

Introduction

Agriculture is a traditional practice involving soil cultivation, crop production, and raising livestock to provide food and resources for humans and animals. The agricultural sector is essential in enhancing food availability and ensuring its security. To meet the rising demand for agricultural productivity, more fertile land is needed, even as climate conditions worsen and pathogenic diseases become more prevalent. Section 1.1 below provides an overview of the role of agriculture in achieving the Sustainable Development Goals (SDGs). Next, section 1.2 explores the overview of food crops i.e., *Oryza sativa* and the impact of various stresses on its yield. It also describes the immune system of *O. sativa* against stresses and the current strategies utilized to target stress tolerance and thus, achieving global food security. The objectives of this thesis work are outlined in section 1.3. Finally, section 1.4 provides an overview of the thesis chapters along with the potential future implications of the research work carried out in this thesis.

1.1 Agriculture and sustainable developmental goals

Over the 1950s, the global population logarithmically increased and will rise by 9.7 billion by 2050 and 10.4 billion in the mid-2080s. [Roth and Galyon, 2024]. With the rising global population, the demand for food and fiber is expected to increase by up to 70% by 2050 [Singh and Trivedi, 2017]. Recent studies suggest that global crop production needs to double by 2050 to meet rising demands [Beltran-Peña *et al.*, 2020].

Access to quality and nutritious food is essential for every individual and directly influences global economic growth. Significant efforts have been made by the scientific community to improve agricultural food production and security for the growing population. In 2015, the United Nations introduced the SDGs as part of its global agenda. This program includes 17 SDGs based on the Millennium Development Goals (MDGs), where Zero Hunger (SDG2) [Viana *et al.*, 2022] was an essential goal. Integrating sustainable agricultural practices is vital for meeting SDG2. By focusing on methods that enhance productivity while protecting the environment, agriculture can drive progress across a broad spectrum of goals globally.

Despite extensive efforts in recent decades to develop strategies for global food security, around one in ten people still faces severe food insecurity (reference needed). Moreover, the growing population, climate change, disease increment and changing diet structure contribute to increasing food insecurity. Thus, achieving food and nutrition security goals demands urgent collective international efforts and effective global food security measures.

1.2 Role of *Oryza sativa* in achieving sustainable developmental goals

Food security can be best achieved by maximizing the productivity of the food or staple crops. The food crops are produced and consumed domestically to feed people daily. Several recent studies report that food crops are critical in achieving global food security, especially in developing countries where small-scale farming practices are a major food source to feed the population of that nation. There are more than 50,000 food crops in the world, however, only a few of them contribute to global food production. Of these, *O. sativa*, wheat and maize are the three major food crops responsible for food production worldwide. *O. sativa* is potentially one of the most essential crops globally, intricately linked to food security, economic

development, employment, cultural identity, and regional stability. Moreover, as a fundamental food source for over 3 billion individuals worldwide, *O. sativa* is a crucial crop for numerous nations, fulfilling food requirements and boosting their economies through exports to other nations. Asia contributes nearly 90% of the world's *O. sativa* production, whereas, outside Asia, *O. sativa* production is less than 9% [Chukwu *et al.*, 2019]. *O. sativa* accounts for approximately 23.3% of the total cultivated land in India and is crucial for the country's food grain production. It has been estimated that by 2050, global demand for *O. sativa* is expected to increase by more than 40% to feed the rapidly growing world population. Although there has been remarkable global growth in *O. sativa* production from 289 million tons in 1968 to 782 million tons in 2018, this substantial increase has not been sufficient to meet the growing demand due to the increasing population at a much faster rate [Jamaloddin *et al.*, 2021]. Moreover, the sustainable production of *O. sativa* has been challenging due to the yield losses caused by various stresses.

Stress in plants can be defined as any external or internal conditions that restrict photosynthesis and reduce the plant's ability to convert energy into biomass [Umar *et al.*, 2021]. The biotic stress could occur due to the presence of living organisms, including insects, pests, viruses, bacteria, fungi, nematodes, arachnids, weeds and competition with other plants. The biotic stress can lead to pathogenic diseases, which is a major limiting factor affecting *O. sativa* production, causing the annual yield loss estimated at 5% [Song and Goodman, 2001]. These diseases infect different parts of *O. sativa*, lowering its productivity. The *Xanthomonas oryzae* (referred to as *Xoo*) and *Magnaporthe oryzae* (known as *M. oryzae*) are the main pathogens infecting *O. sativa*. The bacterial blight (BB) and rice blast (RB) diseases are the pathogenic diseases caused by the *Xoo* and *M. oryzae*, respectively. In addition to biotic stresses, *O. sativa* productivity is also severely affected by various abiotic stresses. These include non-living factors such as cold, heat, drought, flood, salinity, temperature and heavy metals. These stresses negatively affect the overall growth and development of *O. sativa*. Of these abiotic stresses, the drought is the most severe type of abiotic stress which could reduce *O. sativa* productivity by up to 50% annually [Hassan *et al.*, 2023]. The lesser availability of water than optimally required for growth can decrease one-third yield of *O. sativa* cultivars for drought. Furthermore, *O. sativa* shows poor root growth and nutrient uptake in the presence of drought stress [Hassan *et al.*, 2023].

