International Journal of Bio-resource and Stress Management https://pphouse.org/ijbsm.php

Crossref

IJBSM January 2023, 14(1):125-131 Print ISSN 0976-3988 Online ISSN 0976-4038 **Article** AR3198a

Research Article *Natural Resource Management*

DOI: HTTPS://DOI.ORG/10.23910/1.2023.3198a

Evolutionary Relationship and Structural Analysis of Blast Resistance Associated Novel *Osvwa36* and *Osvwa37* Genes in Cultivated and Wild Species of Rice

Suhas Gorakh Karkute, Amitha Mithra Sevanthi and Amolkumar U. Solanke

CAR-National Institute for Plant Biotechnology, LBS Building, Pusa Campus, New Delhi (110 012), India

Open Access

Corresponding amol.solanke@icar.gov.in

0000-0003-2693-0639

ABSTRACT

Rice blast is a dreadful disease that causes enormous losses in rice production worldwide. To develop blast resistant rice cultivars, it is necessary to identify resistance and defence regulator genes and the underlying mechanism of resistance. A novel von Willebrand factor domain A containing genes *OsvWA36* and *OsvWA37* in Tetep cultivar of rice regulate response to *Magnaporthe oryzae* infection and provides significant resistance. Owing to the important role of these genes, their evolutionary relationship has been studied in cultivated and wild species of rice. There is significant diversity in the protein sequence of these genes among the relative wild rice species. The size of *OsvWA36* protein varies from 501 aa to 698 aa whereas size of *OsvWA37* protein varies from 295 aa to 1004 aa. The, *OsvWA36* gene is evolutionarily more conserved than *OsvWA37* among the different rice species indicating its critical role. Besides the global variation in protein sequence, the region of vWA domain is highly conserved among all the species. Interestingly, both the genes in *Oryza barthii* are fused to form a single gene encoding a large protein that indicates their origin in other species from a single gene. The good quality tertiary structures of both OsvWA36 and OsvWA37 proteins in cultivated germplasm cv. Tetep were also generated which can be utilized for protein structural and docking studies.

KEYWORDS: von Willebrand factor domain A, rice, blast, protein model, evolutionary relationship

***Citation* (VANCOUVER):** Karkute et al. Evolutionary Relationship and Structural Analysis of Blast Resistance Associated Novel *Osvwa36* and *Osvwa37* Genes in Cultivated and Wild Species of Rice. *International Journal of Bio-resource and Stress Management*, 2023; 14(1), 125-131. HTTPS://DOI.ORG/10.23910/1.2023.3198a.

Data Availability Statement: Legal restrictions are imposed on the public sharing of raw data. However, authors have full right to transfer or share the data in raw form upon request subject to either meeting the conditions of the original consents and the original research study. Further, access of data needs to meet whether the user complies with the ethical and legal obligations as data controllers to allow for secondary use of the data outside of the original study.

Conflict of interests: The authors have declared that no conflict of interest exists.

RECEIVED on 26th July 2022 RECEIVED in revised form on 15th December 2022 ACCEPTED in final form on 31st December 2022 PUBLISHED on 22nd January 2023

1. INTRODUCTION

Rice is the major crop feeding more than 50% of the global population. It is also an important economic crop for farmers and landless labourers. Asian countries contribute more than 90% to the global rice production (Fukagawa and Ziska, 2019). India is the second largest producer of rice in the world producing 177.6 million tonnes (Kumar et al., 2021). In the year 2019-2020, India exported basmati rice to the tune of US $4.33 billion whereas the non-basmati export was of US$2.01 billion (Anonymous).

Currently, there is annual shortage of rice production and the gap is fulfilled by utilising the reserve stocks. If a similar situation continues without doubling the rice production, the annual shortage is estimated to reach 800,000 tons by 2030 (Asibi et al., 2019). Rice production can be enhanced by increasing the productivity or increasing the cultivated area. Increasing the area is highly difficult task due to unavailability of arable land (Samal and Babu, 2018). Several other challenges exist such as water and labour shortage, climate change, deteriorating soil health, etc. (Nawaz et al., 2022). Another important aspect is to avoid the losses due to environmental factors. It is estimated that the cumulative effect of the several environmental factors cause more than 50% losses in the yield (Anami et al., 2020). Rice blast disease is the major threat to rice production and may cause up to 100% yield loss in case of panicle blast (Khan et al., 2014).

