Computational Analysis of Drought Tolerance Genes in Rice (*Oryza sativa, japonica)*

A Thesis

Submitted to the Orissa University of Agriculture & Technology, Bhubaneswar for the partial fulfillment of the requirement for the award of degree of

MASTER OF SCIENCE
IN
BIOINFORMATICS
BY

IPSITA PANDA

ADM NO:-09BI/14

DEPARTMENT OF BIOINFORMATICS

CENTRE FOR POST GRADUATE STUDIES

ORISSA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

BHUBANESWAR-751003

2016

Name of the Advisor **Sukanta Kumar Pradhan**

CERTIFICATE –I

This is to certify that the thesis entitled**, "Computational Analysis of Drought Tolerance Genes in Rice (*Oryza sativa, japonica)"** submitted in partial fulfillment of the requirements for award of the degree of **Master of Science in Bioinformatics** of Orissa University of Agriculture and Technology, Bhubaneswar is an authentic record of bonafide research work carried out by **MISS IPSITA PANDA (Adm. No. 09BI/14)** under my guidance & supervision. No part of this thesis has been submitted by her for any other degree. I further certify that any help or information received during the course of investigation have been duly acknowledged by her.

Place: Bhubaneswar **(Sukanta Kumar Pradhan)**
Chairman,
Advisory Committee

Date:

CERTIFICATE –II

This is to certify that the dissertation entitled, **"Computational Analysis of Drought Tolerance Genes in Rice (*Oryza sativa, japonica)*"** submitted by **IPSITA PANDA** to the Orissa University of Agriculture & Technology, Bhubaneswar in the partial fulfillment of the requirements for the award of the degree of **Master of Science in Bioinformatics** has been approved by the students advisory committee after an oral examination of the same in collaboration with external examiner.

ADVISORY COMMITTEE

1. Sukanta Kumar Pradhan
   Head of Department — Chairman **……...**
   Department of Bioinformatics
   CPGS, OUAT

2. Dr. K.C Samal — Member **……...**
   Assoc. Professor
   Department of Agril. Biotechnology
   CA, OUAT

3. Mr. S.N. Rath — Member **……**
   Asst. Professor
   Department of Bioinformatics
   CPGS, OUAT

External Examiner **……**

ACKNOWLEDGEMENT

It is my privilege to share my deep sense of gratitude to my advisor Mr. Sukanta Kumar Pradhan, HOD, Department of Bioinformatics, Orissa University of Agriculture and Technology, for his constant guidance, support and help without which the project would not have been successfully completed.

I am thankful to the member of the advisory committee Dr. Kailash Chandra Samal, Assoc. Professor, Department of Agril. Biotechnology, OUAT and Mr Surya Narayan Rath, Asst. Professor, Dept of Bioinformatics, OUAT for their support and encouragement to carry out this work successfully.

I would like to owe my sincere thanks to the Respected Dean, PGF-cum-DRI, CPGS,OUAT for their encouragement.

I would like to express my heartiest and cordial regards to my beloved Bapa, Bou who are real inspiration for me in every step of my life giving me unbound emotional support. I feel great pleasure to express my love to all my sweetest friends for their strong mental support throughout the project period and in these two years.

I feel honoured to be a part of this auspicious university for providing me a healthy atmosphere in these two years. Last but not the least I express my gratitude to God for invaluable inspiration for accomplishment of such a splendid work.

Bhubaneswar **(Ipsita Panda)**

Date:

Name of the Student	:	**IPSITA PANDA**
Admission No	:	**09BI/14**
Title of the Dissertation	:	**Computational Analysis of Drought Tolerance Genes in Rice (*Oryza sativa, japonica)***
Degree for which Dissertation Submitted	:	**Master of Science in Bioinformatics**
Name of the Department & University	:	**Department of Bioinformatics CPGS, O.U.A.T., Bhubaneswar, Odisha**
Year of Submission	:	**2016**
Name of the Advisor	:	**SUKANTA KUMAR PRADHAN Head of Department Department of Bioinformatics CPGS, O.U.A.T., Bhubaneswar**

ABSTRACT

Rice(*Oryza sativa*)is one of the most important crop in the world. Rice, wheat,maize together account for about half of the worlds food production and rice itself is the principal food of half the worlds population. Rice genome consists of 12 pairs of chromosome with genome size of 420MB. Stress resistance and drought tolerance traits are the important among different traits in *Oryza sativa*. In context to the climatic change, identification of the traits bears important as far as rice genomics is concerned.The rice genome has been sequenced for different sub species like *Oryza sativa* Indica and *Oryza sativa japonica*.The whole genome information of *Oryza sativa indica* and *Oryza sativa japonica* are available in Oryza database. Total no of genes are estimated to be 16,940.In this study an attempt has been made to compile and collecting information in an arround abiotic stress in rice. Six no of candidates genes have been studied.as far as its location and function is concerned. The corresponding protein sequences have also been retrieved from the UniProt protein sequence database. The physiochemical parameters, GO(Gene ontology)annotation and functional domain analysis have been carried out to understand its role in stress resistance and drought tolerance. The domains named(cd13960, cd00143, cd13960, Pfam05605, c120571, Pfam14571, Pfam05605, cd02440, PLN02476, Pfam60891, Pfam8160, cd02440, PLN02476, Pfam12705,c100516, cd02440, Pfam08100)present in the protein.The fuction of the protein is depend upon its proper three dimensioal structures. Three dimensioal structures of the above stress resistance protein of *Oryza sativa japonica* have been predicted using online server /tools,to understand arrangement of secondary structure arrangements. This information will helpful for researcher working in rice varietal development.

Key Words: Rice genome, Oryza database, Gene ontology, UniProt

Sukanta Kumar Pradhan **Ipsita Panda**

ADVISOR **AUTHOR**

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Oryza Database	Total a biotic rice stress gene are found.
UNIPROT	Universal protein resources.
	It is a protein sequence database.
PDB	protein data bank
CLUSTAL –X	It is a multiple alignment mode between two or
	More homologue DNA or protein sequence
SWISS MODEL	It is a structural bioinformatics web-server dedicated
	To homology modelling of protein 3D structures
RAPTORX	RaptorX is among the most popular methods for
	Protein structure prediction.
SAVE- Server	It is a statistical technique and its structure quality and distribution is independent on the quality of These known structures.
NCBI	National center for biotechnology information
DT	Drought tolerance
CDS	Conserved domains
GRAMENE	It is a comparative genome database for cereal
	crops
QTL	Quantitative Trait Loci
RFLP	Restriction fragment length polymorphism

CHAPTER-I

1. INTRODUCTION

Rice Genome Annotation Project

Rice is one of the three major staple food crops in the world. Major share of rice is cultivated during kharif season. The rice production in India largely depends on monsoon rains and only 59 per cent rice area has assured irrigation.This indicates that introduction of high-yielding rice varieties coupled with improvements in agriculturalpractices over the last three decades have made a major impact in the form ofincreased rice production.

Rice belongs to the genus *Oryza* which includes approximately 24 species.They are widely distributed growing in different habitats and different soil types.It is primarily a high calorie food which is cultivated mainly for its nutritious grain because of its protein and carbohydrate content.It shows differences in plant growth,yield,pest and disease resistance,stress tolerance and water requirement.Rice grain is classified as short, medium,or long grain sizes. Rice is known to come in a variety of colors, including: white rice, brown rice, black rice, purple rice, and red rice.It is a self pollinated crop.Sowing methods-The seed may be drilled into dry land or sown in nurseries and the seedlings later transplanted into a wet paddy-field.

Morphological characterization is the first step in the classification and evaluation of the germplasm (Smith and Smith, 1989). It is an indispensable tool for selecting varieties or lines based on agronomical, morphological, genetic or physiological characters (Ndour, 1998).The two major rice varieties grown worldwide today are *Oryzasativaindica*and *Oryzasativa japonica*. The two cultivated rice species, *OryzasativaL*. and *O. glaberrimaSteud*., belong to a speciesgroup called *Oryzasativa*complex together with the five wild taxa, *O. rufipogon*(sensulato), *O.longistaminata*Chev. etRoehr., *O. barthiiA. Chev*., *O. glumaepatula*Steud., and *O. meridionalis*Ng.Among these taxa, only *O. rufipogon*produces fertile F_1 hybrids with *O. sativa*and therefore these twospecies are considered to belong to a single biological species. Together with all circumstantial evidence, this suggests that *O. rufipogon*is

the ancestor of *O. sativa*.Also, seedmay be pre-germinated and broadcast into the mud in the paddy-field.It is including plant and cell physiology, molecular biology, genetics, and breeding.Mechanism of drought tolerance and expression of these drought resistance genes in high yielding varieties will help to improve the drought condition.Stress-induced gene expressions are of genes encoding proteins with known enzymatic or structural functions, proteins with as yet unknown functions, and regulatory proteins.*Oryzasativa* contains two major subspecies: the sticky, short grained *japonica* or *sinica* variety, and the nonsticky, long-grained *indica* variety.Salinity stress is one of the major constrains faced by the farmers of India especially at the saline soils and in areas irrigated with brackish water. The harmful effect of salinity occurs due to osmotic stress and specific ion toxicity

Rice belongs to the genus *Oryza* and has two cultivated and 22 wild species. The cultivated species are *Oryzasativa* and *Oryzaglaberrima*.*Oryzasativa* is grown all over the world while *Oryzaglaberrima* has been cultivated in West Africa for the last ~3500 years.They are AA,BB,CC,BBCC,CCDD,EE,HHKK, HHJJ, FF, and GG. There is a fivefold differences in genome size among the species with diploid species having 2n=24 and tetraploid species having 2n=48chromosomes.It is diploid with AA genore Type .O.saival.ssp japonica is the rice variety grown in subtropics and temperate climates. It has been sequenced to 10xcoverage by the members of the International Rice Genome Sequencing Project(IRGSP).O.satival.sspindica is a rice variety grown through the world .it has been estimated genome sizeof 466 Mb.

Molecular markers have become fundamental tools for finger printing establishing phylogenetics, tagging desirable genes, characterization of transformants and study of genome organization. Molecular markers based on DNA sequence are found to be more reliable.Now recently the global production and consumptionof rice is currently at equilibrium. The majority of the population in rice-producing areas, particularly in many Asian and African countries.Rice is the seed of the grass species O*ryzasativa* (Asian rice) or O*ryzaglaberrima*(African rice).Rice,a monocot is normally grown as an annual plant.Drought represents a significant environmental stress for rice production, with 19–23 million hectares of rainfed rice production in South and South East Asia often at risk.Under drought conditions, without sufficient water to afford them the ability to obtain the required levels

of nutrients from the soil, conventional commercial rice varieties can be severely affected – for example, yield losses as high as 40% have affected some parts of India, with resulting losses of around US$800 million annually.The internatioanl rice research institute conducts research into developing drought-tolerant rice varieties, including the varieties 5411 and Sookhadhan, currently being employed by farmers in the Philippines and Nepal respectively.

Soil salinity poses a major threat to rice crop productivity, particularly along low-lying coastal areas during the dry season.For example, roughly 1 million hectares of the coastal areas of Bangladesh are affected by saline soils.These high concentrations of salt can severely affect rice plants' normal physiology especially during early stages of growth, and as such farmers are often forced to abandon these otherwise potentially usable areas. Developed by the International rice research institute, the hybrid variety can utilise specialised leaf glands that allow for the removal of salt into the atmosphere.

Apart from its economic significance, rice has become an important plant forgenetic and genomic studies. Rice is diploid with 24 chromosomes which can bedistinguished individually using cytogenetic techniques. The rice genome is small (about 430 Mb).Rice grain quality affects the nutritional and commercial value of grains.This smallgenome size has contributed to rice becoming the prominent model system for cerealgenomics as well as a model for monocotyledonous plants. An ongoingeffort by International Rice Genome Sequencing Project (IRGSP) to compile acomplete high-quality draft of the rice genome sequence promises to deliver a veryuseful tool for science and rice breeding. The term genome is more than 75 years old and refers to an organism's complete set of genes and chromosomes.Genetic maps of the rice genome have been developed using molecular markers, for example RLFP. About 70% of the RFLP-based maps were developed using rice cDNAs as probes and, of those RFLP markers, about 30% had significant sequence homology to sequences of known genes.A high resolution rice genetic-linkage map has been constructed using EST clones as RFLP probes. Physical maps of the rice genome have been constructed based upon Expressed Sequence Tags (ESTs), Sequence-Tagged Connectors (STCs), bacterial artificial chromosome (BAC), yeast artificial chromosomes (YAC), P1-derived artificial chromosomes (PAC), or shotgun sequence analysis. As more of the

rice genome sequence and mapping of markers becomes available, it becomes critical to identify the functions of thousands of new rice genes.The reverse genetics approach attempts to do this by comparing sequence similarity among plant genes using rice EST markers.Crop domestications are long-term selection experiments that have greatly advanced human civilization. The domestication of cultivated rice (*Oryzasativa* L.) ranks as one of the most important developments in history. The domestication-associated traits are analysed through high-resolution genetic mapping.

A wide range of genetic and archaeological studies have been carried out to examine the phylogenetic relationships of rice.Molecular phylogenetic analyses indicated that *indica* and *japonica* originated independently.

Rice is the first cereal crop to be completely decoded (International RiceGenome Sequencing Project 2005). The high-quality map-based sequence of theentire rice genome is now available in the public domain.

Rice Genomics

In rice plant 12 chromosomes are present. Rice belongs to the genusOryza which includes approximately 24 species. There is a fivefold differences in genome size among the species with diploid species having 2n=24 and tetraploid species having 2n=48chromosomes.It is diploid with AA genotype.*Oryzasativa*.ssp japonica is the rice variety grown in subtropics and temperate climate.Chromosomal location of abiotic stress genes in rice genome are chromosome1- 150genes, chromosome2-73genes, chromosome3-28genes, chromosome4-29 genes, chromosome5-22 genes, chromosome6-94genes, chromosome7-113genes, chromosome8-72 genes, chromosome9-78 genes, chromosome10-69 genes, chromosome11-33 genes and chromosome12-34genes. The stress tolerance gene and disease resistance gene also found in rice plant species.Some selected stress and drought tolerance genes belong to *Oryzasativa* sub species japonica were analysedfrom the gene list table.In chromosome1-2 gene,chromosome2-10 gene, chromosome3-6 gene, chromosome4-5 gene, chromosome5-1 gene, chromosome6-18 gene, chromosome7-13 gene, chromosome8-1 gene, chromosome9-5 gene, chromosome10-2 gene, chromosome11-3 gene, chromosome12-1 stress resistance and drought tolerance genes were retrieved from abiotic rice gene lst table.

OBJECTIVE

1. To identify potential genes and proteins involved in abiotic stress of rice.
2. To study the function the genes and proteins through domain analysis.
3. To study the divergence of the stress resistance protein.
4. To predict the three dimensional structure of the protein involved in stress resistance and drought tolerance.

CHAPTER-II

2. REVIEW OF LITERATURE

Rice is the most vital cereal food crop of India, which occupies about 24 % of gross cropped area of the country. It contributes 42 % of total food grain production and 45 % of total cereal production of the country.During the last five decades the rice production trend has kept in pace with population growth trend. Global demand for rice is rising with the population growth, increasing affluence and changing dietary habits.The UN/FAO forecasts that global food production will need to increase by over 40 % by 2030 and 70 % by 2050 (FAO, 2009). Thus rice production in India as well as in several other Asian countries must be doubled by the year 2025 to meet the requirement of the increasing population.Molecular markers are becoming more functional in enhancing the efficiency in hybrid rice development and crop improvement. A broad spectrum of DNA markers such as RFLP, AFLP, RAPD, ISSR, SSR, SNPs etc., are being extensively used in rice research for genetic diversity analysis, phylogenetic and evolutionary studies, tagging and mapping of genes for quantitative traits of agronomic importance, marker assisted selection (MAS) and marker assisted back crossing (MABC).These are collected and completely sequenced 28,469 full-length complementaryDNA clones from OryzasativaL. ssp. japonica cv. Nipponbare.Through homology searches of publicly available sequence data, we assigned tentative protein functions to 21,596 clones (75.86%). Mapping of the cDNA clones to genomic DNA revealed that there are 19,000 to 20,500 transcription units in the rice genome.