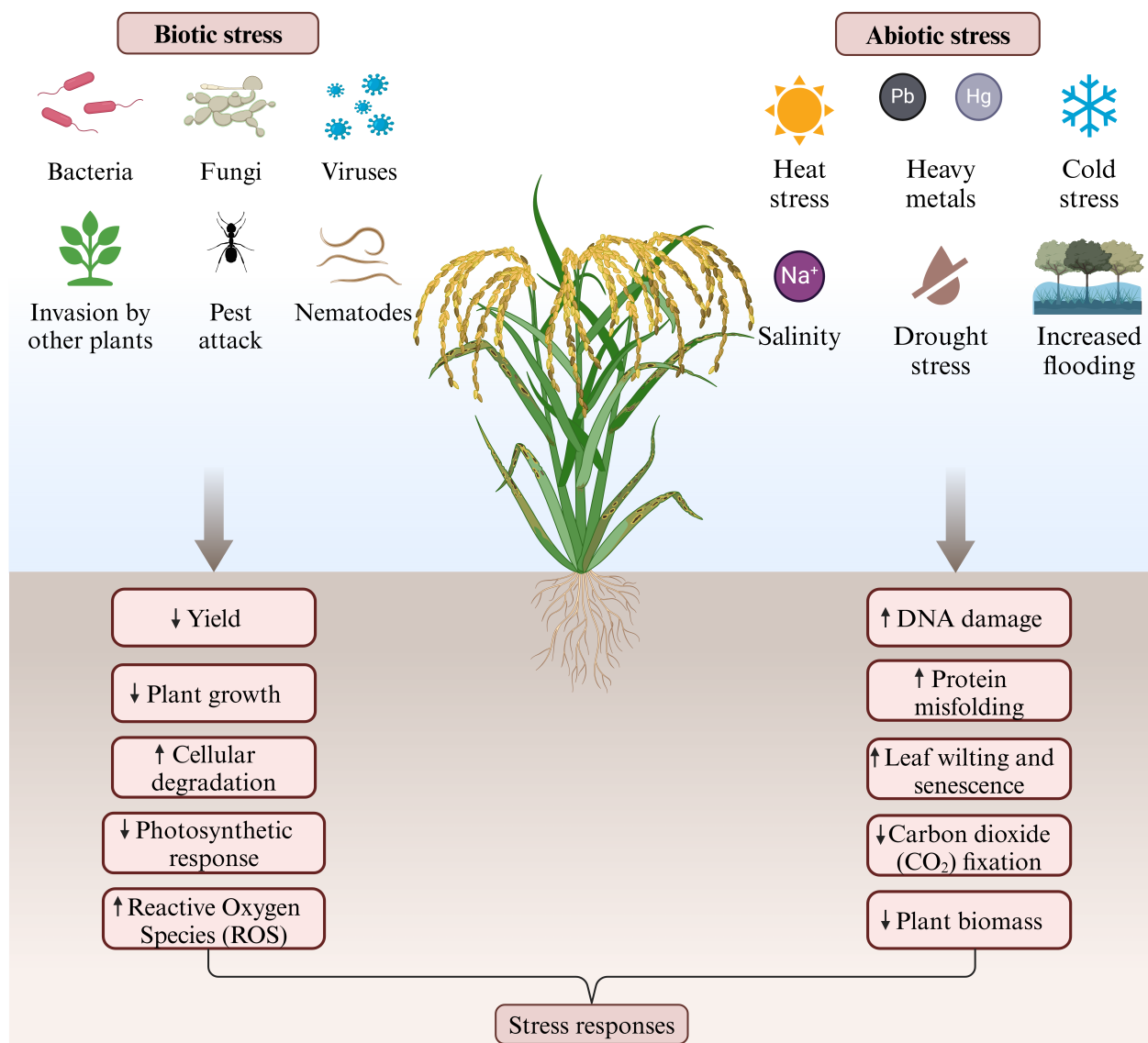


Figure 1.1 : A schematic diagram showing different types of stress and their impact on the overall health of *Oryza sativa*. The right-hand side panel depicts biotic stresses such as bacteria, fungi, viruses, etc. They affect *O. sativa* by declining its yield, growth, and photosynthetic response, and increasing cellular degradation and reactive oxygen species (ROS). In addition, the right-hand side panel highlights abiotic stresses such as heat, cold, and heavy metals including lead (Pb) and mercury (Hg), salinity by sodium ions (Na⁺) and much more. The abiotic stress negatively impacts *O. sativa* by enhancing DNA damage, protein misfolding and declining CO₂ and plant biomass.