Breeders have identified a large number of genes governing blast resistance in rice and several of them have been successfully utilized in breeding programmes (Ning et al., 2020). However, the success is limited by frequent breakdown of R gene mediated resistance (Mentlak et al., 2012) and few genes for panicle blast resistance. As the panicle blast is more severe in terms of yield losses, genes providing broad spectrum durable resistance against panicle blast are highly sought after (Du et al., 2021). Comparative transcriptome analysis of panicle tissues of blast resistant cultivar Tetep and a susceptible cultivar HP2216 at different time points of *M. oryzae* infection, revealed hundreds of differentially expressed genes including two von Willebrand factor domain A containing genes (Kumar et al., 2021). A thorough analysis of the vWA family genes by Karkute at el. (2022) identified *OsvWA36* and *OsvWA37* as the key genes for blast disease resistance among the 40 individual members of the family. *OsvWA36* gene has also been reported to be associated with gall midge resistance in rice (Rawat et al., 2012). On the other hand, *OsvWA37* gene was reported to be associated with resistance to the parasitic weed Striga *hermonthica* as it was induced at different stages of weed infection in rice (Swarbrick et al., 2008). Recently, both *OsvWA36* and *OsvWA37* genes were found to be present in a panicle blast resistance Pb-bd1 locus (Fang et al., 2019). Thus, all these reports suggest the significant role of *OsvWA36* and OsvWA37 genes in rice blast disease response.

Considering the significant role of these genes in blast resistance, it is of utmost importance to analyse the molecular diversity of these genes in different species of rice. Therefore, the present study was carried out to analyse the allelic diversity of these genes in all the 11 species of rice, whose genome sequence data is available on Gramene database (Tello-Ruiz et al., 2021). The function of any protein depends on its 3 dimensional structure and therefore, we have also modelled the 3D structures of these proteins in cultivated germplasm cv. Tetep which is highly resistant to both leaf and panicle blast (Wang et al., 2019).

2. MATERIALS AND METHODS

A study was conducted at ICAR-National Institute for Plant Biotechnology, New Delhi Location/ Lab during 2021.

2.1. Identification of OsvWA36 and OsvWA37 genes from Tetep genome

Complete genome sequence data available at ICAR-National Institute for Plant Biotechnology was utilized to retrieve the sequence of *OsvWA36* and *OsvWA37* genes using BLAST tool in BioEdit. Sequence of these genes in Nipponbare was used as a query for BLAST search. The retrieved sequences were subjected to FGENESH tool (Solovyev et al., 2006) to identify the open reading frame (ORF) and corresponding protein sequence.

2.2. Retrieval of protein and promoter sequences of vWA genes in different rice species

The protein sequences of *OsvWA36* and *OsvWA37* genes in different rice species were retrieved from the Gramene database (Tello-Ruiz et al., 2021) by using the BLAST search tool. The sequences of *OsvWA36* and *OsvWA37* genes in Tetep cultivar were used as a query and searched with default parameters against different species of rice such as *Oryza punctata, Oryza nivara, Oryza sativa Japonica, Oryza barthii, Oryza meridionalis, Oryza rufipogon, Oryza sativa Indica, Oryza glaberrima, Oryza brachyantha, Oryza glumaepatula,* and *Oryza longistaminata.*

2.3. Evolutionary relationship analysis of vWA genes in rice species

All the *OsvWA36* protein sequences of rice species were aligned by using multiple alignment tool MUSCLE (Edgar, 2004) in MEGA-X (Kumar et al., 2018). The phylogenetic tree was constructed based on this alignment by using the maximum likelihood method and JTT matrix-based model with 1000 bootstrap replications. Likewise, phylogenetic

tree for OsvWA37 proteins were generated from all 11 *oryza* species.

