Annotation of Plant Genome: A Case Study of Oryza sativa

Harpreet Kaur¹, Rajan Kesri² and Tanzeel Tufail Mir³
¹Research Scholar, Guru Arjan Dev Institute of Development Studies, 14-Preet Avenue, Majitha Road, PO Naushera, Amritsar-143008, INDIA
²Research Intern, Guru Arjan Dev Institute of Development Studies, 14-Preet Avenue, Majitha Road, PO Naushera, Amritsar-143008, INDIA
³Research Intern, Guru Arjan Dev Institute of Development Studies, 14-Preet Avenue, Majitha Road, PO Naushera, Amritsar-143008, INDIA
¹Corresponding Author: harpreetkaur1182@yahoo.com

ABSTRACT

Rice! A perennial claim crop of the world. Besides satisfying the eagerness of energy rice, has also been known to support the world's trade economy. Hence, being a crop of such crucial importance its examinational study at the genome level will serve in multiplying its production and quality to irrigate the burning crave of humanity. Likewise, the senescence gene of rice is responsible for its age duration. Hence, understanding its property at 360° will help us to modify or to alter its function in a positive portion.

Using Insilco analysis mode, the present study is an attempt to examine various characteristics conformation of senescence causing genes in rice. The two genes chosen were HCP and RR because the interaction between these two led to the onset of senescence in rice. Two genes that are HCP (Histidine-containing phosphotransfer protein 1) and RR (Two-component response regulator) are responsible for attaining the stage of senescence in rice. Understanding their molecular and structural property will be going to let us closer to perform successful adjustments. Moreover, their specific property is also responsible for their specific interaction which led to the generation of such signals that triggers senescence. Therefore, this analysis was aimed to understand the features of the two genes as well as their interaction by the means of computational technique.

Understanding the features, function and flow of gene will lead us to stabilized effective measures in order to get a beneficiary outcome while going for alteration in its characters. As the pure data for the structural conformation of the selected genes are not available so, we have at first, searched the most similar homolog of the query sequence and the search was based on similar sequence homology on the platform of the local alignment tool. And the further analysis was carried out on the base conformation of the most relevant homologs (structure/sequence) found.

We have analyzed the query gene sequence by various dry lab analysis tools to explore its structural and molecular features with the motive to contribute a little knowledge for the sake of further studies to delay senescence in rice plants in order to increase grain productivity.

Keywords- Senescence, Interaction, Signals, Plant Genome Annotation, Plant Ontologies, Plant Gene Family Data Bases, Genome Annotation Pipelines, Functional Annotation, Annotation Repetitive Sequence Functional Description.

I. INTRODUCTION

Annotation is the process of identifying and describing the regions of biological interest within a genome. The location and structure of protein-coding genes are the most common form of annotation, but other types of important sequence annotation include the identification of noncoding RNAs (tRNAs, rRNAs, snoRNAs, miRNAs, siRNAs), repetitive sequences such as transposable elements, and the location of genetic markers. Annotation of plant genomic sequences can be separated into the structural and functional annotation. Structural annotation is the foundation of all genomics as without accurate gene models understanding gene function or evolution of genes across taxa can be impeded. It is dependent on sensitive, specific computational programs and deep experimental evidence to identify gene features within genomic DNA.

Functional annotation describes the biological context of gene sequences. It is highly dependent on sequence similarity to other known genes or proteins as the majority of initial "first-pass" functional annotation on a genomic scale is transitive. Coupling structural and functional annotation across genomes in a comparative manner promotes more accurate annotation as well as an understanding of gene and genome evolution. With the increasing availability of plant genome sequence data, the value of comparative annotation will increase. As with any new field, methodologies are evolving for genome annotation and will improve in the future.

Almost all genome annotation is performed using semi-automated computational pipelines and is subject to some degree of interpretation and error. Therefore, researchers must understand the methods used to create an annotation in order to assess the quality of that annotation.

I. ABOUT ORYZA ATIVA JAPONICA

The majority of the world’s population depends on cereal crops as their primary source of carbohydrates. Among the cultivated cereal crops, rice makes up to 20 percent of the total caloric intake for the human population as a whole.
Oryza sativa Japonica (rice) is the staple food for 2.5 billion people. It is the grain with the second highest worldwide production after Zea mays. Oryza sativa (rice) is a monocotyledonous flowering plant of the family Poaceae and is one of the most important crop plants in the world, providing the principal food source for half of the world’s population. Oryza sativa subsp. japonica is one of three major subspecies of rice, the others being indica and javanica. Oryza sativa subsp. japonica is short-grained and high in amylopectin so that the grains stick together when cooked, which distinguishes it from subsp. indica which is long grained and not sticky. Oryza sativa subsp. japonica is grown in dry fields, mainly in temperate or colder climates.

In addition to its agronomic importance, rice is an important model species for monocot plants and cereals such as maize, wheat, barley and sorghum. O. sativa has a compact diploid genome of approximately 500 Mb (n=12) compared with the multi-gigabase genomes of maize, wheat and barley.

Oryza sativa has a haploid chromosome number of 12, containing 370 Mb with about 36,000 protein-coding genes. Rice was an obvious choice for the first whole genome sequencing of a cereal crop. It is the smallest of the major cereal crop genomes and is the easiest to transform genetically. The cultivar sequenced from the japonica subspecies was Nipponbare (www.thericejournal.com/content/6/1/4/abstract).

Since 2002 when genome assemblies of the two major rice varieties (Oryza sativa L. ssp. Indica 93-11 and Oryza sativa L. ssp. japonica Nipponbare) were published (1,2), efforts to construct better rice reference genomes continue even to this date (3,4). Comparative analyses on rice and other large plant genomes have been promoting the application of genomic research activities to agricultural practice, such as marker-assisted breeding for the improvement of biotic and abiotic stress resistance (5,6). Although there have been a number of databases or web servers constructed for rice and related plant genomes (7–9), a comprehensive database or knowledgebase for general rice genomic information is still necessary, especially when data are still being generated in a fast rate for this much treasured crop.

With the completion of the rice genome (Oryza sativa L. ssp. japonica cultivar Nipponbare) by the international consortium on rice genome sequencing (International Rice Genome Sequencing Project 2005), it has become possible to elucidate the layers of information encoded by the sequence. Analyses of rice full-length cDNAs and the rice proteome are in progress (Kikuchi et al. 2003; Komatsu et al. 2004; Komatsu and Tanaka 2005). Additionally, the construction of integrative annotations for the rice genome, transcriptome, and proteome are being undertaken. In order to standardize the annotation of the genome data for Nipponbare, we organized an international consortium for rice genome annotation, the Rice Annotation Project (RAP), with the aim of allowing more efficient analysis of genomic information and accelerating post-sequencing research activities. It is also hoped that the annotation will provide a comparative data resource for cereal genomics researchers working on other species and contribute to their endeavors.

To cope with the enormous amount of information produced by large-scale sequencing, several automated annotation methods have been developed for the purpose of efficient data processing. However, it is acknowledged that automated annotation alone tends to result in a high proportion of erroneous annotations, and therefore annotation data results should be carefully curated by experts before any public release in order to cut down on the amount of these erroneous annotations. Currently, manual curation remains a necessary process for developing an accurate biological database (Misra et al. 2002; Camon et al. 2003). With this in mind, we brought together a large group of specialists to curate the results of our automated rice gene functional assignment. By bringing individuals from complementary disciplines together, the amount of time required to perform the manual curation was significantly reduced.

There are a large number of full-length cDNAs and expressed sequence tags (ESTs) available for rice and other cereals (Fernandes et al. 2002; Wu et al. 2002; Kiuchi et al. 2003; Gardiner et al. 2004; Lai et al. 2004; Zhang et al. 2004; Jantasuriyarat et al. 2005). This wealth of information is a boon for genome annotation because it provides excellent support for transcribed regions, which, in turn, allows more precise predictions than current ab initio methods can provide. By using the annotation data set based on our curation and mapping of cDNAs to the genome, we were able to approximate the number of genes in the rice genome, classify transcribed sequences by probable function, and identify other features pertinent to the rice genome.