1.3 Functioning of immune system of *Oryza sativa* against biotic and abiotic stresses

O. sativa productivity is extremely threatened by both biotic and abiotic stress conditions. To combat the impact of these stresses, farmers generally utilize various pesticides in the field which have a negative impact on the environment and human health. In addition, various breeding approaches are utilized to reduce the negative impacts of abiotic stresses. However, *O. sativa* has developed a variety of strategies to combat different stresses, including morphological, molecular, cellular, biochemical, and physiological adaptations.

O. sativa consists of a highly functional immune system providing resistance against various stresses. The immunity or resistance against stresses depends on the immune receptors expressed

in every cell. These receptors detect invading signals to initiate pattern-triggered immunity (PTI) or effector-triggered immunity (ETI) ([Saijo and Loo, 2020]). The PTI constitutes the first line of defense, encompassing physical, morphological, and structural barriers. These barriers include the cuticles, wax layer, trichomes, cell walls, epidermis layer and thorns. PTI is triggered by the presence of cell surface-localized pattern recognition receptors (PRRs). These receptor proteins recognize the molecular patterns or structures of pathogens and endogenous damage known as microbe-associated molecular patterns (MAMPs) and danger-associated molecular patterns (DAMPs), respectively. PTI is governed by various resistance (R) proteins, which identify specific receptors from a pathogen known as avirulence (Avr) proteins, leading to hypersensitive responses and programmed cell death in infected and surrounding cells [Iqbal *et al.*, 2021]. The R proteins contain essential domains, including nucleotide-binding sites (NBS) and leucine-rich repeats (LRR), which play a key role in recognizing pathogen effectors either directly or indirectly. The PRR signalling occurs through closely related signalling modules, enabling the functional transfer of recognition specificity through the introduction of receptors alone. PRR pathways display considerable variation in their rate-limiting steps, which in turn affects their responsiveness to both biotic and abiotic stresses. Therefore, the simultaneous activation of numerous PRR pathways is expected to enhance the resilience of overall PTI signalling in response to pathogen attacks and environmental stresses.

To evade PTI, pathogens have developed a wide range of effector proteins that can alter host immune responses, increasing their chances of successful infection and colonization. To overcome this, plants have evolved several intracellular receptors known as effectors that directly or indirectly recognize pathogens. Plants have developed ETI, a second layer of immunity, which relies on highly diverse proteins, such as nucleotide-binding domain and leucine-rich repeat-containing proteins (NLRs), to detect and respond to pathogen effectors [Iqbal *et al.*, 2021]. ETI is more effective as compared to PTI against various biotic stresses. In addition, *R* genes govern a higher degree of defense called systemic acquired resistance (SAR) along with ETI. It provides defense by stimulating resistance to the other plant organs through intra-plant communication. Several pathogenesis-related (*PR*) genes are essential components of innate immune response in plants functioning in SAR. Both PTI and ETI trigger a variety of signalling pathways that involve changes in cytosolic calcium (Ca^{2+}) levels, reactive oxygen species (ROS), mitogen activated protein kinases (MAPKs), hydrogen peroxide (H_2O_2) signaling, and the activation of transcription factors (TFs) such as myeloblastosis (MYB) and APETALA2/Ethylene Responsive Factor (AP2/ERF). Moreover, PTI and ETI activate Ca^{2+} concentration for shorter and longer periods, respectively. The Ca^{2+} concentration in cytoplasmic and various organelles, including endoplasmic reticulum (ER), mitochondria, golgi body, and plant vacuole is elevated by both PTI and ETI through activating Ca^{2+} channels. Finally, once the Ca^{2+} ions enter the cell through its channels, they are sensed by various Ca^{2+} -binding proteins that mediate downstream signaling events. Ultimately, phytohormones such as salicylic acid and ethylene are activated through the downstream signaling pathways triggered by PTI and ETI. Thus, both PTI and ETI provide a broad-spectrum resistance against biotic and abiotic stresses. Figure 1.2 depicts various stress tolerance responses by *O. sativa* against biotic and abiotic stresses.