2.4. *Protein modelling, structure refinement and validation OsvWA36 and OsvWA37 proteins in Oryza sativa cv Tetep*

The Three-Dimensional (3D) structures of OsvWA36 and OsvWA37 proteins were generated by utilizing the I-TASSER (Yang et al., 2015) web tool. This tool determines the C-score, TM-score value and root mean square deviation (RMSD) and provides with the best five predicted structure models of the given protein sequence (Roy et al., 2010). The modelled structure was chosen based on its C-score that range between -5 to 2. The model having higher C-score represents a better model. Additionally, the model structures were refined using the GalaxyRefine webserver (Heo et al., 2013). The structure refinement is the quality improvement process that provides a robust model by using the CASP10 based refinement method (Ko et al., 2012 Nugent et al., 2014). This webserver reconstructs side chains, re-builds unreliable loops, and then repacks them followed by structure relaxation using dynamic simulations (Heo et al., 2013). The refined structures were visualized using BIOVIA Discovery Studio 2020. Further, the Ramachandran plot was created for the refined protein structures by using RAMPAGE server to analyse the overall quality of the structure (Lovell et al., 2003).

3. RESULTS AND DISCUSSION

3.1. *Molecular diversity of OsvWA36 and OsvWA37 genes*

The ORFs of *OsvWA36* and *OsvWA37* genes in Tetep cultivar encoded proteins of 633 and 598 amino acid residues respectively. Both these genes contain a single vWA domain which is important for their functioning through interaction with other proteins (Karkute et al., 2022). The protein sequence alignment revealed a significant diversity in *OsvWA36* and *OsvWA37* genes among the different species of rice (Table 1). Besides sequence variation, the number of amino acid residues also varied from 501 residues in *O. longistaminata*to 633 in *O. sativa* japonica group and Tetep cultivar in case of OsvWA36 protein. Even more diversity was observed in case of OsvWA37 protein. The protein encoded by *OsvWA37* gene in *O. longistaminata*is only of 295 amino acid residues whereas it is of 1004 amino acid residues in *O. meridionalis*. The diversity in the protein sequences of these genes observed in different species reflects their genetic diversity. However, the vWA domain is highly conserved in vWA36 protein (Figure 1a) whereas it is comparatively less conserved in case of vWA37 protein (Figure 1b) among different species of rice. It is important to note that, there is also a significant variation in vWA domain sequence between OsvWA36 and OsvWA37 proteins. One of the interesting observations is that both these genes were

Table 1: Protein level variation of *OsvWA36* and *OsvWA37* gene in different rice species

Sl. No.	Rice species	No. of amino acid residues in *OsvWA36* protein	No. of amino acid residues in *OsvWA37* protein
1.	*Oryza sativa cv Tetep*	633	598
2.	*Oryza punctata*	601	598
3.	*Oryza nivara*	631	561
4.	*Oryza sativa Japonica* group	633	598
5.	*Oryza barthii*	1488	1488
6.	*Oryza meridionalis*	624	1004
7.	*Oryza rufipogon*	602	489
8.	*Oryza sativa Indica* group	631	599
9.	*Oryza glaberrima*	632	593
10.	*Oryza brachyantha*	599	559
11.	*Oryza glumaepatula*	620	620
12.	*Oryza longistaminata*	501	295

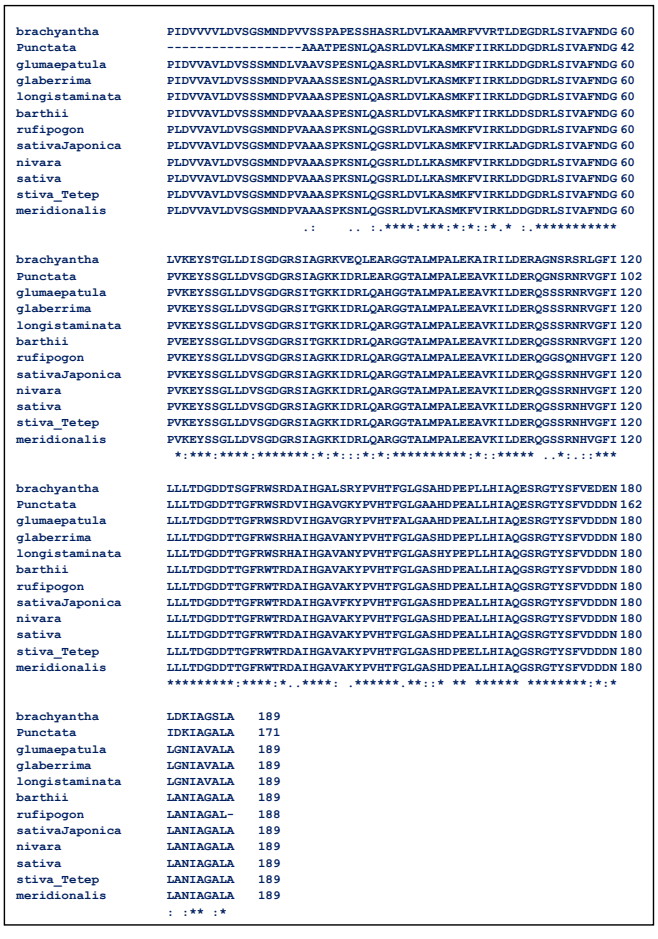

Figure 1: Continue...