Arabidopsis thaliana is one of the most well-studied model organisms. A comparison of rice with the dicotyledon may assist in developing a greater understanding of intrinsic mechanisms among cereals at the molecular level. The use of knowledge accumulated about A. thaliana genes to quantify their counterparts in rice is one example of such a comparative study (Izawa et al. 2003; Yamaguchi et al. 2006). Additionally, clues as to the evolution of these two flowering plants could be obtained. Here we describe a comparative analysis of the genomes and protein sequences of O. sativa and A. thaliana on the basis of manually curated data. This analysis focuses on the number of genes, the composition of functional domains, and patterns of gene duplication.
II. REVIEW OF LITERATURE

Achieving a robust structural and functional genome sequence annotation is essential to provide the foundation for further relevant biological studies. Genome annotation consists of identifying and attaching biological information to sequence features. It represents one of the most difficult tasks in genome sequencing projects (Elsik et al., 2006), particularly today where the advent of high-throughput next generation sequencing (NGS) technologies enables genome sequences to be produced at a high pace. The reality at present is that new genomes are being sequenced at a faster rate than they are being fully and correctly annotated (Cantarel et al., 2008). It took about 7 years and a large community effort to sequence and fully annotate the Arabidopsis thaliana (The Arabidopsis Genome Initiative, 2000) and rice genomes (International Rice Genome Sequencing Project, 2005) at a quality that none of the other genome sequenced after has reached yet. In the past 5 years, the production of plant genome sequences has grown exponentially (for a review see Feuillet et al., 2011). In August 2011, the NCBI Entrez Genome Project web site listed 135 land plant genome sequencing projects including 36 completed or assembled genomes and 101 in progress. Out of the 36 sequenced genomes, 23 have been released in the past 2 years. Among those, only two genomes larger than 1 Gb, maize (Schnable et al., 2009) and soybean (Schmutz et al., 2010), have been sequenced and annotated.

Rice is the progeny of grass family Oryzasesativa commonly known as Asian rice. It is counted under widely preferred energy giving cereal crop, at about 758.8 million tonnes (503.6 million tonnes, milled basis), world paddy production in 2017[1]. And as per FAO now sees India producing 165.5 million tonnes (110.4 million tonnes, milled basis) in 2017, which is about 1 percent more than the 2016 all-time high [1].

As we know rice is a distance progeny of grass family and hence its portion is mainly composed of paddy and leaf, which is being pigmented by chlorophyll completely all over. As we know photosynthesis utilizes water (H₂O), carbon-di-oxide (CO₂), oxygen (O₂) and minerals from soil to convert energy from sunlight into eatable organic components grain. Since the requirement of water and carbon-di-oxide is more in rice as compare to is morphology, which produce a complex carbohydrate cereal (rice). Plant had gone through years of acclimatization by nature which makes them evolve in present days' crop, which are capable enough in energy conversion until they live. Hence, the production of food is limited. Limited! As crop have a specific age of fruiting and limited time to carry out photosynthesis as their genome was pre-programed, while going through a long course of evolution.

In case of rice plant numerous factors induce the response of light toward leaf photosynthesis. The first factor taken into account is elevated leaf temperatures which come up with metabolic imbalances (Pastenes and Horton, 1996b), which have an injurious effect on thylakoid function (Pastenes and Horton, 1996a), photoinhibition gets heightened (Fuse et al., 1993), and raise in photorespiration (Leegood and Edwards, 1996). The second factor to consider is the leaf angle. It has been detected as provoking the upper leaves’ level of light saturation. (Yoshida, 1981b).

The production of rice is also being affected by leaf senescence. The onset of Leaf senescence in rice arouses from the low bottom leaves and moves upward as the growth and development in plants take place. The decline of assimilatory efficiency as leaf senescence goes on investing to limited grain yield (Nooden, 1988b), and delaying the leaf senescence process may tend to lift up the crop productivity. Leaf senescence process that finally leads to organ demise is an endogenously controlled degenerative process. Moreover, the process of age-dependent manner is also being affected by the complex interaction of developmental age relevant factors which are categorized as internal and external factors (Nooden, 1988a; smart, 1994). At the outset of leaf senescence, the activity of photosynthesis decreases more quickly than chlorophyll p; this outcome in lowering of leaf position, which is directly proportional to lowering of the rate of oxygen which was determined by evolution detected on the required content of chlorophyll. The content of the Chlorophyll a/b ratio also falls off with rising senescence. The composition of the antenna system of photosynthesis is also being affected by Senescence. Moreover, it has been evidenced continuously that chlorophyll a is more likely to vanish sooner than chlorophyll b during senescence, and this led to reducing in the chlorophyll a/b ratio (Youn and Ota 1973, Patterson and Moss 1979, Jenkins et al. 1981, Maunder and Brown 1983). Oxygen evolution rates were recognized on the foundation of chlorophyll a, latch to the reaction centre complexes, which stay stable throughout the procedure of senescence. Hence, the shutting down of the process of photosynthesis is firmly in relation to the loss of the reaction centre complexes at the time of leaf senescence of rice seedlings. The inference shows that a decrease in photosynthesis is results of the loss of a functional unit of photosynthesis or by a reduction in the amount of whole chloroplasts. [Relationship between Photosynthesis and Chlorophyll Content during Leaf Senescence of Rice Seedlings: Mariko Kura-Hotta, Kazuhiko Satoh and Sakae Katoh]

It was also detected in the breeding attempt of the last 50 years, in order to delay leaf senescence and stretching the period of dynamic photosynthesis has come up with the rise in the sudden photo assimilation of a raw substrate and led to enhance in the overall grain mass production (Richards 2000; Long et al. 2006).

Rice yield takes place by the deposition grain in cereal, which counts upon the sources of carbon (CO₂) and nitrogen and this is carried out by photo by a photosynthetic dynamic leaf, and those remobilized from
the vegetative tissues (Yang and Zhang 2006). In small-grained cereals such as rice, 10%–40% of the output grain weight was induced by pre-anthesis photo assimilation (Gebbing and Schnyder 1999; Yang and Zhang 2006). But Leaf senescence can turn over the final output of grain in either leading to increment in mass or decrement in mass. Study over modeling work in recent past by tracking nine years of the satellite track record of wheat growth in northern India to control and monitor the rates of wheat senescence by the dissector of it to temperatures higher than 34 °C, shows a continuous acceleration of senescence under high heat climate (Lobell et al. 2012). Thus, acclimating crops under extremal high temperatures will be the requirement of upcoming studies of crop success. It has been evidenced in various stay-green varieties showing delayed leaf senescence results in accruing a number of positive effects, which involves initiating extra root growth, providing extra carbon, and decreasing the time period between anthesis and silking, as reviewed by Davies et al. (2011). Hence, the stages of the arrival of leaf senescence play a major role in crop yield.

There exists a composite kinship between the arrival of leaf senescence and utilization of nitrogen by a plant (Chardon et al. 2010; Masclaux-Daubresse et al. 2010; Masclaux-Daubresse and Chardon 2011). Nevertheless, uncontrolled and non-monitored delay in senescence can cause low output in grain filling. Result of holding up of leaf senescence on grain protein concentration and on grain mass output count upon the presence of nitrogen at the stage of the post-anthesis period (Bogard et al. 2011). Thus, the variation in post-anthesis leaf senescence should be carried out by keeping an eagle eye on genetic and management control.

The two-component His kinases (HK) or two-component system (TCS) is known to play important roles in the regulation of prokaryotic and lower eukaryotic cellular responses to environmental stimuli (Grebe and Stock, 1999; Stock et al., 2000; Urao et al., 2000a, 2000b; Besant et al., 2003). These include antibiotic resistance, chemotaxis, differentiation, and nitrogen metabolism. The existence of a bacterial-type HK in plants was first reported by Chang et al. (1993). Since then, many plants have been documented to possess genes encoding two-component regulators, and their participation in the perception and integration of various extracellular and intracellular signals has been shown (Lohrmann and Harter, 2002; Oka et al., 2002; Grefen and Harter, 2004; Hass et al., 2004; for review, see Stock et al., 2000; Foussard et al., 2001).

In bacterial systems, the typical machinery, such as that of Escherichia coli Env Z protein (an osmosensor), consists of a membrane-localized HK that senses the input signal and a response regulator (RR) that contains a conserved regulatory domain. In this system, the signaling is initiated when the HK, modulated by the environmental stimulus, autophosphorylates the conserved His residue. The phosphoryl group is then transferred to a conserved Asp residue in the RR, which results in modulation of its activity (Fig. 1A).

Autophosphorylation of the HK is a bimolecular reaction between homodimers in which one HK monomer catalyzes the phosphorylation of the His residue in the second monomer (Pan et al., 1993; Surette et al., 1996). Besides directing the forward phosphorylation reaction, many HKs possess a phosphatase activity, enabling them to dephosphorylate their cognate partner (Keener and Kustu, 1988; Lois et al., 1993). The phosphotransfer pathways, which need to be shut down quickly, operate with such bifunctional HKs. More elaborate HKs, also known as hybrid kinases, are typical of eukaryotes and a few prokaryotes and contain multiple phosphodonor and phosphoacceptor sites. Signaling via these pathways uses multiple His phosphotransfers (Hpt) and receiver domains (RDs) or proteins that connect to final RR or other signaling outputs (Fig. 1B). Examples of such HKs include the SLN1/YPD1/SSK1 phosphorelay in yeast (Saccharomyces cerevisiae), which controls the HOG pathways in response to osmotic stress (Wurgler-Murphy and Saito, 1997). Such a multistep phosphorelay reaction is believed to have the advantage of providing multiple regulatory checkpoints for a signal cross talk or negative regulation by certain phosphatases. The existence of multiple phosphorelay reactions in both prokaryotes and eukaryotes suggests that similar mechanisms might be more widely used (Urao et al., 2000a).

