Pontificia Universidad JAVERIANA Cali

ÓMICA

DOCTORATE IN ENGINEERING AND APPLIED SCIENCES

Faculty of Engineering and Sciences

A Comprehensive View of Aluminum Stress in *Oryza sativa* from a Systems Biology Perspective

Advisor: Mauricio Alberto Quimbaya

Co-Advisor: Victor Hugo Garcia-Merchán

Co-Advisor: Diana Carolina Clavijo-Buriticá

Chrystian Camilo Sosa Arango

Dissertation presented in partial fulfillment of the requirements for the degree of Doctorate in Engineering and Applied Sciences

Cali, Valle del Cauca, Colombia

July 2024

Dedicated to people who supported me in this hard journey: In the memory of my beloved mom, Luz Esperanza who passed away without seeing this document finally finished, to my dear father, Aldemar, my dear wife Ashly, and my aunts Ruby, Yoveira, and Dora Enith who provided me with words of encouragement, support, and true love in hard moments.

Acknowledgments

I would like to thank my supervisors Professor Mauricio Quimbaya, Professor Victor Hugo Garcia Merchan, and Dr. Diana Carolina Clavijo Buriticá for their invaluable advice, continuous support, mentorship, friendship, guidance, and motivation during this PhD journey that certainly was challenging at all levels but remarkably at a personal level due to the terrible loss of my mother, and the hard obstacles triggered by this fact.

I would also like to express my immense gratitude to the Pontifical Xaverian University, Cali, and the OMICAS program financed by the World Bank, the Ministry of Science, Technology and Innovation, the Colombian Ministry of Education, the Colombian Ministry of Industry and Tourism, and ICETEX for financing this doctoral training. Thanks, should also go to Leidi Rojas, and the doctoral program in engineering and applied sciences under the leadership of Professor Andres Jaramillo Botero, especially to Andrea Ramirez for providing all possible help and guidance to complete each step of the doctoral program properly.

I am extremely grateful to my doctoral colleagues Jenny Gallo, Camila Riccio, Gustavo Lara, and Nicolas Lopez, for their invaluable friendship and insightful support, help, comments, and suggestions for the development of the GOCompare R package, and different parts of this research project. I would like to thank Jason Suescum for his invaluable help, support, and guidance in the construction of the gene regulatory network model. Furthermore, I want to acknowledge my friends and people who contributed with motivation or advice such as Nora Castañeda, Colin Khoury, Stefan Burkart, Vanessa Giraldo, and Professor Klaas Vandepoele among others.

Lastly, this endeavor would not have been possible without the support of my dear father Aldemar, my mother Esperanza Arango (RIP), my dear wife Ashly Arevalo, my aunts Ruby and Yoveira Ossa, and Dora Enith, and my uncle Mario Who provided me their love, motivation, energy, and support in key moments. I would also like to thank Martha Lucia Vasquez for her support when I arrived again in Colombia to avoid worrying about being in charge of home duties and for her generous help always being available.

Contents

List of abbreviations

Al: Aluminum

ABA: Abscisic acid

AdaSampling: Adaptative sampling

ARATH: *Arabidopsis thaliana*

COBRA: COnstraint-Based Reconstruction Analysis

DFBA: Dynamic Flux balance analysis

ET: Ethylene

FEA: Functional Enrichment Analysis

FBA: Flux balance analysis

GO: Gene ontology term

GO: BP: Biological process Gene ontology term

GO:MF: Molecular Function Gene ontology term

GSEA: Gene Set Enrichment Analysis

GSM: Genome-scale metabolic reconstruction

GWAS: Genome-Wide Association Study

IAA: Indole-3-acetic acid

IBA: Indole-3- Butyric acid

JA: Jasmonic Acid

MSU: Michigan State University

ODE: Ordinary Differential Equations

OSJ: *Oryza sativa* subsp. *japonica*

PULearning: Positive Unlabeled Learning

RAP-DB: The Rice Annotation Project database

RNA-Seq: RNA sequencing

ROS: Reactive Oxygen Species

PPIN: Protein-Protein interaction network

TCA: tricarboxylic acid cycle

TIC: Thermodynamically Infeasible Cycle

Abstract

Rice is one of the most important plant species for human food and for the study of abiotic stresses such as aluminum stress. This stress generates important agronomic yield problems in rice cultivation by affecting the root growth of plants in acidic soils. Rice has been extensively studied in terms of the molecular response to aluminum, this response is complex as it includes different aspects such as detecting the stressor, triggering a transduction signal in the cells, to finally having a physiological plant response. Recently, systems biology has become relevant for stress studies, as it allows the integration of interacting elements at different levels through mathematical modeling to capture emergent properties of these interactions. Currently, there is a lack of knowledge, especially at the metabolic level, regarding the molecular response, that is induced when aluminum is uptaken by rice roots. This information gap includes the lack of knowledge of receptors, transduction signals, molecular mechanisms and specific chemical reactions that inhibit root growth. This dissertation structured a systemic approach to characterize the molecular responses triggered in rice under aluminum stress conditions through: (i) the design of a computational method that implements comparative genomics strategies to obtain functional differences between two contrasting species, with the final goal of identifying unique responses in rice, (ii) The implementation of a machine learning strategy over rice interactome to predict candidate proteins that could be relevant in the response to aluminum stress, and (iii) The use of a multilayer approach that allows modeling the molecular and metabolic response of rice to aluminotoxic conditions. The obtained results represent an advance in the perspective that stress is studied in rice. This research work identified specific functional adaptations at the species level that are linked with rice physiological responses to aluminum stress, similarly, novel putative proteins, not previously described by the literature were detected, and finally, the development and implementation of computational approaches allow the modeling of the molecular response to aluminum at the regulatory and metabolic levels. The results obtained are consistent with those that have been reported in the literature. These results might contribute to plant stress research by the use of the tools generated in this study.

Keywords: Abiotic stress, aluminum stress in rice, comparative genomics, machine learning, plant systems biology, plant stress bioinformatics.

Resumen

El arroz es una de las especies vegetales más importantes para la alimentación humana y para el estudio de estreses abióticos como el inducido por aluminio. Este estrés genera problemas de rendimientos agronómicos en el cultivo de arroz al afectar el crecimiento radicular de las plantas en suelos ácidos. El arroz ha sido ampliamente estudiado en cuanto a la respuesta molecular ante el aluminio, esta respuesta es compleja ya que incluye diferentes aspectos como la detección del estresor, la activación de una señal de transducción en las células, para finalmente tener una respuesta de la planta. Recientemente, la biología de sistemas ha tomado relevancia para los estudios de estrés, ya que permite la integración de elementos que interactúan a diferentes niveles a través de la modelación matemática para capturar las propiedades emergentes de estas interacciones. Actualmente existe una falta de conocimiento, especialmente a nivel metabólico, sobre la respuesta molecular que se induce cuando el aluminio es absorbido por las raíces de la planta de arroz. Este vacío de información incluye el desconocimiento de receptores, señales de transducción, mecanismos moleculares y reacciones químicas específicas que inhiben el crecimiento radicular. Esta tesis ha estructurado un enfoque sistémico para caracterizar las respuestas moleculares desencadenadas en el arroz bajo condiciones de estrés por aluminio por medio de: (i) el diseño de un método computacional que implementa estrategias de genómica comparativa para obtener diferencias funcionales entre dos especies contrastantes, con el objetivo de identificar respuestas únicas en arroz, (ii) la implementación de una estrategia de aprendizaje de máquinas sobre el interactoma del arroz para predecir proteínas candidatas que podrían ser relevantes en la respuesta a aluminio, y (iii) el uso de una aproximación multicapa que permite modelar la respuesta molecular y metabólica del arroz ante condiciones aluminotóxicas. Los resultados obtenidos representan un avance en la perspectiva en que se estudia el estrés en arroz. Este trabajo de investigación identificó adaptaciones funcionales especificas a nivel de especie que están vinculadas con las respuestas fisiológicas del arroz el estrés por aluminio, de igual forma, se identificaron nuevas proteínas putativas, no descritas previamente en literatura, y finalmente, el desarrollo e implementación de enfoques computacionales que permiten modelar la respuesta molecular al aluminio a nivel regulatorio y metabólico. Los resultados obtenidos son consistentes con los reportados en literatura. Estos resultados podrían contribuir a la investigación del estrés vegetal mediante el uso de las herramientas generadas en este estudio.

Palabras clave: Estrés abiótico, estrés por aluminio en arroz, genómica comparativa, aprendizaje automático, biología de sistemas vegetales, bioinformática del estrés vegetal.

List of Figures

List of Tables

Chapter 1

Introduction

Rice (*Oryza sativa* L.) is one of the main staple crops worldwide. It is the main source of carbohydrates for more than half of the world's population (Gao et al., 2008). Specifically, it supplies up to 70% of the calorie intake of poor farmers in Asia and is the main energy food source for people in Latin America (CGIAR, 2017; Fisberg et al., 2021). In a world that needs an increase from 25 to 70% above current production to meet the 2050 crop demand, rice represents an important cultivar in human food security (Hunter et al., 2017). Nevertheless, climate change, a rising human population, and acid soil intensification add important pressures to crop production, intensifying specific abiotic stresses that can harm the food supply. For instance, drastic reductions in important agronomic crops, such as maize, wheat, or rice, are expected by 2050 if adaptation measurements, such as genome editing, new crop varieties, or technology adoption, are not considered (Challinor et al., 2014; Zhu and Shen, 2023).

Acidic soils present a major challenge for agriculture. Soils with pH lower than 5.5, constitute almost 50% of the available arable land (Kochian et al., 2015). They are poor in nutrients and contain heavy metals such as iron (Fe), manganese (Mn), zinc (Zn), molybdenum (Mo), copper (Cu), cadmium (Cd), chromium (Cr), lead (Pb) and aluminum (Al) which have toxic effects on plants and generate losses in agronomic productivity (Kochian et al., 2015; Bojórquez-Quintal et al., 2017; Gallo-Franco et al., 2020). Specifically for aluminum, it constitutes the third most abundant metal on the earth's crust being harmful to plants in acid soils because at low pH, it turns into phytotoxic chemical species, Al^{3+}, $Al(OH)^{2+}$, $Al(OH)4^{-}$ y $[Al(H2O)6]^{3+}$, that causes inhibition in water and nutrient uptake, lipid peroxidation, cytoskeleton modification and overproduction of reactive oxygen species (ROS) that lead to the inhibition of root growth and root meristem cell division as adaptive responses (Bojórquez-Quintal et al., 2017; Rahman et al., 2018; Poschenrieder et al., 2019).

Rice serves as a model organism for studying aluminum stress due to its natural tolerance compared to other cereals (Kochian et al., 2015; Bojórquez-Quintal et al., 2017). Rice comprises two main subspecies: indica, predominantly cultivated across tropical regions and accounts for most of the global production, and japonica, which is grown in both tropical and temperate areas (Campbell et al., 2020). Although genome sequences are available for both subspecies, japonica

exhibits a genetic bottleneck and a lower genetic diversity than indica (Cheng et al., 2019; Goff, 2002; Yu, 2002). Nevertheless, japonica benefits from a more extensive suite of bioinformatics resources, including well-established genome annotations, genome projects, coexpression analysis, metabolite databases or protein-protein interaction networks (Itoh et al., 2007; Kawahara et al., 2013; Sakai et al., 2013; Li et al., 2014; Hong et al., 2019), In contrast, indica lacked a gapless reference genome until recently, which historically limited its use in bioinformatics-driven studies (Du et al., 2017; Li et al., 2021)

Genomic information plays an important role in plant breeding. Currently, quantitative trait locus (QTL) analysis is central to breeding programs that target abiotic stressors (Cobb et al., 2019). Advancements in omics technologies offer the potential to convert breeding into a data-rich science to reduce monetary costs (Cobb et al., 2019). Strategies to investigate and support the development of new rice varieties tolerant to aluminum include genome-wide association studies using high-density SNP markers derived from a core collection of rice landraces (Zhao et al., 2018), the application of gibberellins that suppress aluminum responsive genes (Lu et al., 2024a), or nitric oxide to neutralize the inhibitory effects of aluminum on germination (Peres da Rocha Oliveiros Marciano et al., 2010). Further identification of aluminum tolerance genes is essential for improving productivity in acidic soils(Liu et al., 2022).

Omics technologies such as GWAS, QTL, RNASeq, and proteomics have identified key genes in the responses to the stress, including OsART1, OsNrat1, OsFRDL4, SbMATE, ALS1, FRLD4, STAR1, STAR2, HMG2, STOP1, STOP2, and HMG3, which contribute to the plant adaptation by causing metabolic responses such as reactive oxygen species overproduction, organic acid biosynthesis and release, auxin and ethylene biosynthesis, and putrescine production to induce tolerance mechanisms (Famoso et al., 2011; Arbelaez et al., 2017; Wagatsuma et al., 2017; Tyagi et al., 2020; Xie et al., 2022; Gallo-Franco et al., 2023). The use of rice as a model plant to understand stress tolerance to abiotic factors has been highlighted by its fully sequenced genome (Jackson, 2016), suitable information for systems biology, and comparative genomics strategies (Hong et al., 2019). However, from a systemic perspective, the interactions among genes, proteins, and metabolites organized in a comprehensive molecular portrait to explain stress tolerance mechanisms are still lacking.

Understanding the molecular mechanisms that plants use to cope with aluminum stress is rather complex. For instance, diverse species accumulate the metal in leaves and exclude it from root tissues using organic acids that chelate this toxic element and ameliorate its toxicity in plants

(Brunner and Sperisen, 2013; Kochian et al., 2015; Nagayama et al., 2019). In the presence of different stresses, plants use a common set of genes to deploy a multilevel response that comprises gene regulation, enzyme production and metabolite biosynthesis, to reduce photosynthetic activity and survive toxic aluminum conditions (Shaik and Ramakrishna, 2014; Cohen and Leach, 2019; Jingguang et al., 2020). Consequently, to integrate these different elements, a holistic perspective is necessary and is provided by systemic approaches where multiple omics studies (e.g., transcripts, proteins, and metabolites) serve to obtain insights into abiotic stress responses in a network frame (Cramer et al., 2011).

Some of the factors that hinder a systemic approach to understanding stress responses are: (i) Currently, there is a lack of methods for comparing gene lists compositions to derive relevant biological knowledge (Yu et al., 2012; Wu et al., 2021). (ii) The interaction among proteins that induce the aluminum response in rice is rather unknown, and, some proteins that have relevant functions as triggers or regulators of abiotic stress responses, have not yet been identified in rice. (iii) The molecular mechanisms that lead to root growth inhibition under aluminum stress conditions are unclear, and, how the metabolic reactions that participate in root elongation are altered over time is unknown (Yadav et al., 2024).

This PhD thesis wants to provide a multilayer holistic perspective for studying aluminum stress in rice, through several approaches: (i) a brief literature review regarding the aluminum effect in rice physiology and how this toxic element affects root growth, (ii) an analysis of the functional divergence of rice in comparison with the model plant *Arabidopsis thaliana*, implementing a new method that integrates comparative genomics and graph theory (iii) an approach integrating the current knowledge associated with rice protein interactome and the molecular response elicited after ten days of exposure to aluminum linked to different machine learning approaches to predict undocumented proteins involved in responses to abiotic stress, and (iv) a systems biology approach that includes a dynamic modeling of the principal hormonal cues and signal transductional cascades that regulate aluminum responses (Figure 1.1).

This multilayer holistic approach was developed to address the following main research questions that are complementary to each other:

- From a functional perspective, how similar are rice and *Arabidopsis thaliana* aluminum responses?
- It is possible to identify novel proteins linked with aluminum stress responses in rice by implementing machine learning strategies upon rice interactome?

- It is possible to integrate specific signaling cascades with metabolic processes to model aluminum effect over root growth?

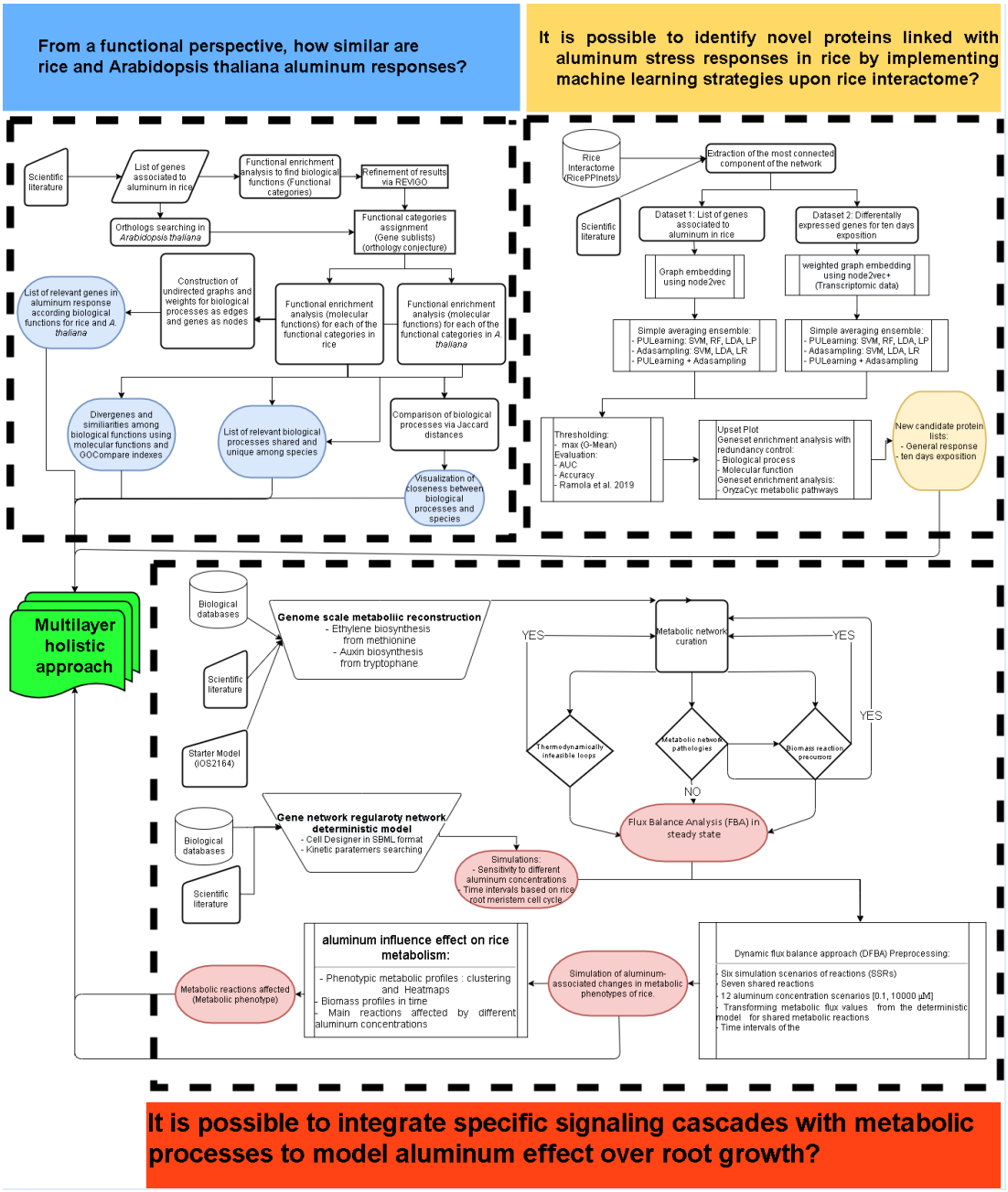


Figure 1.1. Graphical summary of the outline of the doctoral research. Three main sections are highlighted and represent each of the chapters addressing the corresponding research questions and specific objectives.

1.1. Research Hypothesis

- Aluminum stress in rice induces specific metabolic responses that might differ at the species level. Such metabolic responses are orchestrated by a network of interacting proteins and activated transduction cascades that drive differential responses in time. To better understand how the different molecular factors that trigger aluminum responses in rice, a systems biology perspective is needed.

1.2. General objective

- To structure a systemic landscape to portrait the molecular responses associated with aluminum tolerance, implementing integrative and systems biology strategies on the different mechanisms of tolerance that are triggered in rice under conditions of aluminum stress.

1.3. Specific objectives

- To elucidate biological processes associated with aluminum stress responses through comparative genomics strategies on the effects of aluminum on plants from a functional perspective.
- To identify new molecular elements associated with aluminum stress responses, via the analysis of protein-protein interaction networks based on new elements that might regulate the response of rice to aluminum stress conditions.
- To elucidate the aluminum effect over the metabolic response of rice, implementing a metabolic network reconstruction and analysis to generate a mathematical model that relates aluminum stress with the phenotypic response of rice plants.

1.4. Document Outline

This dissertation is organized as follows:

1. Chapter 2 is written as a brief literature review to support the need of a systemic approach to understanding the effect of aluminum over genetic and metabolic responses in rice.
2. Chapter 3 presents the development of a new method to analyze, from a comparative genomics approach, the biological processes associated with the responses to of aluminum stress conditions in *Oryza sativa* and *Arabidopsis thaliana*.
3. Chapter 4 presents a classification analysis via machine learning strategies implemented over the most recent rice interactome, to obtain new protein candidates that might play an important role in rice responses to aluminum stress.
4. Chapter 5 presents a holistic approach based on the modeling of the metabolic network that is associated with the induction of the physiological responses upon aluminum exposure.
5. Chapter 6 provides a general discussion and future perspectives for this study.

Chapter 2

Aluminum response in rice: A brief perspective from comparative genomics, interactomics, and genome-scale metabolic reconstruction

2.1. Abiotic stress response in plants

Because of their sessile habit, plants are subject to ever-changing environmental conditions, which alter their growth, development, morphology, and life history (Lexer et al., 2003; Lexer and Fay, 2005; Mosa et al., 2017). stress in plants can be defined as an unfavorable condition that affects or blocks metabolism, growth, or development, through specific perturbations that alter its homeostasis. In this way, stresses can be classified into biotic and abiotic, depending on whether they come from interactions with other living organisms or have an origin derived from the interaction with the environment.