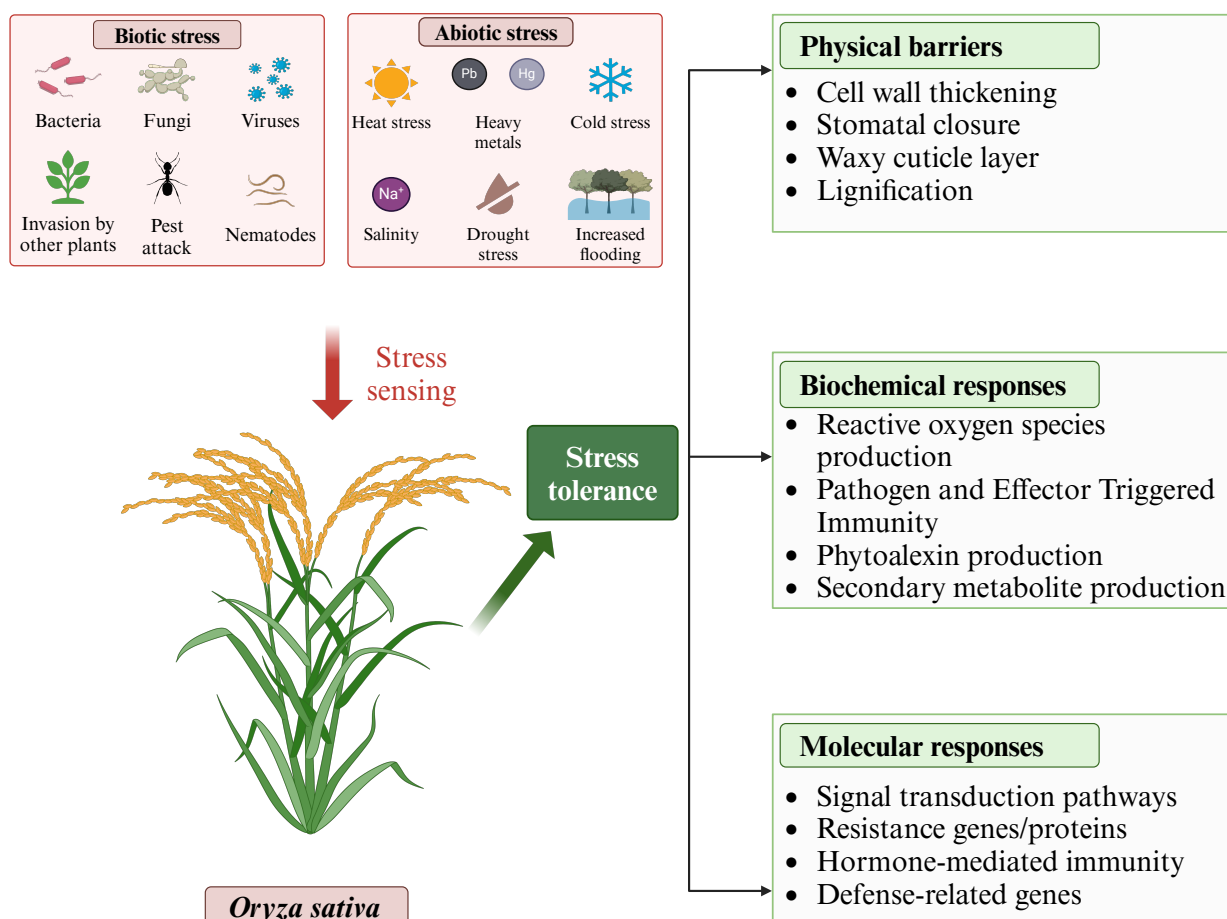


Figure 1.2 : Immune system of *Oryza sativa* consists of various physical barriers and induces biochemical, and molecular responses against biotic and abiotic stresses. The front line of defense occurs in the cell wall, stomata, waxy cuticle layer and lignins. The recognition of pathogens by MAMPs in *O. sativa* triggers immune responses as physical barriers. Further, at biochemical and molecular levels, PTI and ETI are activated which induces ROS, phytoalexin (low molecular weight antimicrobial compounds) and secondary metabolites such as carotenoids, terpenes, etc. In addition, the activation of signal transduction pathways including MAPK, protein kinases, TFs, etc., stress-tolerant and defense-related genes such as *Xa21*, *AP2/ERF* and phytohormones like ethylene, salicylic acid and so on were further triggered by various stresses in *O. sativa*. The immune responses of *O. sativa* not only helps in enhancing tolerance against stresses but also aids in maintaining high productivity.

1.4 Current trends in enhancing stress-tolerance in *Oryza sativa*

Over the past decade, extensive research has focused on identifying genes and proteins that are key in enhancing tolerance against biotic and abiotic stresses. Recently, around 100 *R* genes against *M. oryzae*, the causative agent of RB disease, have been identified, with 26 of these genes successfully cloned. These disease-resistance genes also govern broad-spectrum resistance during RB infection [Liu *et al.*, 2021]. Hence, once an invading pathogen is identified, R proteins, primarily from the NLR family, recognize the secreted pathogen effectors and trigger multiple responses, resulting in resistance. These responses include Ca²⁺ influx, ROS accumulation, MAPK activation, defense hormone signaling activation, and upregulation of defense-related genes. In addition, *O. sativa* has various PRR proteins which recognize PAMPs. These PRRs are the receptor-like kinases (RLKs) having extracellular LRRs and an intracellular kinase domain, and receptor-like proteins (RLPs). Several PRRs, including *Xa21*, *OsFLS2*, *CEBiP*, *OsCERK1*, *LYP4*,

and LYP6, in *O. sativa* have been reported. A study found that the *RLK* gene *Xa21* in *O. sativa* confers a resistance mechanism against *Xoo* [Liu and Wang, 2016].

