

In-silico* Characterization and Comparative Analysis of BLB Disease Resistance Xa genes in *Oryza sativa

Muhammad Anas Ramzan¹

University of Agriculture Faisalabad,
Pakistan.

anasramzan122@outlook.com

Abdul Rehman¹

University of Agriculture Faisalabad,
Pakistan

abdulrehman432256@gmail.com

Ljupcho Jankuloski²

International Atomic Energy Agency,
Vienna

l.jankuloski@iaea.org

Hasib Asghar¹

University of Agriculture Faisalabad,
Pakistan.

hasibasghar7@gmail.com

Dr Muhammad Rashid*

Nuclear Institute for Agriculture and Biology
Faisalabad, Pakistan

drmrashid@niab.org.pk

Abstract

Globally, rice is utilized as staple food, belongs to the family Poaceae. From the past few decades under variable climate conditions, it greatly affected by the bacterial leaf blight (BLB) disease caused by a bacterium *Xanthomonas oryzae* pv. *oryzae* (*Xoo*). Scientists studied causes of the disease and found more than 61 isolates of *Xoo*. About 39 Xa genes were noted that cause race specifically resistance either individually or in pairs against *Xoo*. The canvas on the characterization of these genes is still unrevealed. In this study, amino acid sequences of Xa15, Xa19, Xa20 and Xa21 were mined and used for motif-domain identification, characterization, and comparative analysis. It includes screening of physical and chemical characteristics, sequence comparison to find sequence similarity between them and their phylogenetic comparison with other Xa genes and other species based upon LRR and S_TKc domains to find evolutionary relationship among them. The comparison-based modeling was performed and assessed by different tools to gain better understanding and structural evaluation. The results showed identified domains are specific in function, each domain involved in resistance against biotic and abiotic stresses through regulating different cellular processes. This study also revealed high similarity (>98% sequence identity) between these genes and encode a similar leucine-rich repeat receptor kinase like protein. It will optimize the breeding programs as it will be useful for the selection of effective genes to produce resistance in rice varieties against the specific strains of *Xoo* that will be more effective against BLB than the other Xa genes.

Keywords

BLB Resistance, Xa genes, Xoo, Molecular characterization, Rice, LRR, S_TKc.

Introduction

Rice is the staple food crop and about 2.7 billion people in the world directly or indirectly rely on it (Fairhurst and Dobermann, 2002). It is reviewed that rice is grown by more than 50% of farmers in the world (Rehman *et al.*, 2015) and it covers 9% of the world total cultivated area (Maclean *et al.*, 2002). According to recent studies, 90% of total rice is cultivated in Asia (Denning and Vö, 1995; Khush and Jena, 2009) and in future, the global consumption of rice may increase up to 70% of its total yield (Kubo and Purevdorj, 2004). Recently, the production of rice crop increased in the rest of the world due to its high production and vast usage (Rehman *et al.*, 2015). During 2019-20, according to the United States Department of Agriculture (USDA), the rice production estimate was 499.31 million metric tons in the world as given in the pie chart (Fig. 1).

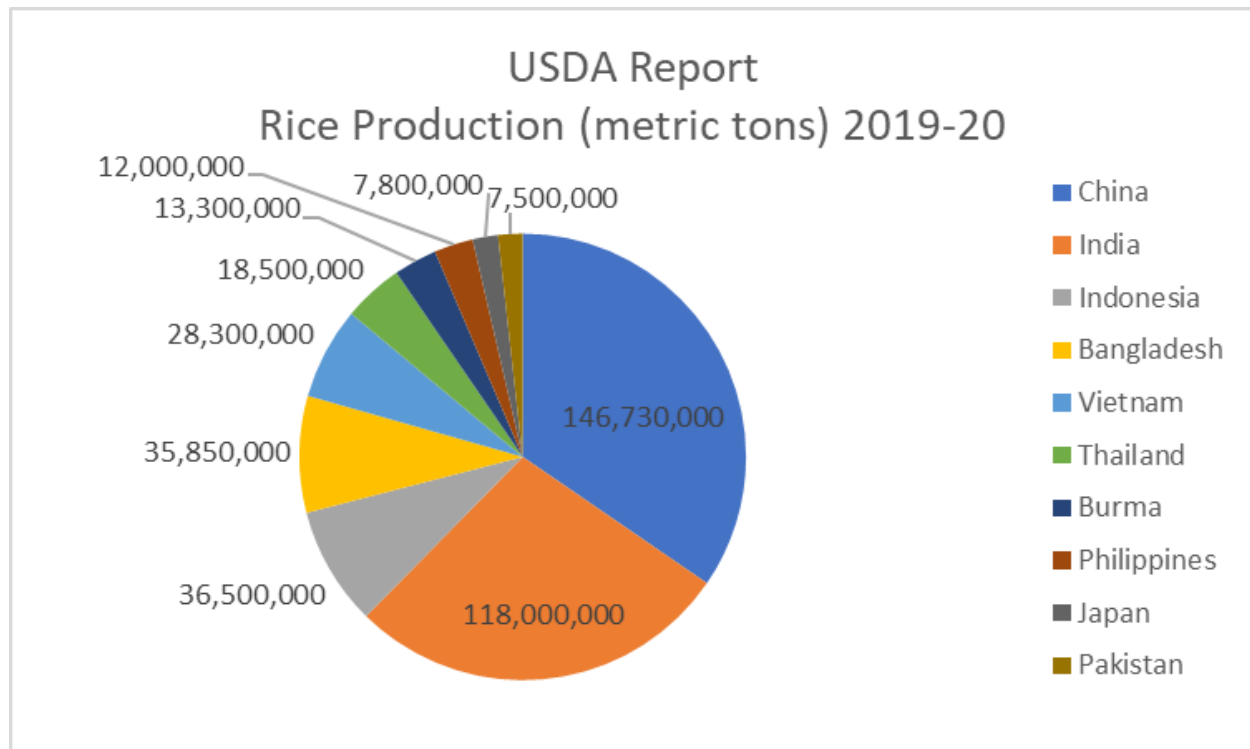


Figure 1: Pie chart of rice production in metric tons 2019-20

Rice is a monocotyledon angiosperm belongs to the family Poaceae and its genus is *Oryza* that has 22 different species. It is estimated that 40,000 rice varieties are known until now (Institute, 1985). The most cultivated varieties are *Oryza sativa* and *Oryza glaberrima*. *Oryza sativa* known as Asian rice (Lemus *et al.*, 2014), divided into *japonica* and *indica* sub species, widely cultivated due to grain stickiness and length (You *et al.*, 2017). The rice production also depends upon favorable weather conditions (Khush and Jena, 2009). Being wide adaptation, it is cultivated as upland and low land rice in the world. It is also an appropriate crop for the countries facing flooded condition due to heavy rain fall. It is estimated that one-third of the earth's fresh water is consumed by the rice crop (Wu *et al.*, 2017). In Pakistan, rice is also considered as the stuff of life and plays a significant role to boost the country exchequer (Rahman *et al.*, 2017). Pakistan is among the largest producers of rice with an average production of 6.7 million tonnes in 2016-17. It also helps in maintaining the momentum of export (Abdullah *et al.*, 2015).

Under conducive environment conditions, several pathogens attack different crop plants that lead them to different harmful diseases. It is estimated that these diseases cause a loss of 16% of the total crop production in the world (K. Reddy, 1979). Among different emerging diseases like sheath blight, brown spot, neck blast, the BLB disease caused by *Xanthomonas oryzae pv. oryzae* (Xoo) is the devastating disease which may appear at seedling, tillering, booting stages of rice plant life cycle under changing climate conditions. The losses of the crop varied from 10% to 90%, if it appears before panicles initiation as it affects the flag leaf of the

plant. Most of the time, genes that cause resistance are dominant genes which response by identifying the unknown proteins produced by pathogen effector genes and the interaction between them are known as gene by gene interaction. Rice, sedges and other grasses mostly affected by this BLB disease caused by *Xoo*. Besides *Oryza*, there are different host of *Xoo* pathogen as *Zoysia japonica*, *Panicum maximum*, *Cynodon dactylon*, *Leersia oryzoides*, *Cyperus difformis* and *Cyperus rotundus* etc. Rice is found as the major host of this bacterium and largely affected by the disease. The *Xoo* belongs to the Xanthomonadaceae family and its genus is *Xanthomonas*. It is a gram negative and non-spore forming rod shaped bacterium with a single polar flagellum (Nas, 2008; Sahu *et al.*, 2018). It enters through the natural opening like hydathodes, stomata and wounds. It can survive in non-growing seasons in rice seeds as well as other living hosts.

The favorable conditions for the spread of this disease include high temperature (25-30 °C) and humidity above 70%. It usually occurs after continuous heavy rain and strong wind which allow the bacteria to spread easily and affect rice plants (Naqvi, 2019). It is mostly observed in rice crops found in high and moderate temperature areas. Its ventures are high under high nitrogen fertilization. Rice cultivars in the tropical areas are highly susceptible to BLB disease. It has a destructive effect on the rice varieties mostly found in Asian countries. Its symptoms include yellowish leaves that roll up and dry, later on lead to its death. It is the one that seriously affects the rice plants and the yield loss depends upon the occurrence of the disease during growth stages (Ghadirnezhad and Fallah, 2014). The death of affected seedlings occurs between two to three weeks. It is unable to chemically control the effect of disease but biological control techniques like bacterial antagonists of pathogens can be used to minimize the BLB with limited usage (Khan *et al.*, 2012).

The farmers are facing many difficulties including decrease in yield that results in very low seed production because most of the plants die at their seedling stage which can cause a high yield loss up to 70%. It is also a big financial loss for the country (Syed-Ab-Rahman *et al.*, 2020). There is a need to develop resistance against the disease for next and improved generations of rice and in future, there are chances of hunger starvation due to this disease.

In 2000, the whole genome of rice was sequenced using shotgun sequencing method which revealed that rice has 12 chromosomes ($2n = 24$) and about 40,000 genes that belong to 15,000 distinct gene families (Song *et al.*, 1995). Rice, like other plant species, has some important genes that fight against different pathogens. After the study of rice genetics, scientists have identified BLB resistant genes named as Xa genes that fight against bacterial species like *Xoo*. According to the recent study, 39 Xa genes harboring resistance to BLB-disease were discovered from the cultivated and wild rice varieties. These resistance genes are named as Xa1 to Xa39 in which some Xa genes are extensively used for breeding in Asia like Xa4, Xa5, Xa13 and Xa21 (Zhang *et al.*, 2015). It is difficult to control the BLB-disease as the pathogens are changing under variable climate conditions but the development of resistance in the host against pathogens is the best way to control this disease.