Abiotic stresses are caused by non-living organisms and mostly are the result of unfavorable environmental conditions (Mosa et al., 2017; Pandey et al., 2017; Madani et al., 2019; Zhang et al., 2021). Furthermore, these stressors cause growth reductions of more than 50% in most plants, and are responsible for severe agronomic losses worldwide (Ben Rejeb et al., 2014). Among abiotic stresses, water deficit, high or low temperature, increased soil salinity, hypoxic conditions, reactive oxygen species (ROS) production, and perceived quantity and quality of light, constitute the most important stressors and currently may be increased due to rapid changes in global climate (Fahad et al., 2017; Mosa et al., 2017). In addition, other types of abiotic stresses are produced by heavy metal ions such as aluminum, cadmium, copper, iron, manganese, zinc, and arsenic affecting plants by (i) stimulating ROS and methylglyoxal production, (ii) interacting with proteins and affecting their catalytic properties, and (iii) displacing important metals with trivalent ions such as iron or manganese (Hossain et al., 2012; Poschenrieder et al., 2019).

The physiological response of plants to abiotic stress such as aluminum is complex and is sensed by primary mechanisms that transduce the stimulus into a biological signal and ultimately into a gene-specific response through complex signaling cascades. Activation of transcription factors, gene methylation, regulation of effector protein localization, and other mechanisms including metabolic responses, have been described as key processes triggered by abiotic stress

responses (Ben Rejeb et al., 2014; Zhang et al., 2018; Gallo-Franco et al., 2020; Lamers et al., 2020). The traditional strategy to study abiotic stresses is to evaluate and quantify the phenotypic response upon various levels of stressful conditions, contrasting tolerant and susceptible genotypes, to characterize the physiological damage caused by stress levels on evaluated plants (Pereira, 2016). This approach is commonly implemented in genomics, transcriptomics, proteomics, metabolomics, and phenomics. Most generated data is freely available to the scientific community. Similarly, the generated information can be also used under the integrative perspective of the systems biology paradigm with the main advantage of using various molecular scales (genes, transcripts, proteins, and metabolites) to explain cell behavior using functional networks and mathematical models that can explain the explored phenomenon from a holistic view (Cramer et al., 2011; Pandey et al., 2020).

2.2. A Systems biology perspective to study plant stresses

The use of a systems biology perspective combines omics technologies with mathematical modeling to obtain a systemic perspective on stress conditions. Plants have evolved to cope with stress conditions providing common responses such as alterations in gene expression, anatomical and morphological modifications, decreased efficiency of photosynthesis, differential nutrient assimilation, cellular metabolic adjustments, and overproduction of stress-related metabolites (Pandey et al., 2020). These responses depend on stress interval and level. For example, at low levels of stressors, metabolic response can enhance metabolic reactions to mitigate harmful conditions and even improve crop growth (hormesis effect). In contrast, plants cannot physiologically respond and die. In the middle zone, plants can adapt to diverse strategies such as stressor escape or avoidance or tolerance induction via genetic reprogramming (Morkunas et al., 2018; Bandurska, 2022). An example of the relevance of systems biology approaches implemented in crops and plant stress is provided by Pazhamala et al., (2021) who discussed the application of systems biology strategies in crop improvement, emphasizing the need to integrate genomics, transcriptomics, proteomics, and metabolomics data to understand the complex biological processes in plants. This integration enables the identification of key genes, pathways, and regulatory networks associated with agronomically important traits. In another review, Cramer et al. (2011) examined the role of systems biology in abiotic stress research in plants. The authors suggested the integration of transcriptomics, proteomics and network analysis to holistically understand plant abiotic responses, pointing out that this approach was able to identify that early downregulation of energy metabolism and protein synthesis are

fundamental processes triggered by stress cues. Thus, a holistic perspective of systems biology is valuable and necessary to better understand the stress responses in plants.

2.3. Rice molecular response to aluminum stress

Rice (*Oryza sativa* L.) is an ideal plant to study the physiology of agronomically important species, being an alternative and complementary model to *Arabidopsis thaliana* which is considered the plant model by default. Rice is a monocotyledonous, diploid, fast-growing plant. It contains a well detailed genetic information, supported by a fully sequenced and well-annotated genome. Additionally, genetic studies are supported by diverse databases such as RAP-DB which has become an important platform for studying plants from a systems biology perspective and analyzing a diverse range of stresses (Liu et al., 2001; Goff, 2002; Yu, 2002; Kawahara et al., 2013; Sakai et al., 2013; Hong et al., 2019). Rice is the grass most resistant to aluminum conditions growing under acidic soil conditions (Famoso et al., 2010). Thus, in acidic soils with low pH (pH <5) (acid soils), aluminum is solubilized, forming the chemical species Al^{3+}, $Al(OH)^{2+}$, Al(OH)4-, and $[Al(H2O)6]^{3+}$ which are usually known as Al^{3+} or aluminum ions. At low levels can enhance plant growth (hormesis effect) but mostly can generate toxic effects, such as root growth inhibition, inhibition of water and nutrient uptake, lipid peroxidation, modification of the cytoskeleton, inhibition of cell division, changes in the cell walls, and the production of reactive oxygen species in the root. All of these factors combined can produce big economic losses for rice producers (Bojórquez-Quintal et al., 2017; Rahman et al., 2018; Poschenrieder et al., 2019; Jingguang et al., 2020).

Owing to the diverse effects of aluminum toxicity in plants, understanding the biological functions and mechanisms associated with aluminum tolerance in rice is crucial for developing strategies to improve crop productivity in acidic soils. For instance, Bojórquez-Quintal et al. (2017) reviewed the mechanisms involved in improving the growth of plants cultivated in acidic soils and discussed the tolerance mechanisms to the toxic effects of aluminum by providing insights into the possible mechanisms and biological functions associated with aluminum tolerance in plants. Huang et al., (2009) investigated the involvement of a bacterial-type ABC transporter in aluminum tolerance in rice reporting the genes STAR1 and STAR2 (sensitive to Al rhizotoxicity1 and 2) as major players that confer tolerance to toxic conditions. Similarly, Kochian and colleagues highlighted the role of two main aluminum-associated tolerance mechanisms: aluminum exclusion, based on the exudation of organic acids such as oxalate, malate, or citrate and aluminum detoxification where

aluminum is taken up across root cell plasma membrane using aluminum transporters. Associated with these mechanisms Aluminum-activated Malate Transporters (ALMT), a family of organic acid transporters associated with UDP-glucose are fundamental for aluminum tolerance (Kochian et al., 2015).

Arbelaez et al., (2017) demonstrated, via RNASeq analyses, that the OsART1 (aluminum resistance transcription factor 1) gene can modulate the response to aluminum for specific rice varieties, upregulating some genes as Multidrug And Toxic Compound Extrusion (MATE) transporters, auxin, glutathione, and reactive oxygen species metabolic pathways. Moreover, Zhang et al., (2019) reported 20 new candidate genes linked with aluminum tolerance in 150 traditional varieties of the crop, including transcription factors NRAT1 (Nramp aluminum transporter 1), ART1 (Aluminum resistance factor 1), and STAR1 (sensitive to Al rhizotoxicity1). Also, Tyagi et al. (2020) identified 40 transcripts associated with signal transduction components, zinc fingers, calcineurin-binding proteins, and cell wall-associated transcripts in the *Oryza sativa* subsp. *indica* aluminum-tolerant variety ARR09. Tyagi and collaborators described a schematic representation where the tolerance of rice to aluminum toxicity is related to a downregulation of aluminum uptake, fructose bisphosphate production, lignin biosynthesis and arginine biosynthesis, similarly, the adaptation process is related to the upregulation of abscisic acid (ABA) and methionine synthesis. In addition, Bhattacharjee et al. (2023) reported that 63 indica rice varieties upregulated genes related to cytoskeletal dynamics, metabolism, and ion transporter dynamics, providing, aluminum tolerance to plants when expressed. Consequently, the elucidation of key molecular elements is translated into tolerance mechanisms that include (i) the release of chelating compounds such as phosphates or organic acids to create aluminum compounds that are not toxic to plants, (ii) redistribution of aluminum to organelles such as vacuoles, (iii) modification of cell wall properties (e.g., pectines), and (iv) pH increment via induction of specific metabolic reactions (Zhao et al., 2013; Kochian et al., 2015; Nagayama et al., 2019; Yan et al., 2022).

2.4. Phytohormones role in the metabolic response to aluminum stress in rice

The exact mechanism of Al-induced root growth inhibition remains elusive; for instance, the specific sensor, and specific phytohormones crosstalks that induce root growth under aluminum stress are unknown (Ofoe et al., 2023). Nevertheless, phytohormones can offer a good starting point, given that auxin in combination with ethylene and cytokinin seem to act synergistically or antagonistically to control plant and root growth under Al stress in plants (Ranjan et al., 2021a).

Phytohormones are signaling molecules that affect central metabolism and chelating molecules in rice (Wang et al., 2023b; Per et al., 2017a). Lu et al., (2024), explored the application of exogenous gibberellin in rice roots to promote reactive oxygen species production and suppress the expression of specific transcription factors such as Al resistance transcription factor 1 (ART1), Nramp aluminum transporter 4 (OsNramp4), and Sensitive to Aluminum 1 (SAL1). In addition, Escobar-Sepúlveda et al., (2017) explored the expression of NAC (NAM ATAF and CUC transcription factors subfamilies) transcription factors in rice plants under aluminum stress, activating abscisic acid production and suggesting that phytohormones such as auxins, cytokinins, gibberellins, abscisic acid, and ethylene, are involved in rice responses to aluminum. Contrastingly, other phytohormones such as brassinosteroids, inhibit the degradation of lipids and overproduce reactive oxygen species under heavy metal conditions (Rajewska et al., 2016; Basit et al., 2022).

Regarding phytohormones and their relationship with root growth, Wu et al. (2014) found that overexpression of the OsPIN2 gene in rice roots reduced the formation of reactive oxygen species, lipid peroxidation, and lignification. In contrast, OsPIN2 increased indole-3-acetic acid production, suggesting an important role for auxin in root elongation during aluminum stress. This was corroborated by Wang et al. (2019) who via CRISPR/Cas9 identified the regulatory role of the OsAUX3 gene in aluminum response by evaluating specific mutant lines. The authors reported that the gene codifies for an auxin carrier transport that when suppressed, inhibits root growth, confirming that auxin plays an important role as a signaling molecule in response to Al stress in rice.

Zhu et al.,(2019) found that putrescine alleviates aluminum toxicity in rice roots by affecting the expression of genes related to ethylene biosynthesis. In addition, Xie et al. (2022), via metabolome analysis of rice cultivars under aluminum stress, found an induction of 1-aminocyclopropane-1-carboxylate, which is associated with ethylene synthesis. In summary, phytohormones are part of the metabolic response of rice, and specifically, auxin and ethylene are involved in root growth and are present in tissues exposed to toxic levels of aluminum (Van De Poel et al., 2015; Kopittke, 2016).

2.5. Auxin and ethylene role in response to aluminum stress and their effect on root growth

The root apex is the main zone that responds to aluminum toxicity (Wang et al., 2023b). Auxin and ethylene crosstalk is important in controlling root growth, as they stimulate cell wall loosening

and affect root thickness. Similarly, the synergistic action of auxin and ethylene mediates root growth inhibition induced by aluminum (Zhou et al., 2020; Wei et al., 2021). Auxin biosynthesis occurs via tryptophan, given that this amino acid is transformed to indole-3-pyruvate and further to indole-3-acetic acid, the main form of auxin, universally present in plants. Moreover, auxin regulates heavy metal stress by activating transcription factors that initiate auxin signaling and its responses such as the regulation of photosynthesis, increased vacuolar sequestration, and enhanced activity of antioxidant enzymes (Mathur et al., 2022). Regarding auxin roles in rice genetics, aluminum upregulates TAA1 (TRYPTOPHAN AMINOTRANSFERASE of ARABIDOPSIS 1) and YUC (tryptophan aminotransferase of Arabidopsis (TAA/YUCCA) transcripts to increase auxin synthesis (Figure 2.1). Together with organic acids, exudation acts synergistically to protect the root growth (Liu et al., 2022). Specifically ,in rice Takehisa et al. (2012), demonstrated that auxin biosynthesis-related genes such as OsASA1, OsASA2, OsASB1, and OsTAA1 and ethylene-related genes are expressed in root apical tissues to control plant growth. Kitomi et al., (2012), used rice mutant plants and GUS (β-Glucuronidase) stainings to demonstrate that auxin signal leads to OsIAA13 degradation and the activation of nine genes involved in lateral root formation. Zhou et al., (2022) showed that ETHYLENE INSENSITIVE3 like1 gene (OsEIL1) directly activates OsTAR2 which is required for auxin biosynthesis. Thus, positive feedback of between ethylene and auxin is present in rice to control root growth. Brumos et al., (2018) tested the role of local auxin biosynthesis in *A. thaliana* root meristem and found that biosynthesis via tryptophan is sufficient for root meristem viability. Also, polar transport from the shoots to roots and local auxin biosynthesis act together to promote plant development.

Ethylene is a gaseous hormone synthesized from S-adenosylmethionine (SAM), which is synthesized from methionine and affects root architecture. It is linked to stress responses in plants by the regulation of auxin synthesis and its distribution (Qin et al., 2019). In support of this regulation, Růžička et al., (2007) using histochemical GUS staining in *A. thaliana* transgenic lines demonstrated that ethylene stimulates auxin biosynthesis and basipetal auxin transport toward the elongation zone to inhibit cell elongation. Yin et al., (2017), suggest that ethylene in rice is involved in seedling growth, flowering, grain filling, grain size, and leaf senescence. Also, the authors proposed that ethylene works as a primary signal for some stresses such as flooding in cereals. In complement Huang et al., (2022), demonstrated that ethylene is accumulated in root tips and induces auxin and abscisic acid (ABA) biosynthesis to modify cell elongation and root radial expansion. Hence, the role of auxin and ethylene crosstalk is corroborated in rice as in *A. thaliana* given that the molecular response to aluminum has conserved molecular elements in plants (Brunner and Sperisen, 2013; Kochian et al., 2015; Kopittke, 2016; Jingguang et al., 2020).

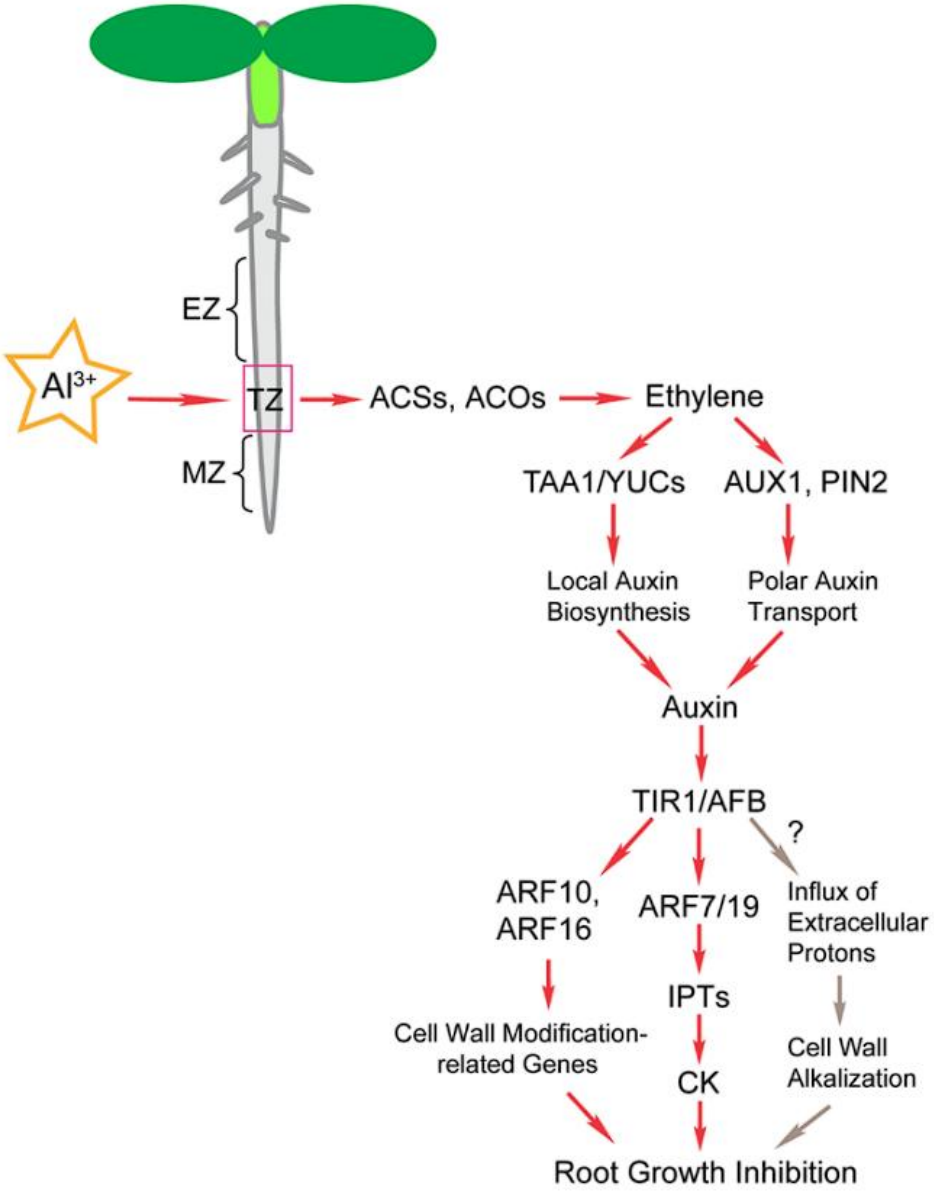


Figure 2.1. Schematic representation of ethylene- and auxin-mediated regulation of root growth inhibition in response to Al stress. The proposed hormone signaling pathway under Al stress was based on recent research on plants. Extracted from Liu et al., (2022).

2.6. Comparative genomics between rice and *Arabidopsis thaliana*

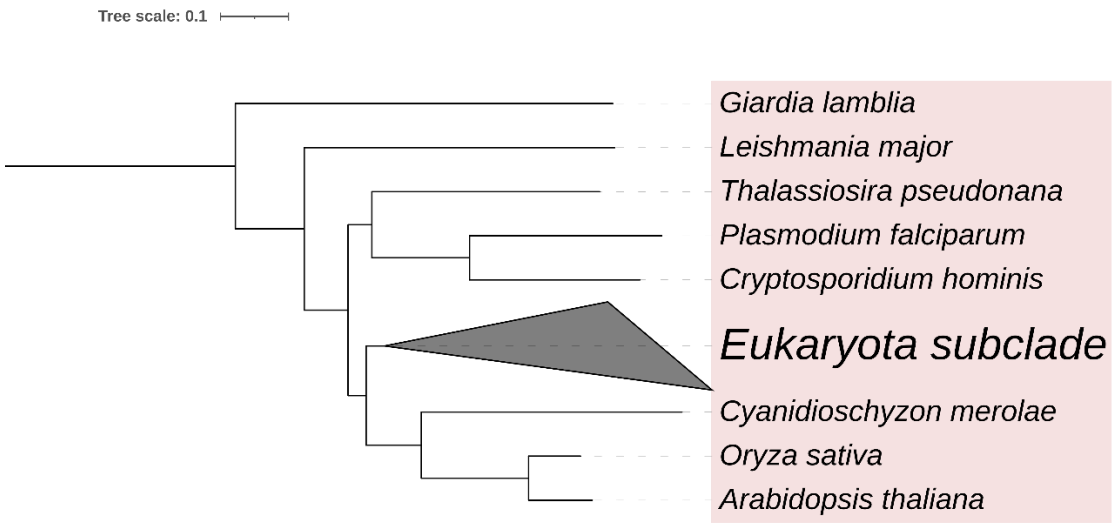


Figure 2.2. Phylogenetic relationships between *Oryza sativa, Arabidopsis thaliana*, and Eukaryota subclade obtained from The Interactive Tree Of Life (https://itol.embl.de) (Letunic and Bork, 2024).

Rice (*Oryza sativa*) and *Arabidopsis thaliana* have been a topic of interest in plant biology research. These two model plant species, particularly rice, have been extensively studied due to their importance in agriculture and suitability for genetic and genomic analyses. *Arabidopsis thaliana*, commonly known as the thale cress, is a key model organism for plant research. Its genome was sequenced and analyzed, providing valuable insights into plant biology. *A. thaliana* has a relatively small nuclear genome, a short generation time, and a single plant can produce thousands of seeds, making it an ideal model for genetic studies (The Arabidopsis Genome Initiative, 2000). On the other hand, rice (*Oryza sativa*) is one of the most important staple crops worldwide. It has a larger genome compared to *Arabidopsis thaliana,* but it has been also extensively studied at the genomic level. Genomic data are available on a large number of data platforms such as RAP-DB, funRiceGenes, RiceNet, or RiceXPro. More than 3000 accessions have been sequenced and massive sequencing projects like the OryzaSNP project, which aimed to discover genetic variation within 20 rice varieties and landraces, have been launched (McNally et al., 2009; Sakai et al., 2013; Hong et al., 2019).

Rice and *A. thaliana* have diverged from 140-200 Million Years Ago (MYA) (Figure 2.2). Rice has a greater diversity of transposable elements, longer nucleotide insertions and produces larger genes than *A. thaliana* (Nakamura et al., 2007). In addition, rice and *A. thaliana* have similar

molecular machineries with approximately 9914 possible orthologs among both species (Itoh et al., 2007), Some examples of this similarity are: (i) The two component system (TCS), in charge of ethylene and cytokinin signaling, phytochrome-mediated responses to light (Pareek et al., 2006), (ii) A similar co-expression network neighborhood, where predicted ortholog genes of species had at least one ortholog gene with conserved regulation showing that there are changes in genes information between plant species, but the regulation of genes is highly similar between rice and *A. thaliana* (Itoh et al., 2007; Netotea et al., 2014), and (iii) the flowering molecular mechanism in rice is conserved between *A. thaliana* (Kobayashi and Weigel, 2007). Thus, comparing the genomic information of different species is valuable for advancing plant physiological research.