Furthermore, a recent study identified 5000 upregulated and 6000 downregulated genes in *O. sativa* following exposure to drought stress. Many of these genes are involved in membrane transport, signalling, and transcriptional regulation [Panda *et al.*, 2021]. Panda *et al.*, 2021 reported that Dehydration-Responsive Element-Binding Protein2 (DREB2)-like gene *OsDRAP1*, *OsNAC5*, *O. sativa* late embryogenesis abundant (*OsLEA-1*), *OsZIP71*, *OsWRKY47*, *OsZIP46* and *OsNAC10* provided drought tolerance by utilizing transgenic approaches [Panda *et al.*, 2021]. Another study revealed that the overexpression of *OsWIH2* in *O. sativa* resulted in higher drought tolerance due to a reduction in water loss, accumulation of ROS and cuticular wax biosynthesis under drought conditions [Gu *et al.*, 2021]. Thus, the drought-tolerance genes can be utilized to enhance drought tolerance in *O. sativa*. However, this potential has not been fully realized, as although several stress-tolerant genes in *O. sativa* have been identified through conventional breeding methods, their functional roles remain largely unknown and require further investigation. Thus, the identification and characterization of stress-tolerant genes through computational approaches along with modern breeding programs is a sustainable solution against various stresses. This approach could contribute to the development of new *O. sativa* varieties with enhanced tolerance against both biotic and abiotic stresses. In addition, novel DT species can be further explored to specifically enhance drought resistance in *O. sativa*.

Prosopis cineraria Druce, commonly known as 'Khejri', is the most important leguminous tree species in the Indian Thar desert. It is native to the semiarid and arid regions of India, Pakistan, Iran and Afghanistan. It is a DT plant that can tolerate extreme dry, saline and arid conditions, due to which it can survive in hot and dry climate conditions [Kaur *et al.*, 2018]. It is a leguminous tree with a well-developed tap root system and the ability to fix atmospheric nitrogen and enhance available calcium and phosphorus [Sudalaimuthasari *et al.*, 2022]. This helps in soil fertility improvement, reducing soil erosion, stabilizing sand dunes and thus, improving crop productivity [Rai *et al.*, 2021]. It has developed a series of regulatory mechanisms comprising stress-responsive TFs and hormonal signaling transduction pathways. These components function together to create an interconnected network where TFs regulate the expression of their target genes by either activating or suppressing transcription [Su *et al.*, 2023]. It has been reported that *P. cineraria* contains several stress-responsive TFs, including the MYB and AP2/ERF superfamily, which are induced in response to drought stress. Further, the AP2/ERF superfamily is classified into the AP2, ERF, DREB, RAV (RELATED TO ABI3 AND VP1), and Soloist subfamilies, based on the presence of at least one AP2 domain.

The DREBs are promising TFs that bind to a specific sequence of six nucleotides (A/GCCGAC) known as dehydration-responsive elements (DRE), enhancing stress tolerance [Su *et al.*, 2023]. The DREB subfamily is involved in the regulation of growth, development and stress responses by transcriptional and post-translational modification. Furthermore, the AP2/ERF superfamily is a large and diverse group of TFs that play crucial roles in plant growth and in conferring tolerance to heat, drought, high salinity, and pathogenic diseases [Faraji *et al.*, 2020]. It is classified into AP2, ERF, DREB, RAV and Soloist subfamilies based on sequence characteristics and the number of AP2-conserved domains. The ERF subfamily can be further divided into DREB (A1-A6), ERF (B1-B6) and Soloist subfamilies. The RAV subfamily contains a unique B3-DNA binding domain and an additional AP2 domain, while the soloist subfamily comprises solely the AP2 domain but lacks extra motifs [Zhu *et al.*, 2023]. Furthermore, with the ongoing climate changes, the growth of *O. sativa* is being threatened by heat and drought stresses. In this context, genomic-based technologies offer great scope for enhancing resilience in *O. sativa* through arid ecosystems. The identification of natural genetic resources and characterization of adaptive genetic variation in *P. cineraria* enhance our knowledge of the genes that are linked to drought. Such genes could act as valuable candidates for improving drought tolerance in *O. sativa*, leading to climate-resilient agriculture practices. Therefore, the development of novel stress-tolerant varieties of *O. sativa* helps not only in overcoming biotic and abiotic stresses but also contributes to achieving SDG2, which aims to end hunger and promote sustainable agriculture. Figure 1.3 illustrates the defense responses of *P. cineraria* against various abiotic stresses.