Xa genes code for specific proteins and have different types of domains in their sequences that proved useful for bacterial resistance. Most of the genes have identical regions and possess the same type of domains due to their sequence and functional similarity. For example, Xa3 and Xa26 genes code for the same type of LRR receptor kinase-like protein. Earlier studies reported that proteins of these genes possess 94% similarity and 92% identity in their sequences (Xiang *et al.*, 2006). It was also observed that recessive xa gene like xa13 also proved resistance ability against BLB disease (Chu *et al.*, 2006).

The detailed study about the characterization of Xa genes needs to be carried out to explore in depth the resistance ability of these genes in order to increase their effect against BLB disease. Some Xa genes also possess high resistance as compared to other genes. Xa genes were introgressed in existing cultivars to improve their resistance against BLB disease attack for sustainable rice production.

Most of Xa genes have been incorporated into the background of susceptible *indica* cultivar IR24. Some Xa genes have been pyramided, either through classical breeding and marker assisted selection or through genetic engineering approaches to develop new plant types and near isogenic lines (Khan *et al.*, 2012). Development of resistant cultivars carrying major resistant genes has been the most effective approach for the control of BLB but for incorporation of the most efficient genes, it is a prerequisite to evaluate the resistant sources against existing races or isolates of *Xoo*.

In Pakistan, PKX1, PKX2, PKX3, PKX4 and PKX5, races of *Xoo* were found where PK and X used for Pakistan and *Xoo* respectively. The International Rice Research Institute (IRRI) developed 26 rice varieties possess different genes or combination of genes against 61 *Xoo* isolates and indicating 14 different rice producing land zones in Punjab, Pakistan. The analysis of disease spread proved that near isogenic lines were prone to one or more *Xoo* isolates (Khan *et al.*, 2012). All the isolates of BLB cannot be controlled by only one gene. For example in *Oryza longistaminata*, Xa21 gene that is present on chromosome 11, works well separately and also in combination for resistance in different isolates (Song *et al.*, 1995). A few isolates of *Xoo* cause virulence and dominate the Xa21. According to IRRI, only single race PKX2 dominate the Xa21 gene and it is the only dominant *Xoo* race in Punjab, Pakistan while Xa21 cause resistance in other *Xoo* races (PKX1, PKX3, PKX4 and PKX5).

In Pakistan, none of the existing cultivated variety exhibited complete resistance to BLB disease. The rice breeders are trying to develop BLB disease resistant rice varieties by incorporating resistant Xa genes developed by IRRI, Philippines. Several Xa genes discovered so far have resistance against BLB disease but still research required for their full characterization.

In the past, many researchers characterized Xa genes. They used different ways to annotate these genes, but full characterization of many genes is still incomplete. Most of the researchers used wet lab for experimentation so there is a need of dry lab i.e. *in-silico* analysis to fully explore the Xa genes using different computational tools to increase the resistance ability of these genes up to 100%. To have complete resistance, the full characterization of the Xa genes

using *in-silico* analysis needs full consideration, so that breeder can incorporate the Xa gene of interest in their breeding rice material.

The objective of study is the sequence and structural analysis of BLB resistance Xa genes to understand and enhance its effect in rice plants and generate useful information that will be helpful for future research. It includes to identify motifs and domains to analysis conserved regions, secondary structure analysis, phylogenetic analysis, homology modeling, Ramachandran plot analysis and comparative analysis using different window-based and web-based tools.

Materials and methods

Sequence retrieval and purification

The amino acid sequences of rice resistance Xa genes downloaded from the UniProtKB database (<https://www.uniprot.org/help/uniprotkb>). Amino acid sequences of some Xa were not available in the UniProtKB so their nucleotide sequences were retrieved from online biological database GenBank (<http://www.ncbi.nlm.nih.gov/genbank>). The MSU (<http://rice.plantbiology.msu.edu>) Rice Genome Annotation Project Database and Resource was used to get information about sequence conformation with the genomic linkage and annotation of Xa genes with other Xa genes that form a semantic network of all rice genes (Kawahara *et al.*, 2013). For the sequences purification and to exclude data redundancy, BLAST Aligning tool for Multiple Protein Sequences

(https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq) was employed to align amino acid sequences according to their similarity with other amino acid sequences (Johnson *et al.*, 2008). Amino acid sequences were required for analysis, so translation of nucleotide sequences was carried out by using EMBOSS Transeq tool (https://www.ebi.ac.uk/Tools/st/emboss_transeq) (Only unavailable amino acid sequences were translated) (Madeira *et al.*, 2019).

Sequence analysis

Motif and domain identification

Multiple Sequence Alignment (MSA) of Xa amino acid sequences was performed using the CLUSTAL-X tool (it employs progressive alignment method using a heuristic approach) (Larkin *et al.*, 2007). Motif identification was carried out using the MEME suite tool (<http://meme-suite.org/tools/meme>) that led to the location identification and graphical representation by sequence logo of the cross-ponding motifs. (Bailey *et al.*, 2009).

For further sequential annotation, the Simple Modular Architecture Research Tool (SMART) (<http://smart.embl-heidelberg.de>) (the Hidden Markov Models for protein domain identification was used to generate detailed results for the domains) (Schultz *et al.*, 1998) and CDD (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) pragmated for domain identification and

purified protein sequences used as the input (Lu *et al.*, 2020). To confirm the resultant domain presence in the sequences, the Pfam database (<https://pfam.xfam.org>) was accessed (El-Gebali *et al.*, 2019). InterProscan (<https://www.ebi.ac.uk/interpro/search/sequence>) was utilized for classification of protein family based on motif conserveness and domain present in the sequences (Jones *et al.*, 2014).

Comparative analysis

All Xa genes were taken as a query one by one and aligned with all Xa genes (as subject sequences) using Align Sequences Protein BLAST tool that showed the sequences similarity percentage among these sequences. For the ease in understanding their similarity, the Sequence Identity Matrix (SIM) was created. The evolutionary relationship also analyzed to take a peek deeply on the evolutionary relationship between all the present Xa genes using MAFFT for alignment (Madeira *et al.*, 2019), FastTree (<https://www.genome.jp/tools-bin/ete>) (Price *et al.*, 2009) for phylogenetic tree construction using approximate maximum likelihood method (Jin *et al.*, 2006) and Interactive Tree of Life (ITol) (<https://itol.embl.de>) to edit and enhance tree topology (Letunic and Bork, 2016).

Phylogenetic analysis

MSA of Xa amino acid sequences was done with the CLUSTAL-X tool that employed progressive alignment method which used a heuristic approach to align the sequences. By using protein MSA, a phylogenetic tree constructed with FastTree tool that used the maximum likelihood method. Finally, ITol tool used for phylogenetic tree annotation, display and manipulation of tree's data like clusters name, text font, tree style, bootstrap value, evolutionary distance, merge clusters into clades, colorful visualization and other useful manipulations (MEGA-X can also be used).

Primary structure analysis

ProtParam tool (<https://web.expasy.org/protparam>) was utilized to calculate the physical and chemical parameters useful regarding the structural analysis in Xa proteins. The calculated parameters include no. of amino acids, molecular weight, formula, theoretical pI, no. of negatively charged residues, no. of positively charged residues, total no. of atoms, extinction coefficient, aliphatic index, GRAVY, and instability index (ProtParam, 2017). To determine the transmembrane helices in Xa amino acid sequences, the TMHMM tool (<http://www.cbs.dtu.dk/services/TMHMM>) was employed that utilize membrane protein topology identification approach based upon the Hidden Markov Model (Krogh *et al.*, 2001).

Secondary structure analysis

In Xa proteins, the prediction of secondary structures was carried out with the help of SOPMA tool (https://npsa-prabi.ibcp.fr/cgi-bin/npsa-automate.pl?page=/NPSA/npsa_sopma.html). The main components of predicted secondary structures include alpha helix, beta-turn, and random coil, etc., (Geourjon and Deléage, 1995). The identification of signal peptides with positions of

their cleavage sites in Xa amino acid sequences was done using the SignalP tool (<http://www.cbs.dtu.dk/services/SignalP>). It is based upon the architectures of recurrent and convolutional neural networks with a random option of condition (Nielsen *et al.*, 1997).

Comparative modeling

Template searching and modeling

The residues of Xa15, Xa19, Xa20 and Xa21 proteins of rice was explored in BLASTP against PDB database (<http://www.rcsb.pdb.org>) to find suitable template for homology modeling to predict 3D structures and functional prediction of these proteins (Berman, 2000). I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>) was also utilized for the searching of well-suited templates for comparative modeling of these Xa proteins (Yang and Zhang, 2015). Different parameters of BLASTP like, highest alignment score and percentage identity, maximum query coverage, and lowest E-value employed for accuracy assurance of the selected templates for Xa proteins. All these steps suggested that 6s6q.2.A a hetero dimer whose A-chain is the most suitable template for Xa15, Xa19, and Xa20, and 4mn8A (A-chain) for Xa21 proteins. Homology modeling based upon target sequence alignment with template sequence whose structure was already available in the PDB database. The sequence alignment of target with template was done by MODELLER9v7 using alignment script (align2d.py). Using the alignment file 15 different 3D structural models were produced by the MODELLER9v7 (Webb and Sali, 2016). The ranking of these predicted 3D models was done according to their discrete optimized protein energy (DOPE) scores and GA3141 score. The lowest DOPE score and highest GA3141 cross-ponds the best modeled structures. For further clear visualization, UCSF Chimera 1.14 was implemented (Chen *et al.*, 2015).

Model quality assessment

The quality assessment along reliability and internal consistency of Xa15, Xa19, Xa20, and Xa21 models was carried out by GNUPLOT (Racine, 2006), Ramachandran plot analysis and a number of other tools. To test the stereo chemical quality of the model by *phi* and *psi* angles of proteins, Swiss model server was used, which identify the number of residues in favorable and unfavorable regions of Ramachandran plot (Spencer *et al.*, 2019). ERRAT (<https://servicesn.mbi.ucla.edu/ERRAT>) tool utilized for the overall quality-factor determination of proteins. It calculated the statistics of unbonded interactions among different types of atoms (Dym *et al.*, 2012). Furthermore, standard bond angles and bond lengths of these models were determined. Structural Analysis and Verification Server (SAVES) (<http://nihserver.mbi.ucla.edu/SAVES>) used to perform the above analysis. The quality validation of 3D predicted models was carried out using MolProbity web server (<http://molprobity.biochem.duke.edu>). This outcome fine detail about any problem in molecular model using atomic contact analysis (Chen *et al.*, 2010). The root mean square deviation (RMSD) between the atoms of the computed models and templates was calculated with the help of I-TASSER server. The updated dihedral-angle diagnostics was done using Ramachandran

plot, built with the help of Swiss model tool of ExPASy server (<https://swissmodel.expasy.org/interactive>) (Waterhouse *et al.*, 2018). For final validation and refinement of the predicted 3D structural models, the ProSA tool (<https://prosa.services.came.sbg.ac.at/prosa.php>) was used to calculate the native protein folding energy of the predicted models (Fig. 2a, 2b, 2c, 2d) and compare the energies of these with the known structural models' energy field (Fig. 3a, 3b) (Wiederstein and Sippl, 2007). MATRAS web server (http://strcomp.protein.osaka-u.ac.jp/matras/matras_pair.html) was used for the pairwise alignment of the structures of suited templates and the computed models to have a better assessment of the conserveness between them and their pairwise 3D structural superimposition (Kawabata, 2003).