Building on these molecular similarities, by comparing the genomes of *Oryza sativa* (rice) and other plants such as *Arabidopsis thaliana*, researchers can identify conserved genes and metabolic pathways, as well as species-specific adaptations (Alföldi and Lindblad-Toh, 2013). This comparative approach can provide valuable insights into the evolution and function of key regulatory modules in plants contributing to the understanding of the genetic basis of important agronomic traits (Liu et al., 2001; Schoof and Karlowski, 2003).

The analysis of the *Arabidopsis thaliana* and rice genomes from a comparative perspective provides a potential opportunity to elucidate conserved molecular elements that can enlighten foundation plant evolutionary biology, crop improvement, and plant stress management (Sakai et al., 2013). For instance, Kim et al., (2020) conducted a genome-wide analysis of the RopGEF gene family in rice (*Oryza sativa*) to identify genes contributing to pollen tube growth by comparing the expression of RopGEF genes in rice and *A. thaliana*, providing insights into the molecular mechanisms underlying pollen tube growth in these two species. Similarly, Rigault et al., (2011), discussed the importance of reference genomes in gymnosperm phyla plant evolution. Although the study did not directly compare *O. sativa* and *A. thaliana*, it emphasized the need for comparative genomics studies to expand plant genome knowledge.

Reinforcing the important role of comparative genomics, rice is one of the most important species within cereals and it has been one of the first available annotated genomes in databases such as GRAMENE (Tello-Ruiz et al., 2021). As a result, the availability of genomic information helps to elucidate unknown genomic aspects of other species due to ortholog gene identification, and given that these homologs tend to conserve their biological and molecular functions among species, that information can be used to enrich genomic annotation of different species, not only

cereals, by comparative genomics approaches (Altenhoff et al., 2012; Gaudet and Dessimoz, 2017).

2.7. Functional enrichment analysis (Gene set enrichment analysis)

Functional enrichment analysis (FEA), also called Gene Set Enrichment Analysis (GSEA) is a widely used approach in crop research to gain insights into the biological processes, molecular functions, cellular components, and metabolic pathways associated with specific sets of genes obtained from diverse omics approaches.

The Gene Ontology (GO) terms are widely used in FEA which provides a standardized vocabulary and hierarchical structure for annotating genes with functional terms, enabling the comparison and integration of data across different species. (Ashburner et al., 2000). By identifying overrepresented Gene Ontology terms, researchers can uncover the underlying biological mechanisms and functions related to a particular set of genes of interest (Subramanian et al., 2005). FEA is a knowledge-based approach for interpreting genome-wide expression profiles and determines, for instance, in transcriptomics experiments, whether a group of differentially expressed genes is enriched for specific pathways or ontology terms by using overlap statistics, such as the cumulative hypergeometric distribution (Huang et al., 2009b).

The GSEA approach is used in conjunction with other resources such as STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database, which provides protein-protein association networks that can be used in functional enrichment analysis by consolidating known and predicted protein-protein association data with the identification of functional modules and pathways associated with a set of genes (Szklarczyk et al., 2019). Even more tools, such as AmiGO, BinGO, DAVID, gProfiler, or topGO, among others, are popular in biological research (Alexa et al., 2006; Huang et al., 2007; Carbon et al., 2009; Kolberg et al., 2020). Specifically, in the context of crop research, FEA is recurrently used in transcriptomics experiments and even there are tools such as AgriGO using specific GO terms for agriculture. Varoquaux et al. (2019) used transcriptomics data from field-droughted sorghum to reveal the biotic and metabolic responses associated with drought stress or Gallo-Franco et al. (2023), used transcriptomics datasets and functional enrichment analysis to obtain insights into aluminum responses in *Oryza sativa* and *Oryza glumaepatula* together, supporting the use of FEA in aluminum stress research. However, to date, functional comparisons with other plant model species have not been available.

2.8. Machine learning use in plant stress research

The implementation of machine learning tools allows the using of the available information provided by a user and the versatility of machine learning approaches together to predict features related to a trait, gene prediction, or protein unknown functions. Ma et al., (2014a) discussed the use of machine learning for Big Data analytics in plants, highlighting its potential for analyzing large-scale datasets and extracting meaningful biological insights. Furthermore, machine learning algorithms have been employed in plant disease diagnosis and identification. Singh et al., (2016) discussed the use of machine learning for high-throughput stress phenotyping in plants, enabling the identification of features from large datasets, facilitating stress phenotyping and the identification of stress-responsive genes and metabolic pathways. Also, the authors provide a model of machine learning applications in plant stress research known as (ICQP), which means: (i) Identification, (ii) Classification, (iii) Quantification, and (iv) Prediction. This IQCP paradigm is extended to other plant stresses, including the hormesis research (benefits to an organism of a stressor such as aluminum). Rico-Chávez et al., (2022) reviewed the use of machine learning approaches in plant stresses, including the data limitations, exploring hardware, and compiling different machine learning applications in plants. Another important reference on the use of machine learning in plants is provided by Mahood et al., (2020) where authors highlighted the potential of machine learning for specific applications of identifying structural features in sequenced genomes, predicting interactions between different cellular components, and predicting gene function and organismal phenotypes.

Machine learning has been extensively studied for gene prediction in rice. For instance, Shaik & Ramakrishna., (2014) used microarray data and a Support Vector Machine algorithm (SVM) to identify abiotic and biotic-related responsive genes in rice. The authors found 1377 differentially expressed genes that are phytohormones-related, such as abscisic acid, jasmonic acid, and ethylene activators; small metabolites like shikimate; transcription factors such as WRKY and MYB, and signaling genes involved in crosstalk cascades between abiotic and biotic stresses. Also, Thanmalagan et al. (2022) identified 128 genes related to drought response, including genes related to growth regulatory hormones such as abscisic acid and hydrogen peroxide, carbohydrate metabolism, and photosynthesis activators. Nevertheless, neither of these approaches considered the possible connections among proteins in protein-protein interaction networks nor co-expression networks-based approaches to predict candidate genes under different stress contexts.

2.9. Graph theory applied to plant stress research

Graph theory has emerged as a powerful tool in various scientific disciplines such as biology and computational biology. It provides a framework for analyzing and understanding complex networks, such as biological networks (Aittokallio, 2006; Klamt et al., 2009; Currie et al., 2014; Mulder et al., 2014; Elmsallati et al., 2018). In recent years, there has been growing interest in applying graph theory-based methodologies to elucidate biological processes, including the response to stress in plants. Biological networks, such as gene regulatory networks, protein-protein interaction networks (PPIN), co-expression networks, or genome-scale metabolic reconstructions (GSM), are represented as graphs where nodes represent biological entities (e.g., genes, proteins, metabolites) and edges represent interactions or relationships between these entities (Uhrig, 2006; O'Brien et al., 2015).

Biological systems networks are highly modular. This means the presence of nodes (e.g., proteins or genes) that connect preferentially with other nodes in the network to form subgroups or functional communities where few nodes have a high number of connections with a high degree value and node statistical distribution of degrees in the form of a power law (Ideker & Sharan, 2008). In this manner, representing a biological phenomenon like plant stress responses is possible in a network such as PPIN since protein interactions are central controllers of all biological processes. A specific stress response creates hubs of proteins where transcription factors, protein kinases, phosphatases, ubiquitin-proteasome, system-related proteins, and redox signaling proteins can be detected by analyzing the structure of the network (Liu et al., 2017a; Vandereyken et al., 2018). A good example of PPIN used in stress is experimentally provided by Seo et al. (2011), who created an interactome of 100 proteins using yeast-to-hybrid (Y2H) techniques to identify an interactome of rice stress responses. The authors showed that a large number of proteins whose molecular function was related to transcription factors had components associated with both biotic and abiotic stress responses, suggesting a combined response of genetic components to various stresses. Additionally, the authors provided a list of ten master regulators of rice stress responses.

Szklarczyk et al., (2019), presented a new version of the STRING database, which provides protein-protein association networks with increased coverage. This resource enables the exploration of PPINs and supports functional discovery in genome-wide experimental datasets. The integration of PPINs with experimental data allows the identification of protein interactions and the inference of functional associations. Therefore, graph representations are useful for

candidate gene or protein predictions and for identifying complex physiological modules linked with stress responses.

Machine learning techniques implemented over interactomes to identify agronomical traits or stress responses have been explored for years (Wen et al., 2002; Ding et al., 2009; Singh et al., 2012; Sharan et al., 2017). For instance, to identify mitogen-activated protein interactions Liu et al., (2017) proposed a predicted rice interactome named RicePPINet consisting of 708819 interactions for 16895 proteins. The authors used the Random Forest method to demonstrate the ability to find candidate genes for disease resistance and drought tolerance. Also, Wimalagunasekara et al. (2023) used a STRING database-derived network and a local neighborhood machine-learning approach, identified a list of 75 candidate proteins linked with root development. In another example of gene prediction and rice stresses, Smet et al. (2023) used transcriptomics-derived data in combination with a random forest algorithm, a co-expression network, and a small subset of transcription factor motifs obtained from the JASPAR database to detect transcriptional response to heat and drought in rice. Also, in a similar work, Das & Maity, (2022) used a co-expression network obtained from microarray experiments and a support vector machine algorithm (SVM), identified hub genes in aluminum stress responses in soybeans.

In order to use interactomes and machine learning together to provide useful information regarding a biological phenomenon such as stress response, it is needed to take into account network properties. Network properties include degree distributions, clustering associations, or network growth models that provide the network scaffold for understanding the application of graph theory in biological systems (Newman, 2003; Barabási and Oltvai, 2004). A good example of the use of biological network properties to get a functional insight is the detection of disease proteins derived from the human interactome. This human interactome has been used from different perspectives: (i) global network topology approaches (probability, or machine learning) or (ii) local neighborhood approaches (Guilt-by-Association, Majority Voting, or Neighbor Counting) (Mulder et al., 2014). These approaches can be extended via methodologies such as N2VKO, which uses a graph embedding, a transformation of topological data into a low-dimensional space matrix, to obtain novel gene predictions (Ata et al., 2018). With this method, it was possible to predict the protein function for 11865 genes in *A. thaliana* under contexts of seed development and water deprivation (plant stress) (Depuydt and Vandepoele, 2021). Nevertheless, local neighborhood approaches are computationally demanding for prediction inference.

The challenge of using global network topology remains in converting that information into a dataset that machine learning algorithms can use properly. To handle this, graph embedding is a

convenient option. Its goal is to find a low-dimensional representation that preserves the network features with the advantage of providing a framework for applications in network alignments (to compare networks and interactomes), community detection (to identify protein groups), and protein link prediction (Grover and Leskovec, 2016; Nelson et al., 2019). In this sense, the approaches node2vec and its counterpart node2vec+, specifically designed for weighted graphs (graphs with weights in their edges), have probed its usability to predict genes even with the lack of labeled data, in this context, genes associated with a functional feature. Both approaches preserve node network features and improve accuracy when nodevec is compared with other machine learning models such Graph Neural Networks (GNN) (Liu et al., 2023). In this regard, it is important to highlight the combination of interactome and machine learning techniques proposed by Clavijo-Buriticá et al. (2023a) where the combination of interactome and positive unlabeled learning (PU-Learning) was useful to obtain protein candidates in *Arabidopsis thaliana*. Thus, novel tools are emerging as the potent software Pecanpy which provides fast, high-quality node embeddings for biological networks of all sizes and densities (Liu & Krishnan, 2021).

In summary, machine learning and PPIN can obtain good results to identify proteins involved in stress responses. Nevertheless, these approaches together do not serve to study the metabolic response caused by aluminum that affects phytohormone biosynthesis or reactive oxygen species generation. Thus, to observe the metabolic response other approaches as provided by the systems biology paradigm are required.

2.10. Characterizing Metabolic Stress Response in Plants

The molecular mechanism of plants to face abiotic stresses is a complex multilevel interaction involving sensing, signal transduction, transcription, translation, and protein modification (Lamers et al., 2020; Zhang et al., 2023). Hence, employing metabolomics, proteomics, genomics, and other omics approaches is mandatory to characterize the complex metabolic networks and regulatory mechanisms underlying plant stress responses. For example, under different abiotic stress conditions Swindel and colleagues through genome-wide association studies (GWAS), and the gene expression responses to stress, found that stress responses were dependent on plant organs, which implies that different *Arabidopsis thaliana* organs respond differently to diverse stresses, but a core of 67 genes always is expressed in response to different stresses (Swindell, 2006). In another study with the model plant, it was shown that common transcriptional units that coordinately trigger abiotic stress responses are highly dependent on abscisic acid (ABA) and reactive oxygen species (ROS) production, implying that some stress responses remain

conserved through different organisms (Ma and Bohnert, 2007). Similarly, rice displays a common core-transcriptome that is based on some phytohormones induction such as abscisic, jasmonic and salicylic acids, and the downregulation of photosynthesis genes, suggesting that rice´s metabolic response to stresses includes an affectation to the central metabolism, modifying ATP production including ethanolic fermentation together with glycolysis (Lakshmanan et al., 2013; Cohen and Leach, 2019). Thus, metabolic responses under stress conditions have a conserved pattern where phytohormones and energy mechanisms are present.

As part of the metabolic response to stress, plants release compounds that act as signals to modulate their response. These common transduction molecules are part of the pathways of (i) reactive species, (ii) calcium-dependent, (iii) phosphorylation/dephosphorylation and kinase cascades, (iv) plant hormones pathways (e.g., abscisic acid, ethylene, auxins, brassinosteroids) (Nykiel et al., 2022). Commonly, phytohormones are the most well-known part of the plant response, and they can interact with each other to produce synergistic or antagonistic actions due to a shared transcriptional machinery. This phenomenon is known as crosstalk and happens with other signal molecules, such as sugars (Fujita et al., 2006; Fàbregas and Fernie, 2021). Additionally, hormones regulate the molecular response of plants in the presence of a contrasting condition. (e.g., plants under an aluminum treatment Vs no aluminum-treated plants), Hormone concentrations in plants remain low, affecting their study; thus, in their biosynthesis, important compounds such as ascorbate, proline, and intermediates of the tricarboxylic acid cycle are required (Fàbregas and Fernie, 2021). Hence, the phytohormone responses to abiotic stresses and their crosstalk provides a metabolic response that is possible to study using different omics technologies.

Different examples of crosstalking are available in literature. For instance, under abiotic stress, *Thymus praecox* can modulate its defenses against pathogens when cadmium is added; its accumulation induces salicylic acid production, which is used against *Erysiphe cruciferous* infection (Morkunas et al., 2018).

Abscisic acid (ABA) is involved in the regulation of seed development and dormancy, including the activation of phaseic acid that interacts with ABA receptors for signaling plasticity in plants (Garay-Arroyo et al., 2012; Weng et al., 2016). ABA acts antagonistically with gibberellins for germination and interacts with jasmonic acids to cope with abiotic stress responses. Also, it can act antagonistically with brassinosteroids, contributing to drought responses in *Arabidopsis thaliana* (Garay-Arroyo et al., 2012; Shu et al., 2018; Cohen and Leach, 2019; Waadt et al., 2022). Another good example of crosstalk is considered for rice and *Arabidopsis thaliana* root growth in

the presence of heavy metals such as aluminum or cadmium, which trigger ethylene biosynthesis. Ethylene is a gaseous hormone that enhances auxin biosynthesis, a hormone involved in root apical meristem growth. Auxin causes metabolic changes limiting the availability of 2-oxoglutarate, succinate, fumarate, and malate, affecting in turn plant growth and crop yield (Kochian et al., 2015; Van De Poel et al., 2015; Jingguang et al., 2020). Thus, crosstalk is observed widely in plant stress responses.

Some examples of the metabolic response caused by plants are provided by Sharma et al., (2019) where authors reviewed the response of the phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress, highlighting the activation of phenylpropanoid metabolism and the role of phenolic compounds in stress tolerance. Another example is provided by Rizhsky et al. (2004) studying the response of *A. thaliana* to a combination of drought and heat stress, revealing the accumulation of sugars and other metabolites involved in stress acclimation. Furthermore, proline, in plant stress tolerance, has been extensively studied. Per et al. (2017) discussed approaches to modulate proline metabolism in plants for salt and drought stress tolerance, highlighting the importance of plant growth regulators and mineral nutrients in enhancing stress tolerance.

Metabolomics and molecular approaches have been employed to uncover metabolic reprogramming in plants under abiotic stress. Kumar et al. (2021) reviewed the application of metabolomics in studying drought stress tolerance in plants for the identification of key metabolites and metabolic pathways involved in stress response. Furthermore, Xie et al. (2022) used metabolome analysis in tolerant and susceptible rice varieties. The authors found that putrescine and glutathione metabolism-related compounds such as L-homoserine, L-methionine, L-ornithine, L-proline, and 1-aminocyclopropane-1-carboxylate, which is associated with ethylene synthesis, are related to aluminum tolerance in rice. Wang et al. (2023) used a targeted metabolomic analysis to suggest that phenylpropanoid biosynthesis is involved in aluminum tolerance. Nevertheless, differential metabolic responses over time are not available in the current literature.

2.11. Genome-scale metabolic reconstruction

A Genome-scale metabolic reconstruction (GSM or GENRE) brings together the biochemical reactions derived from the genome annotation of an organism, a biomass reaction that simulates cellular growth, and a core of exchange fluxes, such as transport, that simulates the extracellular

nutritional environment (Fondi and Liò, 2015). The creation of a GSM is a demanding task that consists of multiple steps divided into four stages: (i) draft reconstruction, based on the genome annotation of the target organism (ii) refinement of reconstruction, for instance, confirmation of reaction directionality, (iii) conversion from reconstruction to mathematical model to verify stoichiometry of the objective function, and (iv) network evaluation to identify and resolve errors, gaps, and inconsistencies in the network (Thiele and Palsson, 2010; Santos et al., 2011). Thus, by integrating genomic information, biochemical knowledge, and computational modeling into a structured format, a GSM combines a network topology in the form of a stoichiometric matrix S, with the metabolites as rows, reactions as columns, and the stoichiometric coefficients in each cell that can be used in modeling approaches such as constraint-based reconstruction analysis methods (COBRA), or Flux balance analysis (FBA). These strategies are suited for studying the metabolism of organisms such as plants from a systemic perspective (de Oliveira Dal'Molin et al., 2010; Thiele and Palsson, 2010).

FBA takes the stoichiometric matrix S, which captures all structural properties of the metabolic network, the pseudo-steady state assumption (no accumulation of internal metabolites compared to flux through the metabolite nodes), and thermodynamical constraints and it uses a linear programming problem (LP) to maximize or minimize a linear combination of reaction fluxes and its general algebraic form is represented in the following equations:

$$Maximize\ z = V_{biomass} \quad \textbf{(Equation 2.1)}$$

$$Subject\ to:$$

$$\sum_{j \in J} S_{ij} v_j = 0, \forall i \in I \quad \textbf{(Equation 2.2)}$$

$$LB_j \leq UB_j, \forall j \in J \quad \textbf{(Equation 2.3)}$$

$$v_j \in \mathbb{R}$$

Where z is the objective function expressed as biomass production represented as the weighted sum of reactions that synthesize biomass precursors $V_{biomass}$ (Equation 2.1). S_{ij}, is the stoichiometric matrix where the stoichiometric coefficient i in the reaction j; $v_j$is a vector of reaction fluxes j (Equation 2.2). LBj y UBj are vectors of lower and upper bounds allowed for the reaction j (Equation 2.3). For irreversible reactions, LBj, and UBj values were 0 and 1000 (Maranas and Zomorrodi, 2016a).

2.12. Plant metabolic reconstructions related to abiotic stresses in rice

For *Oryza sativa*, some genome-scale metabolic reconstructions have been reported. The first model was described by de Oliveira Dal'Molin and collaborators in 2010. This C4GEM model is a modification of the AraGEM model (de Oliveira Dal'Molin et al., 2010) that involves transport reactions between plant organelles to study the metabolism of C4-type plants. This model contains 1755 metabolites and 1588 reactions. Subsequently, two reconstructions for *Oryza sativa* using the RiceCyc database were performed to elucidate the response to light intensity by saturating the rubisco enzyme that acts on the reactions of the dark phase of photosynthesis (Poolman et al., 2013, 2014). Complementarily, Dharmawardhana and colleagues proposed a model of 2190 reactions and 1543 metabolites to study the effect of tryptophan, serotonin, and auxin biosynthesis to analyze photoperiod and biotic stresses caused by these metabolites (Dharmawardhana et al., 2013). Liu and collaborators tested the RiceNETDB database within a genome-scale reconstruction (RiceGEM) to study the plant's central metabolism (Liu et al., 2013).

Another example of rice GSM was implemented by Lakshmanan et al. (2013), who proposed the first metabolic model for rice involving two plant tissues, germinated seeds, and leaves. Lakshmanan also included aerobic and anaerobic conditions to simulate drought and flood stress in rice. This model stated that sucrose cleavage and ethanolic fermentation explain the adaptation of rice to anoxic conditions (Lakshmanan et al., 2013). Subsequently, Lakshmanan et al., (2015), established the iOS2164 model to elucidate aspects of light response based on metabolomic and transcriptomic data. This study revealed that phytohormones, secondary and photosynthetic metabolism are overexpressed under blue light conditions, while carbohydrate degradation was found to be more active under dark environments.

Modeling under the FBA approach has been successfully employed to determine that slow ammonium assimilation produced in the chloroplast and its transport to the cytosol causes an accumulation of this metabolite that is reflected in decreased reaction fluxes for chlorophyll biosynthesis in the plant (Chatterjee and Kundu, 2015).