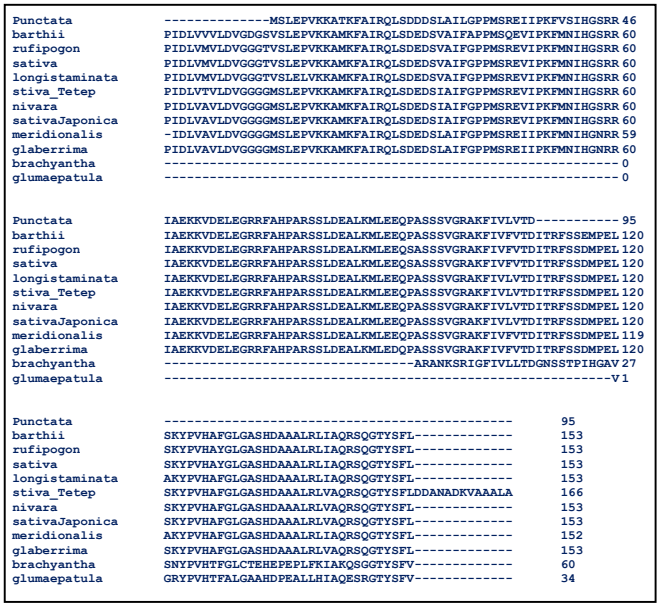

Figure 1: Alignment of vWA domain sequence of *OsvWA36* (1a) and *OsvWA37* (1b) proteins in different species of rice

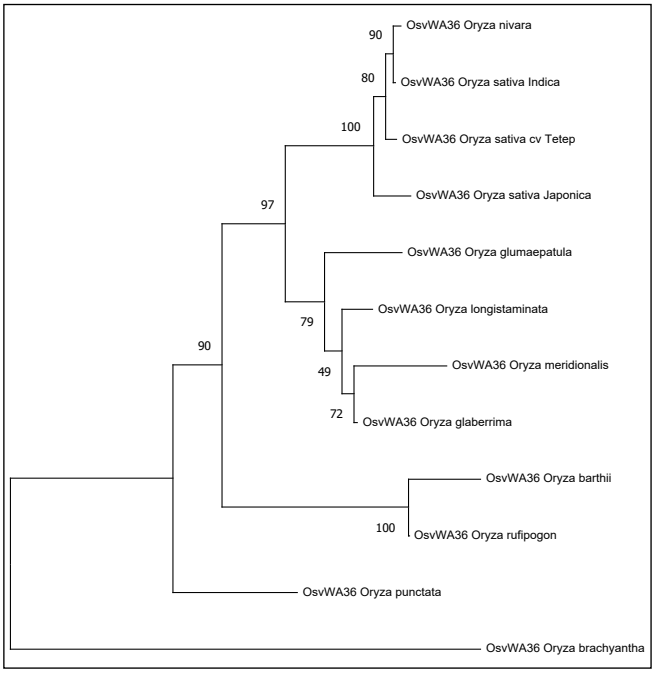

Figure 2: Phylogenetic tree of vWA36 protein in different rice species. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site

not identified as separate genes in *O. barthii*, rather they are combined to form a single gene coding for 1488 amino acid residue protein. Normally, these two genes are located very closely to each other on chromosome 11 in *O. sativa*. This suggests that *OsvWA36* and *OsvWA37* genes might have been originated from a single large gene and got separated during the evolution. This also indicates that the *O. barthii,* a wild species of rice could be one of the oldest species and other species might have originated from it. *O. barthii* is reported to be the progenitor of African rice *O. glaberrima* (Linares, 2002).