Rice (*Oryzasativa*) is an important food crop; it is also a good model for studies of monocot plants because its genome (430 Mb) is small relative to other crop plants of the Poaceae species. Draft sequences of the OryzasativaL. ssp. indica(1) and japonica (2) genomes by the “whole-genome shotgun” sequencing method have been published, as have essentially complete sequences of chromosomes 1 and 4 by a physical mapping (“clone-by-clone”) method In addition to genomic data, full-length cDNA clones are necessary to identify exonintron boundaries and gene-coding regions within genomic sequences and for comprehensive gene-function analyses at the transcriptional (transcriptomic) and translational (protein informatic) levels.it is

describe the collection, grouping, sequencing (6), mapping, and functionalannotation of full-length cDNA (FLcDNA) clones from ssp. japonica (cv. Nipponbare). BLASTN and BLASTX programs (both of which use the Basic Local Alignment Search (Tool) for the gene homology searches between monocot and dicot plant.The mapped the 28,469 FL-cDNA clones to the rice genome sequences—the indica draft genome sequence the japonica draft genome sequence, and the japonica BAC/PAC (bacterial artificial chromosome and P1-derived artificial chromosome) clones.

Mapping results of our japonica originated clones to the indicagenome show that the nucleotide sequences of the gene-coding regions are very similar in these two subspecies. Genome sequence alone could not correctly identify the gene structure, but mapping of cDNA clones and comparison of genome sequences indicate the correct structure of the genes in rice.Phylogenetic analysis was performed with MEGA3.1program by neighbor-joining methodand the bootstrap test was carried out. Genomics research is generating new tools, such asfunctional molecular markers and informatics, as well as new knowledge about statistics and inheritance phenomena that could increase the efficiency and precision of crop improvement.

Genomics-assisted breeding for crop Improvement

Genomics research is generating new tools, such as functional molecular markers and informatics, as well as new knowledge about statistics and inheritance phenomena that could increase the efficiency and precision of crop improvement.In particular, the elucidation of the fundamental mechanisms of heterosis and epigenetics, and their manipulation, has great potential. Nevertheless, marker-assisted breeding andselection will gradually evolve into ‘genomics-assisted breeding’ for crop improvement. During the past few years, functionally characterized genes, EST and genome sequencing projects have facilitated the development of molecular markers from the transcribed regions of the genome. Among the more important and popular molecular markers that can be developed from ESTs are single-nucleotide polymorphisms (SNPs).Considerable progress has been made building infrastructure for applying genomics approaches. These include one-dimensional genetic information (genome sequences), many ESTs and gene knockout populations in several plant species of biological and agronomic importance.

New knowledge and new tools are changing the strategies used in crop plant research and will thus reduce the costs and increase the throughput of the assays. There is a continuing need to integrate disciplines such as structural genomics, transcriptomics, proteomics and metabolomics with plant physiology and plant breeding Bioinformatics is providing the means for integration and structured interrogation of datasets that will facilitate the cross-fertilization of disciplines.Genomics research has successfully unraveled various metabolic pathways and provided molecular markers for agronomic traits. However, the mechanisms of epigenetic phenomena are only beginning to be understood and their potential role in crop improvement is unknown. Similarly, tantalizing bits of information concerning the possible basis of heterosis are gradually emerging. Eventual elucidation of the mechanism of heterosis might be one of the most important contributions of molecular genetics research to crop improvement.

Stressed genomics —bringing relief to rice fields

The bacterial blight disease, detailed knowledge of host–pathogen interactions has enabled a predictive strategy to combine specific genes to provide durable resistance. Large-effect QTLs conferring tolerance to submergence, salinity, and drought have been identified.Environmental stresses are the main constraints for crop productivity.Rice has many of the genetic and genomic tools available in a model organism —complete genome sequence. In dealing with biotic and abitoic stresses, we face two distinct challenges. In developing resistance to biotic stresses, the main problem lies in dealing with the highly variable nature of biological agents, be it pathogen or insect. Predicting which genes would confer durable resistance remains a key challenge. For abiotic stresses, a high degree of genotype _ environment (G _ E) interaction makes assessing the causal relationship between genotype and phenotype difficult.Among the abiotic stresses, drought is undoubtedly the most complicated problem because of the large influe of genotype by environment interactions. For a long time, genetic effects contributing to drought tolerance were considered too small and variable to detect consistent across genotypes and environments.Rice, a first cereal crop whose draft genome sequence from two subspecies (japonica-type cv. Nipponbare and indica-type 93-11) was available in 2002, along with its almost

complete genome sequence in 2005, has drawn the attention of researchers worldwide because of its immense impact on human existence.

Phenotypic and genotypic analysis of drought-resistance traits for development of rice cultivars adapted to rainfed environments:-

Many of the world's rice-growing regions lack adequate irrigation facilities, and drought frequently reduces yield. Information about the type of drought faced in the target region – particularly the timing of the drought (late season terminal drought, early stage vegetative drought, and intermittent drought) and the intensity of the drought – are important in determining the specific plant traits required to improve drought resistance in rice.Drought is a major limitation for rice production in rainfed ecosystems.The ‘‘green revolution’’ in rice improvement has benefited many farmers in irrigated. rice production but has had limited impact on rainfedproduction.Efforts to improve the drought resistance of rice and thus increase yield in drought-prone rainfed areas led to the publicationof a manual for rice drought-resistance breeding. There are a number of advantages to such a system:

(1) natural drought occurrence in rainfed lowland and upland rice in the wet season is unpredictable, thus limiting screening under the desired drought types; (2) managed drought uses resources (e.g., the budget of a drought project of only a few years) more efficiently than waiting for the occurrence of natural drought in the wet season; and (3) there is a high genetic correlation between yield under stress in a managed selection environment and that inthe target environment. Most of the mapping populations were derived from indica _ japonica parents, and it is often the case that favorable alleles for drought-resistance traits are contributed by japonica lines. Considering that indica and japonica ecotypes are grown in different environments and thatmost breeding programs involve locally adapted rice accessions, the results for traits such as yield in indica _ japonica populations have to be interpreted with care.

Gramene: development and integration of trait and gene ontologies for rice

Structural and functional analysis of rice genome

Rice is an excellent system for plant genomics as it represents a modest size genome of 430 Mb. It feeds more than half the population of the world. Draft sequences of the rice genome, derived by whole-genome shotgun approach at relatively low coverage (4–6 X), were published and the International Rice Genome Sequencing Project (IRGSP) declared high quality (>10 X), genetically anchored, phase 2 level sequence in 2002. In addition, phase 3 level finished sequence of chromosomes 1, 4 and 10 (out of 12 chromosomes of rice) has already been reported by scientists from IRGSP consortium. Various estimates of genes in rice place the number at >50,000. Already, over 28,000 full-length cDNAs have been sequenced.Microarray analysis is unraveling the identity of rice genes expressing in temporal and spatial manner and should help target candidate genes useful for improving traits of agronomic importance.Rice genomics obtained a major boost in terms of the sequence drafts that were released by distinct projects. Notwithstanding the value of finished sequence of rice genome being carried out at present only by It is, the data released have already provided valuable information on genome structure and organization Rice sequencers realize that many scientists working in plant breeding, plant molecular genetics, plant molecular biology, and bioinformatics are awaiting the publication of the complete, high-quality rice genome sequence.The IRGSP aims to maintain high sequence quality standards, to publicize all sequence information as soon as possible and to complete the whole genome sequence in the shortest possible time.Intimate collaboration within IRGSP will make this plan realistic and will raise the profile of cereal genomics worldwide. Since 1991, the Rice Genome Research Program in Japan has carried out rice genomics, such as large-scale cDNA analysis, construction of a fine-scale restriction fragment length polymorphism map, and physical mapping of the rice genome with yeast artificial chromosome clones. These studies have made a great impact on research into grass genomes and made rice a model plant for other cereal crop research. Starting in 1998, the Rice Genome Research Program will step into a new stage of genomics—that of genome sequencing. This project eventually should reveal all of the genomic sequence information in the rice plant and be an indispensable aid in understanding the genomics of other grass

species.Gramene(http://www.gramene.org/) is a comparative genome database for cereal crops and a community resource for rice. We are populating and curating Gramene with annotated rice genomic sequence data and associated biological information including mole Gramene will employ three related controlled vocabularies. The specific goal of Gramene is, first to provide a Trait Ontology (TO) that can be used across the cereal crops to facilitate phenotypic comparisons both within and between the genera. Second, a vocabulary for plant anatomy terms, the Plant Ontology (PO) will facilitate the curation of morphological and anatomical feature information with respect to expression,localization of genes and gene products and the affected plant parts in a phenotype.The TO and PO are both in the early stages of development in collaboration with theInternational Rice Research Institute.In recent years, the world of biology has experienced a renaissance in terms of data generation, processing and representation, based on largescale genome sequencing and functional genomics efforts for a number of organisms.

Functional annotation of rice gene products, mutants and phenotypes

A wide array of genetic information is contained in the large reservoirs of crop plant germplasm that have been accumulated over many decades.This information has been enriched by the historical familiarity of the agricultural community with the performance characteristics, crossing histories and environmental adaptation of crop speciesAgricultural researchers have carefully recorded information on mutants, strains, phenotypes, polymorphisms and QTLs and many such studies involve associations between phenotypes and molecular markers.Association studies also hold promise for assessing correlations between specific genetic variants (SSRs, SNPs) and trait differences on a population level. In the rice mutant and phenotype (QTL) database, features associated with mutants and phenotypes will carry an annotation for the modified sequences, linked molecular markers, gene (allele) name, expression patterns of the mutant gene, etc. and their map position on the genetic or the physical maps. One of the specific goals of Gramene is to provide a Trait Ontology (TO) that will enable users to query for candidate genes from a target region on a rice chromosome based on phenotypic comparisons across the grasses.Each curated entry from the protein, mutant and phenotype database will carry an evidence code (http://www.gramene org/plant_ontology/evidence_codes.html).The shared

development and use of the controlled vocabularies such as GO, TO, and the PO, will help in structuring the datasets in Gramene on genes, gene products, mutants, strains, phenotypes, polymorphisms and quantitative trait loci.

Towards an accurate sequence of the rice genome

Several more- or less-elaborated rice genome sequences have been produced recently using different strategies.It has become possible to compare them and to unravel the major features of the rice genome in terms of nucleotide composition, repeats, gene content and variability.These advantages have made rice the next plant species to be targeted for genome sequencing after Arabidopsis thaliana.The official goal of the International Rice Genome Sequencing Project (IRGSP)was to complete the sequencing of the japonica variety 'Nipponbare' by the end of 2008. The consortium initially assembled from ten countries (Japan, Korea, UK, Taiwan, China, Thailand, India, United States, Canada and France) and decided to adopt a clone-by-clone sequencing strategy once a minimumtiling path for each chromosome had been established. This strategy allowed the task of sequencing the Nipponbare genome to be distributed among the participants on a chromosome basis. The availability of a completely assembled high-quality draft of the rice sequence represents very significant progress. Sequencing is the only way to establish a precise catalogue of all the rice genes and to provide the necessary background for their annotation and functional characterization.The availability of a completely assembled high-quality draft of the rice sequence represents very significant progress. Sequencing is the only way to establish a precise catalogue of all the rice genes and to provide the necessary background for their annotation and functional characterization. The elaborated draft provides the necessary background for the positional cloning of agronomically important genes both in rice and in other cereals. The availability of drafts of both an indica and a reference japonica sequence constitutes a tremendous resource, which considerably facilitates positional cloning.Finally, the release of the fully assembled and precise rice genome sequence opens the way for the next steps in functional genomics, the development of DNA chips, proteomics and mutant collections, and the analysis of multigene families.

The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis

Small segments of rice genome sequence have been compared with that of the model plant Arabidopsis thaliana and with several closer relatives, including the cereals maize, rice, sorghum, barley and wheat. The rice genome is relatively stable relative to those of other grasses. Tandem gene duplication/deletion is particularly common, but other types of deletions, inversions and translocations also occur.The many thousands of small genic rearrangements within the rice genome complicate but do not negate its use as a model for larger cereal genomes. Compared to other grasses and cereal crops, rice has a small genome, a large research community, and exceptional agricultural importance.Rice researchers have developed a comprehensive array of physiological, molecular, genetic, and genomic tools that allow the precise characterization of rice genome organization and gene function.The landmark draft sequences of the indica andjaponica rice genomes published in 2002, alongwith the more complete draft sequence that is beingrapidly developed by the International Rice GenomeSequencing Project, have provided a powerful new resource for studies in rice.

Most of the rearrangements that have been detected in comparisons between rice and another cereal have occurred in the other cereal rather than in rice. Hence, rice may contain a relatively stable genome that reflects the ancestral grass genome better than do the genomes of other cereals.

Genome-wide intraspecific DNA-sequence variations in rice

Genome-wide comparative analysis of the DNA sequences of two major cultivated rice subspecies, Oryzasativa L. sspindica and Oryzasativa L. ssp japonica, have revealed their extensive microcolinearity in gene order and content. Intraspecific sequence polymorphisms commonly occur in both coding and non-coding regions.These variations often affect gene structures and may contribute to intraspecific phenotypic adaptations. The past two years have been a time of harvest for rice genome research. In April of 2000, Monsanto announced that it had produced a draft sequence of the rice genome, and that it would share the data with individual researchers and the International Rice Genome Sequencing Project (IRGSP)

consortium.These advances have greatly contributed towards the construction of a minimal tilling path of the large-inserted bacterial artificial chromosome (BAC) or P1-derived artificial chromosome (PAC) clones that are being used to completely sequence the rice genome.Cultivated indica varieties have dispersed throughout the tropics and subtropics from Eastern India, whereas japonica varieties moved northward from Southern China and developed into temperate ecotypes.Indica and japonicaseparated more than 1 million years ago. Intraspecific phenotypic variations, including growth, developmental and environmental adaptations, are apparent between these two cultivated subspecies.As part of the efforts towards these goals, the IRGSP aimed to produce a high-quality genome sequence of the japonica variety 'Nipponbare' by adapting a clone-byclone strategy. Various molecular studies have consistently shown a distinct difference between indica and japonica in the quantification of genomic DNA and repetitive sequence. A set of restriction fragment length polymorphism (RFLP) markers that were used in the construction of both maps allow comparisons between the indica and the indica _ japonica linkage maps.

Various molecular studies have consistently shown a distinct difference between indica and japonica in the quantification of genomic DNA and repetitive sequence. Nevertheless, extensive genomic colinearity has also been demonstrated between rice subspecies by early genetic and recent physical mapping. A high-density genetic linkage map for rice has been constructed with 2275 markers using a single F2 indica _ japonica population. The genome-wide comparative information abou the chromosome organization and sequence polymorphism of two closely related rice subspecies has important implicationsfor the development These intraspecific phenotypic adaptations will impact on molecular rice breeding of new molecular markers for geneticmapping.

Apart from developing new molecular markers for rice molecular breeding, detailed structural and functional analyses of sequence variations between the two major rice subspecies may eventually lead to a molecular understanding of intraspecific variations in phenotype and adaptations.

Impact of high- temperature stress on rice plant and its traits related to tolerance

The predicted 2–4 °C increment in temperature by the end of the 21st Century poses a threat to rice Production. Booting and flowering are the stages most sensitive to high temperature, which may sometimes lead to complete sterility.

Humidity also plays a vital role in increasing the spikelet sterility at increased temperature. Significant variation exists among rice germplasms in response to temperature stress.Plant architecture can play an important role in hightemperature stress tolerance.The reduced evaporation from the anther will ensure swelling of the pollen grains, an important trait for anther dehiscence. The reduced height may also enhance resistance to lodging, thus giving moreresilience to the crop against indirect stresses such as floods, which will become more frequent due to global warming. The future, which will have significant effects onvarious plants. As a C3 plant, rice will certainly benefit from this increase in CO2, mainly through reduced photorespiration. A positive role of CO2 enrichment has also been shown for biomass accumulation, tillering, panicles per plant and grain yield ofrice.So the projected rise in CO_2 concentration will be advantageous in some ways for ricegrowth and development. CO_2 level due to closure of stomata, which in turn, may also reduce the critical air temperature for spikelet sterility. The global rise in temperature will also increase the severity of other environmental stresses such as floods and drought.

Mitigating strategies for the forthcoming warmer climate

The availability of high-density genetic and physical maps, expressed sequence tags (ESTs), genomic sequences and mutant stocks such as T-DNA insertional mutants.The study on the genetic basis of heterosis has received significant attention in recent years. In this study, using a set of introgression lines (ILs) and corresponding testcross F_1 populations, we investigated heterotic loci (HL) associated with six yield-related traits in both Oryzasativa L. subsp. indica and japonica.