### III. METHODS AND METHODOLOGY

#### 1) Selection of Sample Sequence

As we are curious in the processes of senescence in Oryza sativa so, the two query sequence belongs to a protein Histidine-containing phosphotransfer (HSP) which on interaction with the Response Regulator (RR) carry out the process of senescence. The analysis was carried on these two protein sequences to understand their characters. The two sequences are as:

**HSP:**

*spjQ6VAK3*OHP1ORYSJ Histidine-containing phosphotransfer protein 1 OS=Oryza sativa subsp. japonica GN=AHP1 PE=2 SV=1 MAAAALTSNLALVNMMFAMGLLLDDQFQLQMLQDSTAPDFVEVTLYFCDDGERICLRSQLEKPNVFDFDRVSYVHQLKGSASSVGAQKVNTCQFREFCQQRSRDGCLKTLDDLVRTEFYSDLNRNKFQAMLQLEQIQACYPKHI

**RR**

*spjQ6K8X6*ORR23ORYSJ Two-component response regulator ORR23 OS=Oryza sativa subsp. japonica OX=39947 GN=RR23 PE=2 SV=1 MRAAEEKRGVVPAAARRRDQFVGPVRMLAVDADDVPCLVKLLETTLRLRCQHYVHTTNNQAAIALKLMRNRDMFDLVISDVHMPPDMGFKLLLEVGLEMDLVPIMLSVNGETKVLGIGTHACDLYKKPVIEELRNWQHVIRKFKSTRDRLDFYEECNKPPNADSVHVGHHTVCSPDGQRSKPKREYCEESDEEVEQNTQDI
Global protein function prediction was performed in this study through the use of UniProt: the universal protein database, which contains information on protein sequences and their functional annotations. To identify homologs for our query sequence, we performed a homology search using the Blast algorithm against the NCBI database (step 2 above). The NCBI database contains information about proteins whose structure has been analyzed using either NMR or x-ray crystallography. After performing this search, we gained the result on the basis of a similar structure. To perform this search, we used default search parameters as for Blastp, the database was PDB (protein data bank), the matrix was chosen to be BLOSUM62, the threshold value was 21 and the expected value was 10. The results with identity (>49%) and query cover (>93%) were tabulated.

Once again, we have performed Blastp but this time the database was the switch from NR to PDB (protein database). The algorithm followed was Blastp, the matrix was BLOSUM62, the threshold value was 21 and the expected value was 10.

**Tabulating the Sequence**

To carry out further analysis, the sequence of all possible blast hits was replicated down and arranged. After the analysis of the Blast search result (accession no.), we have tabulated the sequence from the database, such as Uniprot and NCBI ([https://www.ncbi.nlm.nih.gov/nuccore/](https://www.ncbi.nlm.nih.gov/nuccore/)). After collecting the sequence results, we have moved forward further analysis.

**3) Performing Multiple Sequence Alignment(MSA)**

The sequence from the Blast was tabulated and was watched for correction such as the elimination of repeated similar sequence and to organize them in a table to increase readability. Afterwards, MSA was performed using Clustalw which uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. The operation was performed by using it in default setting which is enough for the fulfillment of results. Hence, MSA was performed using Clustalw with pairwise alignment parameters for fast/approximate areas K-tuple(word) size:1, window size:5, gap penalty:3, number of top diagonals:5, scoring method: percent whereas for slow /accurate, gap open penalty:10.0, gap extension penalty:0.1. Select weight matrix: BLOSUM (for PROTEIN). Whereas for multiple alignment parameters, gap open penalty:10, gap extension penalty:0.05 and selected weight matrix: BLOSUM (for PROTEIN).

**4) Establishing Phylogenetic Relationship**

To relate different sequence results with each other, we have performed phylogenetic analysis using a web interface Phylogeny.fr, of all the tabulated sequence. This analysis has been carried out with phylogenetic developments among the sequence and create an evolutionary relationship.

**5) Prediction of Structure**

While dealing with protein, the structure is where all our attention lies. As its well know that protein functionality, specificity, mode of operation, and working efficiency are grounded on its structure. Hence every protein is unique in its constructional conformation. So, in a modality of understanding and defining any protein we have performed certain structure predictions.

Hence to study the structure of protein we have initially performed Blastp against PDB database (step 2 above). PDB (protein data bank) contains structural information about those proteins whose structure has been analyzed by using either NMR or x-ray crystallography. So, after performing this we have gained the result on the basis of a similar structure. To perform this, we went with default search parameters as for blastp the database was PDB (protein data bank), the matrix was chosen to be BLOSUM62, the threshold value was 21 and the expected value was 10. The results with identity (>44%) and query cover (>93%) were replicated down.

A secondary structure prediction using GOR4 is performed which has given the structural detail of the amino acid sequence in terms of Alpha helix (Hh), 3₁₀ helix (Gg), Pi helix (Ⅱi), Beta bridge (Bb), Extended strand (Ee), Beta turn (Tt), Bend region (Ss), Random coil (Cc), Ambiguous states (?), and describing their count of presence in the sequence too.

**6) Constructing a 3D Model**

To define a structure there is need of a similar kind of structure(homologs) whose construction is already described. So, to accomplish that we have performed 3D modeling of the sample protein sequence by using SWISS-MODEL. SWISS-MODEL is an online server based on protein structure homology modeling. It read the sequence find the homologs template of its and by talking reference to protein structure database it creates a model of the query protein sequence.

**7) Scanning of Motif**

Motifs have a high level of importance in the protein structure as a sequence motif is a nucleotide or...
amino-acid sequence form that is far-flung or is supposed to have a biological implication. Scanning of the motif of the most relevant hits of blastp (against PDB) was performed using MEME (Multiple Em for Motif Elicitation). MEME discovers novel, un-gapped motifs (recurring, fixed-length patterns) in your sequences (sample output from sequences). MEME splits variable-length patterns into two or more separate motifs.

8) Mapping of Motifs.
We have performed mapping by using PyMol. PyMOLis molecular examination tools are based on virtual visualization and by the application of animations, high-quality rendering, crystallography, and other common molecular graphics activities.

Among the detected motifs we have the most relevant one as per its characters and were marked and coloured on the most relevant protein structure collected from PDB, based on their similar structure homology (blastp). The two structures were 1yvi (For HCP) which is the Chain A, X-RAY STRUCTURE OF PUTATIVE HISTIDINE-CONTAINING PHOSPHOTRANSFER PROTEIN FROM RICE having 99% Query cover and 71% of Ident. The second structure was 51xu (For RR) which is Chain A, Structure of The Dna-binding Domain of Lux Arrhythmo having 8per cent of Query cover and 68per cent of ident. Afterward, the structure around 6 Å was selected and marked with a different colour. This was to show the interaction of the motif toward its neighbors' structure. Polar contact was also being mapped around the region of 6 Å of the selected motif.

9) Creating Protein-Protein Interaction Prediction
By using String
To show the interaction between HCP and RR we have used STRING. Its function is to generate the protein-protein interaction critically, by using direct (physical) as well as indirect (functional) associations. More than 2000 organisms come under STRING coverage. It has necessitated scalable algorithms for transferring interaction information between organisms. It also has a completely redesigned prediction pipeline for inferring protein-protein associations from co-expression data, and an API interface for the R computing environment and improved statistical analysis for enrichment tests in user-provided networks.

In order to predict the Interaction, it uses a different type of input such as protein name, id, or direct fasta sequence of the query protein sample. And by taking the reference of the non-redundant protein database it generates its result.
To carry our work, we have chosen the flavor of multiple protein sequence input and both the sequence (HCP, RR) were pasted on the prescribed portal. The organism selected was Oryzasativa japonica.

By using PPCheck
PPCheck is a web server which can be employed to measure various perspective of interactions between any two given proteins/chains. The measure of interactions is analyzed by using standard energy calculations involving non-bonded interactions like van der Waals, electrostatic and hydrogen bonds. They constitute of these interactions is described as pseudo energy, whose ranges have been standardized using known sets of protein-protein complexes. (Anshul and Sowdhamini, Molecular Biosystems, 2013)
A zip pdb file of the selected structure was grabbed and uploaded to check all possible interaction, performed by PPcheck.

IV. RESULTS AND DISCUSSION

1) Sequence Similarity Search Out Come
At first, the results were searched by using blastp and was performed under the nr (Non-redundant protein sequence) database. The motive of using the nr database was to get results on the basis of similar sequence homology. Once again blastp was performed but this time the database was a switch from nr to PDB (protein databank). The motive of using the PDB database was to get results on the base of similar structure homology.