In another article, Shen et al., (2018) used the integrative metabolic analysis tool (iMAT) to develop sub-models for eleven rice organs based on the iOS2164 model to elucidate metabolic changes in rice grain development, obtaining the first multi-organ genome-scale model in rice. Another metabolic reconstruction in rice tried to observe the salinity and water constraint stresses, mapping 1200 reactions to predict the effect of salinity stress (Wanichthanarak et al., 2020). By

integrating the results of the crop growth simulation model (WOFOST: WOrld FOod STudies), a metabolic model that integrated different plant tissues and a dynamic flux balance analysis demonstrated the metabolic differences that exist under contrasting conditions of water availability (Shaw and Cheung, 2021). In summary, rice has been extensively modeled through FBA to gain insights into the metabolic responses to abiotic stresses. Nevertheless, genome-scale metabolic reconstruction related to aluminum stress in the plant is still absent.

2.13. Dynamic flux balance analysis and its use in plants

FBA is based on a pseudo-steady state assumption, but such an approximation only considers the maximization of an objective function without considering time intervals, the rate of change of flux constraints, or predicting metabolite concentrations. To solve this, the dynamic flux balance analysis (DFBA) was introduced by (Mahadevan et al., 2002). DFBA includes the rates of product and biomass formation to update extracellular concentrations and allows modeling the kinetics of these concentrations as ordinary differential equations (ODEs), accounting for the variation of their concentrations in time (Scott et al., 2018). Thus, DFBA introduces the constraints on the rate of flux change by two formulations: (i) dynamic optimization approach (DOA), where the optimization problem is solved through nonlinear programming (NLP) to obtain the reaction flux and metabolite concentration profiles at discrete points in time intervals, (ii) static optimization approach (SOA) where the optimization problem remains linear but solved over n intervals using the Euler forward method to obtain the flux distribution at a given instant (Mahadevan et al., 2002; Gomez et al., 2014).

The use of DFBA is recent in the literature, some examples are provided by Mahadevan et al., (2002) to model *Escherichia coli* growth. Similarly, Clavijo and colleagues simulated the pyoverdine factor (PVD) biosynthesis in *Pseudomonas aeruginosa* through a systematic approach that combined a DFBA strategy and a gene network deterministic model to obtain bacterial growth outline, and concentration profiles of glucose, PVD, and QS signal molecules (Clavijo-Buriticá, Arévalo-Ferro, et al., 2023b). Another important work is the one performed by Kuriya and collaborators who established a DFBA approach to model the production of shikimic acid from glucose in *Escherichia coli* (Kuriya & Araki, 2020).

Regarding DFBA in plants, Luo et al. (2009) introduced a modified DFBA version called Metabolic Adjustment Dynamic Flux Balance Analysis (MDFBA) to study the robustness of C3 plants photosynthesis under drought stress and high CO_2 concentration conditions. Grafahred and

collaborators applied DFBA in barley (*Hordeum vulgare*) to model plant metabolism during the seed developmental phase of barley plants (Grafahrend-Belau et al., 2013). Another DFBA application has been applied in *A. thaliana* and rice to predict the effect of physiological or environmental perturbation on the growth of *Arabidopsis thaliana* leaves and roots, and the metabolic adaptation to drought exposure (Shaw and Cheung, 2018, 2021). Thus, DFBA has been used in plants and stress research, but only one study has been performed in rice. A detailed description of the methods discussed in this section is provided in Supplementary Table 2.1.

Chapter 3

A functional comparison between *Oryza sativa* subsp. *japonica* and *Arabidopsis thaliana* aluminum tolerance mechanisms

3.1. Link to the previous chapter

The previous chapter presented a brief literature review that supports the advantages of systematic approaches to studying plant abiotic stresses using comparative genomics, machine learning, or systems biology tools in comparison of using traditional research approaches. In this order, this chapter presents a new method inspired by graph theory connecting the elements provided in the literature review, and information obtained in plant species models such as *Arabidopsis thaliana* to obtain specific divergence insights in rice. In consequence, this chapter is related to the specific objective one: To elucidate biological processes associated with aluminum stress responses through comparative genomics strategies to understand aluminum effects over plants from a functional perspective.

3.2. Abstract

Functional enrichment analysis is a cornerstone in bioinformatics as it makes it possible to identify functional information by using a gene list as the source. Different tools are available to compare gene ontology (GO) terms, based on a directed acyclic graph structure or content-based algorithms which are time-consuming and require *a priori* information of GO terms. Nevertheless, quantitative procedures to compare GO terms among gene lists and species are not available. In this chapter a computational procedure is presented, implemented in R, to infer functional information derived from comparative strategies named GOCompare. The package provides a framework for functional comparative genomics starting from comparable lists from GO terms. The program uses functional enrichment analysis (FEA) results and implements graph theory to identify statistically relevant GO terms for both GO categories and analyzed species. Thus,

GOCompare allows finding new functional information complementing current FEA approaches and extending their use to a comparative perspective. It could be allowed to establish hypothesis that might explain a biological process from a functional perspective and narrowing down the possible landscapes to design wet lab experiments. To test this approach, GO terms were obtained for a list of aluminum tolerance-associated genes in *Oryza sativa* subsp. *japonica* and their orthologs in *Arabidopsis thaliana* to detect possible divergences between functional categories (biological processes). Also, a list of possible main driver genes is provided as nodes with high degree values in undirected graphs. GOCompare was able to detect functional similarities for reactive oxygen species and ion binding capabilities which are common in plants as molecular mechanisms to tolerate aluminum toxicity. Consequently, divergences in energy availability modulation. In addition, it is suggested different proteins as key regulators of the aluminum response between species where signaling is important for both species.

3.3. Introduction

Functional enrichment analysis or gene enrichment tools are widely used to identify functional information associated with a list of genes of interest. This type of analysis is based on the existence of gene ontology (GO). GO provides a standard and controlled vocabulary containing information about (i) molecular functions (MF) (ii) biological processes (BP), and (iii) cellular compartments (CC), available for annotated genes in various organisms (Ashburner et al., 2000; Thomas, Paul, D., 2017; The Gene Ontology Consortium, 2019). To derive adequate biological knowledge of GO structure to obtain enriched lists of GO terms, the use of hypothesis tests, such as Fisher's exact test, hypergeometric distribution, or chi-squared, combined with the support of multiple hypothesis correction methods is necessary to decrease noise in the results (Huang et al., 2009b).

As mentioned in the introductory chapter, currently, there are many available tools to obtain information associated with functional enrichment approaches including BinGO (Maere et al., 2005), DAVID (Huang et al., 2007), AmiGO (Carbon et al., 2009), AgriGO (Tian et al., 2017), ShinnyGO (Ge et al., 2020), gProfiler (Kolberg et al., 2020), topGO (Alexa et al., 2006) among others. These are popular resources used to obtain a functional perspective of a given gene list(s). However, enrichment analysis has been considered as an exploratory procedure rather than a statistical solution itself (Huang et al., 2009b; Tipney and Hunter, 2010).

Although there are many available tools for enrichment analysis, the way GO terms are compared is based on semantics or meaning between annotations (Pesquita, Catia, 2017). Thus, the most common approach to compare GO terms is the use of GO-based semantic similarity (SS). It is based on information content (IC), or hierarchical relationships between GO terms structured in a directed acyclic graph (Yu et al., 2010; Mazandu et al., 2016). For semantic similarity analysis, a directed acyclic graph of the ontology terms is required, which increases the computational cost of these analyses, therefore, new hybrid approaches such as the one implemented in GOGO have emerged in order to gain better computational performances (Zhao and Wang, 2018).

Even when SS is a useful approach to obtain a numerical value of the similarity between two GO terms, being successful over the years, there are several problems related to these tools, including but not limited to: non-uniqueness and inconsistencies in information content, unbalanced annotations in GO graph structure, GO term redundancy, term specificity, term similarity, and gene product similarity. These factors hinder the generation of comparative approaches among GO terms (Mazandu et al., 2016; Pesquita, Catia, 2017). Moreover, it is well known that there are pitfalls associated with GO terms such as data aggregation, gene ontology variability over time due to changes in genome annotations, as well as genome annotation incompleteness (Cai et al., 2006).

Some software such as the R package clusterprofiler (Wu et al., 2021), Gorilla (Eden et al., 2009) or goProfiles (Salicrú et al., 2011) provide visual tools to compare functional profiles for several conditions. For example, comparing conditions with *p value*-fit values and gene ratios, or functional profiles based on GO terms frequencies. Nevertheless, none of the available tools compare multiple gene lists through an analytical framework to observe possible patterns based only on the obtained enriched GO terms, which are important for the comparison of gene lists involved in plant abiotic stress as aluminum, as a starting point to describe similar mechanisms across taxa. Thus, in the present chapter, a new method to compare functional gene lists in two species is described and used to obtain insights into functional divergences of aluminum responses in rice and *A. thaliana*.

3.4. Material and methods

3.4.1. A new methodology for functional comparison between species and gene lists (GOCompare)

The package GOCompare was written as an R package in R v4.0 and is available in CRAN. GOCompare was intended as a simpler but robust approach to compare multiple GO terms

obtained for several gene lists (categories) in one or two species through the use of six functions (see Table 3.1). This pipeline is based on the following assumptions: (i) GO terms are present with other GO terms among gene lists (categories); (ii) important GO terms are more frequent among categories and will have greater node weights according to their belonging to a category; and (iii), the presence or absence of a GO term can serve to explore how similar two species and categories are.

Table 3.1. Name and description of the six functions offered in the R package GOCompare. A table with three columns: Function, focus of study, description is provided. The focus of study reflects if the function was designed for one or two species respectively.

Function	Focus of study	Description
mostFrequentGOs	One species	Obtains the most frequent biological processes by category
graphGOspecies	One species	Converts a list of biological processes and categories to networks
compareGOspecies	Two species	Comparison of two lists of biological processes and categories using PCoA and Jaccard distances
evaluateCAT_species	Two species	Evaluation of the importance of functional category for n GO terms using either proportion tests or Chi-square tests.
evaluateGO_species	Two species	Evaluation of the importance of a GO term for n categories using either proportion tests or Chi-square tests
graph_two_GOspecies	Two species	Converting a list of biological processes and categories in two species to a network

Regarding the six implemented functions (see Figure 3.1 for a scheme of how to perform an analysis in GOCompare), the first two (*mostFrequentGOs* and *graphGOspecies*) were designed to compare GO terms in several categories in one species using as input a data frame of GO terms enriched in defined categories (e.g. Pancreatic Cancer in humans, abiotic stress tolerance in plants). The other four functions (*compareGOspecies*, *evaluateGO_species*, *evaluateCAT_species*, and *graph_two_GOspecies*) were designed to compare GO terms in several categories for two species using two input data frames. They were also designed to compare gene lists and their orthologues in other species.

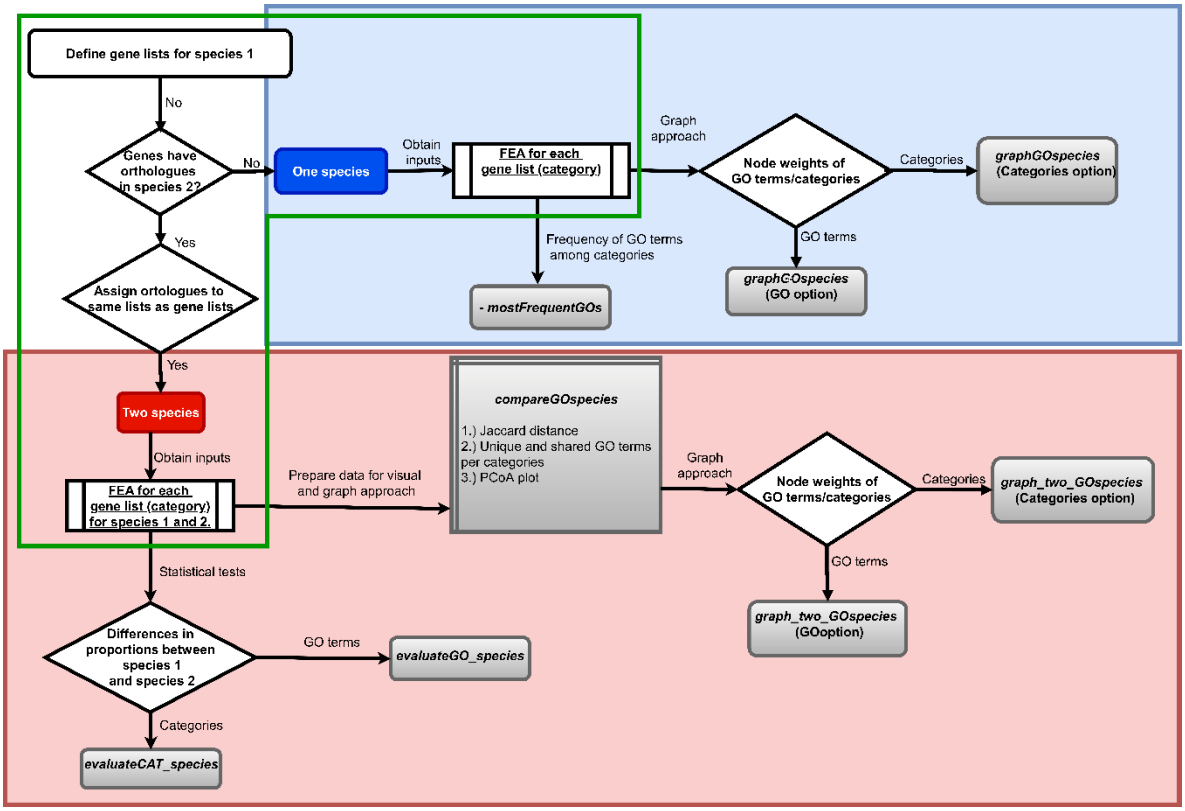


Figure 3.1. Schematic representation of the functions of the R package GOCompare and how to perform an analysis in GOCompare. Functional workflow to perform analysis in GOCompare. Inputs are defined as data frames. Preprocessing step including the functional enrichment analysis (underlined) is indicated in green color. In addition, a table with recommendations for this step is provided in Supplementary Table 3.7). The six functions available in GOCompare are shown in italic font and inside rounded rectangles in gray color. Workflows designed for one species and two species are in blue and red colors respectively.

In the R package, the functions *mostFrequentGOs* and *evaluateGO_species*, obtain the most frequent GO terms in the gene lists (categories). They also perform either the z-proportion test (test="prop") or the chi-squared test (test="chi-squared"), combined with a false discovery rate correction, with the categories and GO terms as the exploratory procedure for the analyses. Hereafter, in terms of multivariate statistical approaches, GOCompare uses Jaccard distances (presence-absence of a GO term) and a principal coordinates analysis (PCoA). These are implemented in the *compareGOspecies* function to provide a visual comparison tool between categories (gene lists), and species.

For visualization and analytical tasks, individuals are considered as the combination of a category and the species. For example, if the user has two categories with GO terms related for two

species, four individuals will be compared: Species1-Category1, Species1-Category2, Species2-Category1, and Species2-Category2. As consequence, the PcoA uses the Jaccard distance which is stored in an object list, and a convex hull for categories of the same species, to easily observe how similar the categories are among them.

In addition, among the functions implemented in the package, the functions *graphGOspecies* and *graph_two_GOspecies*, create two classes of undirected weighted graphs for one and two species, respectively. In both functions, it is possible to use either "Categories" or "GO terms" as nodes. Both functions use CPU-parallelized computing in order to obtain faster results. Therefore, these graphs help to visualize how many GO terms are co-occurring among categories or species. They also help to prioritize which of them are more important, using node and edge weights. The description of how weights are calculated is given below in sections 2.2 and 2.3.

3.4.1.1. Inputs:

The package uses data frames as inputs and each data frame must have two columns to be used: (i) “feature” representing a GO term belonging to a gene list (category) (e.g., genome instability mutation) and (ii) a user-named column with the ID of the GO term, or the name of the GO term to be used for analysis (e.g., "Functional. Category"). These two fields will be used to implement the six functions in the package. In addition, when using the functions for two species, the species names must also be added as characters to be used in the analysis (e.g., "*A. thaliana*"). It is strongly recommended the use data frames with only these two columns to develop the analyses.

In the case of functions related to the comparison of GO terms in two species, users must first run the compareGOspecies function to obtain a list which is the input required for the analysis of the *graph_two_GOspecies* function. The *compareGOspecies* function will provide a list that has four slots: (i) a principal coordinates analysis (PCoA) plotted in ggplot2; (ii) a Jaccard distance matrix used to perform the PCoA; (iii) a list of co-occurring GO terms between species and categories; and (iv) a list of unique GO terms belonging to the respective species and categories. In addition, users can import the functional enrichment results for both species and categories to run the functions: *evaluateGO_species* and *evaluateCAT_species*. These two functions perform a z-proportion test to detect GO terms or categories that are different between two species.

3.4.1.2. Methodological description of the *graphGOspecies* function (undirected-weighted graphs for a species)

To understand a gene list in terms of a functional perspective, two options are provided: (i) the results of the functional enrichment analysis are used as input to create undirected weighted graphs in which the categories are nodes, and the GO terms belonging to the categories are

edges. In this case, this graph allows to calculate the similarity of several categories (e.g., contrast conditions in a transcriptomic experiment). (ii) As a second option, GO terms can be represented as nodes that are linked by edges when different categories share the same GO term. In this case, the graph is useful to prioritize which GO terms are more relevant (more frequent) in explaining biological functionality, and how these GO terms are co-occurring in several categories, as mentioned above.

3.4.1.3. Using Categories to calculate nodes and weights

Consider an undirected graph $G = (V, E)$. Nodes (V) represent groups of gene lists (categories), and edges (E) represent co-occurring GO terms between pairs of categories. More specifically, two categories $u, v \in V$ are connected by an edge $e = (u, v)$ if there is at least one GO term associated to both categories. Take as an example the non-empty set of GO terms of biological processes ($BP \neq \emptyset$) and let $BPu \subset BP$ be the subset of biological processes belonging to category u. Then the edge weights $w(e)$ are defined as the ratio of the number of biological processes co-occurring between two categories ($BPu \cap BPv$) compared to the total number of biological processes available (Equation 3.1). In this case, node weights reflect which categories (e.g., gene lists of contrasting conditions) have more similar common features such as biological processes, molecular functions or cellular compartments, etc. Furthermore, a node weight $K_w(u)$ is defined as the sum of the edge weights where the node u is a participant (Equation 3.2**)**. Thus, the node weight represents how frequently GO terms are reported and expressed in a biological phenomenon or in the context of a series of experiments.

$$w(e) = \frac{|BPu \cap BPv|}{|BP|} \qquad \textbf{(Equation 3.1)}$$

$$K_w(u) = \sum_{v \in V} w(u, v) \qquad \textbf{(Equation 3.2)}$$

3.4.1.4. GO terms as nodes and weights calculation

As another representation, consider an undirected graph $G' = (V', E')$. The nodes (V') represent GO terms and the edges (E') represent categories sharing pairs of GO terms. More specifically, two GO terms $u', v' \in V'$ are connected by an edge $e' = (u', v')$ if there is at least one category with both GO terms associated. Let C be the set of all considered categories and $Cu' \subset C$ be the subset of categories to which the GO term u' belongs. Again, let us take the set of GO terms of BP biological processes as an example. Then, the edge weight $w'(e')$ corresponds to the number

of categories sharing the GO terms u' and v', compared to the total number of GO terms in the ontology of biological processes (Equation 3.3). Similarly to the setting in section 3.4.1.3, a node weight ${K'}_w(u')$ is defined (Equation 3.4), but in this case representing the importance of a GO term in terms of frequency.

$$w'(e') = \frac{|Cu' \cap Cv'|}{|BP|} \qquad \textbf{(Equation 3.3)}$$

$${K'}_w(u') = \sum_{v' \in V'} w'(u', v') \qquad \textbf{(Equation 3.4)}$$

3.4.1.5. Methodological description of the *graph_two_GOspecies* function (undirected-weighted graphs for two species)

The *graph_two_GOspecies* is an analog of the *graphGOspecies* function and has the same options (" Categories " and " GO "). Nevertheless, the way in which the edge and node weights are calculated is slightly different. Since two species are compared, three possible graphs G_1, G_2 and G_3 are available. Graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ represent each of the analyzed species. The third graph $G_3 = (V_3, E_3)$ is a subgraph of G_1 and G_2, which contains the GO terms or categories co-occurring between both species ($V_3 = V_1 \cap V_2$). For the three cases, the edge weights are calculated with equation or for the Categories or GO option, respectively. Nevertheless, node weight $\widehat{K}_w(u)$ for G_3 is calculated as the sum of the weights in G_1 and G_2 for node $u \in V_3$. On one hand, when the “Categories” option is considered, the weights of the nodes provide a measure of the similarity of the categories with respect to the GO terms. On the other, when the “GO” option is considered, the weight of the nodes provides a measure of how a GO term is conserved between two species in a series of categories.

$$\widehat{K}_w(u) = \sum_{v \in V_1} w(u, v) + \sum_{x \in V_2} w(u, x) \qquad \textbf{(Equation 3.5)}$$

3.4.1.6. Methodological description for the combined weight calculation

The weight proposed in (Equation 3.5) can be biased to show high values with categories with highly frequent GO terms co-occurring. Thus, a more impartial node weight was introduced, it was named “combined weight” and symbolized as K_c. The combined weight is based on two complementary measures: (i) the node weight defined above as $\widehat{K}_w$ and (ii) a shared weight K_s computed from a shared weight of edges s. Let N_1 and N_2 be the set of GO terms associated with the edge $e = (u, v)$ for species 1 and 2, respectively. Then, the shared edge weight s is defined

as in (Equation 3.6). The corresponding node shared weight K_s is computed similarly to the other defined node weights, as shown in (Equation 3.7). In addition, normalization min-max was introduced, and calculated according to (Equation 3.8), to apply it to the two measurements that make up the combined weight K_c. More specifically, the combined node weight K_c is defined as the sum of the min-max normalized weights $\widehat{K}_w$ and K_s minus 1 (Equation 3.9). The idea of this combined weight is to find categories with the most frequent GO terms co-ocurring in order to observe functional similarities between two species with a balance of GO terms that coexist between the gene lists (categories) and the two analyzed species. This node weight varies from -1 (categories with GO terms found only in one species and few categories) to 1 (categories with GO terms co-occurring widely between species and among other categories).

$$s(e) = \frac{|N_1 \cap N_2|}{|N_1 \cup N_2|} \quad \textbf{(Equation 3.6)}$$

$$K_s(u) = \sum_{v \in (V_1 \cup V_2)} s(u, v) \quad \textbf{(Equation 3.7)}$$

$$\text{min\_max}(y) = \frac{y - \min(y)}{\max(y) - \min(y)} \quad \textbf{(Equation 3.8)}$$

$$K_c(u) = \text{min\_max}\left(\widehat{K}_w\,(u)\right) + \text{min\_max}(K_s(u)) - 1 \quad \textbf{(Equation 3.9)}$$

3.4.2. Testing GOCompare through aluminum tolerance associated genes in *Oryza sativa* subsp. *japonica* and *Arabidopsis thaliana*

As a case of study, a gene for the GOCompare approach a gene list associated with aluminum tolerance in *Oryza sativa* subsp. *japonica* and its counterpart, the respective orthologues in *Arabidopsis thaliana.* This approach allows observing functional similarities regarding aluminum tolerance mechanisms for biological processes (GO:BP) and molecular functions (GO:MF) in plants. The comparison was performed using the following workflow described below:

- A total of 1084 genes associated with aluminum tolerance *Oryza sativa* subsp. *japonica* were extracted from peer-reviewed articles. This datasets was derived from 13 peer-reviewed articles which used proteomics, transcriptomics, GWAS, QTL, and SSR related-techniques (Fukuda et al., 2007; Famoso et al., 2011; Arenhart et al., 2013, 2014; Wang et al., 2013, 2014; Yang et al., 2013; Zhang et al., 2016, 2019a, 2019b; Arbelaez et al., 2017; Cohen and Leach, 2019; Tyagi et al., 2020). To obtain the genes the following steps were performed: (i) obtain each gene from PDF files using https://extracttable.com/ service

(ii) gene Ids were standardized by obtaining their respective MSU and RAP-DB ID using the R riceidconverter package (iii) Genes with GI accession were transformed into Entrez ID using eFetch service of NCBI (iv) gene ids with GI accession were converted to MSU and RAP-DB IDs as well to obtain MSU and RAP-DB IDs. The dataset and details about references are available at https://data.mendeley.com/datasets/fzxjwrt53m/3 .