3.2. Evolutionary relationship of OsvWA36 gene in different rice species

Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. Thus, the phylogenetic tree was generated showing relationship of *OsvWA36* protein in different wild species of rice along with protein from Tetep cultivar (Figure 2). *O. nivara*, *O. sativa indica, O. sativa japonica* and *O. sativa* cv. Tetep were closely related to each other and thus, formed a single group (group I). The *O. nivara* is the wild progenitor of *O. sativa* present in Asia region (Haritha et al., 2018) and this has been reflected in the phylogenetic tree for *OsvWA36* protein also. Another group (group II) was formed by four species *O. glumaepatula*, *O. meridionalis, O. glaberrima,* and *O. longistaminata.* This second subclade surprisingly consists of species from different geographical regions such as *O. glaberrima,* and *O. longistaminata* from Africa, *O. glumaepatula* from America and *O. meridionalis* from Australia (Mondal and Henry, 2018). These two groups formed a single clade including 8 species. OsvWA36 protein sequences in *O. rufipogan* and *O. barthii* were similar to each other but different from other species and were placed separately in phylogeny. vWA36 in *O. brachyantha* is extremely different from other species.

3.3. Evolutionary relationship of OsvWA37 gene in different rice species

The evolutionary history of *OsVWA37* gene was studied by generating the phylogenetic tree in the similar way as in case of *OsvWA36* gene. Contrary to *OsvWA36* gene, the evolutionary history of *OsvWA37* formed two different clades (Figure 3). The clade I has 8 species including the cultivated species *O. sativa* and *O. glaberrima. OsvWA37* protein in *O. nivara*, *O. sativa* cv. Tetep and *O. sativa* japonica group is completely conserved forming a single group with two more species *O. meridionalis* and *O. glaberrima*. The group 2 of clade I was formed by 3 species such as *O. longistaminata*, *O rufipogan* and surprisingly *O. sativa*indica group. On the other hand, clade II included most diverse species *O. punctata, O. brachyantha, O. barthii*, and *O. glumaepatula.*

3.4. Tertiary structure prediction, refinement and validation analysis

The 3D structure of the proteins was modelled with the

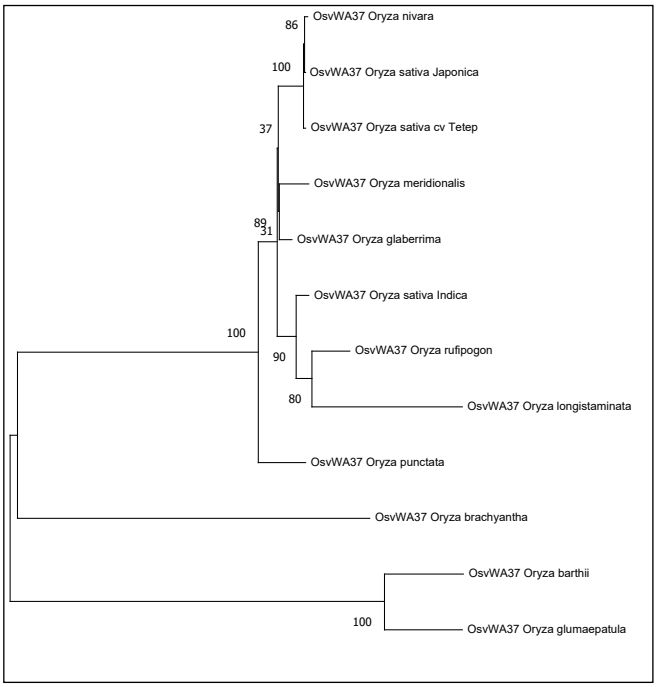

Figure 3: Phylogenetic tree of vWA37 protein in different rice species. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site

help of I-TASSER server. The server carries out 3D modelling based on the consequence of threading template alignment and ranks the confidence of models quantitatively on C-score. The obtained C-scores were -2.24, and -1.75, for the modelled 3D structure of *OsvWA36*, and *OsvWA37* respectively (Figure 4). Further refinement of the models by GalaxyRefine significantly improved the quality parameters (Table 2). Finally, the best refined and polished models were subjected to generate the Ramachandran plot, which indicated the improved percentage of residues in the favored region (Figure 5). The refined models showed Ramachandran plot score of 92.407%, and 92.564%, for the modelled structure of *OsvWA36*, and *OsvWA37* respectively. The vWA domain of the vWA proteins in involved in protein-protein interaction to carry out their cellular function (Yang et al., 2016). The high-quality