Conserved domain was also being detected in both database (nr and PDB) shorting and they are found to be contributing common inference. Rest other output sequences which are not listed below were tabulated and checked for any repetitive sequence and their elimination if present.

The table in green constitutes the output of the nr database, the table in blue represent PDB database while the one in orange represents results for the conserved domain which are common for search in both databases.

For HCP
After performing blastp against nr database we have got 100 blast hits on 100 subject sequences. Among those 100 hits, 13 have alignment score (>=200) and rest other ranging in between (80-200). We have tabulated 10 output. They occupy query cover (>=91per cent) and ident (>=73per cent).

For RR
After performing blastp against nr database we have got 100 blast hits on 100 subject sequences. And all 100 hits have alignment score (>=200). We have tabulated 10 outputs. They occupy query cover (>=99per cent) and ident (>=69per cent).

Table 1: - 10 Relevant Sequences Producing Significant Alignments (Under Nr Database)

| S. No | Description | Max Score | Total Score | Query Cover | E Value | Ident | Accession |
|-------|-------------|-----------|-------------|-------------|---------|-------|-----------|

This work is licensed under Creative Commons Attribution 4.0 International License.
| S No | Description                                                                 | Max score | Total score | Query cover | E value | Ident | Accession          |
|------|-----------------------------------------------------------------------------|-----------|-------------|-------------|---------|-------|--------------------|
| 1.   | PREDICTED: two-component response regulator ORR23 [Oryzasativa Japonica Group] | 1424      | 1424        | 100per cent | 0.0     | 100per cent | XP_015625496.1     |
| 2.   | RecName: Full=Two-component response regulator ORR23                         | 1419      | 1419        | 100per cent | 0.0     | 99per cent    | B8AEH1.1           |
| 3.   | PREDICTED: two-component response regulator ORR23 [Oryzabrachyantha]          | 1188      | 1188        | 99per cent  | 0.0     | 88per cent    | XP_006648050.1     |
| 4.   | PREDICTED: two-component response regulator ORR23 [Brachypodiumdistachyon]  | 996       | 996         | 99per cent  | 0.0     | 75per cent    | XP_003570300.1     |
| 5.   | type-B response regulator ARR12b [Loliumperenne]                             | 977       | 977         | 99per cent  | 0.0     | 74per cent    | ALT32071.1         |
| 6.   | two-component response regulator ORR23-like [Aegilopstauschii subsp. tauschii] | 971       | 971         | 99per cent  | 0.0     | 74per cent    | XP_020170131.1     |
| 7.   | two-component response regulator ORR23 [Setariaitalica]                      | 949       | 949         | 99per cent  | 0.0     | 72per cent    | XP_004954220.1     |
| 8.   | Two-component response regulator ARR12                                       | 942       | 942         | 99per cent  | 0.0     | 69per cent    | EMS58073.1         |
9. Two-component response regulator ORR23 [Dichantheliumoligosanthes] 926 926 99per cent 0.0 69per cent OEL33659.1

10. hypothetical protein PAHAL_A03697 [Panicumhallii] 915 915 99per cent 0.0 71per cent PAN08711.1

Table 3: 6 Relevant Sequences producing significant alignments (under PDB database)

| Description                                                                 | Max score | Total score | Query cover | E value | Ident  | Accession |
|-----------------------------------------------------------------------------|-----------|-------------|-------------|---------|--------|-----------|
| 1. Chain A, X-RAY STRUCTURE OF PUTATIVE HISTIDINE-CONTAINING PHOSPHOTRANSFER PROTEIN FROM RICE, AK104879 | 211       | 211         | 99per cent  | 3e-71   | 71per cent | 1YVI_A    |
| 2. Chain A, Crystal Structure Of Histidine-containing Phosphotransfer Protein, Zmhp2, From Maize | 197       | 197         | 91per cent  | 1e-65   | 71per cent | 1WN0_A    |
| 3. Chain A, Subatomic Resolution Crystal Structure Of Histidine-containing Phosphotransfer Protein Mthpt2 From MedicagoTruncatula | 131       | 131         | 86per cent  | 1e-39   | 48per cent | 4G78_A    |
| 4. Chain A, Crystal Structure of Histidine-containing Phosphotransfer Protein MtHPt1 from Medicagotruncatula | 127       | 127         | 96per cent  | 5e-38   | 48per cent | 3US6_A    |
| 5. Chain B, Crystal structure                                             | 124       | 124         | 90per cent  | 6e-37   | 50per cent | 4EUK_B    |
| 6. Chain A, Crystal Structure Of Histidine-containing Phosphotransfer Protein Ahp2 From Arabidopsis Thaliana | 120       | 120         | 93per cent  | 5e-35   | 44per cent | 4PAC_A    |

Conserved Domain

Table 4: 4 Putative Conserved Domains Detected

| Name | Accession | Description                                                                                                                                                                                                 | Interval | E-value |
|------|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|--------|
| HP   | cd00088   | Participation in signaling is done by the Histidine Phosphotransfer domain through a two-part component system. And in this system the part of phosphoryl donor to a regulator protein is assisted by an autophosphorylating histidine protein kinase; the phosphorylation and dephosphorylation which belongs to conserved aspartic acid residue are responsible for the modulation of response regulator protein; In most of the | 42-135  | 2.7    |

This work is licensed under Creative Commons Attribution 4.0 International License.
Eubacteria, the two-component protein are found in ample amount; there are about 62 two-component protein which participates in number of processes which involves chemotaxis, metabolism, osmoregulation, and transport. This are also found in gram positive and negative pathogenic bacteria and here they are employed in performing the regulation task of basic housekeeping functions moreover, they assure the expression of toxins and other protein which are most relevant to pathogenesis; however, in archaea and eukaryotes, two-component pathways appoint all signaling systems in a very small number; in fungi they are responsible for arbitrate environmental stress responses and, in pathogenic yeast, hyphal development. In Dictostelium and in plants, they participate in crucial processes such as osmoregulation, cell growth, and differentiation; till now the presence of two-component proteins have not been recognised in animals; in most of the prokaryotic arrangements, RR directly effects the output response, which functions as a transcription factor while in other side that is in eukaryotic arrangements, two-component proteins always recognised at the onset of signalling pathways where they port with more conventional eukaryotic signalling schemes such as MAP kinase and cyclic nucleotide cascades.

Pssm-ID: 238041  Cdd Length: 94  Bit Score: 54.31  E-value: 2.74e-10

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|
| 33637786 | VSEVVTLFCDDGERIEELSRQLEK-PNVDFDRVDSYHVQLKGSSASVGAQKVKNCTIQFREFCQQRSRDGCLKTLDDLVRTEFI123 |
| gi:cd00088 | QLAHEL79 |

The histidine-containing phosphotransfer (HPT) domain is one of the fresh protein modules that have an active residue of histidine which is responsible for arbitratess phosphotransfer reactions in the two-component signalling arrangement. A multistep phosphorelay includes the HPT domain has been prescribed for these signaling pathways. The anaerobic sensor kinase (ArcB h) has a HPT domain whose crystal structure has been determined. The HPT domain comprises six alpha helices which is holding a four-helix bundle-folding. The pattern of sequence similarity of the HPT domains of ArcB and few other components in the signaling arrangement can be translated in light of the three-dimensional structure and affirms the decision that the HPT domains have a similar structural motif both in prokaryotes and eukaryotes. In S. cerevisiae ypd1p this domain has been found which depicts to contain a binding surface for Ssk1p (response regulator receiver domain containing protein pfam00072).

Pssm-ID: 307655  Cdd Length: 84  Bit Score: 53.51  E-value: 4.69e-10

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|
| 33637786 | VSEVVTLFCDDGERIEELSRQLEK-PNVDFDRVDSYHVQLKGSSASVGAQKVKNCTIQFREFCQQRSRDGCLKTLDDLVRTEFI123 |
| gi:cd00088 | QLAHEL79 |

This work is licensed under Creative Commons Attribution 4.0 International License.
HPetr COG21 98

| gi 33637786124 YDLRN128 |
|--------------------------|
| Cdd:pfam01627 79 EALRA83 |

HPT (histidine-containing phosphotransfer) domain [Signal transduction mechanisms];

Pssm-ID: 225108 Cd Length: 122 Bit Score: 39.74 E-value: 6.77e-05

10 20 30 40 50 60
*...*...*...*...*...*...*...*...*...*...*...*...*...*

gi 33637786

PDFV5EVVTLCDDGERIIICELSRQLEkpNVDFRVDVSHQLKGSSASVGAQKV<br>KNTCIQFREQCQQR107

Cdd:COG2198

Histidine Phosphotransfer domain; carries an active residue of histidine that arbitrates phosphotransfer reactions. This Domain has been detected only in eubacteria. This alignment is an extension to that show in the Cell structure paper.