- Two filtering steps were considered to ensure the reproducibility of the comparative analysis between *O. sativa* and *A. thaliana*: (i) Only *O. sativa* genes with orthologs in *A. thaliana* were retained for downstream analyses. Orthologous genes for *Arabidopsis thaliana* were obtained using the *gorth* function of the gprofiler2 R package (Kolberg et al., 2020) (ii). Biological_process (GO: 0008150) or molecular_function (GO:0003674) were omitted to avoid the use of the more general GO terms in the ontology.

- For *Oryza sativa,* a functional enrichment analysis for the filtered genes obtained in the previous step was performed, by using the g:SCS multiple test correction method, contained in the gprofiler2 R package (Raudvere et al., 2019; Kolberg et al., 2020). Functional GO-categories based on the BP-ontology were obtained. The background for this first step was the available complete annotated rice genome. To assign genes to functional groups, the following criteria were taken into consideration: (i) Only GO terms with adjusted p values <0.05 were considered as primary functional categories. (ii) To simplify the number of functional categories to be used for the undirected weighted graphs construction, the REVIGO web tool was used with the following parameters: *SimRel* GO was selected as semantics measurement, obsolete GO terms were removed using a suggested mild removing index of 0.7, and *Oryza sativa* belonging to the japonica group were defined as target species. REVIGO prioritizes functional GO terms using their *p*-values and GO term semantics measurements to remove redundant and obsolete terms in the enriched lists (Supek et al., 2011).

- The same functional categories were used for both species. Thus, if a gene O1 was found in the functional category A or B for *Oryza sativa,* its orthologs A1A, and A1B were assigned to the same categories A and B but in *Arabidopsis thaliana.* This was done to functionally compare both species under the assumption that similar genes have similar functions under the orthology conjecture (Altenhoff et al., 2012).

- Functional enrichment analyses (FEA) were performed for each of the functional categories in the two species to obtain lists of enriched molecular functions. As background for all FEA in Oryza sativa was used the list of unique genes extracted from the filtered genes. The same procedure was performed for *A. thaliana* with the unique orthologue genes. Only molecular functions with adjusted p values <0.05 after applying g:SCS multiple test correction method were considered further. Also, a category named UNKNOWN was created to add all genes without any membership to any biological processes and compare the GO terms related to those genes between species and observe possible closeness to other biological processes. These analyses were performed in the gprofiler2 R package.

- To establish common molecular functionalities, hierarchical clustering was performed with the minimum variance method (Ward). Additionally, to detect the optimal number of clusters, the maximum value of the Silhouette coefficient was calculated. This analysis was performed with the immunarchR package (ImmunoMind Team, 2019).

- Undirected graphs were plotted through the igraph R package and layout in circle option. Only node labels whose weights were greater than quantile three were displayed to simplify the visualization. In addition, visNetwork was used to create interactive versions of the GOCompare graphs.

3.4.3. Identification of the possible driving genes for aluminum tolerance in *O. sativa* subsp. *japonica* and *A. thaliana*

Undirected graphs were created for each species according to the gene membership in the previously assigned functional categories found in *Oryza sativa* subsp. *japonica*, and *Arabidopsis thaliana*. To build the gene networks per species, the genes were considered as nodes and as edges were considered as the biological processes where two genes were found. This process was done by using combinatorics. For each of the biological processes, combinatorics were performed to obtain edges lists that were joined to obtain an undirected graph per species and export to Cytoscape using igraph R package. As possible main driver genes for aluminum tolerance per species were considered the genes with degree values greater than percentile 90^{th}. Graphs were made in R v 4.0.2 using the combn function. The node, degree and topological measures were inferred by Cytoscape 3.9.2 (Shannon et al., 2003). All R codes used here are available at https://github.com/ccsosa/Rice_Ara_GO_comparison.

3.5. Results

To validate the power of the designed package A list of 1084 aluminum tolerance related-genes was used. After dataset filtering, a list of unique genes from *Oryza sativa* subsp. *japonica* (OSJ) genes were retained. With this final list, a total of 985 orthologues in *A. thaliana* (ARATH) were obtained.

A total of 27 functional categories were found for the OSJ genes, including a category labeled as unknown to sink all genes without any associated biological process. From the original list of 27 biological processes (hereafter referred as functional categories), only a total of 17 (59.25%) functional categories in OSJ showed enriched molecular functions (FDR <0.05). Whilst, all functional categories in ARATH (100%) exhibited enriched molecular functions.

Given that no differences in proportion tests (false discovery rate <0.05) for functional categories or molecular functions were found using the *evaluateCAT_species*, or the *evaluateGO_species* functions, It was applied and integrate the *mostFrequentGOs*, *graphGOspecies*, *compareGOspecies* (designed for one species), and, *graph_two_GOspecies* functions, in order to detect functional associations that might shed light in the understanding of aluminum responses in plants.

Regarding the Molecular Function (MF) ontology, It was applied the *mostFrequentGOs* function to obtain the most frequent categories for each analyzed species. A list of 17 categories with five of them (metal ion binding, cation binding, catalytic activity, oxidoreductase activity, and antioxidant activity) being associated with oxidoreductase, peroxidase, and antioxidant activity, were retrieved for OSJ. Likewise, for ARATH, a list of 83 categories with 27 of them (upper quartile), were found. Here, the kinase activity function complemented the same molecular processes obtained for OSJ (Supplementary Tables 3.1 and 3.2).

3.5.1. *graphGOspecies function was able to detect similarities in reactive oxygen species (ROS) and binding molecules for Oryza sativa and Arabidopsis thaliana*

OSJ functional categories node weights, ranged from 0.118 to 3.059 (Table 3. ***2***).
Complementarily, node weights for ARATH, ranged from 0 to 0.997 (

Table **3.*3***). In this regard, 29.6% of OSJ nodes were above the median. In contrast, 48.14% of ARATH nodes weighted above its median, suggesting more molecular functions available for ARATH. In Figure 3.2, the functional divergence between the analyzed species is shown, linked with the way how topology in the derived networks is changed. OSJ functional network is

subdivided into three subgraphs related to gibberellin, glucose/monosaccharide, and different metabolic processes (Figure 3.2A). In contrast, ARATH subgraphs are related to gibberellins and unknown processes. Its most connected component is associated with stress responses (Figure 3.2B).

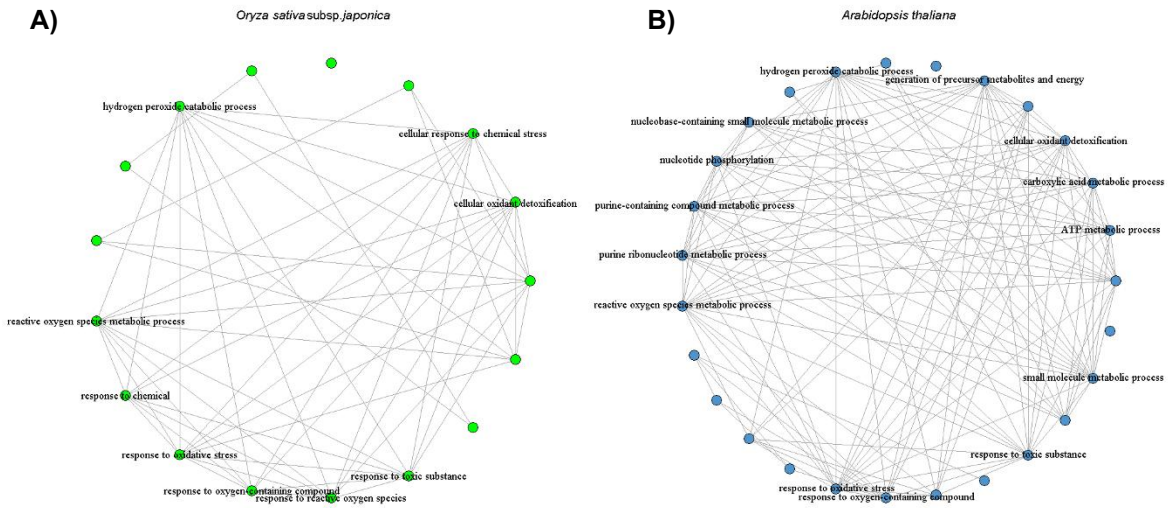


Figure 3.2. Undirected weighted graph representations for 27 functional categories using graphGOspecies function. The nodes and edges represent functional categories and enriched molecular functions respectively. Both species show three subgraphs A) *Oryza sativa* subsp. *japonica* (green) B) *Arabidopsis thaliana* (blue). Only node labels with weight values greater than third quartile are displayed in the graphs. An interactive view is available at https://ccsosa.github.io/GOCompare_interactive_graphs/Functional_cat_individual_graph.html

Table 3. 2. Functional categories descriptions obtained by the *graphGOspecies* for *Oryza sativa* (OSJ). Molecular functions counts, node weights, and degrees are provided for 27 functional categories obtained for the case of study in *Oryza sativa*. Values above medians for molecular functions counts and node weights are underlined.

Functional category	Molecular functions count	Node weight	Degree
hydrogen peroxide catabolic process	9	3.059	10
reactive oxygen species metabolic process	9	3.059	10
cellular oxidant detoxification	8	2.883	10
response to oxidative stress	8	2.883	10
response to toxic substance	6	2.47	9
response to oxygen-containing compound	6	2.057	8
cellular response to chemical stress	5	1.998	9
response to reactive oxygen species	5	1.998	9

Functional category	Molecular functions count	Node weight	Degree
response to chemical	4	1.762	8
carboxylic acid metabolic process	6	0.825	10
small molecule metabolic process	6	0.648	7
generation of precursor metabolites and energy	1	0.177	3
nucleobase-containing small molecule metabolic process	1	0.177	3
glucose metabolic process	1	0.118	2
monosaccharide metabolic process	1	0.118	2
small molecule catabolic process	1	0.118	2
gibberellin metabolic process	1	0	0
aerobic respiration	-	-	-
ATP metabolic process	-	-	-
nucleotide phosphorylation	-	-	-
purine ribonucleotide metabolic process	-	-	-
purine-containing compound metabolic process	-	-	-
response to abiotic stimulus	-	-	-
response to acid chemical	-	-	-
response to inorganic substance	-	-	-
response to temperature stimulus	-	-	-
UNKNOWN	-	-	-

In order to detect similarities among functional categories for both species, an implemented Spearman correlation coefficient was performed with the node weights available in tables 3.2 and 3.3 with values of ρ=0.534, and p =0. 03, it showed that the method was able to detect a moderate correlation between OSJ and ARATH functional categories in terms of molecular functions co-occurrence separately. Graph per species indicates high degree values for four functional categories (upper quartile) in OSJ, and six categories in ARATH (upper quartile). Most of these functional classes were associated with reactive oxygen species production, thus, it is suggested an important role of oxidant activity for OSJ and ARATH aluminum tolerance.

To obtain a description of more frequently molecular functions co-occurring in categories, the *graphGOspecies-GO* option, created an undirected weighted graph of molecular functions as nodes and functional categories as edges. A total of 17 molecular functions were enriched in OSJ and 83 in ARATH. Node weights greater or equal to the upper quartile (2.52) in OSJ were associated with oxidoreductase and antioxidant activity. In contrast, node weights greater or equal to the upper quartile (0.86) in ARATH were associated with binding and phosphofructokinase

activity (Supplementary Table 3.3). Visual graph representation of this network suggests higher connectivity regarding molecular functions for ARATH, with a subgraph related to glucosyltransferases and a graph with more binding functions than OSJ (Figure 3.3B). OSJ exhibited edges linked to oxidoreductase and peroxidase activity, and binding functions (Figure 3.3A). Thus, a possible divergence in molecular function annotations was observed between OSJ and ARATH when the developed method is inquired for species separately.

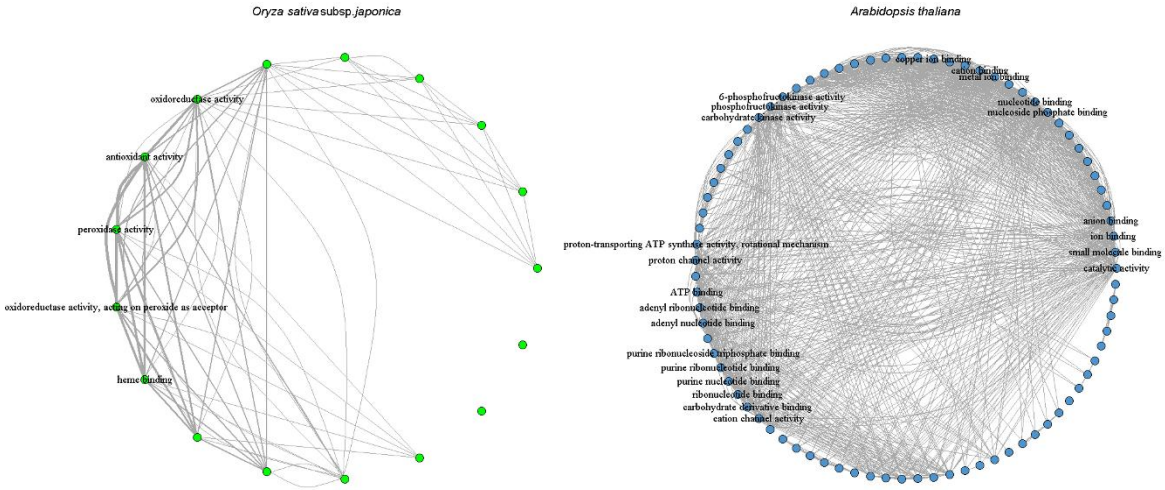


Figure 3.3. Undirected weighted graph representations for enriched molecular functions using *graphGOspecies* function. The nodes and edges represent enriched molecular functions and functional categories respectively. Both species A) *Oryza sativa* subsp*. japonica* (green) B) *Arabidopsis thaliana* (blue). Only node labels with weight values greater than third quartile 3 are displayed. An interactive view is available at https://ccsosa.github.io/GOCompare_interactive_graphs/GO_individual_graph.html.

Table 3.3. Functional categories descriptions obtained by the *graphGOspecies* for *Arabidopsis thaliana* (ARATH). Molecular function counts, node weights, and degrees are provided for 27 functional categories obtained for the case of study in *Arabidopsis thaliana*. The values above medians for molecular function counts and node weights are underlined.

Functional category	Molecular functions count	Node weight	Degree
purine-containing compound metabolic process*	25	1.215	14
purine ribonucleotide metabolic process*	24	1.179	14
small molecule metabolic process	27	0.997	16
carboxylic acid metabolic process	25	0.925	14
ATP metabolic process*	26	0.902	12
hydrogen peroxide catabolic process	9	0.852	16

Functional category	Molecular functions count	Node weight	Degree
reactive oxygen species metabolic process	9	0.852	16
response to oxidative stress	9	0.852	16
cellular oxidant detoxification	8	0.768	15
response to toxic substance	8	0.768	15
generation of precursor metabolites and energy	11	0.768	17
nucleobase-containing small molecule metabolic process	14	0.686	9
nucleotide phosphorylation*	9	0.612	13
response to oxygen-containing compound	8	0.576	12
aerobic respiration*	10	0.444	14
response to chemical	5	0.408	8
response to reactive oxygen species	4	0.36	9
cellular response to chemical stress	5	0.192	9
small molecule catabolic process	5	0.168	10
glucose metabolic process	6	0.096	3
monosaccharide metabolic process*	4	0.096	3
response to abiotic stimulus*	3	0.084	4
response to acid chemical*	2	0.072	3
response to inorganic substance*	2	0.072	3
response to temperature stimulus*	1	0.012	1
gibberellin metabolic process	1	0	0
UNKNOWN*	5	0	0

* Indicates functional categories without enriched molecular functions in *Oryza sativa*.

3.5.2. *graph_two_GOspecies function* identifies reactive oxygen species (ROS) and binding molecules as similar molecular functions in *Oryza sativa* and *Arabidopsis thaliana*

The *compareGOspecies* function was applied to find differences in functional categories for both species through the calculation of Jaccard distances based on the presence and absence of molecular functions associated with a particular biological process. Gibberellin metabolic process, reactive oxygen species, and cellular oxidant detoxification categories clustered together throughout the dendrogram suggesting conserved molecular functions for both species (Figure 3.4).

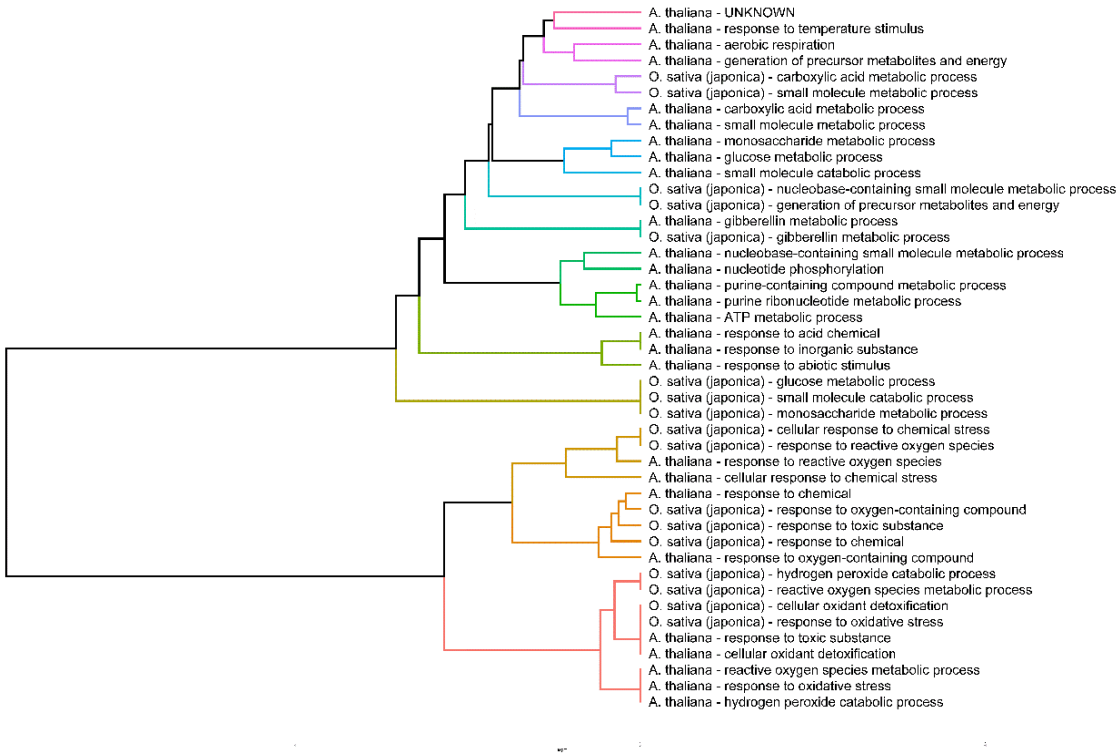


Figure 3.4. Hierarchical clustering dendrogram built with compareGOspecies function. Each dendrogram branch represents a functional category present in *Oryza sativa* or *Arabidopsis thaliana*. Colors represent 16 clusters suggested by inmunarchR R package using Jaccard distances based on the presence of specific molecular functions.