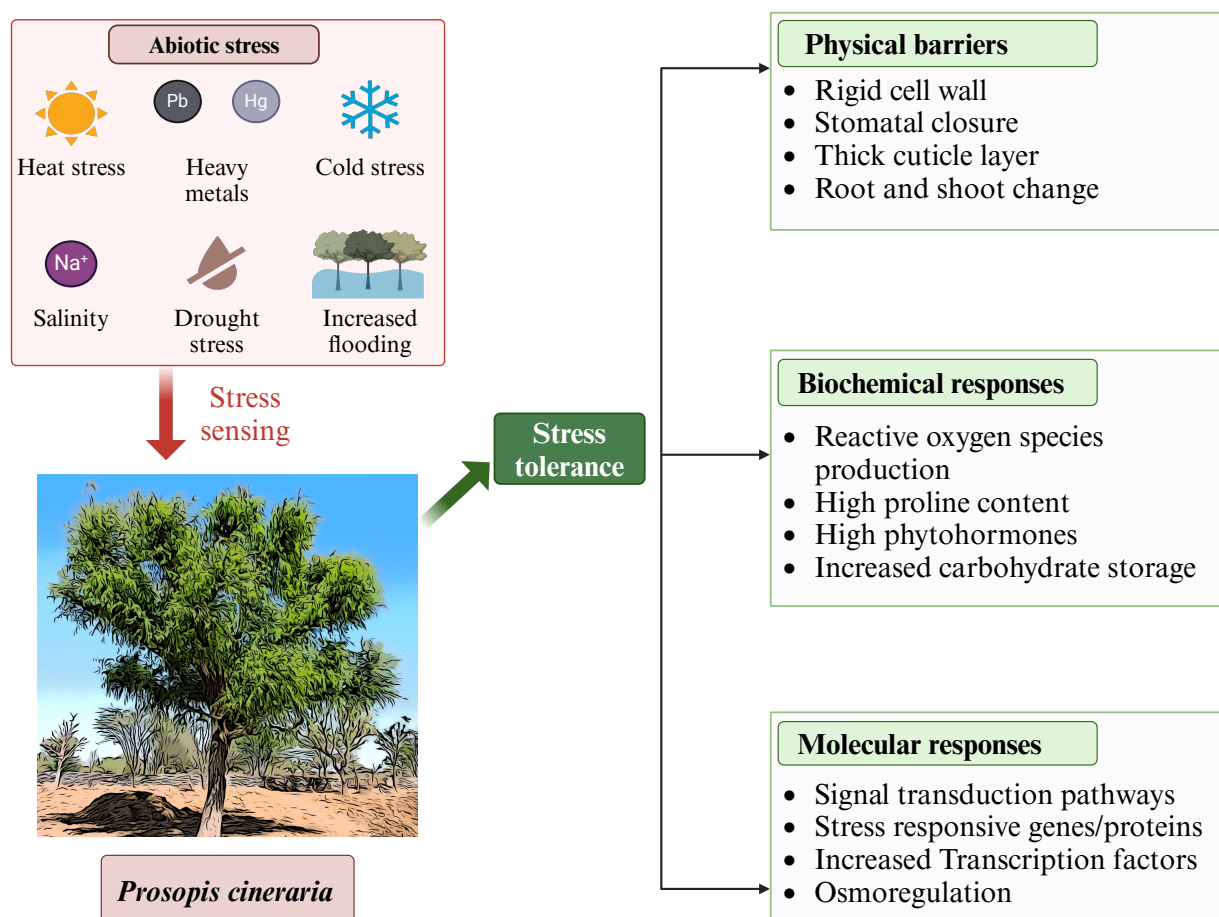


Figure 1.3 : Immune responses of *Prosopis cineraria* against abiotic stresses. *Prosopis cineraria* shows different stress tolerance responses against abiotic stress which include physical barriers, and biochemical and molecular responses. In DT *P. cineraria*, the primary defense is governed by various physical barriers such as cell wall rigidity, stomatal closure, thick cuticle layer and changes in roots and shoots. Moreover, the biochemical responses include the production of reactive oxygen species (ROS), enhancement in proline and phytohormones such as abscisic acid, ethylene and much more. Further, the activation of different signal transduction pathways includes MAPK and calcium signalling, stress-tolerant genes and transcription factors (TFs) such as Dehydration-Responsive Element-Binding Protein (DREB), Ethylene Responsive Factor (ERF) subfamily, etc. and osmoregulation.