Pssm-ID: 197502 Cd Length: 92 Bit Score: 34.53 E-value: 3.21e-03

10 20 30 40 50
*...*...*...*...*...*...*...*...*...*...*...*...*...*

gi 33637786

EVVTLFCDDGERIIICELSRQLEkpNVDFRVDVSHQLKGSSASVGAQKV<br>KNTCIQFREQCQQR107

Cdd:smart00073

For RR we have got 100 blast hits on 100 subject sequence. All this 100 hits have alignment score (>=200).

We have tabulated 10 output which have occupied query cover (>93per cent) and ident (>=44per cent). The outputs are listed below:

Table 5: 10 Relevant Sequences Producing Significant Alignments: (Under Pdb Database)

| s.no | Description                                                                 | Max score | Total score | Query cover | E value | ident  | Accession |
|------|------------------------------------------------------------------------------|-----------|-------------|-------------|---------|--------|-----------|
| 1.   | Chain A, Solution Structure Of Arr10-B Belonging To The Garp Family Of Plant Myb-Related Dna Binding Motifs Of The Arabidopsis Response Regulators | 100       | 100         | 9per cent   | 1e-25   | 69per cent | 1IRZ_A   |
| 2.   | Chain A, Structure Of The Dna-binding Domain Of Lux Arrhythmo                | 87.4      | 87.4        | 8per cent   | 6e-21   | 68per cent | 5LXU_A   |
| 3.   | Chain A, Non-phosphorylated Hemr Receiver Domain From LeptospiraBiflexa      | 59.7      | 59.7        | 15per cent  | 2e-10   | 33per cent | 4Q7E_A   |
| 4.   | Chain A, Solution Structure Of The E.ColiRcsc C-Terminus (Residues 817-949) Containing Phosphoreceiver Domain | 59.3      | 59.3        | 18per cent  | 3e-10   | 32per cent | 2AYZ_A   |
| 5.   | Chain A, Structure Of N-                                                  | 58.9      | 58.9        | 15per cent  | 4e-10   | 31per cent | 3NHZ_A   |
Terminal Domain Of Mtra

| 6. | Chain C, 2.7 Angstrom Structure of a Phosphotransferase in Complex with a Receiver Domain | 58.9 | 58.9 | 17per cent | 6e-10 | 30per cent | 4QPI_C |
| 7. | Chain A, Solution Structure Of The E.ColiRcsc C-Terminus (Residues 700-949) Containing Linker Region And Phosphoreceiver Domain | 60.8 | 60.8 | 18per cent | 7e-10 | 32per cent | 2AYX_A |

Chain A, Solution Structure Of The E.ColiRcsc C-Terminus (Residues 700-949) Containing Linker Region And Phosphoreceiver Domain

| 8. | Chain A, Crystal Structure Of Yycf Receiver Domain From Bacillus Subtilis | 57.0 | 57.0 | 17per cent | 2e-09 | 29per cent | 2ZWM_A |
| 9. | Chain A, Crystal Structure Analysis Of The N-terminal Receiver Domain Of Response Regulator Pmra | 56.6 | 56.6 | 17per cent | 2e-09 | 30per cent | 3W9S_A |
| 10. | Chain A, Crystal Structure Of Unphosphorelated Receiver Domain Of Yycf | 56.2 | 56.2 | 17per cent | 3e-09 | 29per cent | 3F6P_A |

### Multiple Sequence Alignment(MSA) Output.

After the elimination of repeated sequences and those whose sequences are not found in the database (among blastp hits) we are left with 59 sequence material to carry out MSA. The results were replicated down as text material. Among those 58 groups of results, 10 groups with the highest alignment score were chosen, from both HCP as well as from RR MSA calculation.

#### For HCP

Table 7: 10 groups with maximum alignment score (among 59 homolog sequences of HCP)

| Group | Sequences | Score |
|-------|-----------|-------|
| 1     | 2         | 2476  |
| 4     | 4         | 2411  |
| 5     | 5         | 2419  |
| 7     | 2         | 2410  |
| 8     | 2         | 2448  |
| 9     | 4         | 2423  |
| 11    | 6         | 2425  |
| 14    | 2         | 2482  |
| 15    | 3         | 2462  |
| 19    | 2         | 2452  |

#### For RR

Table 8: 10 groups with maximum alignment score (among 59 homologs sequences of RR)

| Group | Sequences | Score |
|-------|-----------|-------|
| 2     | 2         | 11037 |
| 3     | 2         | 10401 |
| 9     | 2         | 11816 |
| 10    | 3         | 10594 |
| 11    | 4         | 10161 |
| 13    | 2         | 11519 |
| 14    | 2         | 11111 |
| 15    | 3         | 11147 |
| 16    | 5         | 10479 |
Fig 1: phylogeny tree of 59 homologs similar sequence of HCP, the green box represents the query sequence.
Fig 2: Phylogeny tree of 59 homologs similar sequence of RR, the green box represents the query sequence.
4. Prediction of structure

Secondary structure prediction is one of the important aspects of the protein study. Hence we have used GOR4 as a secondary structure prediction tool which has predicted the results as follows.

For HCP

| 10 20 30 40 50 60 70 80 90 100 |
|-------------------------------|
MAAAALTSQLNALVNMFAMGLLDDQFQQQLQMLQDSTAPDFVSEVTLCDDGERIICELSRQLEKPNDFDRV |
DSYVHQLKSSASVGAQKVKTNCQF |
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
Sequence length : 688

GOR4 :
- Alpha helix (Hh) : 158 is 22.97 per cent
- 3_10 helix (Gg) : 0 is 0.00 per cent
- Pi helix (Ii) : 0 is 0.00 per cent
- Beta bridge (Bb) : 0 is 0.00 per cent
- Extended strand (Ee) : 120 is 17.44 per cent
- Beta turn (Tt) : 0 is 0.00 per cent
- Bend region (Ss) : 0 is 0.00 per cent
- Random coil (Cc) : 410 is 59.59 per cent
- Ambiguous states (?) : 0 is 0.00 per cent
- Other states : 0 is 0.00 per cent

5. Constructing a 3D model
Every protein has its unique structural model so by using SWISSMODEL we have built a model of the protein.

For HCP

| Model #01 | File | Built with | Oligo-State | Ligands | GMQE | QMEAN |
|-----------|------|------------|-------------|---------|------|-------|

The following model was built (see Materials and Methods "Model Building"):
The template contained no ligands.

Target    MAAAALTSQLNALVNNMFAMGLLDDQ-FQQLQMLQSTDAPFSEVVTLCDDGERIELSRQLEKPNVFDRVDYSY VH
2q4f.1.A - AAAALRTQLTALLSSMFSQGLTDQEFPQLLQDEGGTPGFSEVVTLCDDADRIINEATLLEQPVVFNDVVDAY VH
Target    QLKGSSASVGAQKVKNCTIQFREFCQQRSDGCLKTLVREFTYDRNKFQAMQLEQQIQACYPKH
2q4f.1.A - QLKGSSASVGAQKVFTCMQFRQFCQSDGCLMALAVRNFYDRLRNFQMTQLLEQQIQAYDPKQ

For RR Models

The following models were built (see Materials and Methods "Model Building"):

| Model #02 | File | Built with | Oligo-State | Ligands | GMQE | QMEAN |
|-----------|------|------------|-------------|---------|------|-------|
|           | PDB  | ProMod3 Version 1.1.0. | homo-dimer (matching prediction) | None | 0.08 | -3.30 |

QMEAN -3.30

Cβ -1.99

All Atom -0.98

Solvation 0.08

Torsion -2.95

Fig8: Swiss Model Graph Result
| Template | Seq Identity | Oligo-state | QSQE | Found by | Method | Resolution | Seq Similarity | Range | Coverage | Description |
|----------|--------------|-------------|------|----------|--------|------------|----------------|-------|----------|-------------|
| 3c3m.1.A | 33.90        | homodimer   | 0.15 | BLAST    | X-ray  | 1.70Å      | 0.35           | 26 - 143 | 0.17     | Response regulator receiver protein |

| Ligand | Added to Model | Description |
|--------|----------------|-------------|
| GOL    | × - Not biologically relevant. | GLYCEROL |
| GOL    | × - Not biologically relevant. | GLYCEROL |
| GOL    | × - Not biologically relevant. | GLYCEROL |

Target
MRAAEERKGVVPAARRRDQFPVGMRLAVDDDPVCLKVLETLLLRCQYHTTTTNQAAIALKMLRNRIMDFDLVISDVHMP
3c3m.1.A ------------------------ ILVVDSPMIIVDFVTMLERGGYRPITAFSSEGECLEAL-- NAPPDVLLLDIMME
Target
DMDFGKLLELVGLE---
MDLPVIMLSVGETKTVLGIgACDYYLPVRIELRNQWHVIRKFRSTRDRANLDFYE
3c3m.1.A PMDGWETLERIKTDPA0RDPVLMLTAKPLTPEANEYGSIEDYKLPPTHQLYEAIEHLRHS---