International Rice Genome Sequencing Project: the effort to completely sequence the rice genome

The International Rice Genome Sequencing Project (IRGSP) involves researchers from ten countries who are working to completely and accurately sequence the rice genome within ashort period. The IRGSP works to promote the development of rice and cereal genomics in addition to producing genome sequence data.Rice is a wonderful plant. It feeds about one half of the world's population, mainly in Asia, Africa, and South America.The basic tool that links information in the nucleotide sequence to phenotypic traits throughout the rice genome sequencing project. The first step in understanding rice atthe DNA level is to make a linkage map based on polymorphisms within DNA sequences, such as restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs) and cleaved amplified polymorphic sequences (CAPSs). The japonica cultivar 'Nipponbare' had already been used by the Rice Genome Research Program as a resource for extensive EST sequencing, and the construction of a dense linkage map and a YAC physical map. Rice sequencers realize that many scientists working in plant breeding, plant molecular genetics, plant molecular biology, and bioinformatics are awaiting the publication ofthe complete, high-quality rice genome sequence.The IRGSP aims to maintain high sequence quality standards, to publicize all sequence information as soon as possible and to complete the whole genome sequence in the shortestpossible time.

COMPUTATIONAL STUDY

Computational Systems Biology Study for Understanding Salt Tolerance Mechanism in Rice

Salinity is one of the most common abiotic stresses in agriculture production. Salt tolerance of rice (*Oryzasativa*) is an important trait controlled by various genes. The mechanism of rice salt tolerance, currently with limited understanding, is of great interest to molecular breeding in improving grain yield. In this study, a gene regulatory network of rice salt tolerance is constructed using a systems biology approach with a number of novel computational methods. We developed an improved volcano plot method in conjunction with a new machine-learning method for gene selection based on gene expression data and applied the method to choose genes

related to salt tolerance in rice. The results were then assessed by quantitative trait loci (QTL), co-expression and regulatory binding motif analysis. The selected genes were constructed into a number of network modules based on predicted protein interactions including modules of phosphorylation activity, ubiquity activity, and several proteinase activities such as peroxidase, aspartic proteinase, glucosyltransferase, and flavonol synthase. All of these discovered modules are related to the salt tolerance mechanism of signal transduction, ion pump, abiscisic acid mediation, reactive oxygen species scavenging and ion sequestration. We also predicted the three-dimensional structures of some crucial proteins related to the salt tolerance QTL for understanding the roles of these proteins in the network. Our computational study sheds some new light on the mechanism of salt tolerance and provides a system biology pipeline for studying plant traits in general. Salinity is one of agriculture's most crucial problems in large parts of the world. Soil salinity is a major abiotic stress, which limits rice production in about 30% of the rice-growing area worldwide. Some traditional cultivars and landraces have been identified as tolerant to abiotic stresses,despite their undesirable agronomic traits such as tall plant stature, photosensitivity, poor grain quality and low yield. For example, Pokkali, an Indian landrace, can maintain high K+/Na+ ratio inshoot in a high salinity environment, and it could be a donor of salt-tolerance strains in breeding programs.The high-salinity environment mainly disrupts the ironic and osmotic equilibrium of cells, and as a result, genes in several pathways are activated in response to high sodium concentration. Multiple sources of data can enhance the understanding of salt tolerance. The genetic variations of different rice responses to salt stress may shed some light on the roles of various genes in salttolerance. The availability of rice genome sequencing further paved the way for in-depth study of rice salt tolerance.*Oryzasativa* microarray gene expression data have provided information on regulatory networks of salinity response.

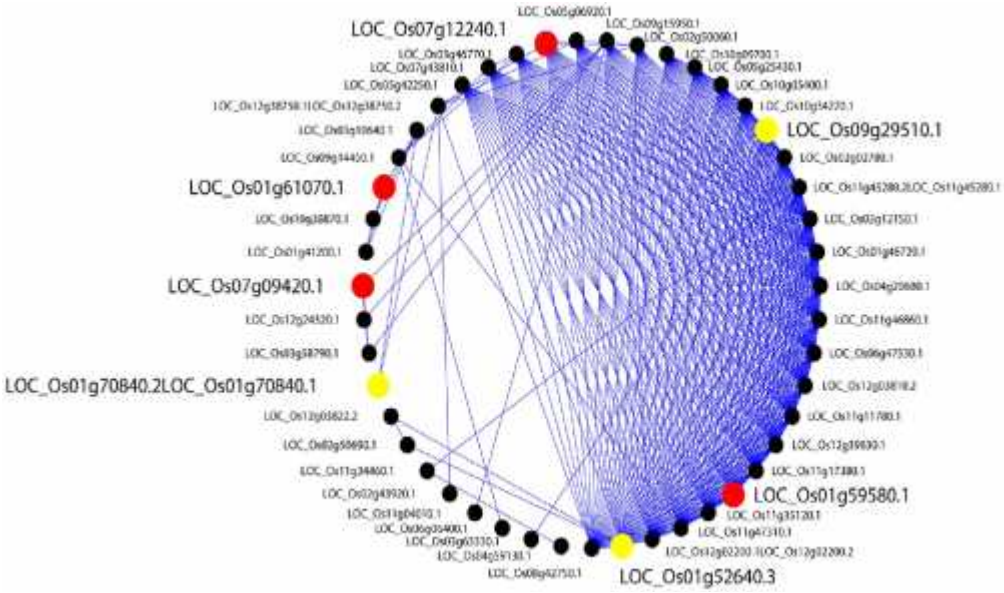


Fig 1: (The largest module in the salt tolerance protein interaction network. Black nodes indicate genes covered by QTLs. Yellow nodesindicate genes covered by extended QTLs.)

Salt tolerance mechanism of root tissue in rice. As there are commonly three samples in just onegenotype of one condition on microarray experiments in contrast to tens of thousands of probe sets, it is a great challenge todetermine feature selection on this small-sample, but highdimension data. Classical statistical feature selection methods, such as t-test assume the samples follow some specific distribution as its hypothesis; however, the limited number of samples narrows the usage of these statistical methods. From the feature selection prospective, the volcano plot method uses two dimensions of fold change and t-test p-value to select genes in microarray analysis. Itis a fast, simple, and widely used methods.

Computational Prediction of Rice (*Oryzasativa*) miRNA Targets:-

Bioinformatic approaches have complemented experimental efforts to inventorizeplant miRNAtargets.The global computational analysis of rice (*Oryzasativa*) transcriptome to generate a comprehensive list of putative miRNA targets.It was observed that more than half (55%) of the targets were conserved between *O. sativaindica*and *O.sativa japonica*. Members of 31 miRNA families were found to possess conserved targets between rice and at least one of other grass family members.MicroRNAs (miRNAs), a class of 22-nucleotide noncoding transcripts, have been shown to play a significant role in plant biology as negative regulatorsm of

gene expression. We also knowthat how the growth and development of rice could be influenced by miRNA-mediated regulation However, despite the availability of whole genome sequences of two subspecies (*indica*and *japonica*), and robust and abundant genomic resources from rice as well as a number of species belonging to the same Poaceae family, a complete repertoire of rice genes regulated by miRNA mediation is yet to be established.

Prediction and analysis of rice miRNA Targets:-

Open access rice sequence data include nucleotide sequence entries, amino acid sequences, and unigenes.Since our work was confined to computational analysis, we wanted to avoid the input that might containpredicted mRNAs and false joining of expressed sequence tags (ESTs). Hence, we opted for the experimentally derived set of rice full-length mapped and annotated cDNA sequences. Out of 242 rice miRNA sequences available in the miRBase database (*24*), the miRanda-based methodology predicted 228 miRNA sequences to have targets among 32,127 full-length cDNA sequences explored.

Conservation of target sequences between rice Subspecies

The cultivated rice (*O. sativa*) is classified into two primary subspecies, *indica*and *japonica*, based onthe morphological and biochemical characters, hybrid sterility, and molecular analyses (*25–28*). Bothsubspecies are the products of separate domestication events from the ancestral species in addition todifferential genome sizes (*indica*466 Mb and *japonica* 389 Mb). *Indica*(tropical) and *japonica* (temperate)have adapted to contrastingly different eco-geography experiencing independent genetic variation for 0.44million years, requiring extensive readjustments in genetic regulatory make-up (*31 , 32*).

For instance, characteristics like photosensitivity, period of cultivation, and grain features greatly differ between *indica*and *japonica* rice cultivars. Hence, *indica*and *japonica* subspecies provide an excellent platformto assess the conservation of the miRNA–target pairsin rice. In this study, homologous *indica*sequences ofevery *japonica* rice miRNA target sequence were obtainedby BLAST analysis.Plants, sessile creatures, need to deal with a variety of stimuli, particularly stress, from the

biotic and abiotic environments, often in a tissue- or stage-specific fashion. These responses are complex but are under stringent regulation (*35–37*). It is therefore plausible that many more hitherto unknown traits are regulated by miRNAs albeit with an effect not as dramatic as observed in the case of transcription factors.

STRESS RESISTANCE GENE:-

Stress resistance genes are classified into 2 types.

1.Biotic

2.Abiotic

ABIOTIC-In ecology and biology,abiotic components are non-living chemical and physical factors in the environment which affect ecosystems.

Example:-water,light,wind,soil,humidity,minerals,gases.

Factors:-Affect the ability of organisms to survive, reproduce; help determine types and numbers of organisms able to exist in environment; limiting factors restrict growth.

Affects:-Individual of a species, population, community, ecosystem, biome, biosphere.

Abiotic resourcfactors are living or once-living organisms in the ecosystem.

BIOTIC:- Biotic factors are living or once-living organisms in the ecosystem. These are obtained from the biosphere and are capable of reproduction.

Factors:-

Living things that directly or indirectly affect organisms in environment; organisms, interactions, waste; parasitism, disease, predation.

Affects:-Individual of a species, population, community, ecosystem, biome, biosphere.

BASIC DISEASE RESISTANCE CONCEPTS:-

1-Non –host resistance

2-Race non –specific resistance

3-Race –specific resistance

4-Basal defence

ABIOTIC STRESSES:-

Situations when enviornmental stimuli that normally influence plant normally influence plant development growth,andproductivity,exceed thresholds (species –specific),damaging the plant.

1-Drought

2-Cold (chilling and freezing)

3-Salt

4-Heavy metals

5-Heat shock

6- Anoxia

7-Nutrient stress

Plant responses to drought stress:-(generally conserved across species)

- Development:-

  Growth reproduction

  Alternations in flowering times

  Increase in the root/shoot ratio

- Morphological adaptations:-

  Stomatalclousure

  Wilty

  Abscission

- Physiological changes:-

  Decrease in the transipiration

  Reduction of water potential

  DROUGHT STRESS:-

  1-Membrane proteins(water,channelproteins,transporter) FUNTIONAL ROTEIN

  2-proteinases(cytoplasm,chloroplast)

  3-protection of macro molecule(chaperons, LEA proteins,)

  4-osmoprotectant synthases(prolineglybetane,sugar)

  5-detoxification enzymes(GST,Esh,SOD)

  6-PI turnover(phospholipase c,PIP5K,DGK,PAP)

  7-Protein phosphatises(PTP)

  8-Protein kinase(MAPK,MPKKK,CDPK,S6K) REGULATORY PROTEIN

  9-Transcription factors(MYC,MYB,Bzip,EREBP/AP2)

CHAPTER-III

3. MATERIALS AND METHODS

DATABASES

PubMed

A database of citations and abstracts for biomedical literature from MEDLINE and additional life Science journals. PubMed Central is a digital archive of full-text biomedical and life Sciences journal literature, including clinical medicine and public health.

Figure 2: Homepage of PubMed

PubMed Central (PMC)

PubMed Central is a free digital database of full-text scientific literature in biomedical and life sciences. It grew from the online EntrezPubMed biomedical literature search system. PubMed Central was developed by the U.S. National Library of Medicine (NLM) as an online archive of biomedical journal articles.

The full text of all PubMed Central articles is free to read, with varying provisions for reuse. Some participating publishers delay the release of their articles on PubMed

Central for a set time after paper publication (often six months).As of January 2013, the archive contains approximately 2.6 million items, including articles, editorials and letters. It appears to be growing by at least 7% per year. As of September 2004, PubMed Central, PubMed, and related NLM services were handling approximately 1,300 hits per second, and supplying 1.3 terabytes of data per day.

Figure 3 : Home page of PubMed Central

NCBI

The National Centre for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. The NCBI is located in Bethesda, Maryland and was founded in 1988 through legislation sponsored by Senator Claude Pepper.

The NCBI houses a series of databases relevant to biotechnology and biomedicine. Major databases include GenBank for DNA sequences and PubMed, a bibliographic database for the biomedical literature. Other databases include the NCBI Epigenomics database. All these databases are available online through the Entrez search engine.

NCBI is directed by David Lipman, one of the original authors of the BLAST sequence alignment program and a widely respected figure in bioinformatics

Figure 4 : Homepage of NCBI

Genbank

The GenBank sequence database is an open access, annotated collection of all publicly available nucleotide sequences and their protein translations. This database is produced and maintained by the National Center for Biotechnology Information (NCBI) as part of the International Nucleotide Sequence Database Collaboration (INSDC). The National Center for Biotechnology Information is a part of the National Institutes of Health in the United States. GenBank and its collaborators receive sequences produced in laboratories throughout the world from more than 100,000 distinct organisms. In the more than 30 years since its establishment, GenBank has become the most important and most influential database for research in almost all biological fields, whose data are accessed and cited by millions of researchers around the world. GenBank continues to grow at an exponential rate, doubling every 18 months.

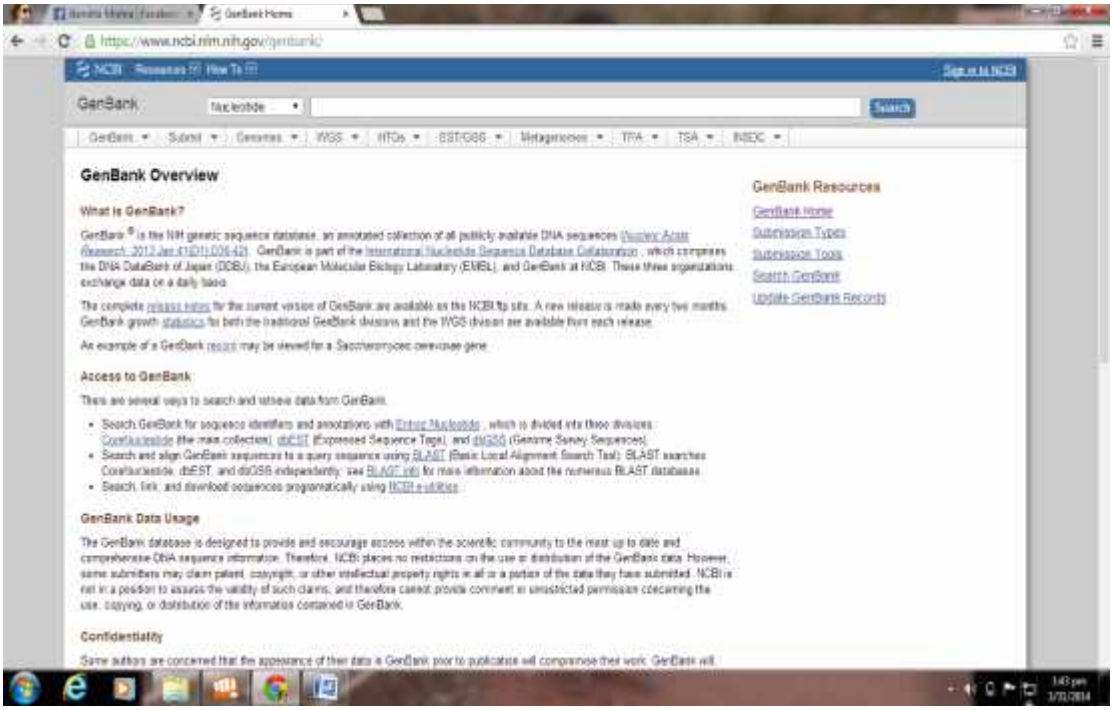


Figure 5 : Homepage of GenBank

UNIPROT KB

UniProtKB/Swiss-Prot is a manually annotated, non-redundant protein sequence database. It combines information extracted from scientific literature and biocurator-evaluated computational analysis. The aim of UniProtKB/Swiss-Prot is to provide all known relevant information about a particular protein. Annotation is regularly reviewed to keep up with current scientific findings. The manual annotation of an entry involves detailed analysis of the protein sequence and of the scientific literature. Sequences from the same gene and the same species are merged into the same database entry. Differences between sequences are identified, and their cause documented (for example alternative splicing, natural variation, incorrect initiation sites, incorrect exon boundaries, frameshifts, unidentified conflicts). A range of sequence analysis tools is used in the annotation of UniProtKB/Swiss-Prot entries. Computer-predictions are manually evaluated, and relevant results selected for inclusion in the entry. These predictions include post-translational modifications, transmembrane domains and topology, signal peptides, domain identification, and protein family classification. Annotated entries undergo quality assurance before inclusion into UniProtKB/Swiss-Prot. When new data becomes available, entries are updated

Figure 6 : Homepage of UniProtKB

PDB (Protein Data Bank)

The Protein Data Bank (PDB) is a repository for the 3-D structural data of large biological molecules, such as proteins and nucleic acids. The data, typically obtained by X-ray crystallography or NMR spectroscopy and submitted by biologists and biochemists from around the world, are freely accessible via the websites of its member organisations (PDBe, PDBj, and RCSB). The PDB is overseen by an organization called the Worldwide Protein Data Bank, wwPDB.