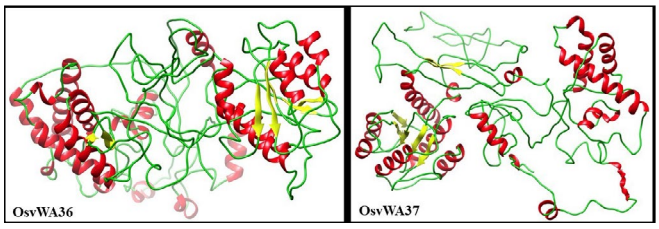

Figure 4: Three dimensional models of the proteins showing prominent secondary arrangements (Red=Helix, Yellow=Sheet, Green=Loop)

Table 2: Comparison of the quality parameters of the 3D model before and after refinement

	OsvWA36		*OsvWA37*	
	Initial model	Refined model	Initial model	Refined model
GDT-HA	1.0000	0.9072	1.0000	0.8980
RMSD	0.000	0.524	0.000	0.585
MolProbity	3.532	2.528	3.166	2.529
Clash score	17.6	23.2	7.3	20.0
Poor rotamers	15.8	1.0	14.4	1.3
Rama favored	65.8	84.3	64.8	85.1

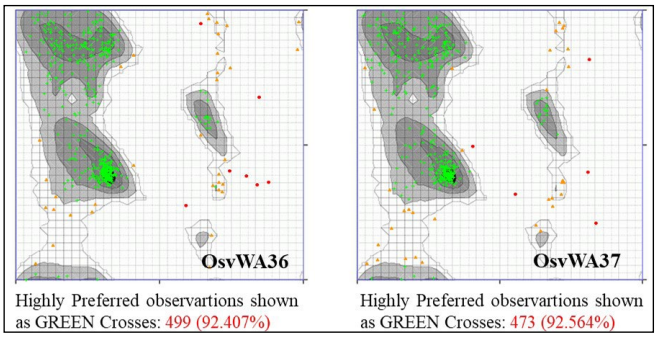

Figure 5: Ramachandran plot statistics showing the % of highly preferred (green marks), preferred (orange marks), and questionable conformations (red marks). The black, dark grey, grey, and light grey represents the highly preferred conformations (Delta≥-2), white with black grid represents preferred conformation (-2>Delta≥-4), and white with grey grid represents questionable conformations (Delta<-4)

structures of the proteins generated in the study will help to carry out the structural characterization and also can be confidently used for the interaction studies. This will be useful to understand and identify the genes interacting with these two key OsvWA proteins and subsequently use them in rice breeding for blast disease resistance.

4. CONCLUSION

The blast disease responsive *OsvWA36* and *OsvWA37*genes have conserved vWA domain among different cultivated and wild species of rice. The genes vary in size of the encoded protein in different rice species. *OsvWA36* gene is comparatively more conserved than *OsvWA37* gene. The vWA domain is involved in protein interactions and therefore to study the interaction of these proteins with other disease responsive proteins, high quality 3 dimensional structures have been developed that can be confidently used for docking studies.

5. FURTHER RESEARCH

The role of these novel *OsvWA36* and *OsvWA37* genes need to be characterized in detail to identify the mechanism of resistance and pathways in which the genes are involved in.

6. ACKNOWLEDGEMENT

The work was supported by ICAR-National Institute for Plant Biotechnology; Indian Council of Agricultural Research (ICAR)-Centre for Agricultural Bioinformatics, IASRI (CABin Scheme).

7. REFERENCES

Anami, B.S., Malvade, N.N., Palaiah, S., 2020. Classification of yield affecting biotic and abiotic paddy crop stresses using field images. Information Processing in Agriculture 7(2), 272–285.

Asibi, A.E., Chai, Q., Coulter, J.A., 2019. Rice blast: A disease with implications for global food security. Agronomy 9(8), 451.

Du, Y., Qi, Z., Yu, J., Yu, M., Cao, H., Zhang, R., Liu, Y., 2021. Effects of panicle development stage and temperature on rice panicle blast infection by *Magnaporthe oryzae* and visualization of its infection process. Plant Pathology 70(6), 1436–1444.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5), 1792–1797.

Fang, N., Wei, X., Shen, L., Yu, Y., Li, M., Yin, C., 2019. Fine mapping of a panicle blast resistance gene Pb-bd1 in Japonica landrace Bodao and its application in rice breeding. Rice 12(1), 1–12.

