In silico studies of stress tolerance genes in non-desert and desert plants

A Thesis submitted by
Vedikaa Dhiman

in partial fulfilment of the requirements for the award of the degree of
Doctor of Philosophy

Indian Institute of Technology Jodhpur
Bioscience and Bioengineering
December 2024

6
Conclusions and Future work

This chapter provides a comprehensive summary of the key achievements outlined in the various chapters of this thesis. In addition, it emphasizes future research directions along with potential extensions and applications of the research work in the area of molecular breeding and biotechnology for the improvement of stress tolerance in *O. sativa*.

6.1 Conclusions

Over the past 50 years, there has been a significant rise in the world population. It is projected to increase by almost 9 billion by the middle of this century. This growing population will require more food with available agricultural resources, with an estimated increase of 70% by 2050. However, there is a significant deficit between the current level of food production and the projected requirement to adequately nourish the global population. In this regard, food crops such as *O. sativa* plays a vital role in food production and security, which is one of the major SDGs. It is a staple food crop consumed by half of the world's population. However, its growth, yield, and overall health are significantly affected by various biotic and abiotic stresses. Biotic stress, such as bacteria, fungi, etc., and abiotic stress, including drought, salinity, and much more, majorly affect *O. sativa* yield. However, it consists of a well-developed immune system governed by stress-tolerant genes or proteins to fight against various stresses. In addition, there are several approaches through which crop productivity can be improved and maintained. These approaches are the application of pesticides in the field, conventional breeding methods to develop novel cultivars, hybridization methods, etc. However, these methods give rise to a few serious issues that need to be addressed to maintain global food production. Pesticide usage leads to poor human health and environmental pollution. Further, few traditional breeding practices are time-consuming, labour-intensive, and costly. Thus, the development of novel stress-tolerant varieties of *O. sativa* through various modern methods is a preferable strategy to achieve global food demand and security for the future. Modern breeding approaches based on multi-omics and computational methods consist of a huge scope in developing novel stress-tolerant varieties. These methods are fast, accurate, precise, less time-consuming, and cost-efficient. This will contribute to precise breeding programs in a short duration of time, resulting in highly stress-tolerant *O. sativa* varieties. Therefore, in this thesis, various *in silico* approaches are utilized to identify and study stress-tolerant genes or proteins in *O. sativa*. This will help in gaining more production and hence, in achieving global food production and security.

Chapter 3 includes the identification of disease-resistance proteins against BB and RB diseases through computational approaches in *O. sativa Japonica*. This work successfully identified five novel disease-resistance proteins using gene network analysis, linking known disease-resistance genes with hypothetically conserved genes. Further, to get more knowledge on these novel proteins, structural and functional enrichment studies were performed. The structural studies included modeling, refinement, and validation of the best 3D structure for the identified five proteins. The functional annotation revealed the involvement of the five proteins in various disease-resistance mechanisms. In addition, phylogenetic analysis depicted the evolutionary relationship between the identified proteins with known disease-resistance proteins. To gain more insights into the expression of the five proteins at the gene level, their gene expression profiling has been performed. It showed differential expression under *Xoo* and *M. oryzae* infection. This study underscores the potential of *in silico* methods for identifying novel disease-resistance proteins, contributing to the development of disease-resistant *O. sativa* varieties.

Further, ML and DL models could provide further insights into novel disease-resistance proteins in *O. sativa* and its evolutionary-related species. The integration of muti-omics datasets, such as proteomics, with ML and DL models provides faster, more accurate and precise information on disease-resistance proteins. This integrated approach will serve as a valuable resource for breeders to develop innovative disease-resistant *O. sativa* varieties. In Chapter 4, a DL-based framework based on protein sequences from *O. sativa* and its related species was developed and analyzed. This study considered various well-known ML and DL models that were trained, tested, validated, and compared to predict disease-resistance proteins in *O. sativa* and its related species. Among the models evaluated, the DL-based MLP model demonstrated the best performance, achieving optimal results across key metrics such as accuracy, recall, and precision. The MLP model was further tested and validated on internal and external datasets. The internal dataset comprised known disease-resistance proteins and the whole proteome of *O. sativa* and its related species. The external dataset consists of the proteins involved in disease resistance, morphological development, and flowering conditions. The MLP model performed accurately and precisely on both the datasets, predicting the disease-resistance proteins efficiently. ML and DL models hold great promise to translate biological knowledge on disease resistance in *O. sativa* and related species for precise breeding. These approaches are faster, accurate, and less time-consuming in identifying the disease-resistance proteins transforming traditional farming practices into a new age of smart farming.

Additionally, while biotic stresses poses significant challenges, abiotic stresses such as drought and heat also severely affect *O. sativa* yield. Furthermore, the interplay between abiotic and biotic stresses is exacerbated by climate change, as disease-causing pathogens are adapting and becoming more resistant. These developments underscore the critical need for integrated strategies to enhance both biotic and abiotic stress tolerance, ensuring sustainable productivity in *O. sativa*.