Figure 7 : Homepage of PDB

COMPUTATION OF PHYSICAL AND CHEMICAL PARAMETERS

ProtParam tool

ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

Figure 8: Homepage of Protparam

SEQUENCE ALIGNMENT TOOLS.

BLAST:

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

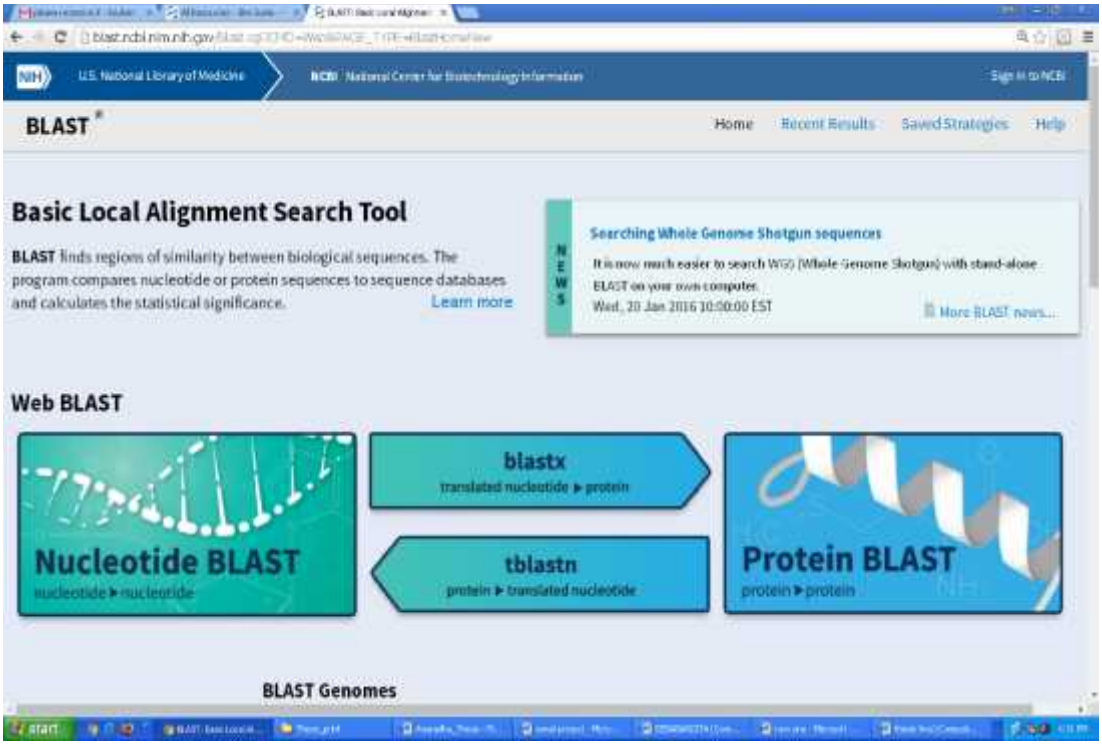


Figure 9: Homepage of Basic Local Alignment Tool (BLAST)

Clustal X

In Bioinformatics Clustal is a series of widely used computer programs for multiple sequence alignment.There have been many incarnations of Clustal that are listed below:

- **Clustal:** The original software for progressive alignment based on a phylogenetic tree.
- **ClustalV:** A rewrite of the original Clustal package that included phylogenetic tree reconstruction on the final alignment for the first time.
- **ClustalW:** command line interface[
- **ClustalX:** This version has a graphical user interface.
- **Clustal Omega:** Command line-only program.

The papers describing the clustal software have been very highly cited, with two appearing in a list of the most cited papers of all time.

The more recent version of the software available for Windows, Mac OS, and Unix/Linux. This program is available from the ClustalHomepageor European Bioinformatics Institute ftp server.

Figure 10: Homepage of Clustal X

DOMAIN ANALYSIS

InterPro is a resource that provides functional analysis of protein sequences by classifying them into families and predicting the presence of domains and important sites. To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium.InterProcombines signatures from multiple, diverse databases into a single searchable resource, reducing redundancy and helping users interpret their sequence analysis results. By uniting the member databases, InterPro capitalises on their individual strengths, producing a powerful diagnostic tool and integrated resource.

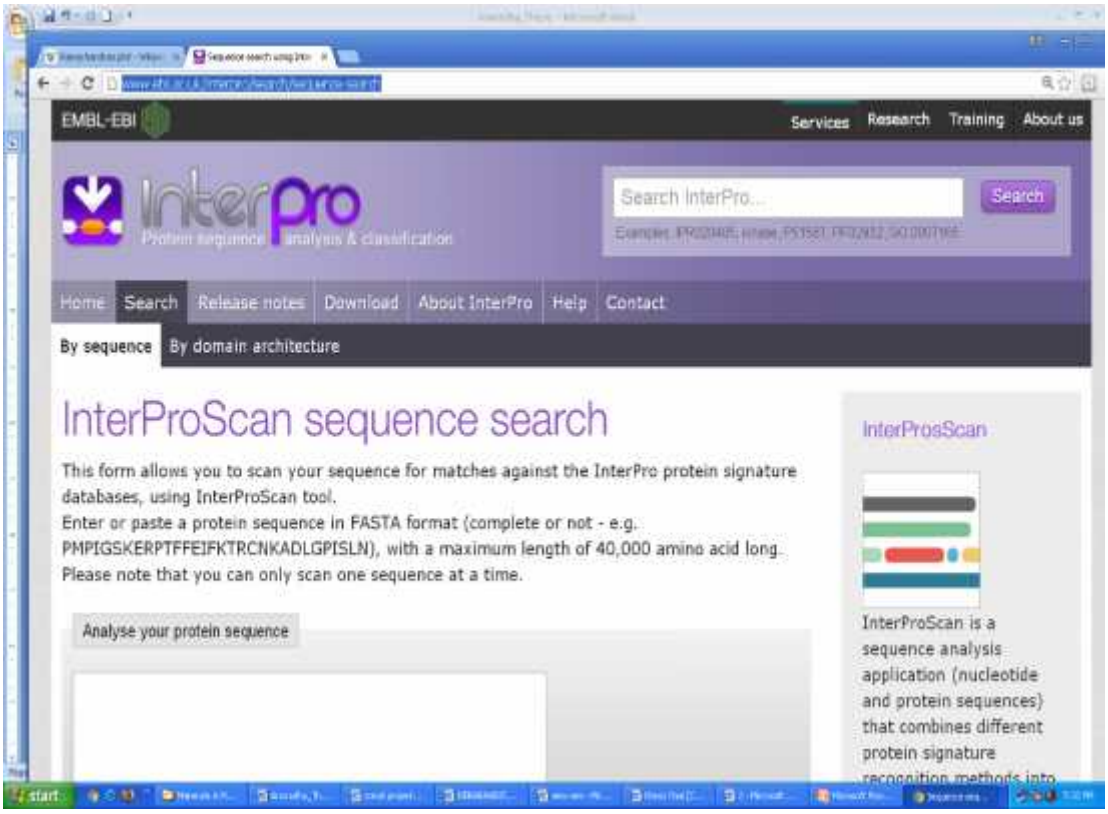


Figure11 : Homepage of InterProScan

PHYLOGENETIC ANALYSIS

MEGA (Molecular Evolutionary Genetics Analysis)

MEGA is an integrated tool for conducting the sequence alignment, inferring phylogenetic trees, estimating divergence times, mining online databases, estimating rates of molecular evolution inferring ancestral sequences & testing evolutionary hypotheses. MEGA is used by biologist in a large number of laboratories for reconstructing for the evolutionary histories of species and inferring the extent and nature of the selective forces shaping the evolution of genes and species.

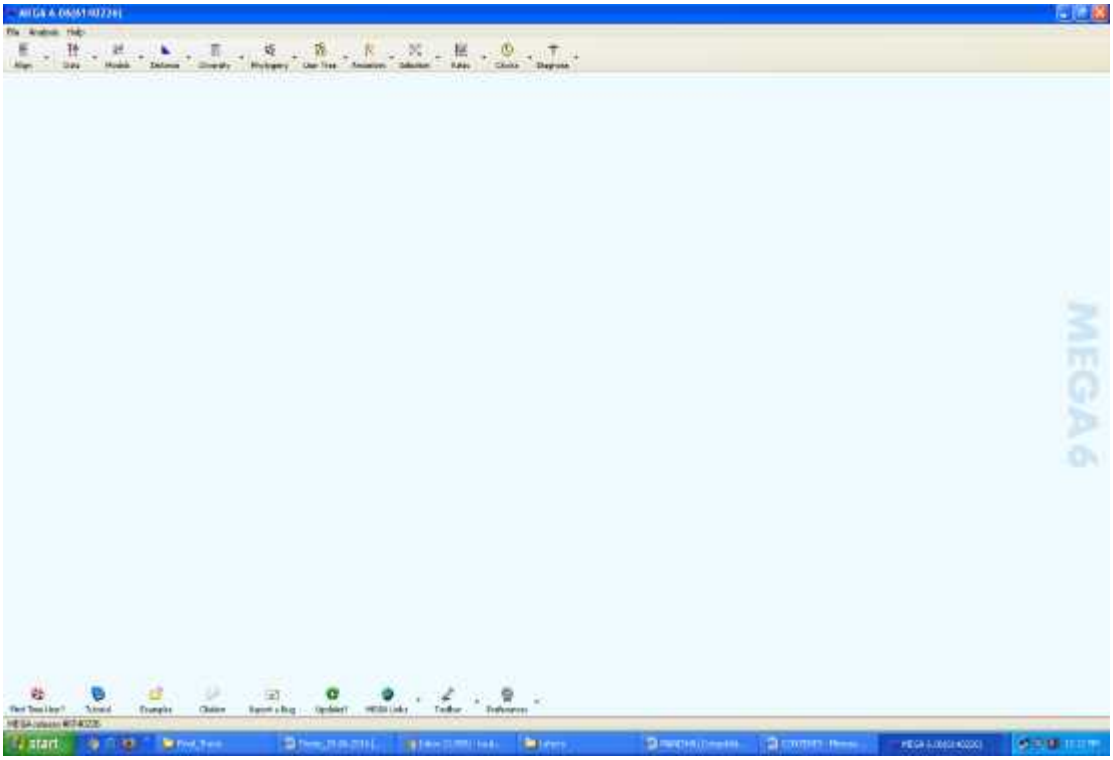


Figure 12: Homepage of MEGA

STRUCTURE PREDICTION AND VALIDATION

SWISS-MODEL

SWISS-MODEL is a structural bioinformatics web-server dedicatedto homology modeling of protein 3D structures. Homology modeling is currently the most accurate method to generate reliable three-dimensional protein structure models and is routinely used in many practical applications. Homology (or comparative) modelling methods make use of experimental protein structures ("templates") to build models for evolutionary related proteins ("targets").Today, SWISS-MODEL consists of three tightly integrated components: (1) The SWISS-MODEL pipeline - a suite of software tools and databases for automated protein structure modelling, (2) The SWISS-MODEL Workspace - a web-based graphical user workbench, (3) The SWISS-MODEL Repository - a continuously updated database of homology models for a set of model organism proteomes of high biomedical interest.[

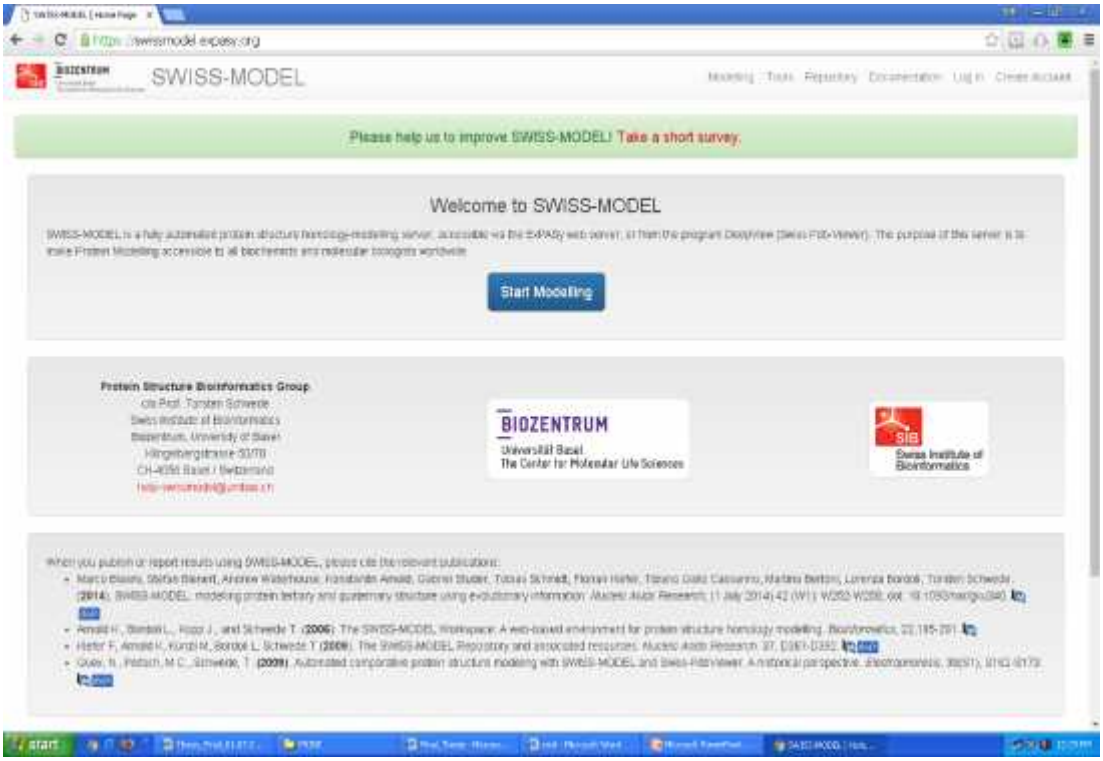


Figure 13: Homepage of SwissModel

RaptorX: a Web Portal for Protein Structure and Function Prediction

This web portal for protein structure and function prediction is developed by Xu group, excelling at secondary, tertiary and contact prediction for protein sequences without close homologs in the Protein Data Bank (PDB). Given a protein sequence, RaptorX predicts its secondary and tertiary structures as well as contact map, solvent accessibility, disordered regions and binding sites. RaptorX assigns the following confidence scores to indicate the quality of a predicted 3D model: P-value for the relative global quality, GDT (global distance test) and uGDT (un-normalized GDT) for the absolute global quality, and RMSD for the absolute local quality of each residue in the model. RaptorX-Binding predicts the binding sites of a protein sequence, based upon the predicted 3D model by RaptorX.

Figure 14: Homepage of Raptorx

Python2.6

Python is a dynamic object-oriented programming language that can be used for many kinds of software development. It offers strong support for integration with other languages and tools, comes with extensive standard libraries, and can be learned in a few days. Many Python programmers report substantial productivity gains and feel the language encourages the development of higher quality, more maintainable code. Python runs on Windows, Linux/Unix, Mac OS X, OS/2, Amiga, Palm Handhelds, and Nokia mobile phones. Python has also been ported to the Java and .NET virtual machines.Python is distributed under an OSI-approved open source license that makes it free to use, even for commercial products.Python is a widely used general-purpose, high-level programming language. Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than would be possible in languages such as C. The language provides constructs intended to enable clear programs on both a small and large scale. Python supports multiple programming paradigms, including object-oriented, imperative and functional programming or procedural styles. It features a dynamic type system and automatic memory management and has a large and comprehensive standard library.