Target
CNKPPNADSHVGHVTGSGPQDGSRPSKKRKEYCSEEEDGEVNTQDI0DPSAPKPRVWVSELRKHFRAVNQLGID
3c3m.1.A -----------------------------------------------
Target
KAVPKRIELMNVEKTLRENSHLQKYRLKRLSAYAVQVSIVAALGGRDPFLHMGGFEGLQGYQAFTSSAALSFT
3c3m.1.A -----------------------------------------------
Target
PHGLNNPRNNPAALGTQGPASISQTMSNHTLSHSDNKYHLSPNQKNGLQGLATSLGTQMQQKWHIHEEDT
3c3m.1.A -----------------------------------------------
Target
DLSTLSGNLQLNSGTLQSVTSSPLLPQELAECTQA0KVSQPSIRTSSVSHEIEAGVGSIGLESRSVQQSTIPS
3c3m.1.A -----------------------------------------------
Target
GFSANGLIIHSFNNTCANKLGTSACAPRSDMVARDTKGGASSFGFAMLPPDTEQYLNFGGNGLQKQKDFDR
3c3m.1.A -----------------------------------------------
Target
TADSLFDLKVWSSPVSSQLASNIGAHAMSQWNNNSNNSNNIGARMIGQATSSGSTDVIPQMKTDVLVSMDMAM
PKNAS3c3m.1.A -----------------------------------------------
Target
DLSIPKLQSELSSSSCSFDGLLNSIVKVEKDDVTFSDDLGCDFYLGGAC1
3c3m.1.A -----------------------------------------------

| Model #01 | File | Built with | Oligo-State | Ligands | GMQE | QMEAN |
|-----------|------|------------|-------------|---------|------|-------|
|          | PDB  | ProMod3 Version 1.1.0. | homo-dimer (matching prediction) | None | 0.08 | -2.36 |
**Fig. 9: Swiss-Model Graph Results**

| Template      | Seq Identity | Oligo-state | QSQE | Found by | Method | Resolution | Seq Similarity | Range | Coverage | Description                  |
|---------------|--------------|-------------|------|----------|--------|------------|----------------|-------|----------|-------------------------------|
| 2zwm.1.A      | 29.41        | homodimer   | 0.18 | BLAST    | X-ray  | 2.04Å      | 0.35           | 25-144| 0.17     | Transcriptional regulatory protein yycF |

**Ligand**

- SO4
- SO4
- SO4
- SO4

**Added to Model**

- X - Not biologically relevant.

**Description**

- SULFATE ION
- SULFATE ION
- SULFATE ION
- SULFATE ION

**Target**

MRAAEERKGVVPAARRRDQFPVGMVRVLAVIDDPVCLKVLETLLRCQYHVTITNQAAIALKMLRENRMDFL VISVHMMP

2zwm.1.A  

Target

DMDFKKLELVLVEMDLPVIMLSNGETKTVLKGITHGACDYLKPVRIEELRNIWQHVRIFKSTRDRANLDFY EECNK

2zwm.1.A  

Target

PNADSDHVHVHTGCSPDQSGPRSKKRKEYCSEEEDEGEVNTQIDDPAPKPRVWVSLHMRKFVAANQL GIDKAV

2zwm.1.A  

Target

PKRILEMNVEKTRENVASHLQKYRLYKRLSAVASQQVSIVAAALGGRDPFLHMGGFEGLQGYQAFSSAALSS FTPHG

2zwm.1.A  

Target

LLNSPRNNPAALGTQGVPAKSIQTMSGSHTLHSINDANKYHSLPGNQKGNLQQLGALTSLGQTQMQQKWIHE ETDCLS

2zwm.1.A  

Target

TILSGNLSNMGSTLQSVTSSPLLPQELAECTQAKIVSQQPSRTSSVSSEHEIQAVGVVSVGLLESRVSQSTIPSFGS
Scanning of motif.

The results which were oriented from the result of blastp against the PDB database were used here. Fasta sequence was grabbed out from PDB database for both the selected structure and was examined under MEME server. After performing both the query sequence analysis by using MEME we have obtained the following results.

For HCP

We have selected the relevant outcome among the 6. 1YVI is the structure selected for further analysis as it occupies 99 per cent Query cover, 71 per cent Identi, and 3e-71 E value.

Table 9: Representation of motif characters

| Motifs | E-value | Sites | Width | Fasta |
|--------|---------|-------|-------|-------|
| M1.    | 4.6e+001| 2     | 13    | >1YVI:A|PDBID|CHAIN|SEQUENCE_site_1 offset= 49 FCDDADRIINEIA >1YVI:A|PDBID|CHAIN|SEQUENCE_site_2 offset= 103 FCQDKSRDGCLMA |
| M2.    | 6.2e+001| 2     | 6     | >1YVI:A|PDBID|CHAIN|SEQUENCE_site_1 offset= 93 VKFTCM >1YVI:A|PDBID|CHAIN|SEQUENCE_site_2 offset= 128 NKFQTM |
| M3.    | 6.4e+001| 2     | 6     | >1YVI:A|PDBID|CHAIN|SEQUENCE_site_1 offset= 140 IQAYDP >1YVI:A|PDBID|CHAIN|SEQUENCE_site_2 offset= 74 VDAYVH |

Table 10: Residue Conformation

| Name   | Freq. | Bg.  |
|--------|-------|------|
| A      | 0.087 | 0.083|
| C      | 0.027 | 0.030|

Figure 10: -Logo of the motif found after MEME analysis
| Motif Site Distribution | ANR: Any number of sites per sequence |
|-------------------------|-------------------------------------|
| Site Strand Handling    | This alphabet only has one strand   |
| Maximum Number of Motifs| 3                                   |
| Motif E-value Threshold | no limit                            |
| Minimum Motif Width     | 6                                   |
| Maximum Motif Width     | 50                                  |
| Minimum Sites per Motif | 2                                   |
| Maximum Sites per Motif | 5                                   |

**For RR**
We have selected the most relevant outcome among the results. 4Q7E is the structure selected for further analysis as it occupies 15 per cent Query cover, 33 per cent Ident and 2e-10E value.

| Table 12: Representation of motif characters |
|---------------------------------------------|
| Motifs | E-value | Sites | Width | Fasta                                                                 |
|--------|---------|-------|-------|------------------------------------------------------------------------|
| M1.    | 3.5e+001| 2     | 6     | >4Q7E:A[PDBID|CHAIN|SEQUENCE_site_1 offset= 43 NLYRPN >4Q7E:A[PDBID|CHAIN|SEQUENCE_site_2 offset= 122 SLTRPH |
| M2.    | 1.1e+002| 2     | 13    | >4Q7E:A[PDBID|CHAIN|SEQUENCE_site_1 offset= 27 QDKYRVWEWAKTIS >4Q7E:A[PDBID|CHAIN|SEQUENCE_site_2 offset= 60 PDGNGFDLAEMIV |
| M3.    | 1.2e+002| 2     | 6     | >4Q7E:A[PDBID|CHAIN|SEQUENCE_site_1 offset= 3 SMKPRI >4Q7E:A[PDBID|CHAIN|SEQUENCE_site_2 offset= 104 IPKPH |
Table 13: Residue Conformation

| Name     | Freq. | Bg.  |
|----------|-------|------|
| A        | Alanine | 0.070 | 0.067 |
| C        | Cysteine | 0.000 | 0.007 |
| D        | Aspartic acid | 0.062 | 0.060 |
| E        | Glutamic acid | 0.116 | 0.107 |
| F        | Phenylalanine | 0.062 | 0.060 |
| G        | Glycine | 0.070 | 0.067 |
| H        | Histidine | 0.016 | 0.020 |
| I        | Isoleucine | 0.047 | 0.047 |
| K        | Lysine | 0.070 | 0.067 |
| L        | Leucine | 0.155 | 0.141 |
| M        | Methionine | 0.016 | 0.020 |
| N        | Asparagine | 0.023 | 0.027 |
| P        | Proline | 0.054 | 0.054 |
| Q        | Glutamine | 0.023 | 0.027 |
| R        | Arginine | 0.078 | 0.074 |
| S        | Serine | 0.031 | 0.034 |
| T        | Threonine | 0.031 | 0.034 |
| V        | Valine | 0.047 | 0.047 |
| W        | Tryptophan | 0.008 | 0.013 |
| Y        | Tyrosine | 0.023 | 0.027 |

Table 14: Motif Site Distribution

| Motif Site Distribution | ANR: Any number of sites per sequence |
|-------------------------|--------------------------------------|
| Site Strand Handling    | This alphabet only has one strand    |
| Maximum Number of Motifs| 3                                    |
| Motif E-value Threshold | no limit                             |
| Minimum Motif Width     | 6                                    |
| Maximum Motif Width     | 50                                   |
| Minimum Sites per Motif | 2                                    |
| Maximum Sites per Motif | 5                                    |

6. Mapping of Motifs.

The detected motif from the above step was marked on their corresponding structure by using PyMol. To show the interaction of the motif with its neighbour structure, we have selected the area around 6Å and was marked with a different colour. Polar contacts were also shown by the dotted line of a different colour.