1.5 Thesis statement

O. sativa, an important staple crop, sustains nearly half of the world's population. As the global population continues to rise, increasing *O. sativa* production is imperative [Dhiman *et al.*, 2024]. However, this goal is hindered by various biotic and abiotic stresses. To minimize reliance on pesticides for managing biotic stresses, developing more disease-resistant varieties of *O. sativa* is essential. In addition, among abiotic stresses, drought poses a significant threat to its cultivation. Incorporating DT genes or proteins from arid-region crops, such as *P. cineraria* from the Indian Thar desert, holds significant potential for developing drought-resilient *O. sativa* varieties.

This thesis aims to conduct a comprehensive study of stress tolerance genes in *O. sativa* in response to biotic and abiotic stresses, using bioinformatics-based approaches. The first objective of this thesis is to use *in silico* methods to identify the disease-resistance proteins in *O. sativa* against BB and RB diseases. It includes performing network analysis, homology modeling of the protein structures, functional annota-

tion, and *in silico* gene expression profiling. The second objective of this thesis is to perform a comparative study of machine learning (ML) and deep learning (DL) models for predicting disease-resistance proteins in *O. sativa* and its closely related species. This objective involves predicting, comparing, and validating the performance of the ML and DL models using both internal and external datasets, including the whole proteome and known disease-resistance proteins in *O. sativa* and its related species. The third objective focuses on drought-responsive transcriptome and genome-wide studies of DT genes in *P. cineraria*. This includes transcriptome profiling, *de novo* assembly, gene network analysis, phylogenetic classification, *cis*-element investigation, DNA polymorphism studies, copy number variation (CNV) evaluation, protein modelling and molecular dynamic (MD) simulations between DT and drought-sensitive (DS) species.

The novel computational frameworks developed in this work explore various approaches, such as 3D structure modeling of unknown proteins, the utilization of protein sequence-based ML and DL models, transcriptome profiling of DT genes, and comparative analysis of Indian and Arabian cultivars of *P. cineraria*. The frameworks proposed in this thesis work aim to address the existing limitations in available studies for stress-related studies in *O. sativa*.

1.6 Scope of thesis

This thesis addresses the challenges of identifying stress-tolerant genes in *O. sativa* through advanced computational approaches. A key focus is on the *in silico* identification and characterization of proteins that confer resistance to BB and RB diseases. In another work, ML and DL models were applied to predict disease-resistance proteins in *O. sativa* and its related species, aiming to provide valuable insights for enhancing disease resistance in *O. sativa* through computational genomics. In addition, the DT genes were identified and analyzed in *P. cineraria* (a plant known for drought tolerance) to transfer these genes into *O. sativa* for enhancing its drought tolerance. The structure of this thesis is outlined in Figure 1.4 and is organized into six chapters. Chapter 1 underscores the importance of staple food crops, particularly *O. sativa*, in achieving SDG2. It also examines the immune mechanisms of *O. sativa* and the role of stress-tolerant genes in mitigating the adverse effects of both biotic and abiotic stresses. Furthermore, this chapter introduces the application of computational approaches to enhance stress tolerance in *O. sativa*, and set out the objectives and expected outcomes of the study. Chapter 2 provides a comprehensive literature review, discussing the limitations of experimental methods in stress tolerance research and highlighting recent advancements in this field. This chapter provides a detailed exploration of bioinformatics-based approaches for identifying genes related to both biotic and abiotic stress tolerance in *O. sativa*.

Chapter 3 focuses on an *in silico* approach for identifying disease-resistance proteins in *O. sativa* to combat BB and RB diseases. This study integrates sequence data from well-known disease-resistance genes associated with BB and RB diseases, alongside hypothetical conserved genes identified from the complete gene set of *O. sativa*. The analysis of these genes encompass gene network mapping, *in silico* protein structure modeling and evaluation, functional annotation, phylogenetic analysis, and *in silico* gene expression profiling. BB and RB are major pathogens, with BB being a bacterial disease and RB caused by a fungal pathogen, both of which severely impact *O. sativa* productivity. Identifying and characterizing disease-resistance proteins against these diseases is crucial for developing resistant rice varieties. Such advancements could reduce pesticide usage, offering significant benefits for human health and environmental sustainability. While some disease-resistance proteins have been identified and utilized, further studies are required. Bioinformatics-based approaches provide a faster, more efficient alternative to traditional methods for identifying and understanding disease-resistance proteins in *O. sativa*.

Chapter 4 presents the *in silico* identification of disease-resistance proteins in *O. sativa*. In this chapter, a comprehensive analysis using ML and DL models is performed to predict disease-resistance proteins in *O. sativa* and related species. Traditional methods for disease identification in *O. sativa*, such as capturing images of diseased leaves and stems, is often time-consuming, costly, and labour-intensive, leading to delays in disease diagnosis. Therefore, developing ML and DL models that utilize protein sequence data is essential

for rapidly identifying disease-resistant proteins. These proteins serve as potential candidates for developing disease-resistant varieties of *O. sativa*.