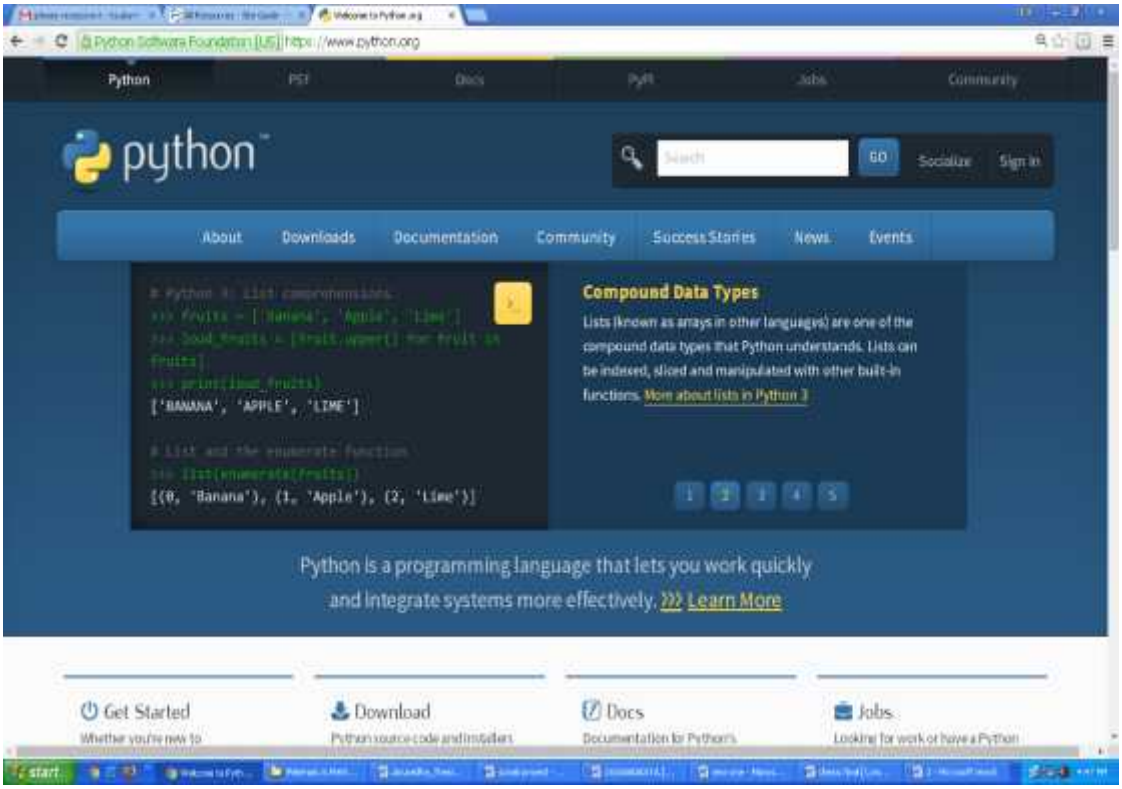


Figure 15: Homepage of python

PyMol

PyMOL is one of a few open-source visualization tools available for use in structural biology. PyMOL is created by Warren Lyford DeLano and commercialized by DeLano Scientific LLC, which is a private Software Company dedicated to creating useful tools that become universally accessible to scientific and educational communities. It can produce high-quality 3D images of small molecules and biological macromolecules, such as proteins. The Py portion of the software's name refers to the fact that it extends, and is extensible by the Python programming language.

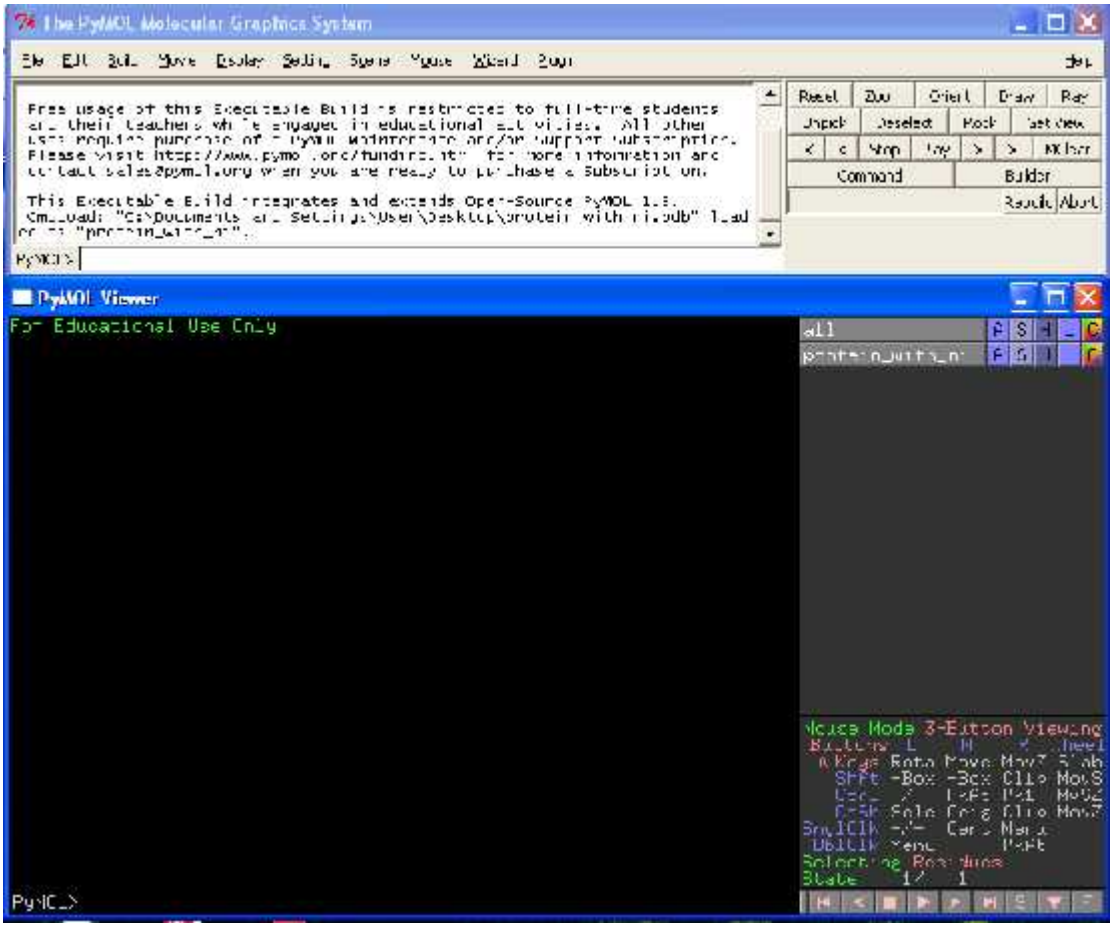


Figure 16: Homepage of PYMOL

RAMPAGE:

A Ramachandran plot (also known as aRamachandran diagram or a [φ,ψ] plot), originally developed in 1963 by G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan,[is a way to visualize energetically allowed regions for backbonedihedral angles ψ against φ of amino acid residues inprotein structure. The figure at left illustrates the definition of the φ and ψ backbone dihedral angles (called φ and φ' by Ramachandran). The ω angle at the peptide bond is normally 180°, since the partial-double-bond character keeps the peptide planar. The figure at top right shows the allowed φ,ψ backbone conformational regions from the Ramachandran et al.

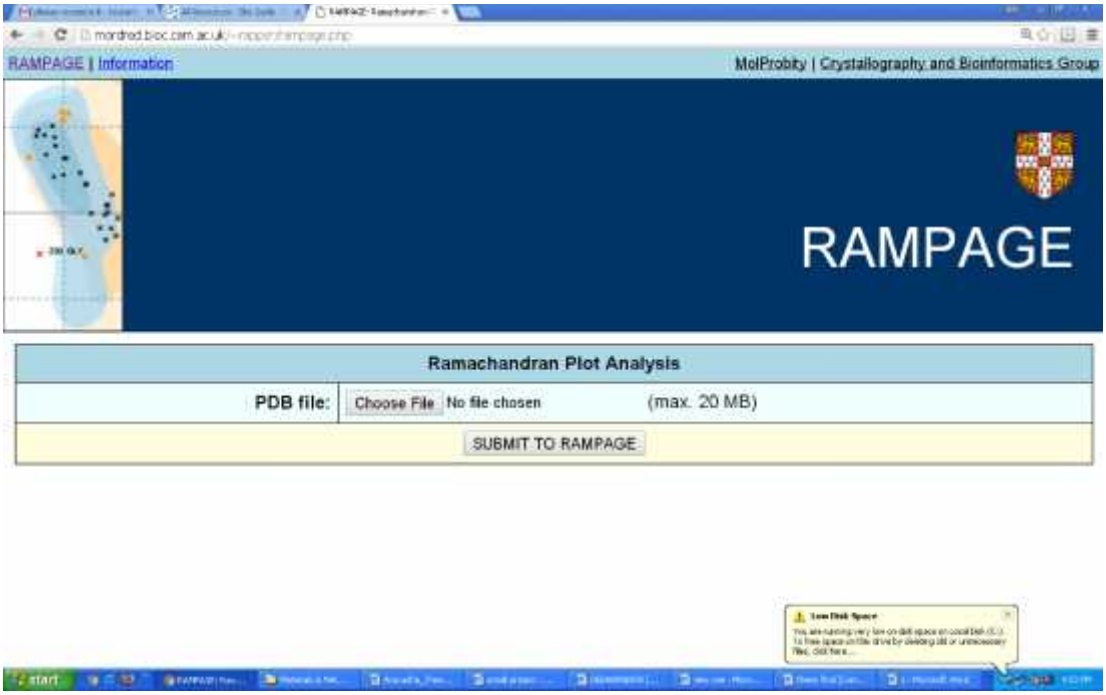


Figure17: Homepage of RAMPAGE

METHODS

LITERATURE ANALYSIS

Review of articles on abiotic stress mechanism in rice

The literature databases like Pubmed/PMC of National Centre Biotechnology Information (NCBI) have been searched to collect the study/experiments conducted in different laboratories across the globe in and around different abiotic stress genes in *Oryzasativa, japonica*.

Genes and Stress response

Retrieval of Primary Protein Sequence of DNMTs from UNIPROT- KB Database

Stress resistance and droughttolerance proteinsinvolved in abiotic stress in *Oryzasativa, japonica*were retrieved fromthe UniProtKB (Universal Protein Resource Knowledge Base) protein sequence database for further analysis and prediction.

PHYSICAL AND CHEMICAL PARAMETERS

Computation of Physicochemical Parameters of DNMTs

ProtParam tool of Swiss Institute of Bioinformatics (SIB) was used to compute various physical and chemical parameters of stress resistance proteins of rice were retrieved from Swiss-Prot. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

FUNCTIONAL ANALYSIS OF STRESS RESISTANCE AND DROUGHT TOLERANCE PROTEINS

Identification of functional domain using InterproScan-EBI

It is very pertinent to mention that the functional delivery of a protein always depend upon the presence or absence of a particular domain/motif. In this study *in silico*approach was followed using InterProScan of EBI to extract specific signature of sequence(domain/motif) responsible for stress resistant and other related functions.

DIVERGENCE ANALYSIS OF STRESS RESISTANCE AND DROUGHT TOLERANCE PROTEINS

Molecular phylogenetic study using MEGA 5.1

The stress resistance and drought tolerance proteinsin fasta format are saved as “FASTA file”. Then open the MEGA 5.1 and .fasta file wasopened, for alignment. From theEditoption all the sequences were selected and was aligned by align by muscle method, then compute .By data option export the alignment file and saved it as mega file. Again mega file was opened for finding out the total, conserved and variable sites in the sequences. Then phylogeny option was chooses and the tree was constructed by maximum likelihood method as they shared low sequence similarity. Then the constructed was saved as pdf or png file.

Analysis
  Analysis ------------------------ Phylogeny Reconstruction
  Statistical Method -------------- Maximum Likelihood
Phylogeny Test
  Test of Phylogeny --------------- Bootstrap method
  No. of Bootstrap Replications --- 1000
Substitution Model
  Substitutions Type -------------- Amino acid
  Model/Method -------------------- Jones-Taylor-Thornton (JTT) model
Rates and Patterns
  Rates among Sites --------------- Uniform rates
Data Subset to Use
  Gaps/Missing Data Treatment ----- Complete deletion
Tree Inference Options
  ML Heuristic Method ------------- Nearest-Neighbor-Interchange (NNI)
  Initial Tree for ML ------------- Make initial tree automatically (Default - NJ/BioNJ)
  Branch Swap Filter -------------- Very Strong
System Resource Usage

STRUCTURE ANALYSIS AND PREDICTION

The function of any protein or enzyme is always dependent upon its correct 3D conformation. The 3D structures of stress resistance and drought tolerance proteinswere searched in PDB, PDBe and PDBj to get their coordinate files(.pdb). The available 3D structures were analysed further, i.e., number of chains, types of domain, location of domains, active sites, binding of ligands, types of interactions using PDBsum. In case of non-availability of structural information, the protein structure was modelled/predicted using the method homology modelling/comparative modelling using SwissModel.

Structure Prediction of Stress resistance and drought tolerance proteins

The stress resistance and drought tolerance proteins of *Oryzasativa*,*japonica*modelled through the SwissModel and Raptorx.

MOLECULAR VISUALISATION OF PREDICTED STRUCTURE

Model structure visualisation using PyMol

Visualization of the predicted structure of stress resistance and drought tolerance proteinswere done using PyMol with the following steps.

- Go to the file
- Open the pdb file
- Click the H button then select cartoon
- Go to C button
- Select “by ss”
- Select “Helix Sheet Loop”

STRUCTURE VALIDATION

Ramachandran Plot Analysis using RAMPAGE

The predicted protein structure of stress resistance and drought tolerance proteinsvalidated using RAMPAGE server

CHAPTER-IV

IV. RESULTS AND DISCUSSIONS

Genes responsible of Stress resistance

PMC/Pubmed literature database were searched and reviewed to understand the different mechanism of stress resistance and its causes. The different genes and their corresponding chromosomal location in *Oryzasativa*,*japonica*have been listed.

Table 1. List abiotic stress genes of *Oryzasativa*,*japonica*

Ch 1	Ch 2	Ch 3	Ch 4	Ch 5	Ch6
OsDi19-3	OsDi19-4	GL3.2	OsHyPRP06	OsDi19-1	OsCEP7
OsDi19-5	OsMULE-JS3	OsCEP1	HyPIGRP06	OsDi19-2	OsRLCK193
OsMULE-JSS1	OsMULE-JS4	OsCEP2	OsHYPRP6	OsDi19-6	OsCrRLCK113
OsMULE-JSS1	OsMULE-IS8	OsCEP3	HyPIGRP6	OsMULE-1S9	OsRLCK195
OsMULE-JS2	OsHyPRP02	OsHyPRPO3	OsHyPRP07	OsRLCK188	OsRLCK196
OsMULE-1S6	HyPIGRP02	HyPIGRPO3	HyPIGRP07	OsRLCK189	OsRLCK198
OsMULE-1S7	OsHYPRP2	OsHYPRP3	OsHYPRP7	OsRLCK156	OsRLCK199
OsPP2C01	HYPIGRP2	HYPIGRP3	HyPIGRP7	OsRLCK190	OsRLCK200
PP2C01	OsPUB55	LTP1	OsHYPRP08	OsRLCK191	OsRLCK201
OsPP2C1	OsRLCK81	OsHyPRP04	HyPIGRP08	OsPUB66	OsRLCK202
PP2C1	OsPUB62	HyPIGRP04	OsHyPRP8	OsPP2C49	OsRLCK203
OsPP1	OsRLCK91	OsHyPRP4	HyPUB76	PP2C49	OsRLCK205
OsPP2C02	OsRLCK73	HyPIGRP4	OsPUB77	OsPP76	OsRLCK206
PP2C02	OsPP2C10	OsPUB50	OsEns-73	ROs1C	OsRLCK208
OsPP2C2	PP2C10	OsRLCK93	OsVTE5	ROsc	OsRLCK209
PP2C2	OsPP18	OsPUB57	DML3b	ROSID	OsRLCK210
OsPP2	OsPP2C11	OsRLCK112	RIM2-M158	ROS1D	OsRLCK211
OsPP2C03	PP2C11	OsPUB59	RIM2-M179	Rim2-M176	OsRLCK212
PP2C03	OsPP19	OsRLCK111	RIM2-M181	Rim2-M202	OsRLCK213

OsPP2C3	OsPP2C12	OsPP2C28	RIM2-M184	OsSTA160	OsRLCK214
PP2C3	PP2C12	PP2s28	RIM2-M190	OsSTA161	OsRLCK215
OsPP7	OsPP21	OsPP43	RIM2-M194	UGT	OsRLCK216
OsPP2C04	OsPP2C14	OsNUG2	RIM2-M199		OsRLCK217
PP204	PP2C14	Rim2-M180	RIM2-M203		OsRLCK218
OsPP2C4	OsPP24	Rim2-M186	CHS		OsRLCK220
PP2C4	OsPP2C15	Rim2-M197	UNP1		OsPUB51
OsPP8	PP2C15	FLS	OsCAF1-3		OsRLCK194
OsPP2C05	OsPP25	GID1	OsCAF1-4		OsPUB63
PP2C05	OsPP2C16	OsCAF1-2			OsPUB64
OsPP2C5	PP2C16	OsCAF1Q			OsPUB65
PP2C5	OsPP26				OsPUB70
OsPP9	OsPP2C17				OsRLCK197
OsPP2C07	PP2C17				OsPUB74
PP2C07	OsPP27				OsVTE2
OsPP2C07	OsPP2C18				OsFbox307
PP2C7	PP2C18				Os-F0034
OsPP11	OsPP28				OsFbox308
OsSTK1	OsPP2C19				Os-F0081
OsRLCK25	PP2C19				OsFbox310
UGT	OsPP29				Os-F0057
UGTT706D1	OsPP2C20				OsFbox311
Rim2-M162	PP2C20				Os-F00115
Rim2-M165	OsPP22C21				OsFbox312
Rim2-M169	PP2C21				Os-F00546
Rim2-M171	OsPP31				OsFbox313
Rim2-M172	OsPP2C22				Os-F0104
Rim2-M175	PP2C22				OsFbox316
Rim2M177-	OsPP33				Os-F0369