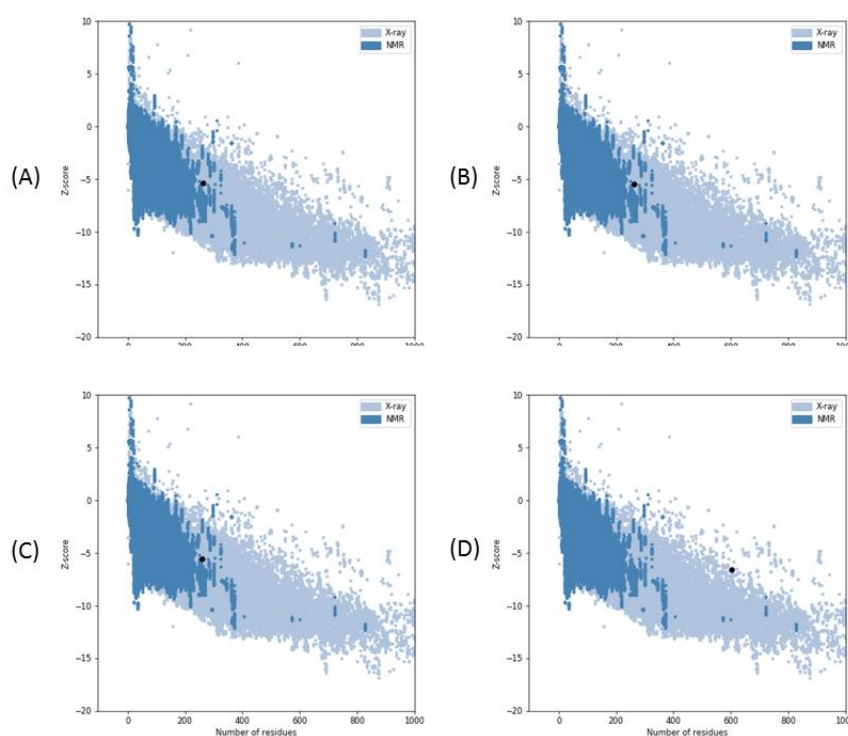


Figure 2: Protein structure analysis (ProSA) of predicted structures of Xa proteins, (A) overall quality of Xa15 model. (B) overall quality of Xa19, (C) overall quality of Xa20 model, and (D) represents overall quality of Xa21 model (blue dotted region show the protein folding energies and the black dot represents the place of predicted model energy value in that region).

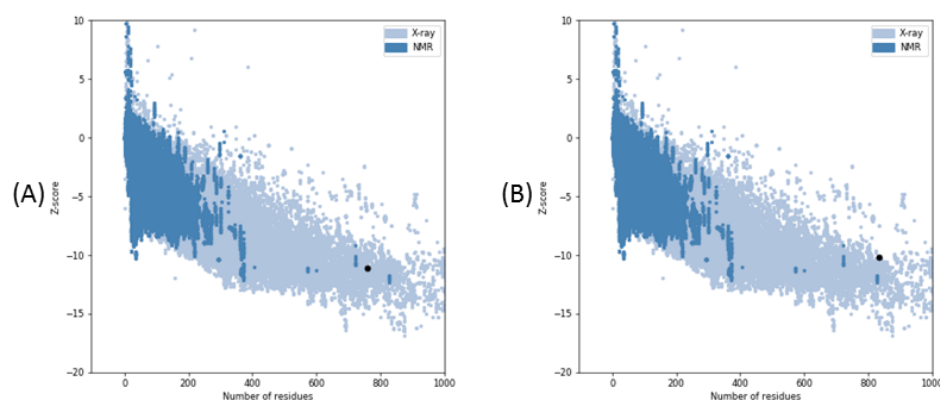


Figure 3: Protein structure analysis (ProSA) of selected templates (A) Image against 4mn8, (B) Image against 6s6q.

Results and discussion

Sequence analysis and domain prediction

The previously published amino acid sequences of Xa15 (363aa), Xa19 (363aa), Xa20 (363aa) and Xa21 (1025aa) proteins from *Oryza sativa* and *indica*, involved in resistance against BLB disease in rice, which were downloaded from UniProt database (<https://www.uniprot.org>). The domain search using SMART tool revealed that all four Xa proteins possess a similar domain leucine-rich repeat receptor-like protein kinase (LRR 8) (Antolín-Llovera *et al.*, 2012). LRR-8 is the 8th family member of clan CL0022 which consist of total 13 family members. Xa15, Xa19 and Xa20, each of these possess five different type of identical domains (LRR) (Yoshimura *et al.*, 1998), on different positions that start from same residue (Leucine) but end on different residues viz, (Leu14-Leu38), (Leu62-Asn85), (Leu86-Ile110), (Leu135-Cys159) and (Leu183-Asp206). Traditionally, LRR protein contains normally one alpha-helix for each beta-strand, variants that involved in the formation of beta-alpha superhelix folds, sometimes contain long loop structures instead of helices that link with successive beta-strands. The toll-like receptors that involved in binding of pathogens with danger associated molecular patterns, possess ten consecutive LRR motifs (Xiang *et al.*, 2006). Xa21 possesses two different type of domains viz, N terminal leucine rich repeat kinase like protein (LRR) beta-alpha subunit domain and C-terminal Serine/Threonine protein kinases (S_TKc) catalytic domain, conserved protein domain that involved in protein kinase's catalytic function, phosphorylation process (works as on/off switch for different cellular processes) and play an important role in the embryonic development and immune system (Cross *et al.*, 2000). Many diseases in human including cancer causes due to abnormal phosphorylation and drugs that affect phosphorylation can treat those diseases. For

SMART results conformation, searches are performed in CDD and Pfam databases. Results of the previously discussed tools and databases are perfectly correlated with that of the results from the InterProscan, which showed that Xa15, Xa19, Xa20 and Xa21 possess the conserved motifs and protein domains of the superfamily (Pkinase-PF00069 and LRV-PF01816) which is directly or indirectly involved in physiological responses, in the nervous and immune system and in the regulation of different cellular processes like apoptosis, cell cycle progression and the phosphorylation that ultimately leads to disease control (Cross *et al.*, 2000). SignalP5.0 predicted that there is no probability of any signal peptide cleavage sites in Xa15, Xa19 and Xa20, but Xa21 have 0.8092 probability of signal peptide cleavage sites at the position of 23 and 24.

MEME suit predicted three different motifs from these Xa amino acid sequences. Xa protein sequences that represented the conserved residues in the sequences using sequence logos (Fig. 4). The maximum selected width of identified motifs was 50 amino acids. The conserveness of residues revealed that Xa15, Xa19, Xa20 and Xa21 are closely related to each other. So, they share a common cluster within a phylogenetic tree. The locations of motifs and their consensus sequences are shown in the figure (Fig. 5).

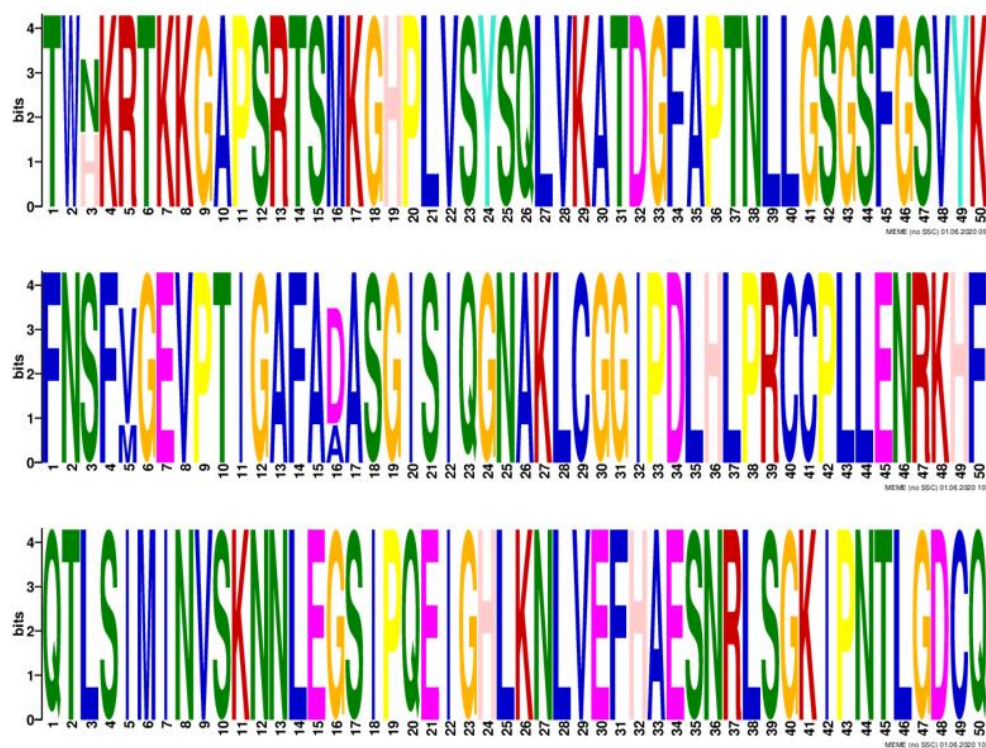


Figure 4: Sequence logo of conserved region among Xa15, Xa19, Xa20, and Xa21 proteins to show the conserveness of the residues at that point, generated by MEME suit.