Interestingly, the hierarchical clustering grouped together the UNKNOWN category with different categories, associated with response to temperature stimulus, aerobic respiration, and generation of metabolic precursor in ARATH. The *graph_two_GOspecies* function showed that the combined node weight values were higher (>0) for functional categories linked to reactive oxygen species metabolism, response to chemicals, and stresses, corroborating the hypothesis of functional similarities between OSJ and ARATH regarding how plants respond to aluminum toxic conditions (Table 3.4).

Table 3. 4. Functional categories node weights result for *Oryza sativa* and *Arabidopsis thaliana* combined data using *graph_two_GOspecies* function. This table represents three weights calculated for functional categories. i.) the GO Term node weight (GO weight). This weight is designed to find categories with more co-occurring molecular functions among the functional categories. ii.) (Shared GO weight) that finds functional categories with more co-occurring molecular functions between ARATH and OSJ, iii) (Combined weight) that ranges from -1 to 1 that combines the two mentioned measurements

Functional category	GO weight	Shared GO weight	Combined weight	Interpretation
reactive oxygen species metabolic process	0.636	7.989	0.8303	Functional category with molecular functions highly similar between species
hydrogen peroxide catabolic process	0.636	7.989	0.8303	Functional category with molecular functions highly similar between species
response to oxidative stress	0.612	8.278	0.8261	Functional category with molecular functions highly similar between species
cellular oxidant detoxification	0.552	9	0.8157	Functional category with molecular functions highly similar between species
response to toxic substance	0.528	8	0.6683	Functional category with molecular functions highly similar between species
response to oxygen-containing compound	0.444	8	0.5415	Functional category with molecular functions highly similar between species
response to chemical	0.348	8	0.3964	Functional category with molecular functions highly similar between species
response to reactive oxygen species	0.36	7.6	0.3701	Functional category with molecular functions highly similar between species
cellular response to chemical stress	0.228	8.2	0.2374	Functional category with molecular functions highly similar between species
small molecule metabolic process	0.541	3.167	0.151	Functional category with molecular functions highly similar between species
purine-containing compound metabolic process	0.674	0	0	Functional category with enriched molecular functions only in Arabidopsis thaliana
purine ribonucleotide metabolic process	0.662	0	1088	Functional category with enriched molecular functions only in Arabidopsis thaliana
ATP metabolic process	0.602	0	-0.1088	Functional category with enriched molecular functions only in Arabidopsis thaliana
carboxylic acid metabolic process	0.481	0.634	-0.2211	Functional category with molecular functions co-occurring among categories but moderately co-occurring in both species
generation of precursor metabolites and energy	0.264	0.617	-0.5508	Functional category with molecular functions co-occurring among categories but moderately co-occurring in both species
nucleotide phosphorylation	0.288	0	-0.5831	Functional category with enriched molecular functions only in Arabidopsis thaliana
nucleobase-containing small molecule metabolic process	0.228	0.75	-0.5904	Functional category with molecular functions co-occurring among categories but moderately co-occurring in both species

Functional category	GO weight	Shared GO weight	Combined weight	Interpretation
aerobic respiration	0.192	0	-0.7281	Functional category with enriched molecular functions only in Arabidopsis thaliana
small molecule catabolic process	0.096	1	-0.762	Functional category with molecular functions co-occurring among categories but moderately co-occurring in both species
glucose metabolic process	0.108	0.75	-0.7717	Functional category with molecular functions co-occurring among categories but moderately co-occurring in both species
monosaccharide metabolic process	0.084	0.75	-0.8079	Functional category with molecular functions co-occurring among categories but moderately co-occurring in both species
gibberellin metabolic process	0.012	1	-0.8889	Functional category with molecular functions co-occurring among categories but moderately co-occurring in both species
response to abiotic stimulus	0.06	0	-0.9275	Functional category with enriched molecular functions only in Arabidopsis thaliana
UNKNOWN	0.06	0	-0.9275	Functional category with enriched molecular functions only in Arabidopsis thaliana
response to inorganic substance	0.048	0	-0.9456	Functional category with enriched molecular functions only in Arabidopsis thaliana
response to acid chemical	0.048	0	-0.9456	Functional category with enriched molecular functions only in Arabidopsis thaliana
response to temperature stimulus	0.012	0	-1	Functional category with enriched molecular functions only in Arabidopsis thaliana

Complementarily, the functional categories with lower values (<0) identified those functional categories found in only one species. Visually, the representation of the combined network for both species is shown in five subgraphs: the first one was related to the functional categories with more molecular functions co-occurring in both species, the second one was linked to purine and respiration processes, the third subgraph was related to response to abiotic, and chemical compounds categories, and the four and fifth subgraph were associated to gibberellin and unknown functional categories respectively. Subgraphs three to five have enriched molecular functions for *Arabidopsis thaliana* exclusively and this is reflected in negative values (red color) for the combined node weights (Figure 3.5).

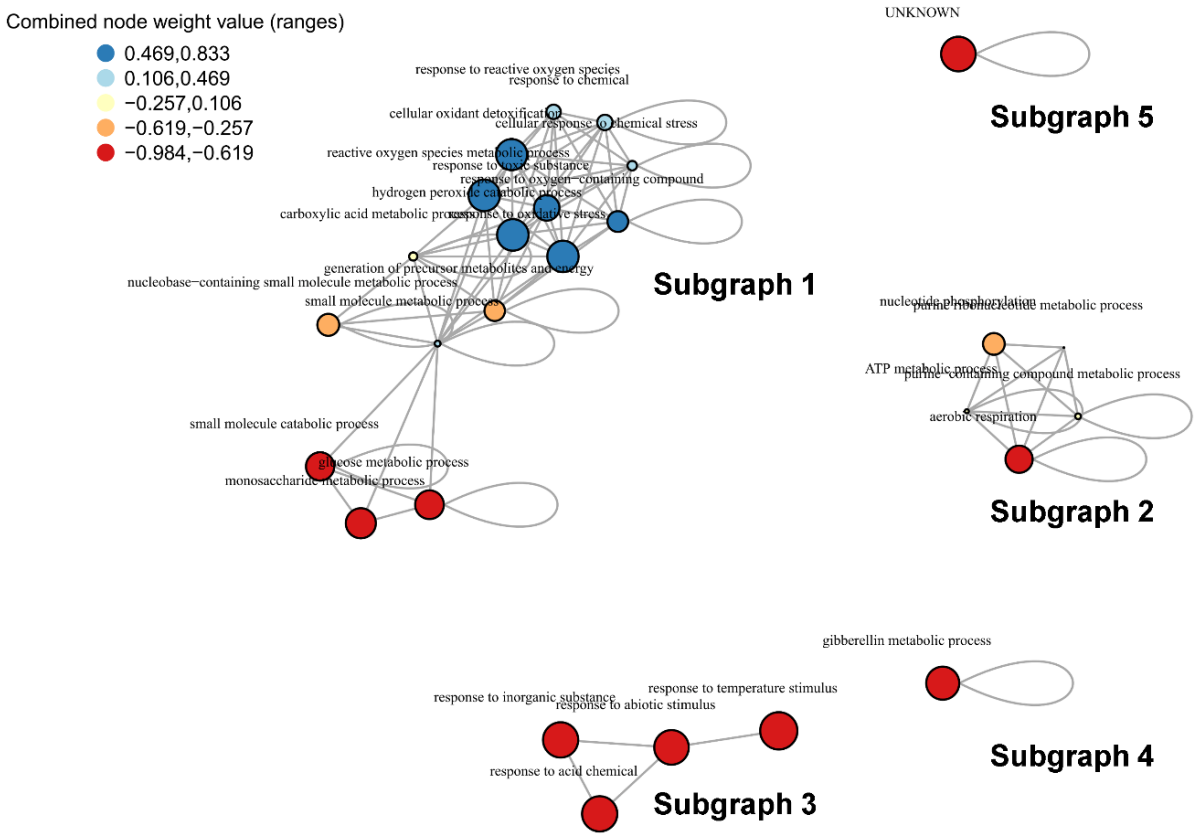


Figure 3.5. Undirected weighted graph representations for 27 functional categories using *graph_two_GOspecies* function. The nodes and edges represent functional categories and enriched GO terms for molecular functions respectively. This graph representation is structured by five subgraphs and is described as subgraphs 1 to 5 in the figure. Node sizes are proportional to the absolute value of the combined node weights. Colors intervals represent five combined node weight ranges. Red (negative values) indicate functional categories only found in Arabidopsis thaliana. Blue (positive values) represent functional categories with co-occurring molecular functions for both species simultaneously. An interactive view is available at https://ccsosa.github.io/GOCompare_interactive_graphs/Categories_graph.html.

To obtain a description of more frequently co-occurring molecular functions associated with biological processes, and to determine if these functions were similar among functional categories the function *graph_two_GOspecies* with the *GO* option was applied. A list of 21 molecular functions with high node weights values (upper quartile ≥1.48) were obtained and associated with molecule-binding activity such as cation or metal ion binding. Similarly, antioxidant, oxidoreductase and phosphofructokinase-related activities were highly frequent (Supplementary Table 3.4). These molecular functions were similar to the ones with high node weight values

individually calculated by the function *graphGOspecies* for rice and *Arabidopsis thaliana,* showing the coherence of the method proposed here (Figure 3.6).

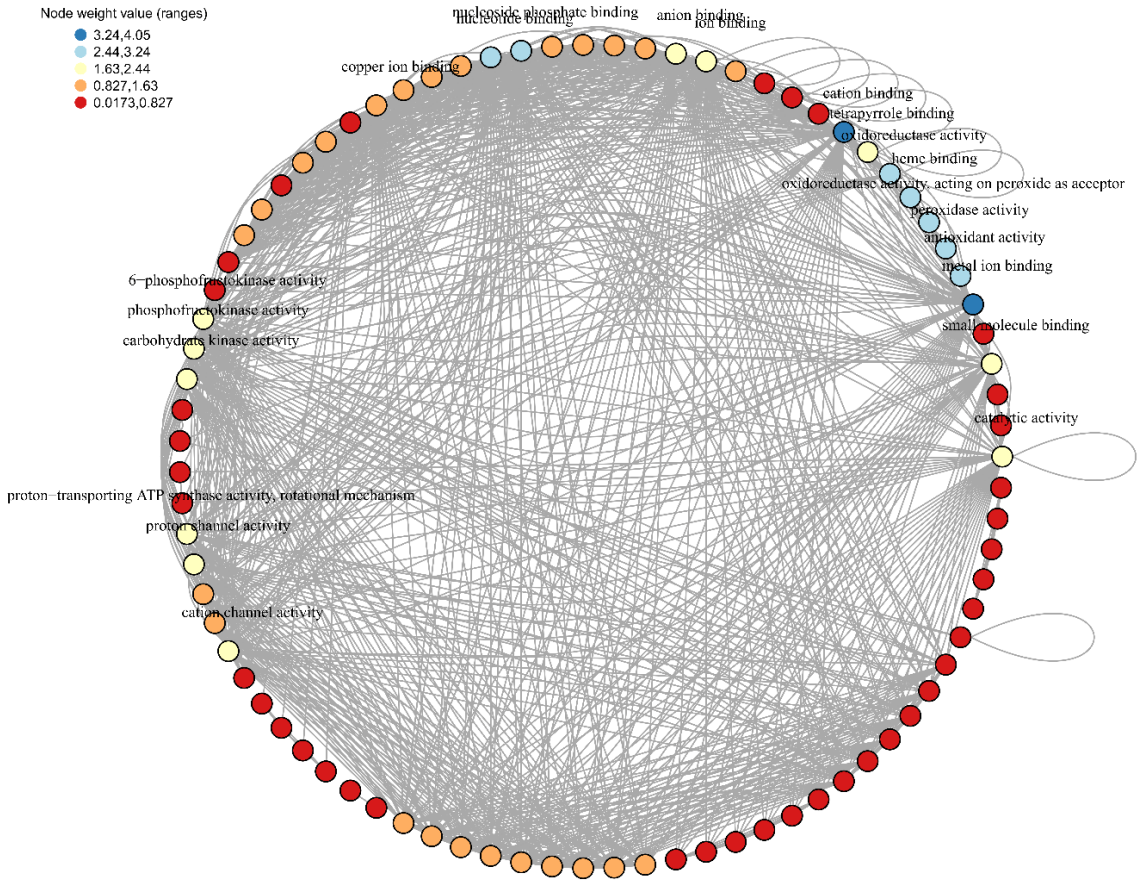


Figure 3.6. Undirected weighted graph representations for 27 functional categories using *graph_two_GOspecies* function. The nodes and edges represent molecular functions and functional categories respectively. This circle graph representation only displays labels for GO weight greater or equal to the upper quartile. Color intervals represent node weight ranges. Red (low values) indicate molecular functions only found in one species or only found in some particular functional categories. Blue (positive values) represent molecular functions co-occurring frequently for both species simultaneously. An interactive view is available at https://ccsosa.github.io/GOCompare_interactive_graphs/GO_graph.html.

3.5.3. A network perspective of possible regulatory genes affecting the aluminum tolerance in *Oryza sativa* subsp. *japonica* and *Arabidopsis thaliana*

An undirected graph with 507 genes as nodes, 140194 edges representing functional groups (GO:BP), two giant components, and an average number of connections of 305 was obtained by functional enrichment analysis for OSJ (Supplementary Figure 3.2). Admittedly, degree distribution showed an average degree of 333 with 283 genes with this value. Also, 51 genes had values greater than or equal to the 90th percentile (degree = 1182) (Supplementary Figure 3.1A).

A total list of genes with their respective degrees and if they are considered as main driver genes is available in the Supplementary Table 3.5.

Among genes with more than 5000 connections which represent values percentile 98th there were two genes (Os02g0169300 and Os06g0247500) that had the highest values and their description was Phosphoglycerate kinase. Os11g0148500 and Os12g0145700 were described as pyruvate kinases, Os01g0817700 and Os05g0482700 were described as 2,3-bisphosphoglycerate-independent phosphoglycerate mutase-like, and the remaining genes were described as Aldolase C-1, Glyceraldehyde-3-phosphate-like, Glucose-6-phosphate, and Glyceraldehyde-3-phosphate dehydrogenase. Therefore, the main role of central metabolism in rice is proposed to be relevant for aluminum tolerance by the analysis (Table 3. 5).

Table 3. 5. *Oryza sativa* subsp*. japonica* genes with degree values greater than 5000 connections (Percentile 90th). Gene description was obtained for RAP-DB project website https://rapdb.dna.affrc.go.jp/

OSJ gene ID	RAPDB Description	Degree
Os02g0169300	Similar to Phosphoglycerate kinase, cytosolic (EC 2.7.2.3)	5366
Os06g0247500	Pyrophosphate-fructose 6-phosphate 1-phosphotransferase (PFP) beta subunit	5052
Os11g0148500	Cytosolic pyruvate kinase	5052
Os12g0145700	Pyruvate kinase	5052
Os01g0817700	Similar to 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	5030
Os05g0482700	Similar to 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	5030
Os01g0905800	Aldolase C-1	5008
Os02g0601300	Similar to Glyceraldehyde-3-phosphate dehydrogenase	5008
Os04g0486600	Cytosolic glyceraldehyde-3-phosphate dehydrogenase 2	5008
Os06g0256500	Similar to Glucose-6-phosphate isomerase	5008
Os08g0126300	Similar to Glyceraldehyde-3-phosphate dehydrogenase	5008

The graph for *Arabidopsis thaliana* (ARATH) possessed 985 genes as nodes, 1309761 edges representing functional groups (GO:BP), one component, and an average number of connections of 483235 (Supplementary Figure 3.1B). Hence a greater number of connections was observed for ARATH genes than OSJ genes compared to OSJ genes. By contrast, the degree distribution showed an average degree of 2659.41 which, is bigger than the number of connections in OSJ (Supplementary Figure 3.1). In addition, 100 genes had values greater than or equal to the 90th percentile (degree = 6332). Moreover, 35 genes possessed 9498 connections (maximum degree value in the network). Also, the ARATH genes graph exhibited greater degree values than OSJ genes graphs as well as more biological processes associated. Thus, it suggests that the roles of the main driver genes for both species would be different. In addition, A total list of genes with their respective degrees is available in Supplementary Table 3.6.

The following genes were found with the maximum value degree (degree = 9498): AT3G57240 and AT3G57260 were described as beta 1,3-glucanase. AT1G07430 and AT5G59220 were described as phosphatases. While AT5G55090, AT4G26890, AT2G32510, and AT1G05100 were described as part of mitogen-activated protein kinases., Eight genes were described as part of P-loop containing nucleoside triphosphate hydrolases superfamily proteins. Also, eight genes were described as P450 cytochrome among other molecular functions. Consequently, these results suggest that aluminum tolerance in *Arabidopsis thaliana* may be associated with a combination of transporters, signaling pathways, and central metabolism, as shown in Table 3. ***6***.

Table 3. 6. Description of ARATH genes with maximum degree value (percentile 98th). Gene description was obtained from https://www.arabidopsis.org/.

Description	ARATH gene ID
beta 1,3-glucanase	AT3G57240//AT3G57260
Cytochrome P450, family 709, subfamily B, polypeptide 2;(source:Araport11)	AT2G46950
Encodes a member of glycosyl hydrolase family 17	AT3G57270
Encodes a member of the group A protein phosphatase 2C (PP2C) family that is responsible for negatively regulating seed dormancy.	AT1G07430
Encodes a member of the MATE (multidrug and toxin efflux family), expressed in roots but not shoots.	AT3G08040
Encodes a member of the PP2C family (Clade A protein phosphatases type 2C). Functions as a negative regulator of osmotic stress and ABA signaling.	AT5G59220
Encodes a mitochondrial ATPase involved in seed and silique development.	AT5G40010

Description	ARATH gene ID
Encodes a putative beta-1,3-endoglucanase that interacts with the 30C02 cyst nematode effector. May play a role in host defense.	AT4G16260
Encodes a root citrate transporter which together with the root malate transporter ALMT1 is the primary mechanism of aluminum tolerance.	AT1G51340
Encodes protein phosphatase 2C. Negative regulator of ABA signaling. Expressed in seeds during germination. mRNA up-regulated by drought and ABA.	AT3G11410
Highly ABA-induced PP2C protein 3	AT2G29380
MAPKKK15, mitogen-activated protein kinase kinase kinase 15	AT5G55090
MAPKKK16, mitogen-activated protein kinase kinase kinase 16	AT4G26890
MAPKKK17, mitogen-activated protein kinase kinase kinase 17	AT2G32510
MAPKKK18, mitogen-activated protein kinase kinase kinase 18	AT1G05100
Member of CYP709B	AT2G46960
Member of CYP709B The mRNA is cell-to-cell mobile.	AT4G27710
P-loop containing nucleoside triphosphate hydrolases superfamily protein;(source: Araport11)	AT3G28510//AT3G28520//AT3G28540//AT3G28570//AT3G28580//AT3G28600//AT3G28610//AT5G40000
Putative cytochrome P450	AT3G14610//AT3G14620//AT3G14630//AT3G14640//AT3G14680
¨Putative cytochrome P450 brassinosteroid binding	AT1G17060
Putative cytochrome P450 The mRNA is cell-to-cell mobile.	AT3G14650//AT3G14660//AT3G14690

Genes are split by //.

3.6. Discussion

A new set of tools is described here (GOCompare), it allows the identification of relevant biological processes and molecular functions that can be co-occurring or exclusively present when contrasting GO categories. The R package works as a complement to functional enrichment analysis that makes it possible to explain biological phenomena using a straightforward pipeline whose inputs and requirements can be obtained easily by most R users, regardless of their expertise in the programming language.

Moreover, this methodology allows one to explore GO term analyses without introducing complex algorithms, detecting optimal hyperparameters, or using previous information obtained from gene ontology projects. It also allows the use of the functional enrichment analyses obtained by the

user's favorite tools. Thus, GO Compare offers a toolkit that can be complementary to current resources used by bioinformaticians in fields such as transcriptomics or comparative genomics (Thomas, Paul, D., 2017; The Gene Ontology Consortium, 2019).

The simplicity of the methodology in comparison with semantic similarity approaches (SS) is evident, given that GOCompare method only uses undirected weighted graphs without considering the hierarchical structure of the directed acyclic graphs, or information content measures such as SS. GOCompares can complement these SS analyses to provide graphic comparisons based on the use of presence-absence approaches such as Jaccard distance. The present approach could be similar to goProfiles, because its algorithm defines functional profiles based on annotations of GO terms (Salicrú et al., 2011) or comparable to Metascape web resource (Zhou et al., 2019) which permits comparative analysis of multiple independent gene lists.

The use of GOCompare functions based on graphs theory to explore functional enrichment analysis results, holds several advantages over the canonical ways of interpreting functional enrichment: i) it allows a better visualization of interactions among GO terms or functional categories. ii) it can detect differences in genome annotation, and iii) it gives a better numerical and visual interpretation of big datasets in one workflow.

Taking advantage of GO terms cooccurrence, and implementing graphical and numerical outputs, GOCompare allows a straightforward interpretation of functional enrichment results. This is not present in tools such as BinGO that provides a directed graph based on GO-hierarchy and *p* values based node size (Maere et al., 2005). Even when graphs or enrichment maps are obtained in other tools such as ShinnyGO, they are not extended to several gene lists. The benefits from these graphical and numerical outputs were shown in the case of study, where ROS and binding-related biological processes were highlighted as important functions for aluminum tolerance in plants, given the weights provided by the software.

The network output, derived from the *graphGOspecies* and *graph_two_GOspecies* functions, can evidence genome annotations differences. In the analyzed case of study, the number of molecular functions enriched for the 27 detected biological processes was unbalanced towards more abundant and available ARATH annotations. These differences are visible in graphs with more densely connected functional categories for *A. thaliana*, while OSJ networks were more disconnected. Similarly, with the implementation of the node combined weight, most of the ARATH

functional categories had negative values indicating a trend to be exclusively present in one species.

GOCompare provides output for users to graphically observe patterns in big datasets. These outputs can provide insights that statistical approaches or descriptive statistics cannot easily deliver.