Rim2-M185	OsPP2C23				OsFbox317
Rim2-M187	PP2C23				Os-F0038
Rim2-M191	OsPP34				OsFbox318
Rim2-M196	OsPP2C24				Os-F0416
Rim2-M205	PP2C24				OsFbox319
Rim2-M206	OsPP35				Os-F0429
Rim2-M207	OsPP2C25				OsFbox320
	PP2C25				Os-F0613
	OsPP36				OsFbox322
	OsPP2C26				Os-F0520
	PP2C26				Os-F0520
	OsPP37				OsFbox323
	OsPP2C27				Os-F0105
	PP2C27				OsFbox324
	OsPP40				Os-F0220
	OsGIRL1-1				OsFbox326
	OsGIRL1-2				Os-F0327
	OsGIRL1-3				OsFbox327
	DML3a				Os- F0126
	UGT				OsFbox328
	Rim2-M163				Os –F0254

OsFbox001	Rim2-M198				DHER
OsFbox1	CHS				UDP-GT
Os-f0015	OSK3				OsENOD93-1
OsFbx1					OsSTA163
OsFbox002					OsSTA164
OsFbox2					OsSTA165

OsFbox003					OsSTA166
OsFbox3					OsSTA167
Os-f0209					OsSTA168
OsFbx3					OsSTA169
OsFbox004					OsSTA170
OsFbox4					OsSTA172
Os-F0387					OsSTA173
OsFbox005					OsSTA174
OsFbox5					OsSTA175
Os-F0747					OsSTA176
OsFbx4					OsSTA177
OsFbox006					OsSTA178
OsFbox6					OsSTA179
Os-F0083					OsSTA180
OsFBX5					OsSTA181
OsSTA12					UGT
OsFbox007					
OsFbox7					
Os-F0351					
OsFbox008					
OsFbox9					
Os-F0016					
OsFbox010					
OsFbox10					
Os-F0156					
OsFBX6					
OsFbox011					
OsFbox11					
Os-F0196					

OsFBX7					
OsFbox012					
OsFbox12					
Os-F0539					
OsFBX8					
OsFbox013					
OsFbox13					
Os-F0590					
OsFBX10					
OSGT47A					
OsFbox014					
OsFbox14					
Os-F0467					
OsFBX11					
OsFbox015					

Ch7	**Ch8**	**Ch9**	**Ch10**	**Ch 11**	**Ch12**
OsFbox337	OsCEP6	OsCEP5	OsHYPRP09	OsMULE-JS5	OsDi19-7
Os-F0302	OsRLCK244	OsFbox467	HYPIGRP09	Os-F0627	Os-F0684
OsFbox338	OsRLCK245	Os-F0397	OsHYPRP9	Os-F0723	Os-F0691
Os-F0537	OsRLCK247	OsFbox468	HYPIGRP9	Os-F0031	Os-F0700
OsFbox339	OsFbox445	Os-F0398	OsHYPRP10	Os-F0079	Os-F0763
Os-F0636	Os-F0550	OsFbox469	HYPIGRP10	Os-F0390	Os-F0799
OsFbox340	OsFbox446	Os-F0272	OsHYPRP11	Os-F0253	Os-F0686
Os-F0124	Os-F0095	OsFbox470	HYPIGRP11	Os-F0690	UNP2
OsFbox341	OsFbox447	Os-Fo504	OsHYPRP12	Fbox-12-Pseus0	PROL-26
Os-F0215	Os-F0225	OsFbox471	HYPIGRP12	Os-F0761	Pro13a-6
OsFbox342	OsFbox448	Os-F0538	OsHYPRP14	Fbox-7	Pro13a.5
Os-F0418	Os-F0428	OsFbox472	HYPIGRP14	Os-F0758	Pro1-24

OsFbox343	OsFbox449	Os-F0599	OsHYPRP15	Os-F0718	Pro13a.3
Os-F0141	Os-F0345	OsFbox473	HYPIGRP15	Os-F0650	Pro1-25
OsFbox344	OsFbox450	Os-F0455	OsHYPRP16	Adh3	Pro1-28
Os-F0183	Os-F0185	OsFbox473	HYPIGRP16	RZ53	Pro1-29
OsFbox345	OsFbox451	Os-F0632	OsHYPRP17	NiFS	Pro1-30
Os-F0495	Os-F0189	OsFbox474	HYPIGRP17	OsFbox594	Pro13a.4
OsFbox346	OsFbox452	Os-F0634	OsHYPRP18	Os-F0295	Qsh
Os-F0078	Os-F0193	OsFbox475	HYPIGRP18	Fbox-1	qDTH-12
OsFbox347	OsFbox453	Os-F0457	Os-F0581	OsFbox592	PMS3
Os-F0072	Os-F0339	OsFbox476	Os-F0711	Os-F0128	LDMAR.
OsFbox348	OsFbox454	Os-F0147	Os-F0772	Fbox-4	Qalcr-12-1
Os-F0035	Os-F0207	OsFbox477	OOs-F0742	OsFbox591	OsGSTZI
OsFbox349	OsFbox455	OsFbox478	Os-F0753	Os-F0435	GSTZ-1
Os-F0252	Os-F0173	Os-F0354	Os-F0664	Fbox-5	OsGSTZ2
OsFbox350	OsFbox456	OsFbox479	Os-F0778	Rim2-M166	GSTZ-
Os-F0212	Os-F0543	Os-F0399	Os-F0771	Rim2-M200	(RrgF12)
OsFbox351	OsFbox457	OsFbox480	OsPUB49	CHI	PsaL
Os-F0298	Os-F0575	Os-F0388	OsRLCK307		OsNek2
OsFbox352	OsFbox458	OsFbox481	OsPUB54		OsISC8
Os-F0525	Os-F0530	Os-F0362	OsRLCK306		RHL
OsFbox353	OsFbox459	OsFbox482	OsPUB58		UNP2
Os-F0501	Os-F0074	Os-F0167	OsPUB61		
OsFbox354	OsFbox460	OsFbox483	OsRLCK285		
Os-F0278	Os-F0200	Os-F0233	OsPUB67		
OsFbox355	OsFbox461	OsFbox484	Rim2-M159		
Os-F0349	Os-F0025	Os-F0373	Rim2-M160		
OsFbox356	OsFbox462	OsFbox485	Rim2-M161		
Os-F0118	Os-F0019	Os-F0149	Rim2-M164		
OsFbox357	OsFbox463	OsFbox486	Rim2-M167		

Os-F0033	Os-F0230	Os-F0355	Rim2-M168		
OsFbox358	OsFbox464	OsFbox487	Rim2-M170		
Os-F0098	Os-F0048	Os-F0548	Rim2-M173		
OsFbox359	OsFbox465	OsFbox488	Rim2-M174		
Os-F0431	Os-F0016	Os-F0562	Rim2-M178		
OsFbox336	OsFbox466	OsFbox489	Rim2-M192		
Os-F0293	Os-F0587	Os-F0053	Rim2-M193		
OsFbox335	OsPUB68	OsFbox490	Rim2-M195		
Os-F0487	OsPUB69	Os-Foo91	FLS		
OsFbox334	Os-F0199	OsFbox491	OsSTA240		
Os-F0361	Os-F0765	Os-F0580	OsCAF1-14		
OsFbox333	Os-F0586	OsFbox492	OsCAF1F		
Os-F0159	Os-F0795	Os-F0175	OsCAF1-15		
OsFbo322x	Os-F0790	OsFbox493	OsCAF11		
Os-F0267	Os-F0779	Os-F0049	OsCAF1-16		
OsFbox331	Os-F0662	OsFbox494	OsCAF1E		
Os-F0273	Os-F0488	Os-F0637	OsCAF1-17		
OsFbox330	OsSTA203	Os-F0709	OsCAF1M		
Os-F0521	OsSTA204	Os-F0736	OsCAF1-18		
OsFbox329	OsSTA205	Os-F0715	OsCAF10		
Os-F0529	OsSTA206	Os-F0740	OsCAF1-19		
Rim2-M204	OsSTA209	Os-F0710	OsCAF1P		
Rim2-M201	OsSTA210	Os-F0789	OsCAF1-20		
Rim2-M189	OsSTA213	Os-F0661	OsCAF1R		
Rim2-M188	OsSTA214	Os-F0448	OsCAF1-21		
Rim2-M183	OsSTA215	Os-F0689	OsCAF1N		
Rim2-M182	OsSTA216	Os-F0783	OSK2		
OsVTE3	OsSTA217	OsPUB52	UGT		
Os-F0300	OsSTA218	OsRLCK280			

Os-F0606	OSTPS29	OsPUB56			
Os-F0793	OSK15	OsRLCK281			
Os-F0764		OsSTA219			
Os-F768		OsSTA220			
OsRLCK243		OsSTA221			
OsRLCK242		OsSTA222			
OsRLCK241		OSK5			
OsRLCK240		OSK20			
OsRLCK239					
OsRLCK238					
OsRLCK237					
OsRLCK236					
OsRLCK235					
OsRLCK234					
OsRLCK233					
OsRLCK232					
OsRLCK231					
OsRLCK227					

CHS					
OsSTA184					
OsSTA185					
OsSTA186					
OsSTA187					
OsSTA189					
OsSTA190					
OsSTA191					
OsSTA194					
OsSTA195					
OsSTA196					
OsSTA197					
OsSTA198					
OsSTA200					
OsSTA201					
UGT					

Table 2. List of stress and drought tolerant genes

Genes	Function
OsVTE3	Important role in enhanced plays role in salt,and drought and cold tolerance gene
OsPP2C10	Overexpression of OsPP2C10 helps in enhanced cold tolerance,drought tolerance and salt tolerance.
OsVTE2	Over expression of transgenic lines to salinity to access functions of tocopherol in rice.
OsDi19-4	Key message The OsDi19 proteins functioned as transcription factors and played crucial roles in response to abiotic stress. Overexpression of OsDi19-4 in rice increased drought tolerance by enhancing ROSscavenging activity. Overexpression of one stress-responsive gene, *OsDi19-4*, in rice resulted in significantly increased tolerance to drought stress compared with the wild type plants. Obviously increased ROS-scavenging ability was detected in the*OsDi19-4*-overexpressing plants under normal and drought stress conditions.
OsDi19-3	Overexpression of OsDi19-4 in rice increased drought tolerance by enhancing ROSscavenging activity. Overexpression of one stress-responsive gene, *OsDi19-4*, in rice resulted in significantly increased tolerance to drought stress compared with the wild type plants.
OMT	Characterization and expression of the OMT gene in salt tolerant and sensitive barley .a c dna encoding an O-methyl transferase OMT was isolated from salt tolerant barley by substraction hybridization with cDNAs of the salt sensitive barley roots as a driver of cDNA.

Table 3. List of stress and drought tolerant proteins

ACCESION NO	FUNCTION	LENGTH	ORGANISIMS	PROTEIN NAME	GENE NAME	SUBCELLULAR LOCATION AND KEY WORD
Q0D576	Involved in the synthesis of tocopherol (vitamin E). Catalyzes the condensation of homogentisate and phytyldiphosphate to form dimethylphytylhydrquinone	379	*Oryzasativa subsp. japonica (Rice)*	Probable homogentisatephytyltransferase 2, chloroplastic	HPT2	Plastids>chloroplastthylakoid membrane Multi pass membrane protein Integral membrane of protein Chloroplast ,membrane ,plastids ,thylakoid
Q67UX7	a protein-serine/threonine phosphate + H_2O = [a protein]-serine/threonine + phosphate.	348	*Oryzasativa subsp. japonica (Rice)*	Probable protein phosphatase 2C 10	Os02g0149800	Transmembrane, Trans membrane helix
B7FA90	Involved in the synthesis of tocopherol (vitamin E). Catalyzes the condensation of homogentisate and phytyldiphosphate to form dimethylphytylhydroquinone	404	*Oryzasativa subsp. japonica (Rice)*	Probable homogentisatephytyltransferase 1, chloroplastic	HPT1	Plastids>chloroplastthylakoid membrane Multi pass membrane protein Integral membrane of protein Transit peptide, Transmembrane, Transmembrane helix Chloroplast ,membrane ,plastids ,thylakoid
Q6H6E6	(N____A)	245	*Oryzasativa subsp. japonica (Rice)*	Protein DEHYDRATION-INDUCED 19 homolog 4	DI19-4	(N_A)
Q5QMP3	(N____A)	246	*Oryzasativa subsp. japonica (Rice)*	Protein dehydration induced-19 homolog 3	DI19-3	(N_A)

Q9XGP7	Catalyzes the stepwise methylation of tricetin to its 3'-mono- and 3',5'-dimethyl ethers. No 3',4',5'-trimethylated ester derivatives are produced. Can use caffeoyl-CoA, 5-hydroxyferulic acid, luteolin, tricetin, quercetin, myrcetin and 7,8-dihydroxyflavone as substrates, but not naringenin, apigenin or kaempferol. The 2,3-double bond and the O-dihydroxyl group of the substrate are both required for catalytic activity of the enzyme	252	*Oryzasativa subsp. japonica (Rice)*	Tricin synthase1	ROMT-15	Cellular component of nucleus
Q6ZD89	Methylates OH residues of flavonoid compounds	368	*Oryzasativa subsp. japonica (Rice)*	Flavone 3'-O-methyltransferase 1	ROMT-9	Cellular component of cytosol

Q7F8T6	Catalyzes the stepwise methylation of tricetin to its 3'-mono- and 3',5'-dimethyl ethers. No 3',4',5'-trimethylated ester derivatives are produced. Can use caffeoyl CoA, 5-hydroxyferulic acid, luteolin, tricetin, quercetin, myrcetin and 7,8-dihydroxyflavone as substrates, but not naringenin, apigenin or kaempferol. The 2,3-double bond and the O-dihydroxyl group of the substrate are both required for catalytic activity of the enzyme	292	*Oryzasativa subsp. japonica (Rice)*	:- Tricin synthase2	ROMT-17	
Q01P69	(N____A)	375	*Oryzasativa subsp. japonica (Rice)*	Uncharacterized protein	Acid_7651	
Q7XXD4	It is a binding site. S-adenosyl-L-methionine; via carbonyl oxygen. S-adenosyl-L-methionine; via amide nitrogen.	375	*Oryzasativa subsp. japonica (Rice)*	Probable inactive methyltransferase Os04g0175900	Os04g0175900	Cellular component of cytosol

Table 4. Catalytic activity, pathways,molecular function and biological function

ACCESSION NO	**FUNCTION**	**CATALYTIC ACTIVITY**	***PATHWAYS(TOCOPHEROL BIOSYNTHESIS)***	**MOLECULAR FUNCTION**	**BIOLOGICAL PROCESS**
Q0D576	Involved in the synthesis of tocopherol (vitamin E). Catalyzes the condensation of homogentisate and phytyldiphosphate to form dimethylphytylhydrquinone	Phytyldiphosphate + homogentisate = diphosphate + 2-methyl-6-phytylbenzene-1,4-diol + CO_2	This protein is involved in the pathway tocopherol biosynthesis, which is part of Cofactor biosynthesis. View all proteins of this organism that are known to be involved in the pathway tocopherol biosynthesis and in Cofactor biosynthesis	prenyltransferase activity	vitamin E biosynthetic process
Q67UX7		[a protein]-serine/threonine phosphate + H_2O = [a protein]-serine/threonine + phosphate		metal ion binding protein serine/threonine phosphatase activity	
B7FA90	Involved in the synthesis of tocopherol (vitamin E). Catalyzes the condensation of homogentisate and phytyldiphosphate to form dimethylphytylhydroquinone	Phytyldiphosphate + homogentisate = diphosphate + 2-methyl-6-phytylbenzene-1,4-diol + CO_2	his protein is involved in the pathway tocopherol biosynthesis, which is part of Cofactor biosynthesis. View all proteins of this organism that are known to be involved in the pathway tocopherolbiosynthesis and in Cofactor biosynthesis.	homogentisatephytyltransferase activity	
Q6H6	**(N-A)**	**(N-A)**	**(N-A)**	**(N-A)**	**(N-A)**