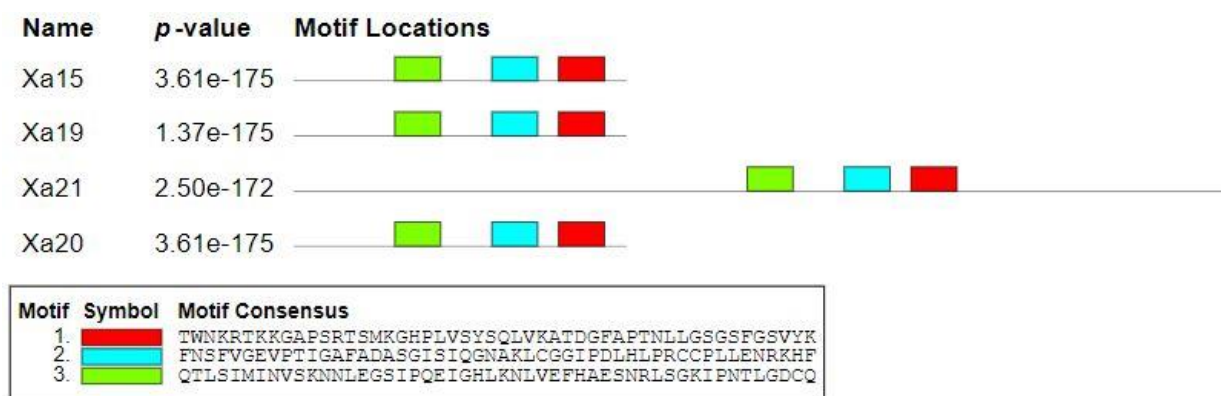


Figure 5: Consensus sequences in Xa15, Xa19, Xa20 and Xa21 and their positions. The color shows the motif of a specific sequence and sequence against that color motif is defined in the bottom of the image.

Primary structure analysis

The ProtParam tool employed to compute primary structural parameters of the Xa proteins like molecular weight, that is 38.85 kDa for Xa15, 38.93 kDa for Xa19, 38.83kDa for Xa20 and 111.33 kDa for Xa21 protein. The isoelectric point (pI) refers to the pH at which net charge on protein is zero and the surface of protein is covered with charge. At pI, proteins behave as compact and stable. The protein Xa15, Xa19 and Xa20 had same pI of 9.30, indicating its basic nature ($pI > 7.0$) and Xa 21 have pI of 7.35, indicates its neutral nature ($pI \sim 7.0$). The aliphatic index (AI) is defined as the relative volume of a protein occupied by aliphatic side chains such as alanine, valine, isoleucine, and leucine. It is considered as a positive factor for the increase of thermal stability of globular proteins. Aliphatic index of rice Xa15, Xa19 and Xa20 protein was very high (i.e., 115 and 109 in case of Xa21), indicating that these proteins may be stable for a wide range of temperature (Ikai, 1980). The instability index provides approximation about the stability of protein in in-vitro environment. A protein possesses instability index smaller than 40 is considered as stable while value greater than 40 considered that the protein may be unstable (Guruprasad *et al.*, 1990). It was calculated that the instability index of these protein is about 21.91, 22.45, 22.52 and 32.79 respectively, which indicated the stable nature of the protein structure. The GRAVY index of Xa15 is 0.107, for Xa19 is 0.108, for Xa20 is 0.106 and for Xa21, it is 0.049. These were on positive side and show their slightly hydrophobicity nature ($GRAVY > 0$), indicating its low affinity for water. Extinction coefficient is the amount of light absorption by the protein at a specific wavelength. In case of rice protein extinction coefficient of Xa15, Xa19 and Xa20 is 23170, and for Xa21, it is 87415.

Secondary structure analysis

In order to find the structural features of the Xa proteins, its secondary structures predicted using SOPMA tool. A high proportion of random coils and alpha helix were found in the protein structures. The results of predicted secondary structures showed that the proportion of extended strands (β folds), α helices, β turns, and random coils accounted for Xa15; 14.88, 34.44, 4.13, and 46.56%, for Xa19; 13.22, 34.71, 4.41 and 47.66%, for Xa20; 16.25, 32.51, 4.68 and 46.56%,

for Xa21; 14.54, 35.32, 4.29 and 45.85% of the secondary structure, respectively. TMHMM tool predicted the presence of one transmembrane helix of 22.72 amino acids in Xa15, one transmembrane helix of 22.80 amino acids in Xa19, one transmembrane helix of 22.69 amino acids in 20 at location 267-289 and one transmembrane helix of 23.09 amino acids in 21, at location 652-674.

Comparative analysis with other Xa genes

The comparative analysis of Xa proteins show the sequence similarity between these proteins as shown in Percentage Sequence Identity Matrix (PSIM) Table 1.

Table 1: Percentage Sequence Identity Matrix (PSIM)

	Xa1	Xa4	Xa5	Xa10	Xa13	Xa15	Xa19	Xa20	Xa21	Xa23	Xa25	Xa26	Xa27
Xa1	100	-	-	-	-	31.17	31.17	31.17	32.47	-	50.00	23.85	45.45
Xa4	-	100	-	-	-	28.00	26.09	28.00	24.11	-	-	27.56	-
Xa5	-	-	100	-	-	62.50	62.50	62.50	20.45	-	-	-	-
Xa10	-	-	-	100	-	-	-	-	-	35.78	-	-	-
Xa13	-	-	-	-	100	-	-	-	-	-	56.48	53.85	27.69
Xa15	27.74	60.87	62.50	-	-	100	99.17	99.72	98.07	-	-	36.76	-
Xa19	27.74	60.87	62.50	-	-	99.17	100	98.90	98.35	-	33.33	36.76	-
Xa20	27.74	60.87	62.50	-	-	99.72	98.90	100	97.80	-	-	36.49	-
Xa21	28.68	28.15	62.50	-	-	98.07	98.35	97.80	100	-	35.48	34.85	-
Xa23	-	-	-	35.78	-	-	-	-	-	100	-	-	-
Xa25	50.00	-	-	-	56.48	-	33.33	-	35.48	-	100	-	-
Xa26	23.49	27.56	-	-	53.85	36.76	36.76	36.49	37.22	-	-	100	27.59
Xa27	45.45	-	-	-	27.69	-	-	-	-	-	-	66.67	100

The phylogenetic analysis showed the relationship between Xa proteins. The results revealed that the Xa19, Xa21, Xa15 and Xa20 fall in the same cluster because they are more closely related to each other due to higher sequence similarity between them. They are also closely related to Xa26. So, fall in the same clade. Meanwhile, Xa13 and Xa25 have the same cluster due to their closely relatedness. Xa5 also makes clade with Xa13 and Xa25 due to low evolutionary distance from these Xa proteins. Xa10 and Xa23 make their separate clade due to weak relationship with other proteins. Xa27, Xa1 and Xa4 also have some relationship with the lower clade (Fig. 6).

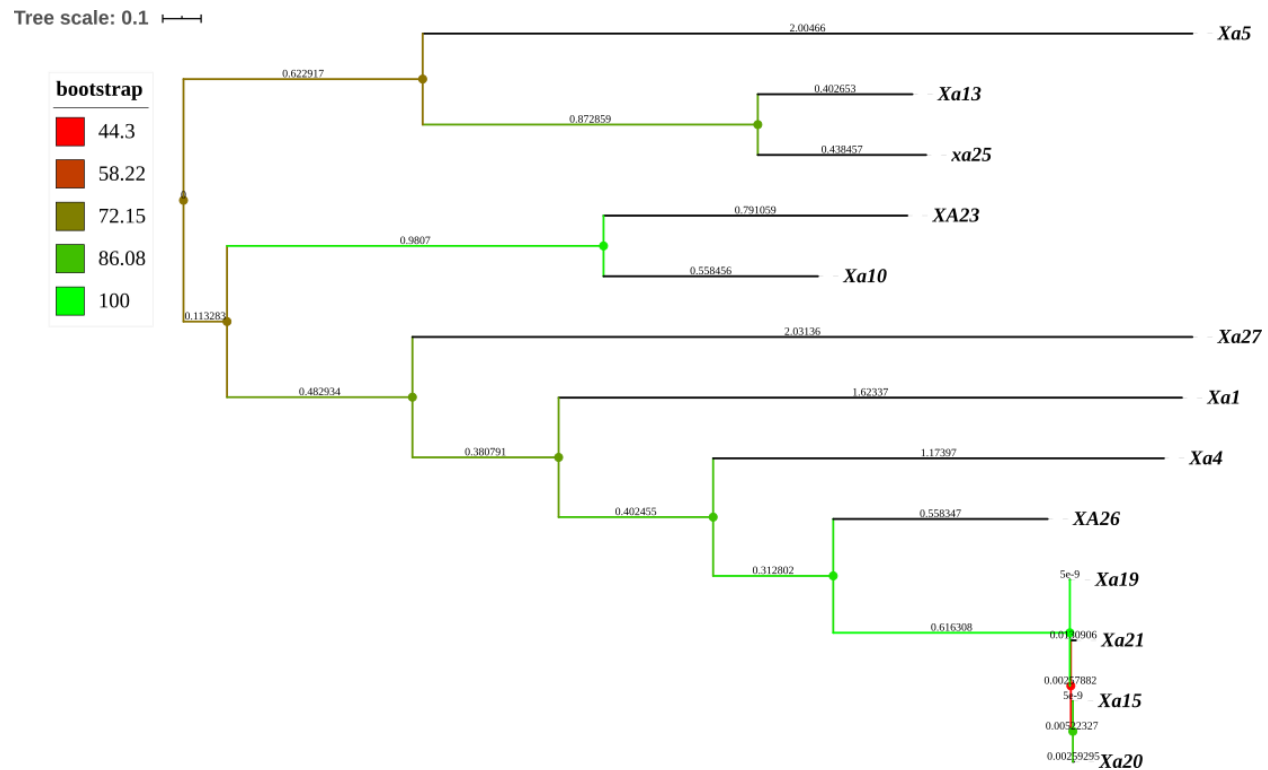


Figure 6: Neighbor joining based evolutionary tree inferred from all available Xa proteins with the help of FastTree tool. (color of the branches represents the percentile bootstrap value according to the define parameters and Tree scale 0.1 shows the substitution at per base position).

Molecular evolutionary analysis

For comparative sequence analysis of rice Xa proteins, BLAST tool against non-redundant (nr) database results that these Xa genes are closely related to the family members of the leucine rich repeats (LRR, CL0022) and Serine/Threonine protein kinases catalytic domains (S_TKc, CL33413) from various species of plants that are evolutionary related to eudicots and other monocot species. The sequences from the BLAST results were selected based upon the highest score, identity percentage (>95%), Query coverage (>90%) and lowest E-value (near to Zero) with responding to our query. Sequences from 11 different species including Xa proteins for LRR domain and 9 different species related to S_TKc domain including Xa21 protein were compiled for multiple sequence alignment using MAFFT tool (Fig. 7). The MSA revealed that the leucine rich repeats and serine/threonine protein kinase are highly conserved in these species during over all evolutionary pathway.