For HCP(1YVI)

M1(>1YVI:A|PDBID|CHAIN|SEQUENCE_site _1 offset= 49) was selected from the above scanned motifs as it occupies width 13 and have e-value 4.6e+001.

This work is licensed under Creative Commons Attribution 4.0 International License.
Figure 12: 1YVI cartoon structure; motif is marked with red, the selected region around 6Å was marked with yellow and the polar covalent contacts are shown with green dotted line; while the non-selected upper region is coated with blue and inner loops are with grey.

For RR(4Q7E), M2(>4Q7E:A|PDBID|CHAIN|SEQUENCE_site_1 offset= 27QDKYRVEWAKTIS) was selected from the above scanned motifs as it occupies width 13 and have e-value 1.1e+002.

Figure 13: 4Q7E cartoon structure; motif is marked with red, the selected region around 6Å was marked with yellow and the polar covalent contacts are shown with green dotted line; while the non-selected upper region is coated with blue and inner loops are with grey.
7. Creating protein-protein interaction prediction

STRING output

LAB EXPERIMENTS

Relevant datasets in Oryza sativa Japonica: (none).

Relevant information transferred from other organisms:

| Protein-protein interaction (bind) | Saccharomyces cerevisiae: YPD1, YPD3, SLN1 |
|-----------------------------------|---------------------------------------------|
| Protein-protein interaction (dig) | Saccharomyces cerevisiae: YPD1, YPD3, SLN1 |
| Detected by biochemical and two hybrid assays |                                           |
| Protein-protein interaction (grid) | Arabidopsis thaliana: ARR4, AHP1           |
| Detected by patmc/MI0004 (affinity chromatography technology assay) |                         |
| Protein-protein interaction (grid) | Arabidopsis thaliana: ARR4, AHP1           |
| Detected by patmc/MI0006 (two hybrid) assay |                               |
| Protein-protein interaction (grid) | Arabidopsis thaliana: ARR4, AHP1           |
| Detected by patmc/MI0006 (pull down) assay |                              |
| Protein-protein interaction (grid) | Arabidopsis thaliana: ARR4, AHP1           |
| Detected by patmc/MI0018 (two hybrid) assay |                                |
| Protein-protein interaction (grid) | Arabidopsis thaliana: ARR4, AHP1           |
| Detected by patmc/MI0018 (pull down) assay |                             |
| Protein-protein interaction (grid) | Arabidopsis thaliana: ARR4, AHP1           |
| Detected by patmc/MI0018 (two hybrid) assay |                              |

[Transcated after 10 items]

Figure 14: Relevant Information Transferred From other Organisms

Figure 15: Network Stats
Figure 16: Gene co-expression

Figure 17: Interaction Relationship Shown with Various Coloured Line

Figure 18: Tabular description to define colour and figure arrangements of figure 17
Figure 19: Gene Co-Occurrence Description
**Table 15: Interactions and Total Stabilizing Energy**

| Interaction                  | Energy (kJ/mol) |
|------------------------------|-----------------|
| Hydrogen Bond Energy         | -12.93          |
| Electrostatic Energy         | 19.14           |
| Van der Waals Energy         | -278.14         |
| Total Stabilizing Energy     | -271.93         |
| Number of interface residues | 102             |
| Normalized Energy per residue| -2.67           |

| Interaction                              | Count     |
|------------------------------------------|-----------|
| No. of Short Contacts                    | 5         |
| No. of Hydrophobic Interactions          | 8         |
| No. of van der Waals Pairs               | 10118     |
| No. of Salt Bridges                      | 1         |
| No. of Potential Favourable Electrostatic Interactions | 4         |
| No. of Unfavourable Electrostatic Interactions | 5         |

**Table 16: Potential Hydrogen Bonds**

| Residue-1 | Residue-2 |
|-----------|-----------|
| Res Num   | Res Name  | Chain | Atom Name |
| 33        | LEU       | A     | O         |
| 53        | ASP       | A     | OD2       |
| Res Num   | Res Name  | Chain | Atom Name |
| 102       | ARG       | B     | NH1       |
| 81        | GLN       | B     | NE2       |

| Distance (A) | |
|--------------|---|
| 2.99         | |
| 2.55         | |

**Table 17: Potential Hydrophobic Interactions**

| Residue-1 | Residue-2 |
|-----------|-----------|
| Res Num   | Res Name  | Chain | Atom Name |
| 42        | PHE       | A     | CB        |
| 46        | VAL       | A     | CB        |
| 49        | LEU       | A     | CB        |
| 77        | ALA       | A     | CB        |
| 77        | ALA       | A     | CB        |
| 49        | LEU       | A     | CB        |
| 77        | ALA       | A     | CB        |
| 77        | ALA       | A     | CB        |
| 78        | TYR       | A     | CB        |

| Distance (Å) | |
|--------------|---|
| 6.44         | |
| 4.53         | |
| 5.17         | |
| 6.70         | |
| 6.34         | |
| 4.59         | |
| 5.47         | |
| 6.84         | |

**Table 18: Potential Salt Bridges**

| Residue-1 | Residue-2 |
|-----------|-----------|
| Res Num   | Res Name  | Chain | Atom Name |
| 53        | ASP       | A     | OD1       |
| Res Num   | Res Name  | Chain | Atom Name |
| 56        | ARG       | B     | NH2       |

| Distance (Å) | |
|--------------|---|
| 3.85         | |

**Table 19: Potential Favourable Electrostatic Interactions**

| Residue-1 | Residue-2 |
|-----------|-----------|
| Res Num   | Res Name  | Chain | Atom Name |
| 45        | GLU       | A     | CB        |
| 45        | GLU       | A     | CB        |
| 74        | LYS       | A     | CB        |
| 80        | HIS       | A     | CB        |

| Distance (Å) | |
|--------------|---|
| 4.48         | |
| 8.77         | |
| 4.62         | |
| 8.61         | |
### Table 20: -Potential Unfavourable Electrostatic Interactions

| Res Num | Res Name | Chain-1 | Atom Name | Res Num | Res Name | Chain-2 | Atom Name | Distance Å |
|---------|----------|---------|-----------|---------|----------|---------|-----------|------------|
| 45      | GLU      | A       | CB        | 73      | ASP      | B       | CB        | 4.68       |
| 45      | GLU      | A       | CB        | 76      | ASP      | B       | CB        | 6.89       |
| 53      | ASP      | A       | CB        | 53      | ASP      | B       | CB        | 8.67       |
| 73      | ASP      | A       | CB        | 45      | GLU      | B       | CB        | 4.85       |
| 76      | ASP      | A       | CB        | 45      | GLU      | B       | CB        | 7.05       |

### Table 21: -Potential Short Contacts

| Res Num | Res Name | Chain-1 | Atom Name | Res Num | Res Name | Chain-2 | Atom Name | Distance Å |
|---------|----------|---------|-----------|---------|----------|---------|-----------|------------|
| 39      | THR      | A       | HG1       | 73      | ASP      | B       | OD1       | 2.08       |
| 39      | THR      | A       | HG23      | 73      | ASP      | B       | OD1       | 2.20       |
| 49      | LEU      | A       | HD21      | 81      | GLN      | B       | OE1       | 1.93       |
| 53      | ASP      | A       | OD2       | 81      | GLN      | B       | HE22      | 1.68       |
| 73      | ASP      | A       | OD1       | 40      | PRO      | B       | HD2       | 1.99       |

For RR(4Q7E)

### Table 22: -Interactions and Total Stabilizing Energy

| Interaction Type | Energy (kJ/mol) |
|-----------------|-----------------|
| Hydrogen Bond Energy | -24.29         |
| Electrostatic Energy  | -61.61         |
| Van der Waals Energy   | -211.23        |
| Total Stabilizing Energy | -297.13        |
| Number of interface residues | 93             |
| Normalized Energy per residue | -3.19         |
| No. of Short Contacts | 14            |
| No. of Hydrophobic Interactions | 4          |
| No. of van der Waals Pairs | 9419       |
| No. of Salt Bridges | 4            |
| No. of Potential Favourable Electrostatic Interactions | 12          |
| No. of Potential Unfavourable Electrostatic Interactions | 9           |

### Table 23: -Potential Hydrogen Bonds

| Res Num | Res Name | Chain-1 | Atom Name | Res Num | Res Name | Chain-2 | Atom Name | Type of H-Bond | Distance (D-A) Å |
|---------|----------|---------|-----------|---------|----------|---------|-----------|---------------|------------------|
| 89      | ARG      | A       | NE        | 109     | GLU      | B       | OE2       | SS            | 2.81             |
| 89      | ARG      | A       | NH2       | 109     | GLU      | B       | OE1       | SS            | 2.80             |
| 109     | GLU      | A       | OE1       | 89      | ARG      | B       | NH2       | SS            | 2.83             |
| 109     | GLU      | A       | OE2       | 89      | ARG      | B       | NE        | SS            | 2.83             |