Haritha, G., Malathi, S., Divya, B., Swamy, B.P.M., Mangrauthia, S.K., Sarla, N., 2018. *Oryza nivara* Sharma et Shastry. In The Wild *Oryza* Genomes. Springer, Cham 207–238

Heo, L., Park, H., Seok, C., 2013. Galaxy refine: protein structure refinement driven by side-chain repacking. Nucleic Acids Research 41(Web Server issue), W384–388.

Karkute, S.G., Kumar, V., Tasleem, M., Mishra, D.C., Chaturvedi, K.K., Rai, A., Sevanthi, A.M., Gaikwad, K., Sharma, T.R., Solanke, A.U., 2022. Genome-Wide Analysis of von Willebrand Factor A (vWA) gene family in rice for its role in imparting biotic stress resistance with emphasis on rice blast disease. Rice Science 29(4), 375–384.

Khan, M.A.I., Bhuiyan, M.R., Hossain, M.S., Sen. P.P., Ara, A., Siddique, M.A. Ali, M.A., 2014. Neck blast disease influences grain yield and quality traits of aromatic rice. Comptes Rendus Biologies 337(11), 635–641.

Ko, J., Park, H., Heo, L., Seok, C., 2012. Galaxy WEB server for protein structure prediction and refinement. Nucleic Acids Research 40 (Web Server issue), W294–297. https://doi.org/10.1093/nar/gks493

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35(6), 1547–1549.

Kumar, V., Jain, P., Venkadesan, S., Karkute, S.G., Bhati, J., Abdin, M.Z., 2021. Understanding rice-*Magnaporthe Oryzae* interaction in resistant and susceptible cultivars of rice under panicle blast infection using a time-course transcriptome analysis. Genes 12(2), 301.

Linares, O.F., 2002. African rice (*Oryza glaberrima*): history and future potential. Proceedings of National Academy of Sciences 99(25), 16360–16365.

Lovell, S.C., Davis, I.W., Arendall, W.B., Bakker, P.I.W.D., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins 50(3), 437–450. https://doi.org/10.1002/prot.10286.

Mentlak, T.A., Kombrink, A., Shinya, T., Ryder, L.S., Otomo, I., Saitoh, H., Talbot, N.J., 2012. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. The Plant Cell 24(1), 322–335.

Mondal, T.K., Henry, R.J., 2018. The wild *Oryza* genomes. Springer. ISBN: 978-3-319-71997-9

Nawaz, A., Rehman, A.U., Rehman, A., Ahmad, S., Siddique, K.M., Farooq, M., 2022. Increasing sustainability for rice production systems. Journal of Cereal Science 103, 103400.

Ning, X.I.A.O., Yunyu, W., Aihong, L., 2020. Strategy for use of rice blast resistance genes in rice molecular breeding. Rice Science 27(4), 263–277.

Nugent, T., Cozzetto, D., Jones, D.T., 2014. Evaluation of predictions in the CASP10 model refinement category. Proteins 82(2), 98–111. https://doi.org/10.1002/prot.24377

Rawat, N., Naga, N.C., Meenakshi, S.R., Nair, S., Bentur, J.S., 2012. A novel mechanism of gall midge resistance in the rice variety Kavya revealed by microarray analysis. Functional and Integrative Genomics 12(2), 249–264.

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: A unified platform for automated protein structure and function prediction. Nature Protocols 5(4), 725–738.

Samal, P., Babu, S., 2018. The shape of rice agriculture towards 2050. Conference, July 28-August 2, 2018, Vancouver, British Columbia 277550, International Association of Agricultural Economists.

Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D., 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biology 7(1), 1–12.

Swarbrick, P.J., Huang, K., Liu, G., Slate, J., Press, M.C., Scholes, J.D., 2008. Global patterns of gene expression in rice cultivars undergoing a susceptible or resistant interaction with the parasitic plant Striga hermonthica. New Phytologist 179(2), 515–529.

Tello-Ruiz, M.K., Naithani, S., Gupta, P., Olson, A., Wei, S., Preece, J., 2021. Gramene 2021: harnessing the power of comparative genomics and pathways for plant research. Nucleic Acids Research 49(D1), D1452–D1463.

Yang, H., Li, Y., Hua, J., 2006. The C2 domain protein BAP1 negatively regulates defense responses in Arabidopsis. Plant Journal 48(2), 238–248.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nature Methods 12(1), 7–8.