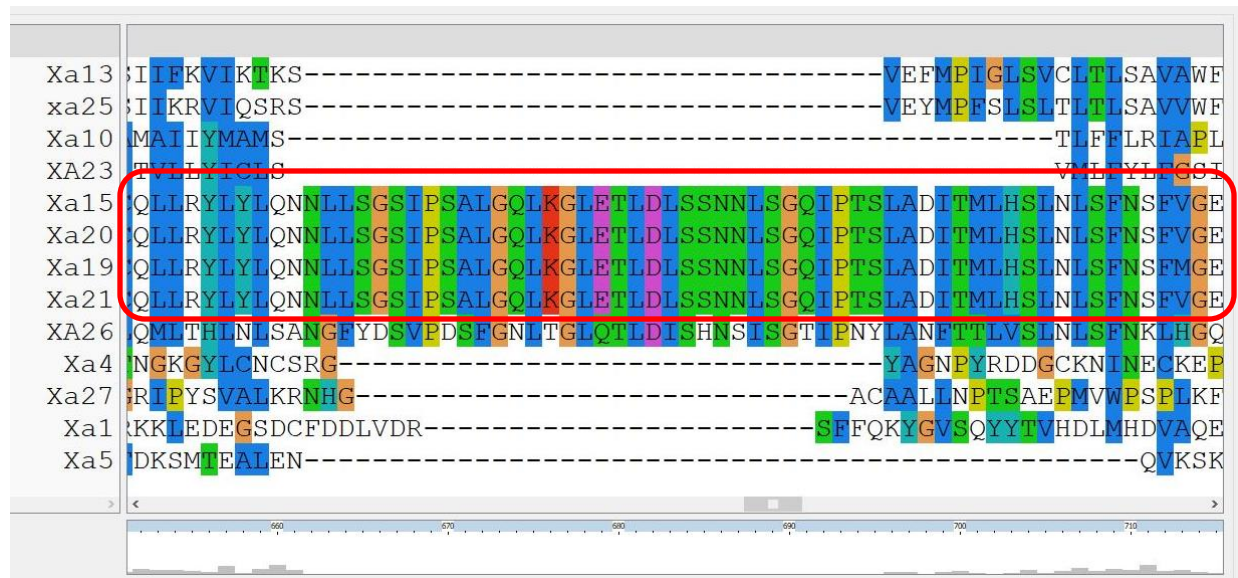


Figure 7: Multiple sequence alignment of all available Xa proteins using MAFFT aligning tool (view of more conserved region)

The evolutionary analysis of these species was carried out using the approximately maximum likelihood method in FastTree that generate their mid rooted evolutionary tree. This tree divided the species into two major groups (cluster I and cluster II) on the basis of specific divergence among them with strong bootstrap values within their node (green color for highest and red for lowest value). The LRR cross ponding tree consists of one group, cluster II for different species of *Oryza* genus (*Oryza indica*, *Oryza japonica*, and *Oryza longistaminata*) and one group clusterII for species other than rice which include eudicots,(i.e. *Kingdonia uniflora*, *Nelumbo mucifera*, *papaver somniferum*, *Aquilegia coerulea*, and *thalictrum thalictroides*) (Fig. 8). The S_TKc cross-ponding tree consists only monocot related species that are further divided into two major groups, cluster I for *Oryza* species (*OS indica*, *OS japonica*, and *Oryzaongistaminata*) and other cluster II for other species (i.e. *Brachypodium distachyon*, *Panicum miliaceum*, *Aegilops tauschii* and *Triticum turgidum*) (Fig. 9). All LRR and S_TKc sequences from *OS japonica* was clustered together with *OS indica* which indicates, *indica* and *japonica* are highly similar in contrast to other members of the tree.

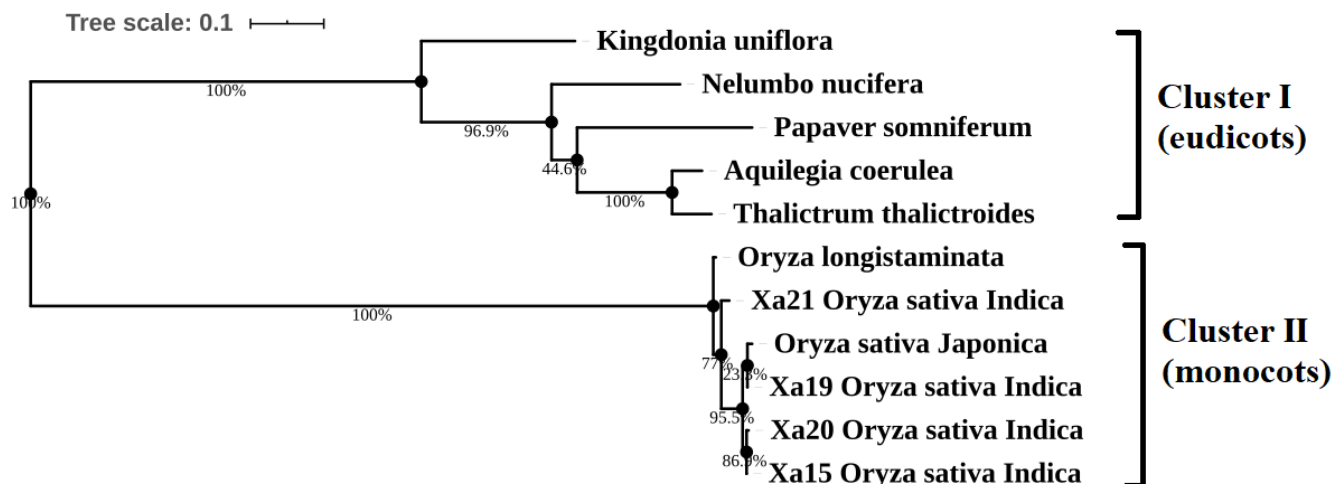


Figure 8: LRR domain based evolutionary tree among rice and other species having same domain using NJ approach. (percentage on the branches represents the percentile bootstrap value. Tree Scale 0.1 shows the substitution per base position. Cluster-I is for all eudicots species and Cluster-II is for all monocots species including rice).

Tree scale: 0.01

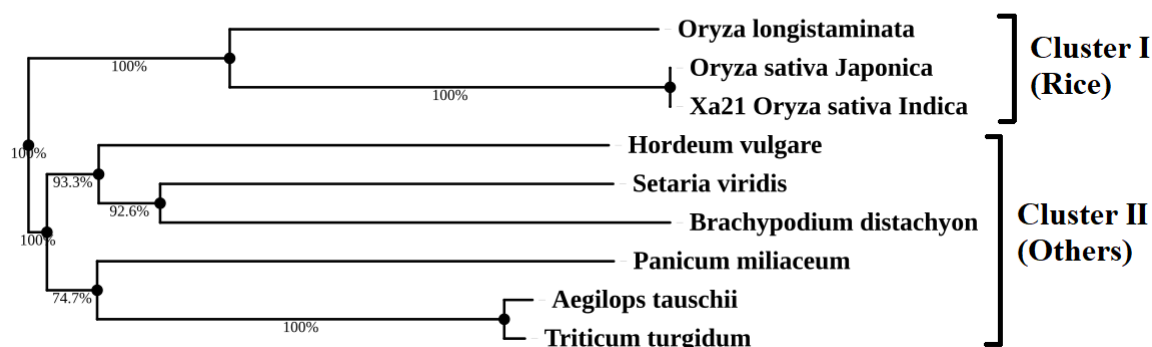


Figure 9: S_TKc domain based evolutionary tree among rice and different species have same domain using NJ approach. (percentage on the branches represents the percentile bootstrap value. Tree Scale 0.1 shows the substitution per base position. Cluster-I is for rice species and Cluster-II is for all other species).

Comparative modeling of Xa proteins

For 3D structural prediction of proteins, comparative modeling also known as homology modeling, considered as one of the best approaches that outcomes structural models having high precision and accuracy as compared to other similar approaches. It is taken as a method of choice for the proteins that have closely related sequence similarity due to their evolutionary relationship. This technique used the idea that if two proteins possess sequence similarity then they also have structural similarities. For a precise and reliable alignment between target and

template high sequence identity is preferred. 5gr8A, 5gr9B, 6s6q.2.A, 5hyxB, 5gijB, 4mn8A, 4j0mA and 3rgxA were predicted top templates with the help of I-TASSER. 6s6q.2.A that have highest similarity with Xa15, Xa19 and Xa20 and 4mn8A with Xa21 as shown in Table 2. Swiss model server search also suggested that crystal structures of the 6s6q.2.A (chain-A) and 4mn8A (chain-A) in *Arabidopsis thaliana* with a resolution of 2.95 Å and 3.06 Å respectively, was the best suited templates for these Xa proteins. Sequences of these templates were pairwise aligned with the Xa proteins sequences by using CLUSTAL-X and Seaviewer tool was utilized to visualize the alignment. Based on the target and template alignment, MODELLER9v7 generated 20 rough models of queried proteins sequences using 6s6q (for Xa15, 19 and 20) and 4mn8 (for Xa21) as templates. From these 20 models, the model having lowest DOPE score and highest GA3141 score was considered to be thermodynamically stable and chosen for further validation and refinement. The secondary structure of selected predicted models (Fig.11) and selected templates (Fig 12) were visualized using chimera tool and compared which conclude that, the predicted model have some conserved pattern with their template used for model building. The secondary structure comparison between the predicted models and selected templates showed both of comprised of helices and beta sheets that shared strong homology across the entire length. The identity of each alpha and beta sheets between target and template is shown in Table 3.