Chapter 5 describes drought tolerance transcriptome profiling and genome-wide investigation of APR/ERF superfamily genes from the desert tree *P. cineraria*. This plant is highly tolerant to abiotic stress including drought and a potential source of robust orthologs of tolerance genes that can be incorporated in *O. sativa* to improve its drought tolerance. Gene network analysis revealed the upregulation of various DT genes in *P. cineraria* under drought conditions. A comparative genome-wide investigation, including phylogenetic analysis, gene and protein structure study, and promoter and *cis*-acting element assessment, between DT and DS species was performed. The amino acid variation positions in the DBD of AP2/ERF proteins demonstrated that *P. cineraria* has more polarity as compared to *O. sativa Japonica* and *Indica*. This results in stronger binding between AP2/ERF and GCC box/DRE. A higher number of DBD-DNA hydrogen bonds were found in *P. cineraria*, resulting in strong DNA-binding affinity, possibly contributing to its higher drought tolerance than many DS species. Moreover, *P. cineraria* showed high copy numbers of *AP2/ERF* genes, further indicating their involvement in drought tolerance. In addition, variants such as SNPs, MNPs, etc., were identified between the Indian and Arabian cultivars of *P. cineraria*, indicating the importance of AP2/ERF factors in the plants' adaptation to varied conditions of drought stress. Therefore, the study indicated that AP2/ERF proteins of *P. cineraria* might have a stronger interaction with the GCC box/DRE than *O. sativa*, leading to a higher transcription of downstream target genes conferring drought tolerance.

6.2 Scope of Future Work

The research presented in this thesis opens several promising avenues for further exploration and development in breeding programs aimed at enhancing stress-tolerant traits in *O. sativa*. Some key areas for future work include the following:

1. Long-term disease resistance is an essential trait of breeding programs in *O. sativa*. However, achieving durable resistance remains challenging due to the adaptive nature of pathogens like *Xoo* and *M. oryzae*. These pathogens evade the plant's immune system by mutating their effector genes, which are the targets of disease-resistance genes or proteins in *O. sativa*. To address this, Chapter 3 utilizes *in*

silico approaches to identify and study novel disease-resistance proteins working against *Xoo* and *M. oryzae*.

Future research could focus on identifying and analyzing quantitative trait loci (QTLs) associated with the proteins identified in this study. These QTLs might have several disease-resistance genes that can be used to develop novel cultivars of *O. sativa*. Furthermore, extensive investigations can be performed on the identified novel disease-resistance proteins: 1) phylogenetic classification of the identified proteins, 2) promoter and cis-acting element analysis, 3) real-time PCR analysis of these proteins under BB and RB infections, 4) gene structure analysis, 5) CNV analysis and, 6) synteny assessment to explore evolutionary relationships between the identified disease-resistance proteins in *Japonica* species and related *O. sativa* species. In addition, the proteins similar to our identified novel disease-resistance proteins, and whose structural and functional information is less explored, can be studied. These proteins can be investigated through various physiochemical and stereochemical characteristics. Such proteins may play a role in the resistance mechanism against BB and RB diseases, as several proteins in *O. sativa* exhibits similar features to our identified disease-resistance proteins.

2. Several ML and DL-based models have been reported using image datasets to identify diseases in *O. sativa*. However, there is a noticeable scarcity of models leveraging multi-omics datasets. Multi-omics approaches have the potential to provide better insights for breeders to develop novel disease-resistant varieties of *O. sativa*. Thus, Chapter 4 presents a DL-based framework developed based on the proteomics dataset that efficiently predicts disease-resistance proteins in *O. sativa* and its related species.

   The DL-based MLP model developed in this study has a scope for further extension by exploring it in other subspecies of *O. sativa* and economically important members of the *Poaceae* family, such as wheat, sugarcane, barley, and maize. The model holds great potential to utilize multi-omics data to enhance precise breeding programs, providing broad-spectrum resistance against various pathogens. Additionally, the disease-resistance proteins identified by the model can be further explored through the following approaches: 1) conducting an in-depth disease-resistance transcriptome analysis of the genes associated with the identified proteins,2) performing a phylogenetic study on the identified proteins, (3) examining the involvement of these proteins in different signaling pathways such as phytohormones and, 4) developing a graphical user interface (GUI)-based, web accessible, open-source database. This database will serve as a bioinformatic resource having disease-resistance gene or protein data of *O. sativa* and its related species. It can be utilized by the scientific community as a reference platform to analyze resistance genes or proteins and supporting breeding experiments.

3. TFs are the key regulators that control abiotic stress signaling by activating or repressing transcriptional activities in DT plants. However, the functions of only a limited number of TFs specifically involved in drought response have been elucidated. Therefore, Chapter 5 addresses this gap by conducting transcriptome-based profiling and genome-wide studies of the AP2/ERF superfamily in *P. cineraria*. The AP2/ERF orthologs of *P. cineraria* will help in developing DT varieties in *O. sativa*.

   The current genome-wide study of *AP2/ERF* genes in *P. cineraria* opens a wide scope for the investigation of these genes in other abiotic stresses, such as salinity, cold, and extreme heat. Further research could be conducted for a better understanding of the evolution of the AP2/ERF superfamily in *P. cineraria* and other tolerant desert plants. This could include: 1) a comparative study on structural and functional aspects of these genes between different natural ecotypes and cultivars of *P. cineraria*, 2) in vitro validation of the stronger DNA binding by the AP2/ERF orthologs of *P. cineraria* compared to *O. sativa* and other DS species by biochemical experiments and ectopic expression of these contrasting orthologs *in planta* followed by expression analysis of the downstream target genes of these TFs under drought stress and, 3) comparative analysis of the DNA-binding affinity of AP2/ERF orthologs of other DT species, particularly *Vigna unguiculata*, and other desert legumes, to identify

evolutionary adaptations enhancing drought tolerance.