In addition, GOCompare helps users to simplify enrichment analysis and extends functional characterization of biological phenomena to the field of comparative genomics. Additionally, GOCompare goes beyond the current available programs by: i) creating undirected-weighted graphs to interpret GO terms as node weights, ii) implementing Jaccard distances to identify functional similarities, and iii), providing statistical testing for specific GO categories composition.

3.6.1. Case of study: GOCompare identifies functional similarities between *Oryza sativa* and *A. thaliana* for ROS detoxification mechanisms associated with aluminum toxicity

A straightforward pipeline to functionally explain biological phenomena based on functional enrichment analysis was introduced. From this perspective, GOCompare was able to identify common functionalities and molecular mechanisms used by two evolutionarily distanced species in aluminum responses.

The different implemented functions were able to identify ROS detoxification mechanisms and metal transporters as key processes associated with aluminum responses, given: i) evaluateCAT_species and evaluateGO_species did not show any differences in proportion, and thus, it was not possible to identify differences in molecular functions or biological processes, ii) for each species, the mostFrequentGOs function always showed antioxidant and ion binding related molecular functions with high frequencies among categories, suggesting a possible similarity in these functions, iii) the graphGOspecies module, showed high node weight values for previously mentioned biological processes, and a moderate correlation among them was observed, supporting the hypothesis of similarities between functional enrichment results, iv) the compareGOspecies function-based dendogram, clearly presented a cluster of biological processes related to ROS signalling, or detoxification mechanisms grouped together for both species, v) the graph_two_GOspecies module provided evidence of functional categories with high combined node weight values for those biological processes. It means that the biological processes mentioned above are present in both species more frequently. Moreover, node weights related to oxidoreduction or antioxidant activity exhibited the same behavior as the associated

biological processes. Finally, the graphGOspecies function individually supports the hypothesis of functional similarities for ROS, ion transport, and ROS detoxification for both species.

The results presented here are in agreement with previously described results for aluminum-tolerant species, where exclusion mechanisms such as metal ion transport are key to maintaining aluminum out of the root apical meristem in plants (Bojórquez-Quintal et al., 2017).

Results are resembling the response to other abiotic stresses such as heat, cold, drought, and salinity, which comprise complex signaling pathways that start with a stimulus perception that can be driven by the increase of molecules such as Reactive Oxygen Species (ROS) (Das and Roychoudhury, 2014; Lamers et al., 2020). ROS can work as signaling molecules that can regulate plant growth and cellular division among other biological processes (Ranjan et al., 2021b). Thus, the presence of ROS and antioxidant-related functional categories is in agreement with reported processes in *A. Thaliana,* rice, and higher plants, where the presence of cations such as iron (Fe) or (Al) can increase ROS and antioxidant molecules (Bojórquez-Quintal et al., 2017; Daspute et al., 2017; Awasthi et al., 2019).

In contrast, GOcompare was able to detect processes that are preferentially exhibited for either *Arabidopsis thaliana* or rice. These differences are of high interest given that they represent molecular alternatives to face aluminum tolerance that have diverged during evolution. One important fact about aluminum tolerance is that the main molecular mechanism used by plants is exclusion. This mechanism uses metal ion transports to handle trivalent elements such as aluminum (Al), chromium (Cr), arsenic (As), and antimony (Sb). Transporters allow the exclusion of these aluminum ions from the root and, in combination with chelating molecules such as organic acids (OA), or pectins, operate to detoxify plants from heavy metals (Liu et al., 2014; Bojórquez-Quintal et al., 2017; Poschenrieder et al., 2019; Kar et al., 2021).

Organic acids are produced from the tricarboxylic acid cycle (TCA). TCA has overrepresented enriched biological processes in *A. thaliana* but not in *O.sativa.* This could be explained given that *Arabidopsis* is one of the most aluminum-sensitive species in *Brassicaceae* due to the lack of *Brassicaceae* R or C genomes which is present in *Raphanus sativus* or *Brassica oleracea* that support the presence of many aluminum-responsive genes (Huang et al., 2002). Probably the main mechanism to face aluminum toxicity in *A. thaliana* is based on organic acids exudation and thus, a core of genes and associated biological functions linked to this mechanism have evolved and diversified in the model plant, but not in rice. In contrast, *Poaceae* species such as rice exhibit

a complex diversity of aluminum tolerance mechanisms including but not limited to: i) pectin methylation to change plasma membrane properties, ii) secretion of organic acids, iii) exudation of phenolic compounds, iv) nitrogen release to change soil conditions, v) rhizospheric alkalization, vi) ammonium preference that facilitates tolerance to low pHs and the potential of modifying central metabolism down-regulating photosynthesis and up-regulating hormonal genes (Zhao et al., 2013; Ma et al., 2014b; Kochian et al., 2015; Cohen and Leach, 2019; Wei et al., 2021). The potential diversity in detoxification mechanisms were demonstrated for rice in the variety of molecular functions exclusively associated with this cultivar.

GOCompare resulted in a novel and informative platform for comparative genomics analyses that allow the identification of relevant biological processes and molecular functions that can be shared or exclusively present when contrasting GO categories.

3.6.2. Possible caveats with the user inputs

It is important to highlight that functional enrichment results derived from GOCompare implementation, are dependent on GO terms that serve as inputs for the software. Redundancy is present in gene annotations since hierarchical (i.e. “parent-child) relationship among GO terms exists. As a consequence, a gene can acquire a GO term parent and child dependencies at the same time (Jantzen et al., 2011). Since the objective of this work is to provide a novel framework for functional enrichment analysis, to postulate an alternative to compare GO term levels is beyond the aim of the present study.

Therefore, it is strongly recommend curating to users to follow high standards in order to generate enriched categories that will be used as inputs for GOCompare, ensuring reproducibility and confidence in the derived results (Wijesooriya et al., 2022). The use of specific tools such as REVIGO (Supek et al., 2011), GO-Trimming (Jantzen et al., 2011) may improve the accuracy of the obtained results since they can refine authorship bias, and variation of GO term frequencies among species (Altenhoff et al., 2012). Hence, a recommendation table is provided in order to suggest a trustable pipeline prior GOCompare implementation (Supplementary Table 3.7).

Another topic to take into consideration at the moment of using GOCOmpare is that Gene ontology results are under Simpson´s paradox (Gaudet and Dessimoz, 2017), this means that varying gene lists that will be implemented in functional enrichment analyses can lead to different GO term composition and node weights variability. For example, the approach was tested by removing 1%, 5%, 10% and 25% of *Oryza sativa* genes and performing the same approach here

described. Indeed, by comparing node presence by Jaccard similarity index, the higher the percentage of removed genes is, the lower Jaccard indexes for analyzed categories are.

Most of the functional enrichment analysis softwares do not take into account the acyclic graph structure. An alternative to avoid this issue could be the use of algorithms that consider DAG topologies to enhance the GO term inference (Alexa et al., 2006). Nevertheless, DAG structure-based hypothesis must be carefully interpreted given that the familywise error rate (FWER) is preferred over false discovery rate (FDR) methods for multiple test correction given that GO terms lack independence derived from DAG structure (Meijer and Goeman, 2016). Nevertheless, even with the caveats mentioned above, the GOCompare approach has enough power to provide insights into a biological problem as it was shown by the conclusions derived from this case of study. Therefore, GOCompare can be implemented to design better-wet lab experiments in order to validate a particular hypothesis.

3.6.3. *Oryza sativa* subsp. *japonica* and *Arabidopsis thaliana* driver genes reveal different strategies to achieve the aluminum tolerance

The presence of kinases related to central metabolism is evident in genes with high degree values. For instance, this was observed by the presence of Phosphoglycerate kinase proteins (PGK) in rice. These genes covered more connections in the rice gene graph than in the *Arabidopsis thaliana* graph, indicating a possible more active role of central metabolism for aluminum tolerance in rice than in *Arabidopsis thaliana*. This agrees with the reported in the subspecies japonica where tolerant rice varieties use modulation of energy availability to combat aluminum toxicity (Wang et al., 2013; Tyagi et al., 2020). Moreover, PGKs are involved in the efficiency of photosynthesis by affecting the ATP consumption required in the Calvin Belson cycle (Rosa-Téllez et al., 2018). Therefore, the electron transport chain in photosynthesis is affected as consequence and this agrees with the presence of the core genes upregulated for diverse abiotic stress for rice (Cohen and Leach, 2019) where the photosynthetic efficiency is a trait affected. Likewise, evidence for modulation of energy availability can be corroborated by the use of molecules such as a sucrose synthase to enhance ethanolic fermentation under anaerobic conditions allowing the rice to survive under drought stress (Lakshmanan et al., 2013). Thus, energetic modulation can explain the high-degree values of these genes in the graph.

In contrast to the major *Oryza sativa* subsp. *japonica* driver genes suggested in this study, cytochrome P450 and proteins of the Mitogenic Active Protein Kinase (MAPK) pathway. Specifically, with aluminum tolerance in *Arabidopsis thaliana,* P450 gene expression has been related to the interplay between abiotic and biotic stresses, including the release of reactive

oxygen species in the presence of some metals such as copper and aluminum and protection against reactive oxygen species for plant species (Narusaka et al., 2004; Das and Roychoudhury, 2014). Furthermore, the cytochrome P450 genes are part of the biosynthesis of secondary metabolites and phytohormones, such as gibberellin metabolism, and are an active part of plant stress tolerance mechanisms while CYP family proteins such as cytochrome P450 have been found in all domains of life (Narusaka et al., 2004; Xu et al., 2015; Pandian et al., 2020; Aslam et al., 2021) which is coherent with a conserved mechanism to handle environmental stresses.

Likewise, the presence of other proteins such as protein kinases in both species is evident in the *A. thaliana* results. They contribute to the crosstalk between biotic and abiotic stress, the regulation of cellular processes in conjunction with calcium-binding proteins such as calcineurin that act in signaling and stress sensing in plants (Zhang et al., 2018; Vemanna et al., 2019). Another protein with a high degree of value was beta 1,3-glucanases. These proteins are involved in callose deposition (an aluminum sensitivity signal) and have been described in a wide spectrum of plants such as *Sorghum* bicolor and *A. thaliana,* where symplastic communication is affected by callose modulation by beta 1,3-glucanases, and modulation of these by the transcription factor SbSTOP1 in *A. thaliana* leads to increased aluminum sensitivity (Levy et al., 2007; Gao et al., 2019). Hence, the results suggest an active role of signaling in both species as the aluminum tolerance mechanisms; nevertheless, it is important to highlight that differences in gene network can be related to differences in genome annotations because: (i) the results shown previously in the GOCompare combined node weight, (ii) there is no correlation between degree and betweenness centrality for the prioritized and no prioritized genes, and (iii) There are median differences for betweenness centrality (p-value < 0.01) with higher centrality values for *A. thaliana* no prioritized genes, and higher centrality values for *O. sativa* subsp. *japonica* prioritized genes.

3.7. Conclusions and future perspectives

This research presents a new software, which is designed under the R package programming language used in biology (Tippmann, 2015). This approach provides six functions that are flexible and can be considered as a useful alternative to complement the classic and available enrichment tools. The use of GOCompare functions based on graphs theory to explore functional enrichment analysis results holds several advantages over the canonical ways of interpreting functional enrichment, including a better visualization of interactions among GO terms or functional categories, the detection of differences in genome annotation, and it delivers a better numerical and visual interpretation of big datasets in one workflow. It is important to consider that this approach would be helpful to obtain better insights in the functional analysis of complex datasets,

facilitating comparative genomics approaches and allowing to test preliminary hypotheses to design better wet lab experiments.

The GOCompare test with the case study comparing *Oryza sativa* subsp. *japonica* and *Arabidopsis thaliana* is consistent with the literature and the idea of a basal aluminum tolerance mechanism across plants based on the exclusion of aluminum species, where the exudation of organic acids and metal ion transporters is important since they remain conserved across the species. Moreover, results show that *Oryza sativa* subsp. *japonica* acquired a diversity of mechanisms different from exclusion, and this covers the ability of energy availability modulation to achieve the reported tolerance such as the detoxification of reactive oxygen species. On the other hand, *Arabidopsis thaliana* aluminum tolerance found here would be based mainly on aluminum exclusion mechanisms and possible main driver genes could be different between both species and observed in more connected genes for rice than *Arabidopsis thaliana* where Cytochrome P450 genes and MAPK presence are high in comparison with rice possible main genes found.

it is clear that aluminum mechanisms are functionally conserved between species but also a clear metabolic divergence was found, and it needs to be considered to fully understand metabolic insights over aluminum tolerance of rice. Therefore, the gene list here obtained will be used for modeling the possible aluminum effect over the metabolic phenotype of rice.

Chapter 4

Integration of transcriptomics information into protein-protein interaction network to predict new candidates proteins related to aluminum molecular mechanism tolerance in *Oryza sativa* subsp. *japonica*

4.1. Link to the previous chapter

Functionally there are aspects of the aluminum stress response that are shared between species as the metabolism of reactive oxygen species metabolism and response. But also, rice possesses divergent mechanisms present in its central energy metabolism. Nevertheless, it is described that even with the current knowledge of genes and proteins implicated in the tolerance to aluminum stress in rice, there is a gap regarding what proteins can be implicated in the response to aluminum stress and that can bring options to be studied in breeding programs. This chapter complements the previous chapter by providing possible new candidate proteins, taking into consideration the protein-protein relationships and the current proteins known as participants in the process. Thus, this chapter is related to the specific objective two: To identify new molecular elements associated with aluminum stress responses, via the analysis of protein-protein interaction networks, to postulate new elements that might regulate the response of rice to aluminum stress conditions.

4.2. Abstract

Rice is one of the most important crops worldwide, but it is affected by abiotic stress, such as aluminum stress, that leads to agronomical yield losses in acidic soils. Even if rice is a model species for cereals and possesses a wide range of bioinformatics resources and an extensive list of genes and proteins participants in stress, a holistic perspective of unknown protein participants has not been attempted using interactome and experimental omics data. Thus, the purpose of this chapter was to predict candidate proteins using the most recent rice interactome (RicePPInets) and transcriptomics integration data for two rice cultivars transcript counts. To test the influence of this integration on protein prediction, a strategy of five stages including four prediction scenarios with and without transcriptomics integration with the interactome using graph

embedding algorithms node2vec (unweighted graph), and node2vec+ (weighted graph with transcriptomics integration) were used to classify new protein candidates with two training datasets (i) Dataset 1: An extensive list of proteins previously described in aluminum tolerance in rice and (ii) Dataset 2: A differentially expressed genes list obtained for ten days of aluminum exposure. Additionally, a single averaging of seven classification models from Positive Unlabeled Learning (PULearn) and Adaptive Sampling (AdaSampling) was performed. The results of this workflow revealed a good performance of the classification models as well as more similarities to the training dataset 1 and an increase in the probability for the positive class when the transcriptomics integration was employed. A new screen of 3100 and 3057 proteins for dataset 1 and 2 is provided as a result. Also, the overlapping of results for weighted and unweighted predicted protein for each dataset suggests that transcriptomics integration helps to refine the prediction results by helping to enrich biological processes, molecular functions, and metabolic pathways that are not observed in the unweighted interactome. Also, a list of 37 proteins for all the scenarios provided was involved in aluminum stress response observed in other plants such as biosynthesis of ethylene, and cytokinins phytohormones, auxin transport, and changes in cell wall support the accuracy of the modeling approach used in this chapter. Therefore, it is indicated that the aluminum stress is complex, and it must be observed from a metabolic perspective as well to elucidate how metabolic pathways act together to ensure tolerance to aluminum stress.

4.3. Introduction

The *Poaceae* family (cereals) is an active part of food security, requiring accelerated plant breeding to adapt to the nutritional values, fiber, and carbohydrate requirements of the growing human population (Poole et al., 2022). Different species that belong to the family constitute the biggest proportion of carbohydrates ingest worldwide (Bandumula, 2018). Likewise, rice has been considered one of the model species for abiotic stress in the family, and thus, it has been extensively studied using bioinformatics tools and a wide repertoire of systems biology strategies to improve not only rice but plant breeding of other staple crops (Hong et al., 2019; Jia et al., 2021)

Several approaches have been used to obtain insights into genes and proteins involved in the phenotypic responses of aluminum tolerance in rice. These strategies vary from multi-omics methods, using contrasting conditions (aluminum treatment vs control conditions) to the evaluation of designed populations via GWAS (Fukuda et al., 2007; Famoso et al., 2010, 2011; Zhang et al., 2019b; Tyagi et al., 2020). These studies have provided lists of genes that suggest complex mechanisms acting synergistically, including modulation of central metabolism,

phytohormones crosstalk, photosynthesis, and aluminum transporters involved in described exclusion and tolerance mechanisms that finally lead to the inhibition of root growth (Kochian et al., 2015; Meng et al., 2019; Zhang et al., 2019b; Gallo-Franco et al., 2020; Huang et al., 2022; Sosa et al., 2023). Nevertheless, the identification and characterization of novel proteins linked with Al responses in rice, using systemic approaches like protein-protein interaction networks (interactomes) analyses, has not been previously implemented.

The use of specific interactomes provides information about the physical interactions among proteins, allowing an understanding of biological systems as an integrated core immersed in a holistic view of complex phenomena such as abiotic stress responses in plants (Cusick et al., 2005; Vandereyken et al., 2018; Hawe et al., 2019). As a model species, rice has available interactomic information related to agronomical traits or abiotic stress phenotypes (Liu et al., 2017a; Muthuramalingam et al., 2021). Additionally, the information accessibility to RNA-Seq experiments allows its integration into the interactome via Pearson's Correlation Coefficient to obtain a biological-specific context interactome, an interactome enriched with different omics layers to study a biological phenomenon as plant abiotic stress (Hawe et al., 2019), providing additional information to predict candidate proteins related to specific features such as a disease or stress processes via machine learning and graph theory approaches (Clavijo-Buriticá et al., 2023a).

Machine learning approaches to predict candidate proteins related to plant traits are not new. Some examples are found in *Arabidopsis thaliana* (Depuydt & Vandepoele, 2021; Clavijo-Buriticá, et al., 2023a), or agronomical traits in rice (Liu et al., 2017). Most of these approaches consist of the use of a training dataset of previously described proteins, an algorithm that converts the network into a numeric space (embedding) (Liu et al., 2023), the use of the edges to explore node connections (Tripathi et al., 2019; Depuydt and Vandepoele, 2021), and the use of algorithms that predict the new candidate proteins and an evaluation based on a metric as accuracy or area under curve (AUC).

Most of the training datasets used for predicting new proteins are considered as a positive class (proteins associated with a specific trait but unlabeled for other traits), which means that binary or multiclass algorithms are not accurate enough to be used, since negative information is not present to use with the machine learning algorithms. Therefore, new modeling strategies such as Positive Unlabeled Learning (PULearning) (Mordelet and Vert, 2014) and Adaptative sampling (AdaSampling) (Yang et al., 2019), take advantage of sampling strategies with the combination of well-established predictive algorithms (e.g. Random forest, logistic regression, KNN) to fill up

the gap of predict with a unique class. Nonetheless, these strategies produce overfitting in their predictions that create obstacles to creating generalizable models that can distinguish observed data from unseen data (Yang et al., 2019; Ying, 2019; Li et al., 2022). Though, strategies of ensembling different machine learning algorithms reduce overfitting in prediction significantly (Fernandes and Vinga, 2016).

To date, there is no information about what possible proteins are involved in the aluminum response in rice, based on the current knowledge from separated omics approaches found in literature. In this way, complementary information from interactomes such as RicePPInets (Mishra et al., 2022) can be useful to obtain unseen proteins involved in the plant response. Thus, the aim of this chapter is to predict new candidate proteins associated with aluminum tolerance in rice, implementing a machine-learning strategy based on previously collected information from literature and RNA-Seq experiments.

4.4. Methodology

A strategy of five stages was considered as follows (Figure 4.1): (i) preprocessing including graph embedding, omitting and including transcriptomic data. (ii) a classification stage using Positive Unlabeled Learning (PULearning) and Adaptive sampling (AdaSampling). (iii) Ensemble stage to avoid overfitting, (iv) Evaluation and thresholding, (v) predicted protein description using Upset plots to detect protein overlaps and gene set enrichment analyses.

The described scheme (Figure 4.1) was implemented to obtain four prediction scenarios. All the constructed scenarios were designed to predict new candidate proteins with strong associations with the evaluated phenotype, but they differ in the datasets that were considered to develop the prediction scenarios:

Known aluminum response:

- Scenario 1 (Dataset 1 and node2vec): it uses the interactome embedding representation based on previously reported genes or proteins found in scientific literature.

- Scenario 2 (Dataset 1 and node2vec+): it uses the interactome embedding representation with integrated transcriptomics data based on previously reported genes or proteins found in scientific literature.

Rice response at ten-day of aluminotoxic conditions exposure:

- Scenario 3 (Dataset 2 and node2vec): it uses the interactome embedding representation based on reported genes obtained from differentially expressed genes derived from RNA-seq experiments.

- Scenario 4 (Dataset 2 and node2vec+): it uses the interactome embedding representation with integrated transcriptomic data based on reported genes obtained from differentially expressed genes derived from RNA-seq experiments.