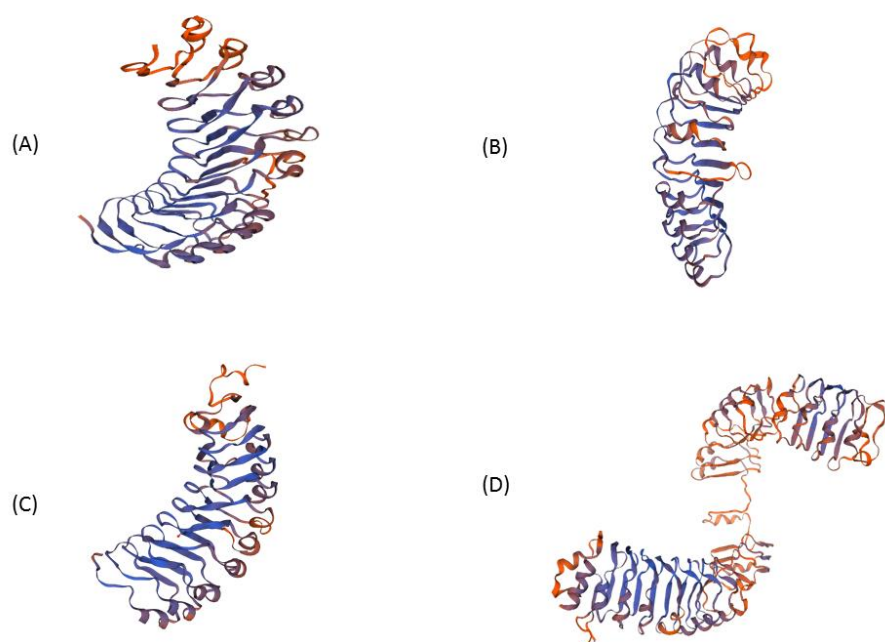
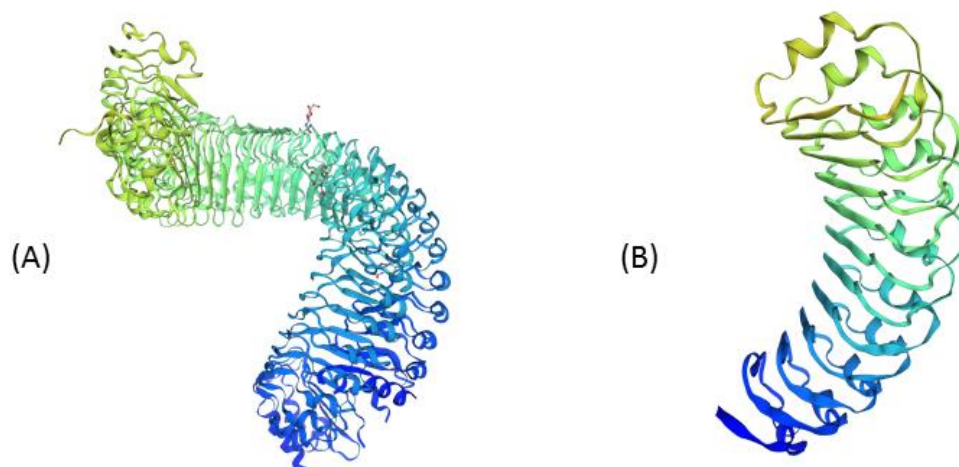


Figure 10: 3D view of secondary structures of homology based predicted Xa proteins models: (A) Snip of Xa15 model, (B) Snip of Xa19 model, (C) Snip of Xa20 model, and (D) Snip of Xa21 model using chimera.

Table 2: Selected templates for homology modeling from BLAST against PDB database

Sr. No.	Templates	Total score	Identity (%)	Query coverage (%)	E value	Resolution (Å°)
1	4mn8_A (for Xa15)	1013	36.54	69	7e-21	3.06
2	6s6q_A	1414	40.96	73	1e-40	2.95
3	5hyx_B	696	38.13	84	1e-39	2.60
4	5gr8_A	1233	38.55	98	6e-36	2.59
5	4mn8_A (for Xa19)	1007	36.54	69	1e-40	3.06
6	6s6q_A	1408	40.96	73	2e-40	2.95
7	5hyx_B	819	38.13	84	2e-39	2.60
8	5gr8_A	1226	38.55	98	9e-36	2.59
9	4mn8_A (for Xa20)	1123	36.21	69	1e-40	3.06
10	6s6q_A	1326	40.56	73	2e-40	2.95
11	5hyx_B	694	38.13	84	2e-39	2.60
12	5gr8_A	1353	38.55	98	7e-36	2.59
13	4mn8_A (for Xa21)	917	32.69	58	6e-92	3.06
14	6s6q_A	1073	36.79	66	6e-84	2.95
15	5hyx_B	267	33.86	56	2e-77	2.60
16	5gr8_A	463	32.28	57	7e-73	2.59

**Figure 11: 3D view of secondary structures of the proteins selected as templates for comparison based homology modeling: (A) Image of 4mn8(Chain-A), and (B) Image of 6s6q (Chain-A).****Table 3: Secondary structure comparison of selected templates and predicted models**

Sr. No.	Target/ template	Number of amino acid residues (%)				Total No. of amino acids
		Turn	Helix	Strand	Coils	
1	Xa15	4.13	34.44	14.88	46.56	363
2	Xa19	4.41	34.71	13.22	47.66	363
3	Xa20	4.68	32.51	16.25	46.56	363
4	Xa21	4.29	35.32	14.54	45.85	1025
5	6s6q	1.39	38.24	9.50	50.87	863
6	4mn8	1.28	40.54	10.36	47.83	782

Model assessment and validation

The conserveness of the secondary structure revealed the reliability and robustness of predicted models computed by MODELLER9v7 based on the alignment between target and template. The graphical view of that alignment shown with the help of GNUPLOT (Fig. 12).

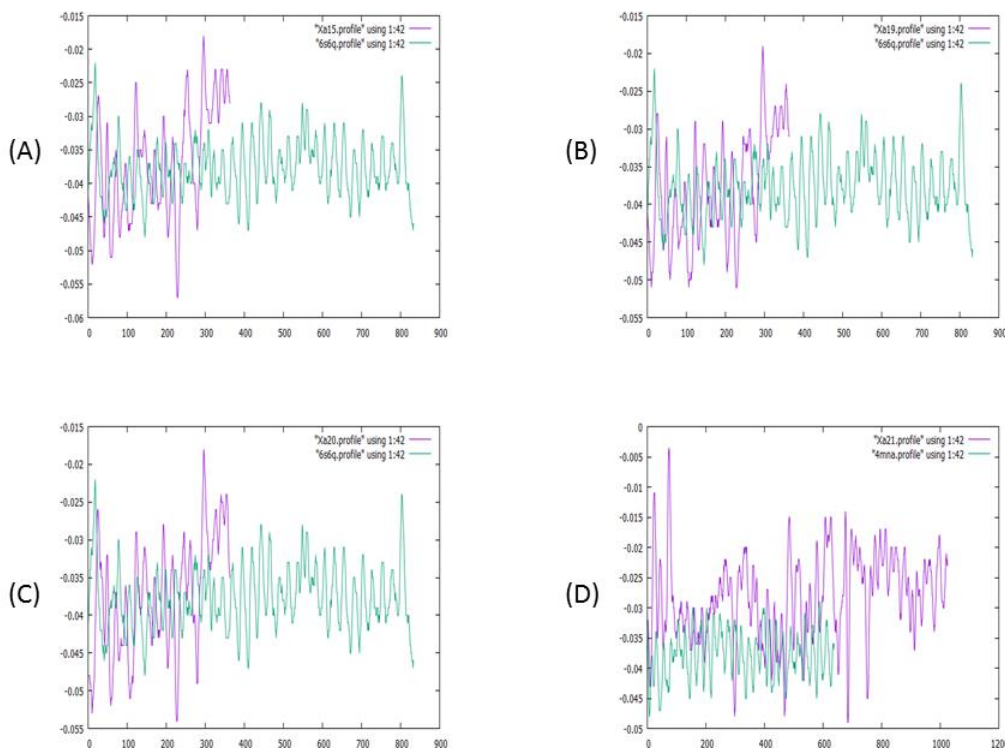


Figure 12: GNU plots of predicted models of Xa15, Xa19, Xa20 and Xa21 proteins respectively created by GNU plotter (purple colored curve indicate the cross-ponding Xa protein and green colored curve indicates the selected template).

To check the reliability of dihedral angles and stereo chemistry of the model for backbone confirmation the Swiss model on ExPasy online server (<https://swissmodel.expasy.org/interactive>) was interpreted and computed the residues fall in the available zone of Ramachandran plot (Fig. 13). The Ramachandran plot analysis of the models showed that the 88.08% residues of Xa15 (Fig. 13a), 88.08% residues of Xa19 (Fig. 13b), 86.72% residues of Xa20 (Fig. 13c) and 92.67% residues of Xa21 (Fig. 13d) fell in the most favored regions. 11.98% residues of Xa15 and Xa19, 13.28% residues of Xa20 and 7.33% residues of Xa21 fall in the additional allowed regions, and no residue of them was found in the

generously allowed and disallowed region. The residues percentage in computed 3D models were nearly similar with the Ramachandran plot of templates (Fig. 14) as shown in Table 4, which strongly recommend the computed structural models as precise and reliable for its backbone conformation. Similarly, the VERIFY-3D (<https://servicesn.mbi.ucla.edu/Verify3D>) tool determined more than 80% residue score > 0.2 which shows the atomic-model (3D) compatibility with its own residues (1D), (Lüthy *et al.*, 1992).

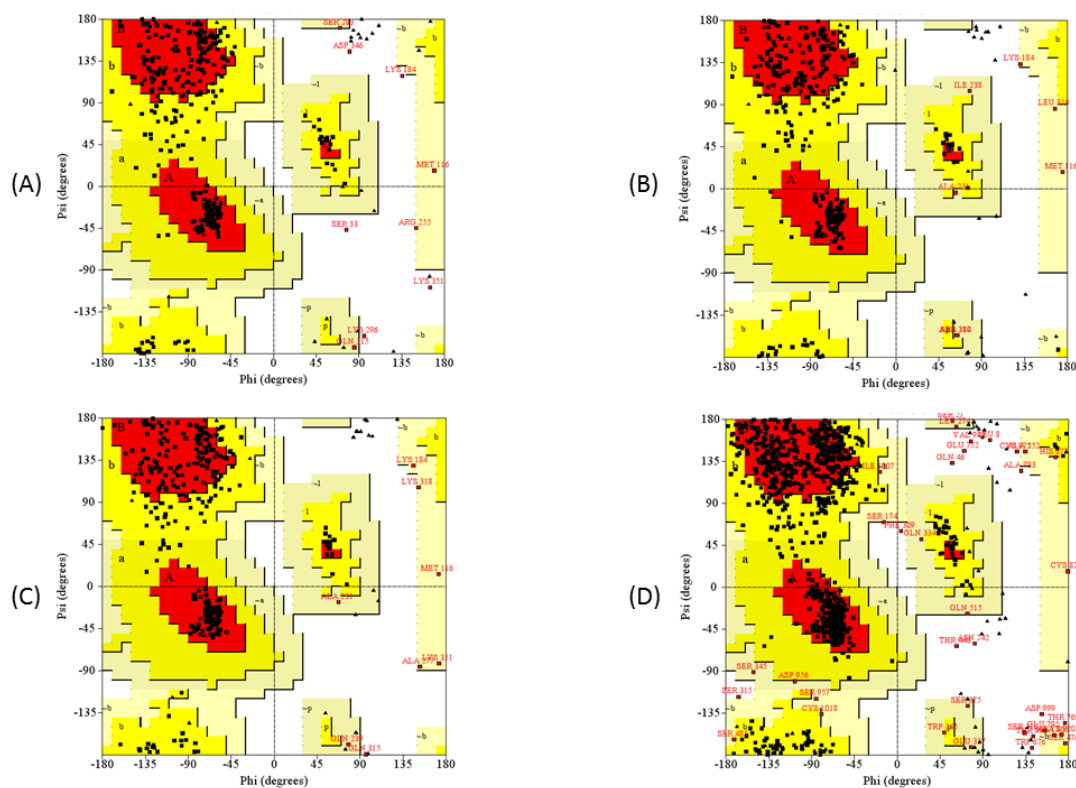


Figure 13: Ramachandran plot of predicted structures of proteins: In (A) Ramachandran plot of Xa15 protein, (B) Ramachandran plot of Xa19 protein, (C) Ramachandran plot of Xa20 protein, (D) Ramachandran plot of Xa21 protein.

