-D419.
17. Moore, K. L., Tosi, P., Palmer, R., Hawkesford, M. J., Grovenor, C. R., & Shewry, P. R. (2016). The dynamics of protein body formation in developing wheat grain. Plant biotechnology journal, 14(9), 1876-1882.
18. Osborne, T. B. (1924). The vegetable proteins. Longmans, Green and Company.
19. Paris, R., Petruzzino, G., Savino, M., De Simone, V., Ficco, D. B., & Trono, D. (2021). Genome-Wide Identification, Characterization and Expression Pattern Analysis of the γ-Gliadin Gene Family in the Durum Wheat (Triticum durum Desf.) Cultivar Svevo. Genes, 12(11), 1743.
20. Pedrazzini, E., Mainieri, D., Marrano, C. A., & Vitale, A. (2016). Where do protein bodies of cereal seeds come from? Frontiers in plant science, 7, 1139.
21. Prior, P., Ailloud, F., Dalsing, B. L., Remenant, B., Sanchez, B., & Allen, C. (2016). Genomic and proteomic evidence supporting the division of the plant pathogen Ralstonia solanacearum into three species. BMC genomics, 17(1), 1-11.

22. Radhika, V., & Rao, V. (2015). Computational approaches for the classification of seed storage proteins. Journal of food science and technology, 52(7), 4246-4255.
23. Sharma, G., Upadhyay, A. K., Biradar, H., & Hittalmani, S. (2019). OsNAC-like transcription factor involved in regulating seed-storage protein content at different stages of grain filling in rice under aerobic conditions. Journal of genetics, 98(1), 1-9.
24. Shewry, P. R., & Casey, R. (1999). Seed proteins. In Seed proteins (pp. 1-10). Springer, Dordrecht.
25. Shewry, P. R., & Halford, N. G. (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. Journal of experimental botany, 53(370), 947-958.
26. Shewry, P. R., & Tatham, A. S. (1990). The prolamin storage proteins of cereal seeds: structure and evolution. Biochemical journal, 267(1), 1.
27. Shewry, P. R., Napier, J. A., & Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. The plant cell, 7(7), 945.
28. Shewry, P. R., Tatham, A. S., & Halford, N. G. (2001). Nutritional control of storage protein synthesis in developing grain of wheat and barley. Plant Growth Regulation, 34(1), 105-111.
29. Singh, N. K., & Shepherd, K. W. (1985). The structure and genetic control of a new class of disulphide-linked proteins in wheat endosperm. Theoretical and Applied Genetics, 71(1), 79-92.
30. Singh, N. K., Shepherd, K. W., Langridge, P., Clem Gruen, L., Skerritt, J. H., & Wrigley, C. W. (1988). Identification of legumin-like proteins in wheat. Plant molecular biology, 11(5), 633-639.
31. Souza, P. F. (2020). The forgotten 2S albumin proteins: Importance, structure, and biotechnological application in agriculture and human health. International Journal of Biological Macromolecules, 164, 4638-4649.
32. Takaiwa, F., Ogawa, M., & Okita, T. W. (1999). Rice glutelins. In Seed proteins (pp. 401-425). Springer, Dordrecht.
33. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. Molecular biology and evolution, 38(7), 3022-3027.
34. UniProt: the universal protein knowledgebase in 2021. Nucleic acids research 49, no. D1 (2021): D480-D489.

35. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., Beer, T. A., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research, 46(W1), W296-W303.
36. Y. Wakasa, F. Takaiwa, Seed Storage Proteins, Brenner's Encyclopedia of Genetics (Second Edition), Academic Press, 2013, Pages 346-348.



Ouriginal by Turnitin

## Document Information

| | |
|---|---|
| **Analyzed document** | chandni_msc_thesis_plag.docx (D142483214) |
| **Submitted** | 2022-07-28 07:44:00 |
| **Submitted by** | Atul Kumar Upadhyay |
| **Submitter email** | atul.upadhyay@thapar.edu |
| **Similarity** | 5% |
| **Analysis address** | atul.upadhyay.thapar@analysis.urkund.com |

## Sources included in the report

| | | | |
|---|---|---|---|
| SA | **Tamil Nadu Agricultural University / Plagiarism check Sonamane Chetan Sitaram 2018600313 (1).docx**<br>Document Plagiarism check Sonamane Chetan Sitaram 2018600313 (1).docx (D96043866)<br>Submitted by: jskennedy@tnau.ac.in<br>Receiver: jskennedy.tnau@analysis.urkund.com | 88 | 3 |
| SA | **Chhatrapati Shahu Ji Maharaj University, Kanpur / Seed protein chapter Chapter 14.doc**<br>Document Seed protein chapter Chapter 14.doc (D104241008)<br>Submitted by: answer.dev@gmail.com<br>Receiver: answer.dev.csjmu@analysis.urkund.com | 88 | 1 |
| SA | **Jawaharlal Nehru Krishi Vishwavidyalaya, Jabalpur / 1636054396148_GAURAV FINAL THESIS.pdf**<br>Document 1636054396148_GAURAV FINAL THESIS.pdf (D118062784)<br>Submitted by: gauravhoney0@gmail.com<br>Receiver: dir.biotech.jnkv@analysis.urkund.com | 88 | 1 |
| W | URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6141389/<br>Fetched: 2021-01-24 13:08:47 | 88 | 3 |
| W | URL: https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/seed-storage-proteins<br>Fetched: 2020-01-31 12:22:33 | 88 | 5 |

## Entire Document

1. INTRODUCTION As we all know that plants are the principal source of human diet and animal feed. A major part of our human diet contains cereals and legumes. Around 70% of human food comprises cereals and legumes and the remaining 30% comes from animals.

| 65% | **MATCHING BLOCK 1/13** | SA | Plagiarism check Sonamane Chetan Sitaram 20186 ... (D96043866) |
|---|---|---|---|

Protein content in seeds ranges from -10% (in cereals) to -40% (in certain legumes and oil seeds) of dry weight, making them a significant source of dietary protein (