E6					
Q5QMP3	**(N-A)**	**(N-A)**	**(N-A)**	**(N-A)**	**(N-A)**
Q9XGP7	Catalyzes the stepwise methylation of tricetin to its 3'-mono- and 3',5'-dimethyl ethers. No 3',4',5'-trimethylated ester derivatives are produced. Can use caffeoyl-CoA, 5-hydroxyferulic acid, luteolin, tricetin, quercetin, myrcetin and 7,8-dihydroxyflavone as substrates, but not naringenin, apigenin or kaempferol. The 2,3-double bond and the O-dihydroxyl group of the substrate are both required for catalytic activity of the enzym	2 S-adenosyl-L-methionine + tricetin = 2 S-adenosyl-L-homocysteine + 3',5'-O-dimethyl tricetin			
Q6ZD89	Methylates OH residues of flavonoid compounds	S-adenosyl-L-methionine + 3'-hydroxyflavone = S-adenosyl-L-			

		homocysteine + 3'-methoxyflavone.			
Q7F8T6	Catalyzes the stepwise methylation of tricetin to its 3'-mono- and 3',5'-dimethyl ethers. No 3',4',5'-trimethylated ester derivatives are produced. Can use caffeoyl CoA, 5-hydroxyferulic acid, luteolin, tricetin, quercetin, myrcetin and 7,8-dihydroxyflavone as substrates, but not naringenin, apigenin or kaempferol. The 2,3-double bond and the O-dihydroxyl group of the substrate are both required for catalytic activity of the enzym	2 S-adenosyl-L-methionine + tricetin = 2 S-adenosyl-L-homocysteine + 3',5'-O-dimethyl triceti			
Q01P69	*Enzyme and pathway databases*				
Q7XXD4	It is a binding site. S-adenosyl-L-methionine;			O-methyltransferase activity S-	aromatic compound

	via carbonyl oxygen. S-adenosyl-L-methionine; via amide nitrogen.			adenosylmethionine-dependent methyltransferase activity	biosynthetic process

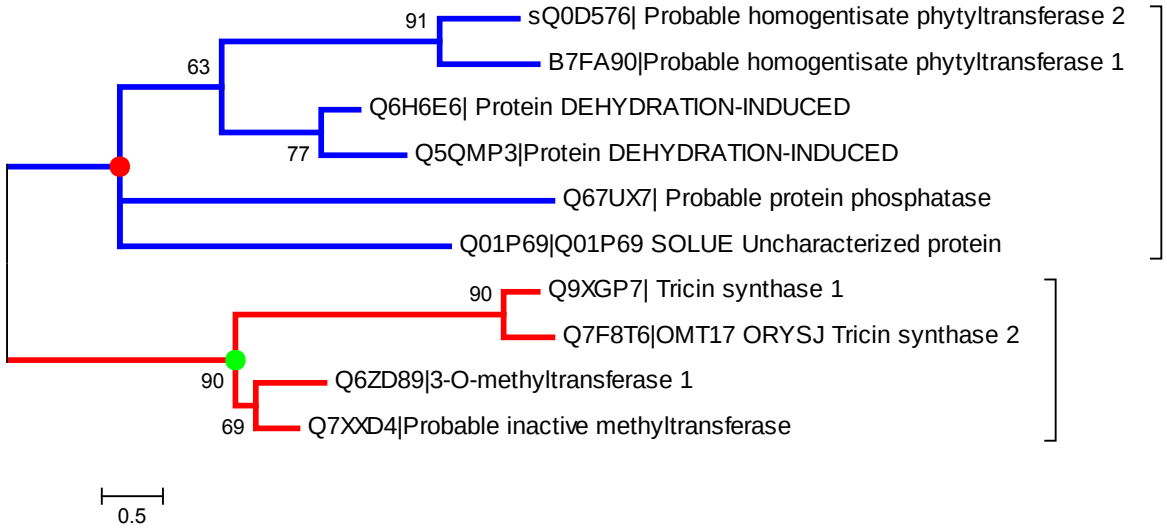


Figure : Phylogenetic analysis of stress resistant proteins of *Oryzasativa, japonica*

Table 5. List of Domains

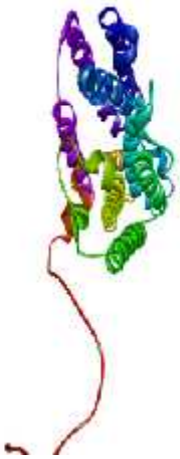
ACCESSIN NO OF PROTEIN	**ACCESSION DOMAIN**	**INTERVAL**	**NAME OF DOMAIN**	**FUNCTION OF DOMAIN**
Q0D576	Cd13960	87-377	*PT_UbiA_HPT1*	Tocopherolphytyltransferase Tocopherolpolyprenyltransferase (TPT1), also known as homogentisatephytyltransferase 1 (HPT1), tocopherolphytyltransferase, or VTE2, catalyzes the first step in the biosynthesis of the tocopherol forms of vitamin E, which involves the prenylation of homogentisate using phytyldiphosphate (PDP) as the prenyl donor. Prenyltransferases (PTs) catalyze the regioselective transfer of prenyl moieties

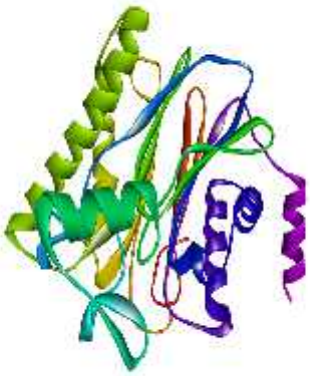
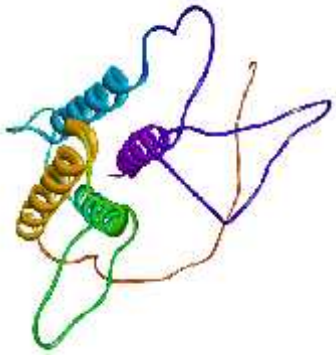
				onto a wide variety of substrates and play an important role in many biosynthetic pathways.
Q67UX7	Cd00143	81-328	*PP2Cc*	Serine/threonine phosphatases, family 2C, catalytic domain; The protein architecture and deduced catalytic mechanism of PP2C phosphatases are similar to the PP1, PP2A, PP2B family of protein Ser/Thrphosphatases, with which PP2C shares no sequence similarity.
B7FA90	Cd13960	111-401	*PT_UbiA_HPT1*	Tocopherolphytyltransferase Tocopherolpolyprenyltransferase (TPT1), also known as homogentisatephytyltransferase 1 (HPT1), tocopherolphytyltransferase, or VTE2, catalyzes the first step in the biosynthesis of the tocopherol forms of vitamin E, which involves the prenylation of homogentisate using phytyldiphosphate (PDP) as the prenyl donor. Prenyltransferases (PTs) catalyze the regioselective transfer of prenyl moieties onto a wide variety of substrates and play an important role in many biosynthetic pathways.
Q6H6E6	Pfam05605 C120571	58-111 132-242	*zf-Di19* *Di19_C*	1.Drought induced 19 protein (Di19), zinc-binding This family consists of several drought induced 19 (Di19) like proteins. Di19 has been found to be strongly expressed in both the roots and leaves of Arabidopsis thaliana during progressive drought. This domain is a zinc-binding domain 2.Stress-induced protein Di19, C-terminal C-terminal domain of Di19, a protein that increases the sensitivity of plants to environmental stress, such as salinity, drought, osmotic stress and cold. the protein is also induced by an increased supply of stress-related hormones such as abscisic acid ABA and ethylene. There is a zinc-finger at the N-terminus, zf-Di19, pfam05605.
Q5QMP3	Pfam14571 Pfam05605	132-243 60-113	*Di19_C* *zf-Di19*	1.Stress-induced protein Di19, C-terminal C-terminal domain of Di19, a protein that increases the sensitivity of plants to environmental stress, such as salinity, drought, osmotic stress and cold. the protein is also induced by an increased

				supply of stress-related hormones such as abscisic acid ABA and ethylene. There is a zinc-finger at the N-terminus, zf-Di19, pfam05605. 2.Drought induced 19 protein (Di19), zinc-binding This family consists of several drought induced 19 (Di19) like proteins. Di19 has been found to be strongly expressed in both the roots and leaves of Arabidopsis thaliana during progressive drought. This domain is a zinc-binding domain
Q9XGP7	Cd02440 PLN02476	90-193 2-250	*AdoMet_MTases PLN02476*	1.S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I; AdoMet-MTases are enzymes that use S-adenosyl-L-methionine (SAM or AdoMet) as a substrate for methyltransfer, creating the product S-adenosyl-L-homocysteine (AdoHcy). There are at least five structurally distinct families of AdoMet-MTases, class I being the largest and most diverse. Within this class enzymes can be classified by different substrate specificities (small molecules, lipids, nucleic acids, etc.) and different target atoms for methylation (nitrogen, oxygen, carbon, sulfur, etc.).
Q6ZD89	Pfam00891 Pfam8100	107-345 31-89	*Methyltransf_2* *Dimerization*	1.O-methyltransferase This family includes a range of O-methyltransferases. These enzymes utilize S-adenosyl methionine 2.Dimerization domain This domain is found at the N-terminus of a variety of plant O-methyltransferases. It has been shown to mediate dimerization of these proteins.

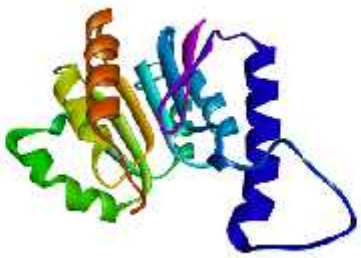
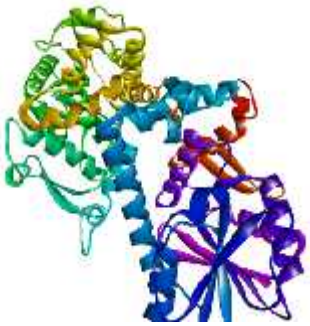
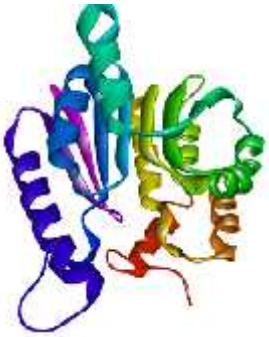
Q7F8T6	Cd02440 PLN02476	133-233 74-290	*AdoMet_MTases* PLN02476	1.S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I; AdoMet-MTases are enzymes that use S-adenosyl-L-methionine (SAM or AdoMet) as a substrate for methyltransfer, creating the product S-adenosyl-L-homocysteine (AdoHcy). There are at least five structurally distinct families of AdoMet-MTases, class I being the largest and most diverse. Within this class enzymes can be classified by different substrate
Q01P69	Pfam12705 C100516	65-103 29-85	*PDDEXK_1 Mrr_cat*	1.PD-(D/E)XK nuclease superfamily Members of this family belong to the PD-(D/E)XK nuclease superfamily 2.Restriction endonuclease Prokaryotic family found in type II restriction enzymes containing the hallmark (D/E)-(D/E)XK active site. Presence of catalytic residues implicates this region in the enzymatic cleavage of DNA.
Q7XXD4	Cd02440 Pfam08100	212-307 39-91	*AdoMet_MTases Dimerization*	1.S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I; AdoMet-MTases are enzymes that use S-adenosyl-L-methionine (SAM or AdoMet) as a substrate for methyltransfer, creating the product S-adenosyl-L-homocysteine (AdoHcy). There are at least five structurally distinct families of AdoMet-MTases, class I being the largest and most diverse. Within this class enzymes can be classified by different substrate specificities (small molecules, lipids, nucleic acids, etc.) and different target atoms for methylation (nitrogen, oxygen, carbon, sulfur, etc.). 2.Dimerization domain This domain is found at the N-terminus of a variety of plant O-methyltransferases. It has been shown to mediate dimerization of these proteins

Table 6. 3D structure of stress resistant proteins of *Oryzasativa, japonica*

Accession no(gene name)	**Model structure**
Q0D576 (HPT2)	

Q67UX7 (Os02g0149800) Q6H6E6(DI19-4)	 

Accession no(gene name)	**Model structure**
Q5QMP3(DI19-3)	

Q9XGP7(ROMT-15)	
Q6ZD89(ROMT-9)	
Q7F8T6(ROMT-17)	

Accession no(gene name)	**Model structure**
Q01P69(Acid_7651)	

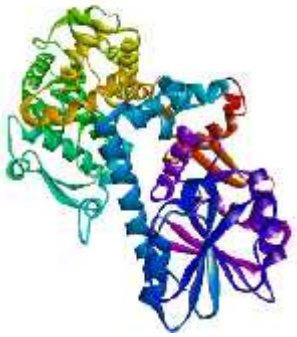
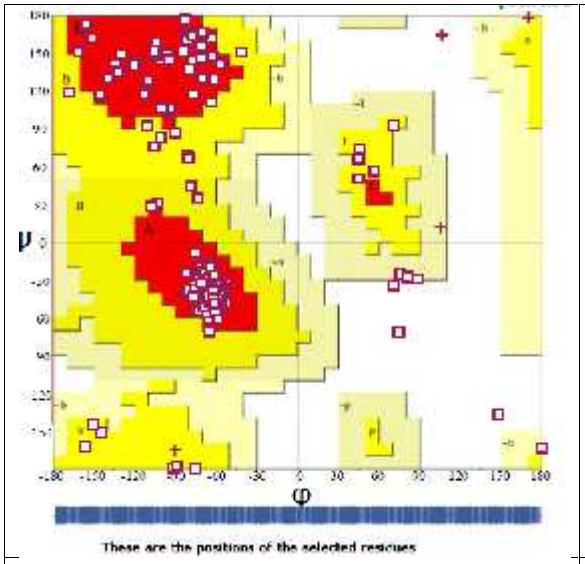
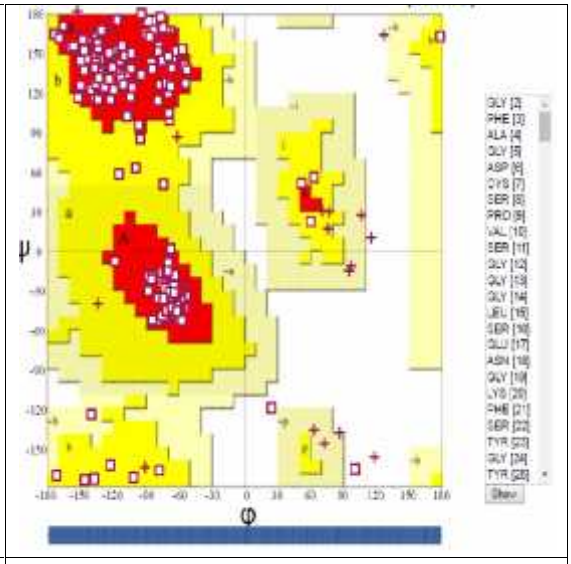
B7FA90(HPT1)	
Q7XXD4 (Os04g0175900)	

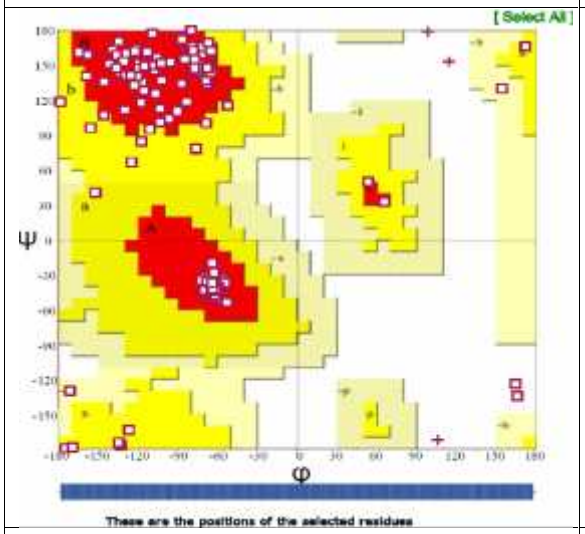
Table 7. Ramachandran plot analysis of predicted 3D structure of stress resistant proteins of *Oryzasativa, japonica*