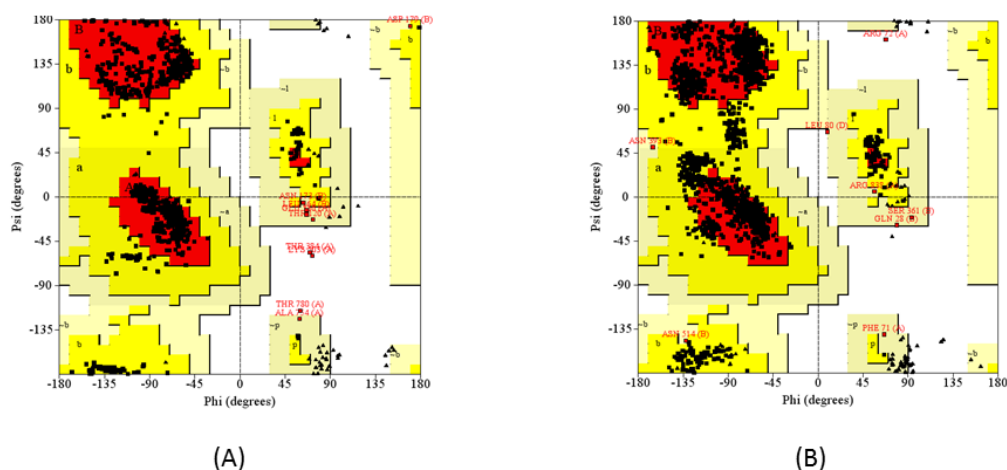


Figure 14: Ramachandran plot of selected templates (4mn8 and 6s6q respectively)

This result indicates that the proposed models are reliable. WHATCHECK (<https://servicesn.mbi.ucla.edu/WHATCHHECK>) server used to accessed the coarse packing quality of the protein, planarity, collision with symmetry axis and to compute the anomalous bond length, anomalous bond angles, proline puckering, and distribution of omega angles of the model protein, indicate that it is acceptable and of good quality. PROVE (<https://servicesn.mbi.ucla.edu/PROVE>) program computed the Z-score mean, Z-score Standard deviation and Z-score RMS of the models to estimates the average magnitude of the volume irregularities. The Z-score RMS values of the predicted models were 1.65 for all Xa15, Xa19, Xa20 and xa21, respectively which is positive and near to 1 and indicates the moderate resolution of our predicted models (Z-score RMS value in positive direction or near to one represents the structures have good resolution). The results of MolProbity server (<http://molprobity.biochem.duke.edu/index.php>) showed that our computed models had 0% bad bonds and only 1% residues had bad angels (a good model have both values close to 0%) that further confirmed their reliability. To ensure the Comparative model quality the computed model structures were compared with the selected templates structure that was predicted using X-ray diffraction technique (2.95Å), using atoms RMSD assessment and the superimposition.

Table 4: Ramachandran plot result elucidation

Sr. #	Protein Name	Model length (No. of Residues)	Amino-acid Residues in			
			Most Favored regions	Additional Allowed regions	Generously Allowed regions	Disallowed regions
1.	Xa15	262	75.0%	53	2	0
2.	Xa19	262	75.0%	53	2	0
3.	Xa20	262	75.0%	53	2	0
4.	Xa21	602	80.9%	92	5	2
5.	6s6q.2	857	74.2%	189	4	1
6.	4mn8.1	971	82.8%	134	6	3

MATRAS server (http://strcomp.protein.osaka-u.ac.jp/matras/matras_pair.html) computed pairwise alignment of 3-D structures of the predicted models and selected templates that provides the key elements of secondary structure and proves the strongly conserveness (Reliability of folds and Superfamily were 93.9% and 64.4% respectively) in the alignment where there exists a sequence similarity of ~35% and normalized S_dis score (28.5%). The DRMS similarity (deviation of CA trace) between the structures of predicted models and selected template was 0.93Å that denotes, the predicted models are reasonably good and similar to the selected templates as in Figure 15.

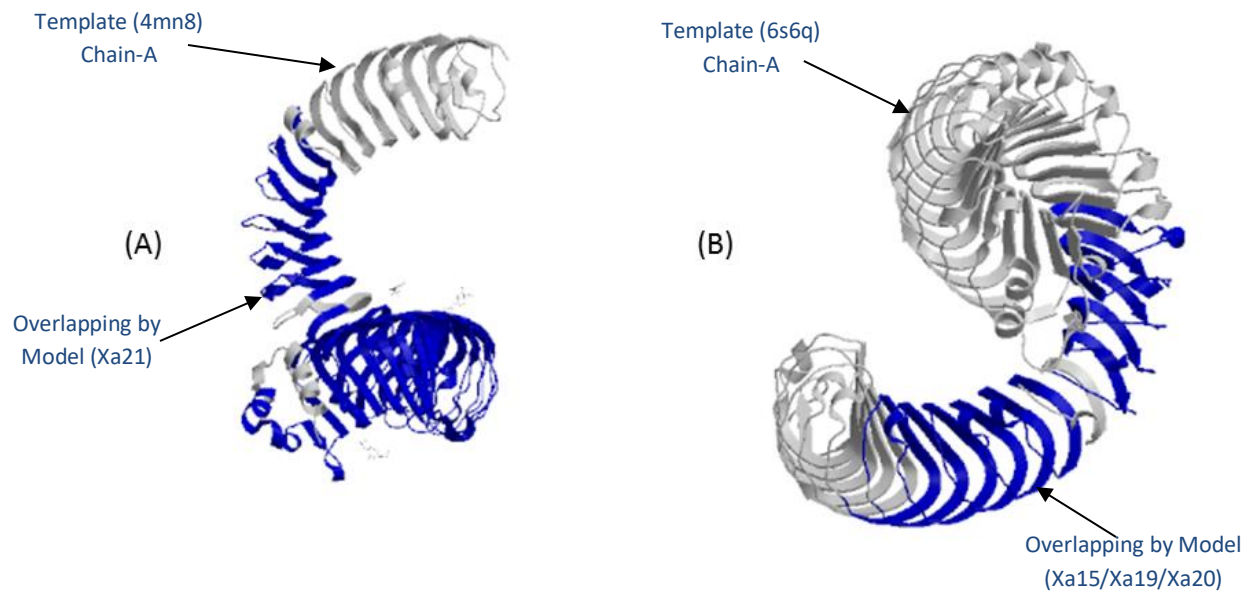


Figure 15: The 3D secondary structure alignment of selected templates and predicted models. (A) Structural alignment of Xa21 with 4mn8 (Chain-A). (B) Structural alignment of Xa (15/19/20) proteins with 6s6q (Chain-A). (Blue color represents the overlapping region on template by predicted models' structures and grey represents the structure of selected templates).

Conclusion

The dominant genes Xa15, Xa19, Xa20 and Xa21 belong to family of Xa gene in rice and encode beta-alpha kinase protein subunits and provide complete resistance race specifically against the pathogen *Xoo*. The characterization of these Xa genes revealed that these are highly similar (more than 95%) with each other and possess similar domains of leucine rich repeat kinase like proteins. Xa21 is greater in sequence length (1025) and have a patch of 363 amino acids similar to the Xa15, Xa19 and Xa20. In the addition, a patch of amino acids sequence possesses a small subunit of serine/threonine kinase protein (S_TKc). These kinase-like proteins are involved in kinase's catalytic functions and phosphorylation process. The phosphorylation process work as an on/off switch for the different cellular processes and play an important role in the resistance causing in many species including rice. The abnormality in this process leads to the damaging or lacking the immunity of the cells against any disease or pathogen attack. Because of the involvement of these proteins in phosphorylation process, the presence of these genes in any cultivar of the crop will leads to the regularity in this process which ultimately cause the resistance against pathogens. These phosphorylated amino acids have recently been identified to be present in human cell extracts and fixed human cells using a combination of antibody-based analysis. The *Homo sapiens* also possess the abnormality in the phosphorylation process which leads to cause cancer. The leucine rich repeat kinase like protein is the member of family LRV which directly involved in catalytic activity of cell like physiological responses and regulation different cellular processes, that are directly or indirectly involved in the resistance

process and the presence of both domains may lead these Xa genes to cause resistance against bacterial leaf blight disease in rice and these domains in any other species will also help for causing resistance against pathogens. The prediction of 3-D structures of these genes was done using the comparative based modeling method (homology modeling) carried out by using LRR and S_TKc based templates (6s6q and 4mn8) by comparing them with Xa genes and used due to their highest similarity with them. The modeled structures were observed homologous to the templates that used for structure prediction computationally, concluded that these Xa may also possess the same biological processes in rice. Accuracy of these predicted models was computed using different model assessment and validation techniques like Ramachandran plot analysis, target-template structural alignment, back-bone conformation of the models, compatibility of atomic model with their own amino acid sequences, and the coarse packing quality of the protein. The phylogenetic analysis showed the evolution of these kinase proteins in different species of monocots and eudicots clades. The evolution shows the diversity of these kinase proteins in regard to the biological process with the passage of time and due to segregation of the genes in these species results in the wide spread of the functional abilities.

The domains identification and the characterization will lead us to a specific domain that involve in resistance biologically. The *Xoo* races are changing continuously and we can use these identified specific regions to predict new genes and pairing arrangements of genes for the upcoming races.