Chapter 5 explores the transcriptomic response of *P. cineraria* to drought stress, to identify genes associated with drought tolerance. This study involved the analysis of gene expression patterns, focusing on upregulated and downregulated drought tolerance genes, along with the *de novo* assembly of the *P. cineraria* transcriptome and a genome-wide investigation on DT genes. Nucleotide variations, including single nucleotide polymorphisms (SNPs), multi-nucleotide polymorphisms (MNPs), insertions, deletions, and INDELs, were identified and compared to the Arabian cultivar of *P. cineraria*. Additionally, comparative analyses of CNVs, 3D structure modeling, molecular docking, and MD simulations were performed to explore the differences between DT and DS species. Furthermore, structural differences and hydrogen bond interactions analysis within DT genes were examined across *P. cineraria*, *O. sativa Japonica*, and *Indica*. The identified DT genes hold potential as promising candidates for improving drought tolerance in *O. sativa*.

Chapter 6 presents the conclusions of this thesis work and outlines several potential avenues for future research in developing stress-tolerant varieties of *O. sativa*. This thesis emphasizes the application of *in silico* methods to identify novel stress-tolerant genes in *O. sativa*, highlighting their significant advantages over traditional approaches. Additionally, a comprehensive biological analysis of the findings has been conducted to demonstrate their relevance and alignment with the study's objectives.

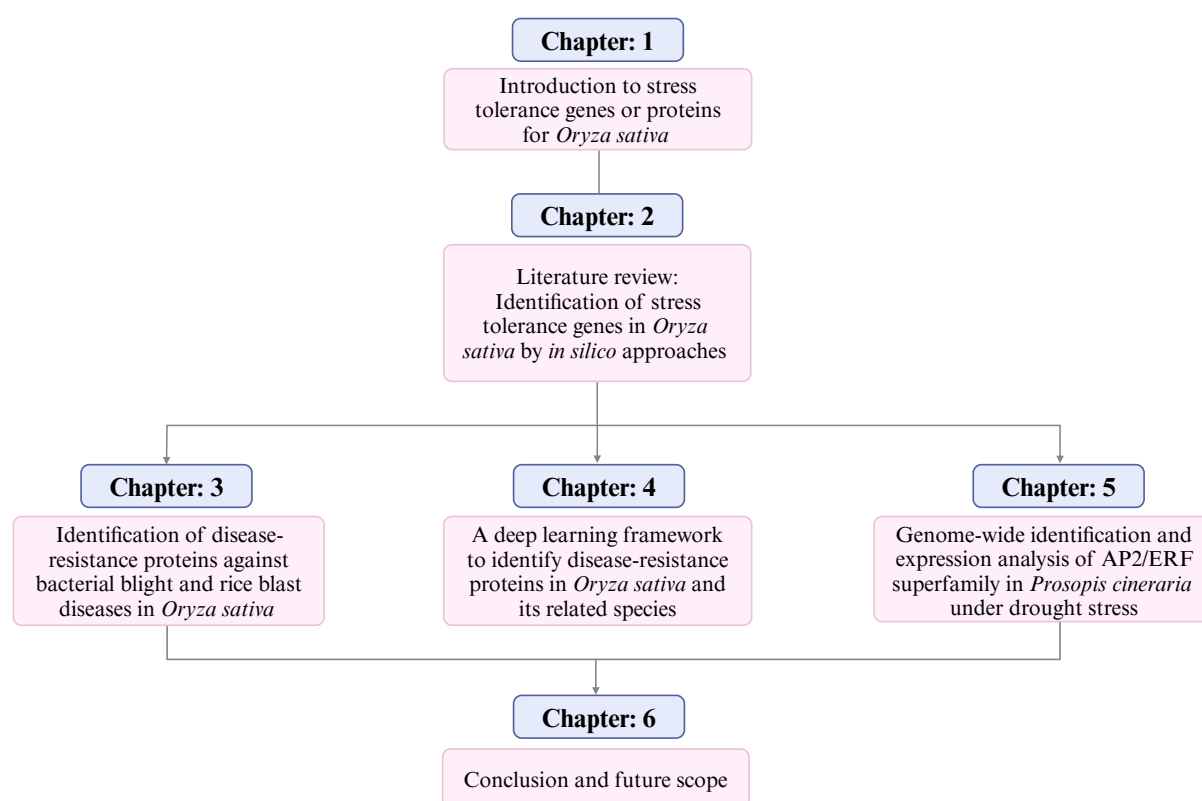


Figure 1.4 : A schematic representation of the thesis structure organized in chapters.

Abbreviations: APETALA2 (AP2); Ethylene-Responsive Factor (ERF)