### Table 24: -Potential Hydrophobic Interactions

| Res Num | Res Name | Chain-1 | Atom Name | Res Num | Res Name | Chain-2 | Atom Name | Distance Å |
|---------|----------|---------|-----------|---------|----------|---------|-----------|------------|
| 86      | ALA      | A       | CB        | 105     | PHE      | B       | CB        | 5.34       |
| 93      | PHE      | A       | CB        | 112     | ILE      | B       | CB        | 4.74       |
Table 25: Potential Salt Bridges

| Res Num | Res Name | Chain-1 | Atom Name | Res Num | Res Name | Chain-2 | Atom Name | Distance Å |
|---------|----------|---------|-----------|---------|----------|---------|-----------|------------|
| 89      | ARG      | A       | NH2       | 109     | GLU      | B       | OE1       | 2.80       |
| 89      | ARG      | A       | NH2       | 109     | GLU      | B       | OE2       | 3.82       |
| 109     | GLU      | A       | OE1       | 89      | ARG      | B       | NH2       | 2.83       |
| 109     | GLU      | A       | OE2       | 89      | ARG      | B       | NH2       | 3.78       |

Table 26: Potential Favourable Electrostatic Interactions

| Res Num | Res Name | Chain-1 | Atom Name | Res Num | Res Name | Chain-2 | Atom Name | Distance Å |
|---------|----------|---------|-----------|---------|----------|---------|-----------|------------|
| 89      | ARG      | A       | CB        | 109     | GLU      | B       | CB        | 5.60       |
| 91      | ARG      | A       | CB        | 109     | GLU      | B       | CB        | 9.48       |
| 94      | GLU      | A       | CB        | 106     | HIS      | B       | CB        | 9.64       |
| 94      | GLU      | A       | CB        | 108     | LYS      | B       | CB        | 7.30       |
| 99      | GLU      | A       | CB        | 113     | ARG      | B       | CB        | 9.58       |
| 99      | GLU      | A       | CB        | 116     | ARG      | B       | CB        | 9.86       |
| 106     | HIS      | A       | CB        | 94      | GLU      | B       | CB        | 9.45       |
| 108     | LYS      | A       | CB        | 94      | GLU      | B       | CB        | 7.11       |
| 109     | GLU      | A       | CB        | 89      | ARG      | B       | CB        | 5.44       |
| 109     | GLU      | A       | CB        | 91      | ARG      | B       | CB        | 9.21       |
| 113     | ARG      | A       | CB        | 99      | GLU      | B       | CB        | 9.50       |
| 116     | ARG      | A       | CB        | 99      | GLU      | B       | CB        | 9.75       |

Table 27: Potential Unfavourable Electrostatic Interactions

| Res Num | Res Name | Chain-1 | Atom Name | Res Num | Res Name | Chain-2 | Atom Name | Distance Å |
|---------|----------|---------|-----------|---------|----------|---------|-----------|------------|
| 89      | ARG      | A       | CB        | 106     | HIS      | B       | CB        | 8.00       |
| 89      | ARG      | A       | CB        | 113     | ARG      | B       | CB        | 8.98       |
| 91      | ARG      | A       | CB        | 106     | HIS      | B       | CB        | 9.84       |
| 94      | GLU      | A       | CB        | 109     | GLU      | B       | CB        | 8.30       |
| 99      | GLU      | A       | CB        | 99      | GLU      | B       | CB        | 9.97       |
| 106     | HIS      | A       | CB        | 89      | ARG      | B       | CB        | 7.83       |
| 106     | HIS      | A       | CB        | 91      | ARG      | B       | CB        | 9.63       |
| 109     | GLU      | A       | CB        | 94      | GLU      | B       | CB        | 8.01       |
| 113     | ARG      | A       | CB        | 89      | ARG      | B       | CB        | 8.97       |

Table 28: Potential Short Contacts

| Res Num | Res Name | Chain-1 | Atom Name | Res Num | Res Name | Chain-2 | Atom Name | Distance Å |
|---------|----------|---------|-----------|---------|----------|---------|-----------|------------|
| 74      | ASP      | A       | O         | 9       | GLU      | B       | OE2       | 2.04       |
| 76      | PRO      | A       | CG        | 9       | GLU      | B       | OE1       | 1.92       |
| 76      | PRO      | A       | CD        | 9       | GLU      | B       | OE1       | 2.44       |
| 76      | PRO      | A       | HG2       | 9       | GLU      | B       | OE1       | 1.73       |

This work is licensed under Creative Commons Attribution 4.0 International License.
V. CONCLUSION

Insilco analysis of Sequence and structure (homologs) of HCP and RR protein of rice was carried out. Various similar homologs sequence and structure were found from blast and analysis such as creating phylogeny by using PHYLOGENY.FR server to understand the evolutionary relation of rice HCP with other plants and deriving the best alignment score of the various searched results were carried out by using ClustalW. Moreover, we have constructed the 3d model of HCP and RR by using SWISSMODEL and prediction of the structure was also being carried out by GOR4 secondary structure prediction tools. We have also performed interaction based analysis by using STRING and PPCHECK which define various levels of its structure while dealing with its interacting molecule.

No doubt, this analysis does not completely describe the functionality of the senescence of rice which is mediated by interaction in HCP and RR protein but they are going to illuminate the path of understanding of senescence in Rice up to a certain extent. As the role of senescence is very wide in rice and what we have performed is just small piece fragments.

REFERENCES

[1] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

[2] Gish, W., & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nature Genetics*, 3(3), 266–272. https://doi.org/10.1038/ng0393-266

[3] Madden, T. L., Tatusov, R. L., & Zhang, J. (1996). Applications of network BLAST server. In *Methods in Enzymology* (Vol. 266, pp. 131–141). Elsevier. https://doi.org/10.1016/S0076-6879(96)66011-X

[4] Altschul, S. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

[5] Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology*, 7(1−2), 203–214. https://doi.org/10.1089/10665270050081478

[6] Zhang, J., & Madden, T. L. (1997). PowerBLAST: A New Network BLAST Application for Interactive or Automated Sequence Analysis and Annotation. *Genome Research*, 7(6), 649–656. https://doi.org/10.1101/gr.7.6.649

[7] Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics*, 24(16), 1757–1764. https://doi.org/10.1093/bioinformatics/btn322

[8] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 421. https://doi.org/10.1186/1471-2105-10-421

[9] Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct*, 7(1), 12. https://doi.org/10.1186/1745-6150-7-12

[10] Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

[11] Goujon, M., McWilliam, H., Li, W., Valentin, F., Squiracito, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, 38(Web Server), W695–W699. https://doi.org/10.1093/nar/gkq313

[12] Latysheva, N. S., & Babu, M. M. (2016). Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Research*, 44(10), 4487–4503. https://doi.org/10.1093/nar/gkw282

[13] Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., & Gascuel, O. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, 36(Web
Two Rice Authentic Histidine Phosphotransfer Proteins, OsAHP1 and OsAHP2, Mediate Cytokinin Signaling and Stress Responses in Rice. Plant Physiology, 165(1), 335–345. https://doi.org/10.1104/pp.113.232629

[25] Sakai, H. (2001). ARR1, a Transcription Factor for Genes Immediately Responsive to Cytokinins. Science, 294(5546), 1519–1521. https://doi.org/10.1126/science.1065201

[26] Schrodinger, L.L.C., 2017. The PyMol molecular graphics system, (v.2.0). Schrödinger, LLC, NEW YORK.

[27] Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J., & von Mering, C. (2017). The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Research, 45(D1), D362–D368. https://doi.org/10.1093/nar/gkw937

[28] Szklarczyk, D., Franceschini, A., Wyder, S., Forsslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., & von Mering, C. (2015). STRING v10: Protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research, 43(D1), D447–D452. https://doi.org/10.1093/nar/gku1003

[29] Chen, X., Yang, J.-R., Guan, N.-N., & Li, J.-Q. (2018). GRMDA: Graph Regression for MiRNA-Disease Association Prediction. Frontiers in Physiology, 9, 92. https://doi.org/10.3389/fphys.2018.00092

[30] Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Mingeze, P., Bork, P., von Mering, C., & Jensen, L. J. (2012). STRING v9.1: Protein–protein interaction networks, with increased coverage and integration. Nucleic Acids Research, 41(D1), D808–D815. https://doi.org/10.1093/nar/gks1094

[31] Sukhwal, A., & Sowdhhamini, R. (2013). Oligomisation status and evolutionary conservation of interfaces of protein structural domain superfamilies. Molecular BioSystems, 9(7), 1652. https://doi.org/10.1039/c3mb25484d