Acknowledgement

Authors acknowledge University of Agriculture Faisalabad, Pakistan, Nuclear Institute for Agriculture and Biology (NIAB) Faisalabad, Pakistan and International Atomic Energy Agency, Vienna for their support and conductance.

References

- Abdullah, M., J. Li, S. Ghazanfar, J. Ahmed, I. Khan and M.N. Ishaq. 2015. Where Pakistan Stands Among Top Rice Exporting Countries, an Analysis of Competitiveness. *J. Northeast Agric. Univ. (English Ed.* 22:80–86.
- Antolín-Llovera, M., M.K. Ried, A. Binder and M. Parniske. 2012. Receptor Kinase Signaling Pathways in Plant-Microbe Interactions. *Annu. Rev. Phytopathol.* 50:451–473.
- Bailey, T.L., M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li and W.S. Noble. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37:W202–W208.
- Berman, H.M. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- Chen, J.E., C.C. Huang and T.E. Ferrin. 2015. RRDistMaps: a UCSF Chimera tool for viewing and comparing protein distance maps. *Bioinformatics* 31:1484–1486.
- Chen, V.B., W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson and D.C. Richardson. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.*

66:12–21.

- Chu, Z., B. Fu, H. Yang, C. Xu, Z. Li, A. Sanchez, Y.J. Park, J.L. Bennetzen, Q. Zhang and S. Wang. 2006. Targeting xa13, a recessive gene for bacterial blight resistance in rice. *Theor. Appl. Genet.* 112:455–461.
- Cross, T.G., D. Scheel-Toellner, N. V. Henriquez, E. Deacon, M. Salmon and J.M. Lord. 2000. Serine/Threonine Protein Kinases and Apoptosis. *Exp. Cell Res.* 256:34–41.
- Dehury, B., M. Sahu, K. Sarma, J. Sahu, P. Sen, M.K. Modi, G.D. Sharma, M.D. Choudhury and M. Barooah. 2013. Molecular Phylogeny, Homology Modeling, and Molecular Dynamics Simulation of Race-Specific Bacterial Blight Disease Resistance Protein (xa 5) of Rice: A Comparative Agriproteomics Approach. *Omi. A J. Integr. Biol.* 17:423–438.
- Denning, G.L. and T.X. Võ. 1995. *Vietnam and IRRI: a partnership in rice research: proceedings of a conference held in Hanoi, Vietnam, 4-7 May 1994*. Int. Rice Res. Inst.
- Dym, O., D. Eisenberg and T.O. Yeates. 2012. ERRAT.
- El-Gebali, S., J. Mistry, A. Bateman, S.R. Eddy, A. Luciani, S.C. Potter, M. Qureshi, L.J. Richardson, G.A. Salazar, A. Smart, E.L.L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S.C.E. Tosatto and R.D. Finn. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res.* 47:D427–D432.
- Fairhurst, T. and A. Dobermann. 2002. Rice in the Global Food Supply. *Better Corps Int.* 16:3–6.
- Geourjon, C. and G. Deléage. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11:681–684.
- Ghadirnezhad, R. and A. Fallah. 2014. Temperature effect on yield and yield components of different rice cultivars in flowering stage. *Int. J. Agron.* 2014:846707.
- Guruprasad, K., B.V.B. Reddy and M.W. Pandit. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng. Des. Sel.* 4:155–161.
- Ikai, A. 1980. Thermostability and Aliphatic Index of Globular Proteins. *J. Biochem.* 88:1895–1898.
- Institute, I.R.R. 1985. *International rice research: 25 years of partnership*. Manila (Philippines) IRRI.
- Jin, G., L. Nakhleh, S. Snir and T. Tuller. 2006. Maximum likelihood of phylogenetic networks. *Bioinformatics* 22:2604–2611.
- Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis and T.L. Madden. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36:W5–W9.
- Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A.F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez and S. Hunter. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240.
- K. Reddy, A.P. 1979. Relationship Between Nitrogen Fertilization, Bacterial Leaf Blight

Severity, and Yield of Rice. *Phytopathology* 69:970.

Kawabata, T. 2003. MATRAS: a program for protein 3D structure comparison. *Nucleic Acids Res.* 31:3367–3369.

Kawahara, Y., M. de la Bastide, J.P. Hamilton, H. Kanamori, W.R. McCombie, S. Ouyang, D.C. Schwartz, T. Tanaka, J. Wu, S. Zhou, K.L. Childs, R.M. Davidson, H. Lin, L. Quesada-Ocampo, B. Vaillancourt, H. Sakai, S.S. Lee, J. Kim, H. Numa, T. Itoh, C.R. Buell and T. Matsumoto. 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4.

Khan, J.A., H.M.I. Arshad, K. Saleem, A.F. Sandhu, S. Hasnain and M.M. Babar. 2012. Evaluation of resistance genes in rice against local isolates of *Xanthomonas oryzae* pv. *oryzae* in Punjab Province of Pakistan. *Arch. Phytopathol. Plant Prot.* 45:1826–1839.

Khush, G.S. and K.K. Jena. 2009. Current Status and Future Prospects for Research on Blast Resistance in Rice (*Oryza sativa* L.). *Advances in Genetics, Genomics and Control of Rice Blast Disease*. Springer Netherlands, Dordrecht. pp.1–10.

Krogh, A., B. Larsson, G. von Heijne and E.L.L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes¹¹ Edited by F. Cohen. *J. Mol. Biol.* 305:567–580.

Kubo, M. and M. Purevdorj. 2004. The Future of Rice Production and Consumption. *J. Food Distrib. Res.* 35:128–142.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.

Lemus, C., A. Angelis, M. Halabalaki and A.L. Skaltsounis. 2014. γ -Oryzanol. *Wheat and Rice in Disease Prevention and Health*. Elsevier. pp.409–430.

Letunic, I. and P. Bork. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44:W242–W245.

Lu, S., J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, M. Yang, D. Zhang, C. Zheng, C.J. Lanczycki and A. Marchler-Bauer. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48:D265–D268.

Lüthy, R., J.U. Bowie and D. Eisenberg. 1992. Assessment of protein models with three-dimensional profiles. *Nature* 356:83–85.

Macleán, J.L., D.C. Dawe, G.P. Hettel and B. Hardy. 2002. *Rice Almanac: source book for the most important economic activity on earth* J.L. Macleán, D.C. Dawe, B. Hardy, and G.P. Hettel (eds.). CABI, Wallingford.

Madeira, F., Y.M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R.N. Tivey, S.C. Potter, R.D. Finn and R. Lopez. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47:W636–W641.

Naqvi, S.A.H. 2019. Bacterial Leaf Blight of Rice: An Overview of Epidemiology and Management with Special Reference to Indian-Sub-Continent. *Pakistan J. Agric. Res.*

32:359.

- Nas, M. 2008. Rice diseases series (part 1)-bacterial leaf blight. *Agric.*
- Nielsen, H., J. Engelbrecht, S. Brunak and G. von Heijne. 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int. J. Neural Syst.* 8:581–599.
- Price, M.N., P.S. Dehal and A.P. Arkin. 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* 26:1641–1650.
- ProtParam, E. 2017. ExPASy-ProtParam tool.
- Racine, J. 2006. gnuplot 4.0: a portable interactive plotting utility. *J. Appl. Econom.* 21:133–141.
- Rahman, M.A., A. Saboor, I.A. Baig, U. Shakoore and H. Kanwal. 2017. An investigation of the impact of climate change on rice crop in Pakistan: A multivariate analysis. *Pakistan J. Agric. Sci.* 54:561–566.
- Rehman, A., L. Jingdong, B. Shahzad, A.A. Chandio, I. Hussain, G. Nabi and M.S. Iqbal. 2015. Economic perspectives of major field crops of Pakistan: An empirical study. *Pacific Sci. Rev. B Humanit. Soc. Sci.* 1:145–158.
- Sahu, S.K., P. Zheng and N. Yao. 2018. Niclosamide Blocks Rice Leaf Blight by Inhibiting Biofilm Formation of *Xanthomonas oryzae*. *Front. Plant Sci.* 9:408.
- Schultz, J., F. Milpetz, P. Bork and C.P. Ponting. 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci.* 95:5857–5864.
- Song, W.Y., G.L. Wang, L.L. Chen, H.S. Kim, L.Y. Pi, T. Holsten, J. Gardner, B. Wang, W.X. Zhai, L.H. Zhu, C. Fauquet and P. Ronald. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* (80-.). 270:1804.
- Spencer, R.K., G.L. Butterfoss, J.R. Edison, J.R. Eastwood, S. Whitelam, K. Kirshenbaum and R.N. Zuckermann. 2019. Stereochemistry of polypeptoid chain configurations. *Biopolymers* 110:e23266.
- Syed-Ab-Rahman, S.F., L.C. Carvalhais and D. Omar. 2020. Development of plant-based emulsion formulations to control bacterial leaf blight and sheath brown rot of rice. *Heliyon* 6:e03151.
- Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46:W296–W303.
- Webb, B. and A. Sali. 2016. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* 54:5.6.1–5.6.37.
- Wiederstein, M. and M.J. Sippl. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35:W407–W410.
- Wu, X.H., W. Wang, C.M. Yin, H.J. Hou, K.J. Xie and X.L. Xie. 2017. Water consumption, grain yield, and water productivity in response to field water management in double rice systems in China. *PLoS One* 12:e0189280.

- Xiang, Y., Y. Cao, C. Xu, X. Li and S. Wang. 2006. Xa3, conferring resistance for rice bacterial blight and encoding a receptor kinase-like protein, is the same as Xa26. *Theor. Appl. Genet.* 113:1347–1355.
- Yang, J. and Y. Zhang. 2015. Protein Structure and Function Prediction Using I-TASSER. *Curr. Protoc. Bioinforma.* 52:5.8.1-5.8.15.
- Yoshimura, S., U. Yamanouchi, Y. Katayose, S. Toki, Z.-X. Wang, I. Kono, N. Kurata, M. Yano, N. Iwata and T. Sasaki. 1998. Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proc. Natl. Acad. Sci.* 95:1663–1668.
- You, K.Y., L.L. You, C.S. Yue, H.K. Mun and C.Y. Lee. 2017. Physical and Chemical Characterization of Rice Using Microwave and Laboratory Methods. *Rice - Technology and Production*. InTech. pp.81–99.
- Zhang, F., D.-L. Zhuo, F. Zhang, L.-Y. Huang, W.-S. Wang, J.-L. Xu, C. Vera Cruz, Z.-K. Li and Y.-L. Zhou. 2015. Xa39 , a novel dominant gene conferring broad-spectrum resistance to *Xanthomonas oryzae* pv. *oryzae* in rice. *Plant Pathol.* 64:568–575.