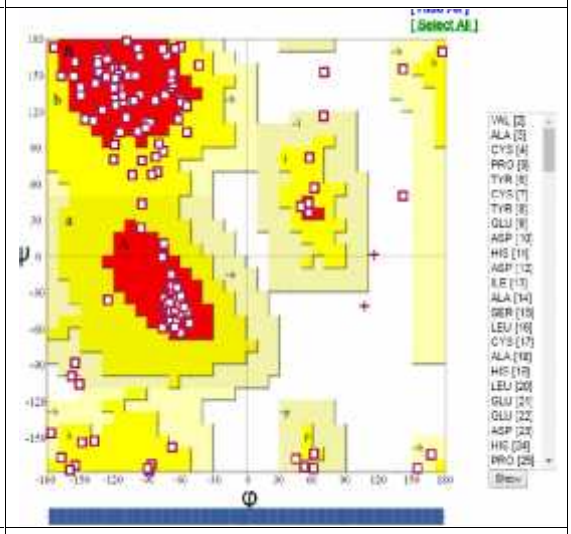
Q0D576



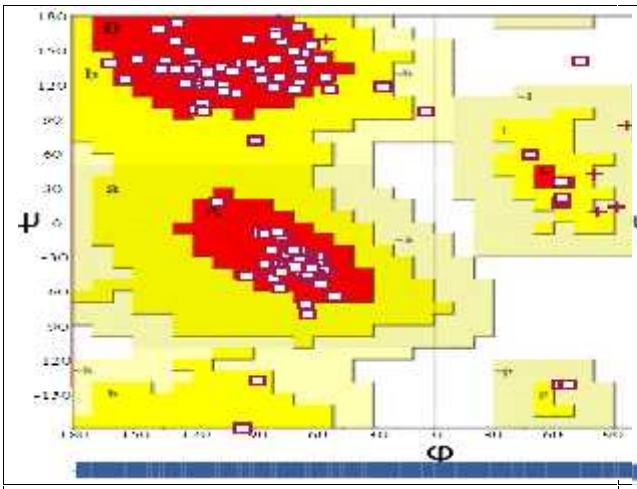
Q67UX7



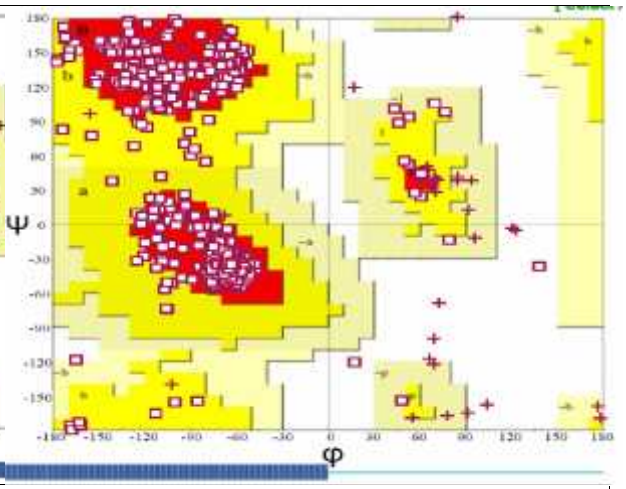
Q6H6E6



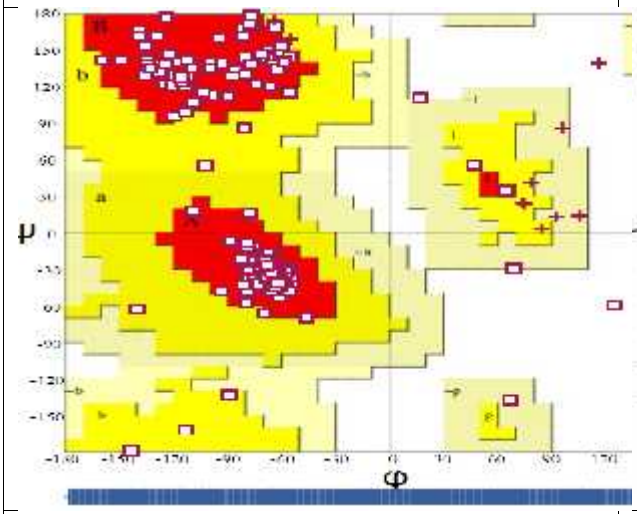
Q5QMP3



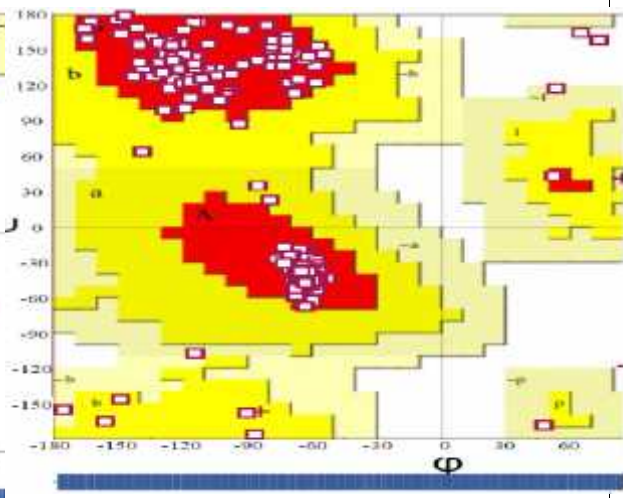
Q9XGP7



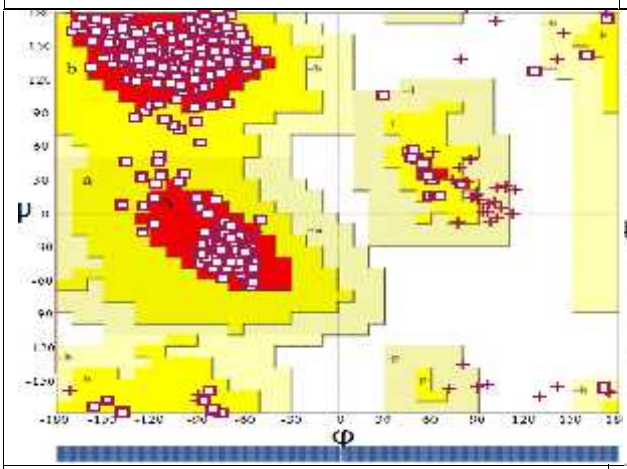
Q6ZD89



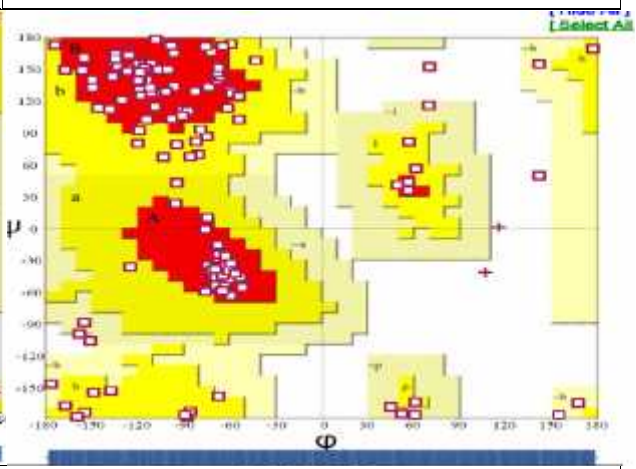
Q7F8T6



Q01P69



Q7XXD4



B77FA90

DISCUSSION:

Stress resistance and drought tolerance traits are the important among different traits in *Oryza sativa*. In context to the climatic change, identification of the traits bears important as far as rice genomics is concerned.The rice genome has been sequenced for different sub species like *Oryza sativa* Indica and *Oryza sativa japonica*.The whole genome information of *Oryza sativa indica* and *Oryza sativa japonica* are available in Oryza database.Rice genome consists of 12 pairs of chromosome.The rice genomeis well mapped and well characterized and its smallest of the major serial crop genome and itestimated to 400-430mb.Total no genes are estimated to be 16,940.In this study an attempt has been made to compile and collecting information in an arround abiotic stress in rice.six no candidates genes have been studied.As far as location and function is concerned. The corresponding protein sequence have also been retrieved from the uniprot protein sequence database.the physiochemical parameters GO(Gene ontology)annotation andfunctional domain analysis have been carried out to understand as role in Stress resistance and drought tolerance.the domains named(cd13960,cd00143,cd13960,Pfam05605,c120571,Pfam14571,Pfam05605,cd02440,PLN02476,Pfam60891,Pfam8160,cd02440,PLN02476,Pfam12705,c100516,cd02440,Pfam08100)present in the protein.The fuction of the protein is depend upon its structure proper three dimensioal structures .Three dimensioal structures from the above Stress resistance protein of *Oryza sativa japonica* have been predicted using online server /tools,to understand of secondary structure arrangement.

CONCLUSION:-

Rice*(Oryza sativa)*is one of the most important crop in in the world.rice, wheat,maize together account for about half the worlds food production and rice itself is the principal food of half the worlds population .Rice genome consists of 12 pairs of chromosome with genome size of 420 MB .Various plant breeding of approaches have been carrIed out eariler production and productivity of rice .But in recent years due to climate change the production and productivity have been affected most due to decrease in rain fall and other climatic factors.Molecular approaches have been taken as an important area to understand role of different stress related traits in rice.After the genome sequencing of rice various genes have been identified and studied .Its role and different models in the study an attempt has been made,To know the different types of genes and proteins responsible for stress resistance and drought tolerance. This information will helpful for researcher working in rice varietal development.

REFFERENCE:-

1. . Nam-SooJwaa, Ganesh Kumar Agrawalb,1, Shigeru Tamogamic, Masami Yonekurad,OksooHane, Hitoshi Iwahashif, RandeepRakwalb,f,**Role of defense/stress-related marker genes, proteins and secondary metabolites in defining rice self-defense mechanisms** Plant Physiology and Biochemistry 44 (2006) 261–273

2. Takuji Sasaki and Benjamin Burr† **International Rice Genome Sequencing Project: the effort to completely sequence the rice genome** Current Opinion in Plant Biology 2000, see front matter © 2000 Elsevier Science Ltd.

3. Takuji Sasaki **Rice genome analysis understanding the genetic secret of rice plant** Breeding science 53:281-289(2003)

4. Nguyen Thi Lang1,Nguyen Quang Cao Binh1, ChauThanhNha, Bui Chi Buu2 **A Candidate gene Response to Drought Stress Condition in rice**(Oryzasativa).

5.. Kazuo Shinozaki1,and Kazuko Yamaguchi-Shinozaki2,**3 Gene networks involved in drought stress response and Tolerance** Journal of Experimental Botany, Vol. 58, No. 2, pp. 221–227, 2007Integrated Approaches to Sustain and Improve Plant Production underDrought Stress Special Issue Advance Access publication 30 October, 2006.

6. TAKUJI SASAKI **The rice genome project in Japan** Proc. Natl. Acad. Sci. USA Vol. 95, pp. 2027–2028, March 1998 Colloquium Paper

7. Pankaj Jaiswal1, Doreen Ware3, Junjian Ni1, Kuan Chang1, Wei Zhao3, Steven Schmidt3, Xiaokang Pan3, Kenneth Clark3, Leonid Teytelman3, Samuel Cartinhour2, Lincoln Stein3 and Susan McCouch1* **Gramene: development and integration oftrait and gene ontologies for rice Comparative and Functional Genomics**Comp FunctGenom 2002; 3: 132–136.

8. TakujiSasakiand Benjamin Burr† **International Rice Genome Sequencing Project: the effort to completely sequence the rice genome Current Opinion in Plant Biology** 2000, **3**:138–141

9. Takuji Sasaki1, Jianzhong Wu1, Hiroshi Mizuno1, Baltazar A. Antonio1,and Takashi Matsumoto1 Takuji**The Rice Genome Sequence as an IndispensableTool for Crop Improvement** Biotechnology in Agriculture and Forestry 62.© Springer-Verlag Berlin Heidelberg 2008

10. Juexin Wang1,2, Liang Chen1, Yan Wang1, Jingfen Zhang2, Yanchun Liang1, Dong Xu1,2 **A Computational Systems Biology Study for Understanding Salt Tolerance Mechanism in Rice** Published June 7, 2013

11. JibanMitra**Genetics and genetic improvement of drought resistance in crop plants**REVIEW ARTICLES- CURRENT SCIENCE, VOL. 80, NO. 6, 25 MARCH 2001.

12. Renata Pereira da Cruz1, Sandra Cristina Kothe Milach2 and Luiz Carlos Federizzi3**Inheritance of rice cold tolerance at the germination stage Genetics and Molecular Biology,** 29, 2, 314-320 (2006)

13. .Wangxia Wang Æ BasiaVinocur Æ Arie**Altman Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance**Planta (2003) 218: 1–14

14. Akihiko Kamoshita a,, R. Chandra Babu b, N. ManikandaBoopathi b, ShuFukai c **Phenotypic and genotypic analysis of drought-resistance traits for developmentof rice cultivars adapted to rainfed environments** Field Crops Research 109 (2008) 1–23

15. Nam-SooJwaa, Ganesh Kumar Agrawalb,1, Shigeru Tamogamic, Masami Yonekurad,OksooHane, Hitoshi Iwahashif, RandeepRakwalb,f,**Role of defense/stress-related marker genes, proteins and secondary metabolites in defining rice self-defense mechanisms** Plant Physiology and Biochemistry 44 (2006) 261–273

16. Hei Leung Stressed genomics —**bringing relief to rice fields**

17. David J. Mackill, Abdelbagi M. Ismail, Alvaro M. Pamplona, Darlene L.Sanchez, Jerome J. Carandang and Endang M. SeptiningsihInternational Rice Research Institute, Metro Manila, Philippines **Stress Tolerant Rice Varieties forAdaptation to a Changing Climate.**

18. A K H I L E S H K. T Y A G I, J I T E N D R A P. K H U R A N A , P A R A M J I T K H U R A N A , SAURABH RAGHUVANSHI, A N U P A M A G A U R , A N I T A K A P U R , V I K R A N T G U P T A , D I B Y E N D U K U M A R , V . R A V I, S H U B H A V I J, PARUL KHURANA a n d S U L A B H A S H ARMA **Structural and functional analysis of rice genome** Indian Academy of Sciences

19. Takuji Sasaki1, Jianzhong Wu1, Hiroshi Mizuno1, Baltazar A. Antonio1,and Takashi Matsumoto1 Takuji**The Rice Genome Sequence as an IndispensableTool for Crop Improvement** Biotechnology in Agriculture and Forestry 62.© Springer-Verlag Berlin Heidelberg 2008

CURRICULUM VITAE(C.V)

MISS.IPSITA PANDA

Mobile:-9583365555,9938835742

Email ID:-pandaipsita94@gmail.com

CARRER OBJECTIVE:-

To secure a challenging position in Bioinformatics environment with committed and dedicated people to contribute my Bioinformatics and computer skills in the field of biological research to serve our nation.

ACADEMIC QUALIFICATION:-

Discipline	Institution	Board /University	Year of Passing	% of Marks	Div
M.Sc. (Bioinformatics)	C.P.G.S,OUAT, Bhubaneswar, Odisha	Orissa University of Agricultural &Technology	Cont.		
B.Sc.(Botany)	Nimapara Autonomous College, Nimapara	Utkal University	2014	70.33%	1st
+2 Science	Govt. Junior Science College , Malkangiri	Council of Higher Secondary Education	2011	52.33%	2nd
10 th	Govt. Girls high School , Malkangiri	Board of Secondary Education Odisha	2009	73.33%	1st

BIOINFORMATICS SKILLS:-

Area of skills	skill
Bioinformatics Tools	BLAST,FASTA,EMBOSS,Clustal w,Translate tool Expasy,Modeller,Rasmol,CN3D,,DISCOVERY STUDIO ,MEGA
Bioinformatics Database	NCBI,GENBANK,SWISS-PROT,EMBL,DDBJ,PDB,SWISS PROT

COMPUTER SKILLS:-

Area of skill	Skill
Operating system	MS-DOS,WINDOWS95/98/2000/XP,Red Hat Linux9
Database and RDBMS	MySQL,MS-ACCESS,PHP
Programming language	C,VB.NET,DATA STRUCTURE,JAVA,Perl,HTML

AREA OF INTEREST:-

Biological Database Management,Molecular biology,Bio-chemistry,Genetics,Proteomics,Molecular modelling,Drug discovery&Drug designing,Genetic Engineering.

PERSONAL INFORMATION:-

Name	:	IPSITA PANDA
DATE OF BIRTH	:	04.06.1994
GENDER	:	FEMALE
Marital status	:	UN –MARRIED
Nationality	:	INDIAN
Language Proficiency	:	ENGLISH,ODIYA

PRESENT ADDRESS : MS.IPSITA PANDA
AT:SaradaDevi Girls, Hostel,
OUAT BBSR,
PO:Bhubaneswar
Dist:Bhubaneswar
STATE:ODISHA
PIN:751003

DECLARATION:

I hereby declare that the above mentioned information is correct up to my knowledge and i bear the responsibility for the correctness of the above mentioned particulars.

Place:Bhubaneswar

Date: **Ipsita Panda**