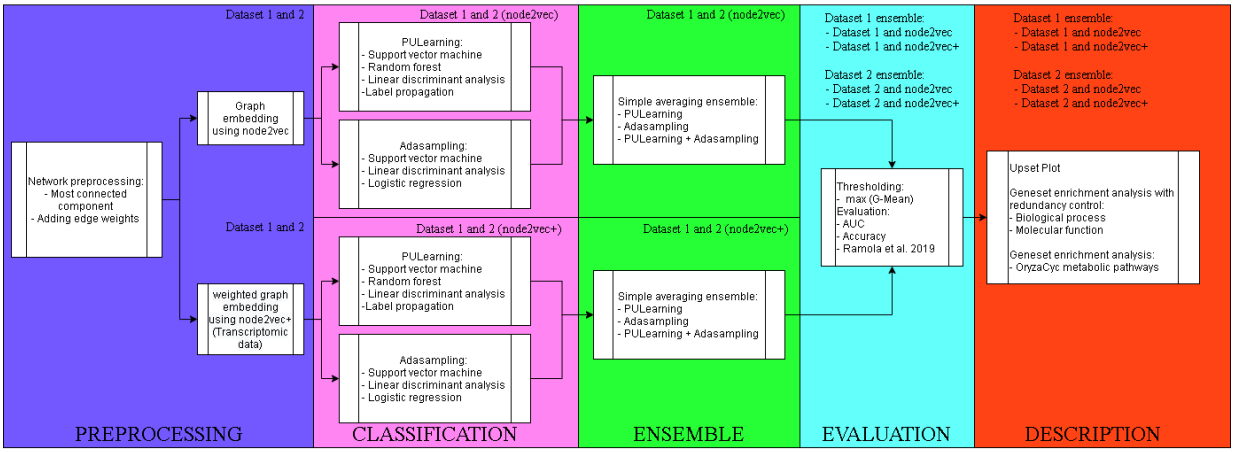


Figure 4.1. Followed strategy to predict novel rice proteins involved in aluminum tolerance mechanisms. Five stages are established for the prediction: Preprocessing (blue), Classification (violet), Ensemble (green), Evaluation (light blue), and predicted proteins description (red).

4.4.1. Preprocessing stage

4.4.1.1. Datasets obtention

The interactome obtained from RicePPInets (Mishra et al., 2022) was used as the main input to predict candidate proteins associated with aluminum tolerance in rice. This interactome reflects 17421 proteins with MSU annotation (nodes) and 759851 edges (protein connections). Complementarily, as positive class to detect new candidate proteins two datasets were chosen: (i) Dataset 1: a list of 1084 rice genes (*Oryza sativa* subsp. *Japonica*) that according to scientific literature are associated with aluminum tolerance, used previously to test GOCompare R package (Sosa et al., 2023), and available at https://data.mendeley.com/datasets/fzxjwrt53m/3. (This dataset represents the rice known response to aluminum stress) (ii) Dataset 2: differentially expressed genes obtained from Gallo-Franco et al., (2023) with absolute Log2 fold-change values

≥ 1.0. (This dataset represents the specific stress response at ten days of exposure to aluminotoxic conditions).

4.4.1.1.1. Differentially expressed genes from two contrasting rice cultivars

A total of twelve RNA sample sequences were generated via Illumina Hiseq technologies from two contrasting rice cultivars, Azucena (aluminum tolerant) and BGI9311 (aluminum susceptible) were used as datasets. These samples represent three biological replications subjected to aluminum stress conditions (ten days of exposure to Al_3Cl-100 uM) and three biological repeats under no-stress condition (controls), per cultivar. Only raw sequences with Phred score >20 were aligned to the Os-Nipponbare-Reference-IRGSP-1.0 genome (Kawahara et al., 2013) using STAR v2.5 (Dobin et al., 2013) and genes counts were obtained using FeatureCounts v.2.0.3 software (Liao et al., 2014). Further, differentially expressed genes were obtained via DESeq2 R package (Love et al., 2014) using the Benjamini-Hochberg (BH) method for multiple testing correction. Only genes with absolute Log2 fold-change values ≥ 1.0 $|LFC \geq 1|$ (after comparing aluminum and control conditions for both cultivars were labeled as dataset2.) This LFC value was chosen to ensure the reduction of the number of false positive differential expressed genes in the analysis and this value is widely used in trasnscriptomics analysis (Montgomery and Mank, 2016). The raw counts matrix used as input for DESeq2 was also implemented for the transcriptomics data integration with the studied interactome as edge weights and it is explained in the next section. A full description of the methodology used to obtain the transcriptomics data is available at Gallo-Franco et al., (2023).

4.4.1.2. Network preprocessing

To avoid the use of unconnected components for graph embedding, only the most connected component of the interactome was kept, being used for classification purposes via Networkx Python module v3.1. For the identification of novel proteins involved in Al tolerance, the graph embedding procedure was performed. As result, a vector space that represents the neighborhood of nodes and network´s structure was obtained.

4.4.1.3. Edge weights obtention

- Protein ids from interactome data were converted from MSU to RAP annotation via Riceidconverter v1.1.1 R package. This step was performed to assign the correct edge weights, as the interactome, and transcriptomics data were described in different IDs systems: MSU for the interactome and RAPDB for the transcriptomics data respectively

- For each protein, its respective id was identified in the raw count matrix. In case of multiple hits of RAPDP proteins for a node (e.g. gene 1 and gene 2 have the id LOC_A), the average of counts for the proteins was considered.
- For each node pair, the Pearson correlation coefficient (PCC) using the count matrix obtained in FeatureCounts was calculated and added as the edge weight. In case of missing ids or no coincidence between RAP and MSU ids (e.g. an MSU id A had a possible RAPDB id associated), a value of zero was added as the edge weight. No PCC thresholds were used to filter the edge weights. The code used is available at https://github.com/ccsosa/Aluminum_genes_classification/tree/master/SCRIPTS**.**

4.4.1.4. Network Embedding preprocessing

A total of two approaches were used for the graph embedding: (i) the node2vec algorithm (Grover and Leskovec, 2016) was used using the p (return parameter) and q (in-out parameter) equal to one to obtain a vector of 128 features (Clavijo-Buriticá, et al., 2023a). (ii) The algorithm node2vec+ (Liu et al., 2023) was used with the same hyperparameters above described over the weighted interactome. Node2vec+ incorporates edge weights in the embedding, thereby, allowing the integration of the PCC calculated previously to consider the effect of the integration of transcriptomics information in the classification of new protein candidates related to aluminum tolerance. Both procedures were performed with the Pecanpy Python module v2.0.8 (Liu and Krishnan, 2021) with the DenseOTF option. As a preprocessing step for classification, a Z-score normalization was performed for each of the two embedded graphs.

4.4.2. Classification stage

4.4.2.1. Positive Unlabeled Learning

Since no additional datasets were available providing information about non- participant proteins in the aluminum tolerance in rice, a Positive Unlabeled-Learning approach was used. This approach lacks negative examples in the training set and employs bagging (bootstrap aggregating) techniques to iteratively discriminate positive and unlabeled samples (Mordelet and Vert, 2014). To identify new candidate proteins a set of four classification models were used. The two embeddings obtained via node2vec and node2vec+ algorithms were selected as inputs. Four models were executed due to their popularity for classification purposes: (i) Support vector machine, (ii) Random Forest, (iii) Linear discriminant analysis, and (iv) Label propagation. Each model was iterated 10000 times using the maximum number of samples to ensure all the positive

class labels for training. Finally, Python modules pulearn 0.0.7 and scikit-learn 1.2.2 were used to run the analyses. The code used is available at https://github.com/ccsosa/Aluminum_genes_classification/tree/master/SCRIPTS.

4.4.2.2. Adaptative sampling strategy

Adaptive sampling was considered as a second approach for the prediction of new candidate proteins. This method can work for PULearning and progressively reduces the risk of mislabeling by iteratively updating the training dataset and as a consequence creating more generalizable models (Yang et al., 2019). The following models were used as they were the unique available algorithms in the current AdaSampling R package: (i) Support vector machine, (ii) Logistic regression, (iii) Linear discriminant analysis. Similar to the PULearning approach, each classification algorithm was run 10000 times. These analyses were performed in the AdaSampling R v1.3. The code used is available at https://github.com/ccsosa/Aluminum_genes_classification/tree/master/SCRIPTS.

4.4.3. Ensemble stage

An ensemble approach was considered by averaging the probabilities for each gene. This strategy was considered because it reduces the bias of a single mode while decreasing its variance gets smaller. Similarly, to avoid overfitting a simple averaging ensemble was preferred over a weighted ensemble (Fernandes and Vinga, 2016; Zhou, 2021). Thus, three ensembles were obtained for each of the embedded graphs: (i) an ensemble considering the four algorithms used with PULearning, referred as PULearning ensemble (ii) another ensemble using the three algorithms implemented in AdaSampling referred as AdaSampling ensemble, and (iii) a last one ensemble averaging the seven algorithms together henceforth named single averaging ensemble.

4.4.4. Thresholding and evaluation

To obtain final candidate proteins, the probability value that maximizes the geometric mean (G-MEAN) of true positive rate (TPR) and True negative rate (TNR) was used as the threshold (De Diego et al., 2022). For the different ensembles applied to the evaluated embeddings, values greater or equal to the G-MEAN were considered as candidate proteins. To evaluate ensemble performance, the modified metrics of Area Under Curve (AUC), accuracy, balanced accuracy, F value, and Matthews Correlation Coefficient were implemented (Ramola et al, 2018).

4.4.5. Functional description of predicted proteins stage

Upsets plots were generated to show the number of shared proteins for the different analyzed datasets: (i) Prediction scenarios 1 and 2, (ii) prediction scenarios 3 and 4 and (iii) the four prediction scenarios combined. The Upset plots were generated using UpsetR v1.4.0 R package.

To validate whether the found results had concordance with the proteins used for training, Gene set enrichment analyses were performed to find functional similarities among datasets . The analyses were done using gprofiler2 v0.2.1 (Kolberg et al., 2020) with a g:SCS multiple comparisons correction procedure and using as background genes available at RicePPInets database (Mishra et al., 2022). Only enriched biological processes (GO:BP), and molecular functions (GO:MF) retained after a false discovery rate (FDR) < 0.05 implementation, were considered for the analysis. In the case of biological processes and molecular function the following procedure was performed to reduce the redundancy in the results:

- For each enriched GO:BP or GO:MF term, the children's terms associated with it were derived from the GO.db R package v3.14.0.
- The children's terms were the data subset representing the parental term present in the enriched GO:BP or GO:MF.
- For each of the children's terms obtained for a particular GO:BP or GO:MF hierarchy, the false discovery rate was calculated.
- The parental term was discarded when a lower FDR was present for the children's terms.

To identify common metabolic pathways described in rice. A one-sided Fisher's exact test, retaining proteins with a false discovery rate (FDR) < 0.05. For the procedure a file with the described metabolic pathways associated with rice, was downloaded from OryzaCyc (Schläpfer et al., 2017; Hawkins et al., 2021, p. 15). As background, the available genes linked to the RicePPInets database were considered (Mishra et al., 2022). These analyses were performed in the R package bc3net v1.0.4. The implemented codes to reduce terms redundancy and gene sets enrichments analyses are available at: https://github.com/ccsosa/reducereduntGO.

Finally, the predictions were added to the interactome and plotted in Cytoscape v3.10. (Shannon et al., 2003). To identify topological differences among predicted, positive class labels and unlabeled proteins in the interactome, degree and betweenness centrality were calculated and represented by boxplots using the ggpubr v.0.6.0 R package (Hansen et al., 2020).

4.5. Results

The filtered interactome comprised 17228 proteins in the MSU id format, integrating 759605 interactions, indicating a loss of 1.10% of the original proteins present in the RicePPInets interactome. For dataset 1, out of 1084 proteins related to aluminum tolerance, 782 were mapped to the interactome. To train the model, 133 differentially expressed genes with $|LFC \geq 1|$ obtained from transcriptomic experiments (dataset 2) were used (see Supplementary Figures 4.1A and 4.1B).

4.5.1. Models evaluation shows a good performance with single averaging ensemble (PULearning +AdaSampling)

For scenario 1, the single average ensemble approach (PULearning +AdaSampling) showed an Area Under Curve (AUC) values ranging from 0.611 (AdaSampling) to 0.971 (PULearning) and from 0.556 (AdaSampling) to 0.990 (PULearning) in scenario 2. Averaging PULearning and Adasampling together (single averaging) had values above 0.9 for both scenarios indicating a good performance of the single averaging approach (PULearning +AdaSampling) (Table 4. 1). In addition, similar evaluation metrics were obtained for dataset 2, evaluating differentially expressed genes (DEG) with $|LFC \geq 1|$ values (

Table 4. ***2***).

Table 4. 1. Obtained metrics for dataset 1 prediction scenarios. Asterisk represents traditional AUC obtained using sci-kit learn Python module. G-Mean = Geometric mean of sensitivity and specificity for the threshold value.

	Embeed interactome without edge weights (Scenario 1)			Embeed interactome with transcriptomics integration (Scenario 2)		
Metrics	**PULearning**	**AdaSampling**	**Average**	**PULearning**	**AdaSampling**	**Average**
AUC *	0.973	0.712	0.918	0.990	0.704	0.939
AUCcr	0.971	0.611	0.908	0.990	0.556	0.932
Accuracy	0.999	0.829	0.987	1.000	0.824	0.994
Balanced accuracy	0.995	0.785	0.971	0.997	0.806	0.986
F	0.993	0.698	0.955	0.997	0.753	0.979
MCC	0.832	0.308	0.694	0.889	0.262	0.732
ALPHA	0.045	0.258	0.112	0.015	0.334	0.098
Threshold	0.612	0.550	0.522	0.652	0.491	0.534
G-Mean	0.973	0.712	0.918	0.990	0.704	0.938

For both datasets, AdaSampling had lower evaluation metric values than PULearning suggesting poorer performance for the AdaSampling approach and an overfitting trend for the PULearning strategy. However, the averages for both modeling strategies (single averaging ensemble: PULearning +AdaSampling) had intermediate evaluation metrics, representing values above 0.9 for AUC, accuracy, balanced accuracy, and F value suggesting good confidence in the results, even better than those using only AdaSampling or PULearning for prediction (see Supplementary Figures 4.2 and 4.3).

Table 4. 2. Obtained metrics for dataset 2 prediction scenarios. Asterisk represents traditional AUC obtained using sci-kit learn Python module. G-Mean = Geometric mean of Sensitivity and Specificity for the threshold value.

	Embeed interactome without edge weights (Scenario 3)			Embeed interactome with transcriptomics integration (Scenario 4)		
Metrics	**PULearning**	**AdaSampling**	**Average**	**PULearning**	**AdaSampling**	**Average**
AUC *	0.938	0.714	0.902	0.958	0.701	0.916
AUCcr	0.932	0.638	0.889	0.955	0.616	0.909
Accuracy	0.994	0.846	0.982	0.998	0.832	0.985
Balanced accuracy	0.983	0.771	0.957	0.992	0.758	0.951
F	0.971	0.643	0.927	0.986	0.628	0.914
MCC	0.784	0.338	0.701	0.842	0.313	0.761
ALPHA	0.094	0.211	0.120	0.068	0.222	0.078
Threshold	0.570	0.547	0.515	0.581	0.544	0.539
G-Mean	0.941	0.714	0.906	0.962	0.701	0.920

4.5.2. Small similarities are evident between the datasets 1 and 2 for the prediction of new candidate proteins

For the first simulation scenario, list of 744, 4256, and 1840 new candidate proteins were predicted after applying PULearning, AdaSampling and averaging both strategies respectively. Similarly, for the second scenario, 249, 5505, and 1621 novel candidate proteins were predicted using PULearning, AdaSampling and single averaging ensemble both approaches respectively. Thus, a lower number of proteins were predicted when transcriptomic data were added to the interactome. Similarly, for the third scenario, a list of 1602, 3605, and 2050 new candidate proteins were predicted for PULearning, AdaSampling, and their average, and finally, a list of 1169, 3794,

and 1338 novel candidate proteins were predicted for PULearning, AdaSampling, and their average respectively, for the scenario 4 (see Supplementary Figure 4.1 to observe the predicted proteins on the interactome used).

Jaccard indexes were calculated to identify similarities among positive class datasets and predicted new candidate proteins. The dataset 1 and the predicted new candidate proteins using transcriptomic data had the highest value of the Jaccard index, with the most similar dataset predicted across the comparisons (0.302). Likewise, when using dataset 2, the most similar predicted dataset was obtained when the interactome was complemented with weights obtained from transcriptomic data. Thus, it is suggested that the addition of experimental information can influence the predictions obtained (Figure 4.2).

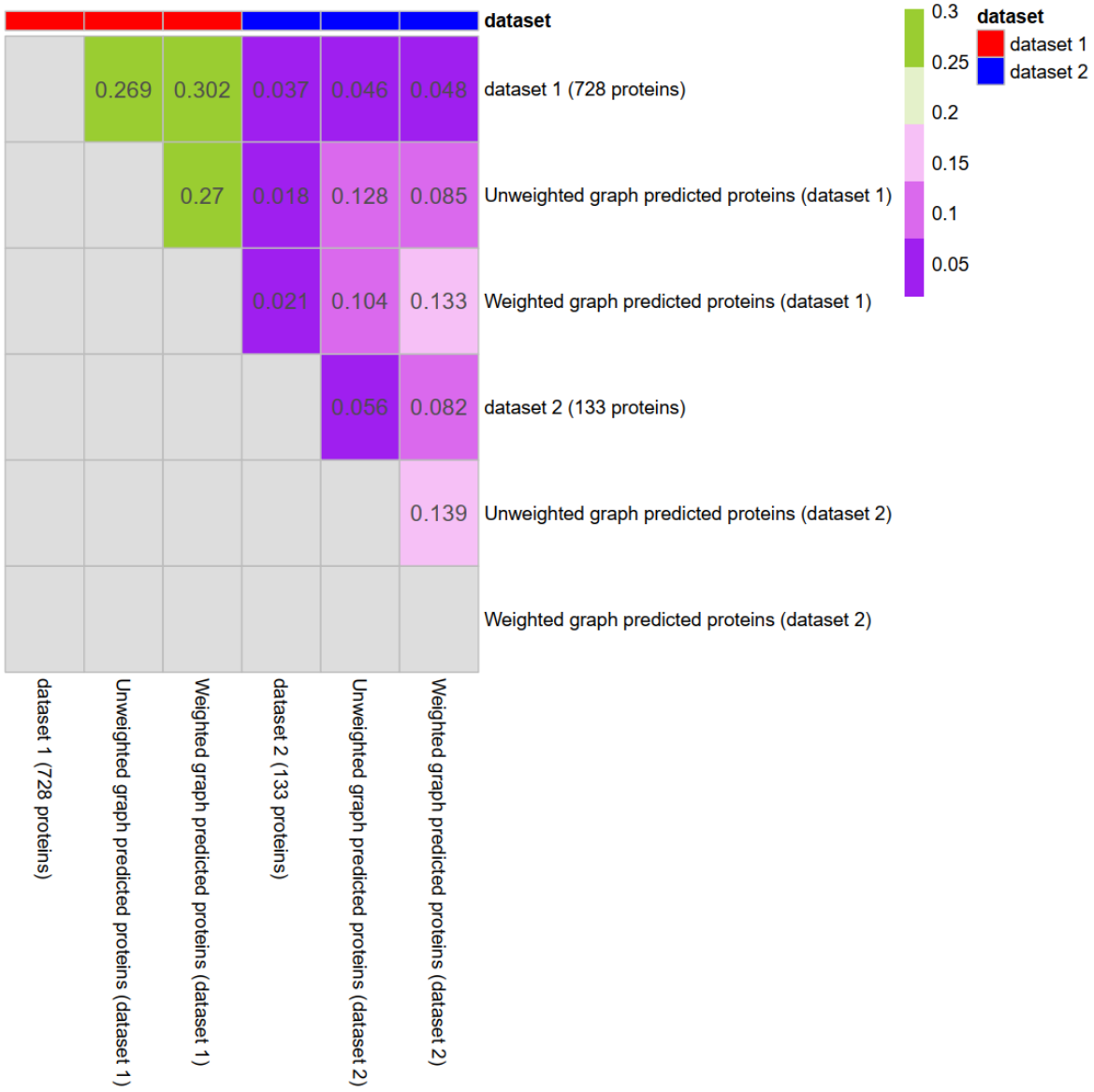


Figure 4.2. Jaccard index heatmap (close to one value indicate perfect overlap for shared proteins in two datasets). In red color (dataset 1, related positive class and predicted datasets), in color blue (dataset 2, related positive class and predicted datasets). Diagonal values are not displayed for visualization purposes.

4.5.3. Transcriptomics integration helps to predict the positive class label and contributes to predicting new candidate proteins for the dataset 2

To compare the impact of the transcriptomic integration data on the prediction of positive class proteins and new candidate proteins, violin plots were constructed for the three single-averaging ensembles (PULearning, AdaSampling, and the single average ensemble). The positive class used for the training models showed higher average predicted values for the dataset 1. In addition, differences in the ensembles for unweighted and weighted interactomes (p-value < 0.05) show

that the integration of additional data can influence the derived results. In contrast, dataset 2 did not show differences between unweighted and weighted interactomes, but, like those found in the dataset 1, the positive class used for model training showed differences in the average predicted values for dataset 2 (Supplementary Figure 4.4 and 4.5).

4.5.4. New candidate proteins predicted using unweighted and weighted interactome at the same time

In dataset 1, 1479 novel proteins were found in the unweighted interactome. Similarly, 1260 new candidate proteins were exclusively detected in the weighted interactome integrating transcriptomic data (Figure 4.3A). A total of 672 proteins were common to the positive class in dataset 1 at observe scenarios 1 and 2. This indicates that 92% of matches with the positive class in dataset. In contrast, 361 new candidate proteins were common to scenarios 1 and 2 and 331 novel proteins were common to scenarios 3 and 4 (Figure 4.3A and B). These proteins can be considered as novel high-confidence predictions because they are not present in the positive class for either dataset (Table 4. 3 and Supplementary Table 4.2).

A moderate confidence category was created for the predicted proteins that were present in only one scenario (unweighted or weighted interactomes). For the first dataset, a list of 1479 and 1260 proteins were predicted for the single averaging ensemble using unweighted and weighted interactomes as inputs, respectively. Similarly, for the second analyzed dataset, lists of 1719 and 1007 proteins were predicted for the single averaging ensemble using the unweighted and weighted interactomes, respectively (Table 4. 3, Figure 4.3A and 4.3B).

Table 4. 3. Summary of predicted protein categories for both datasets inferred via single averaging of the PULearning and AdaSampling models. Parentheses represent the percentage of total proteins in the interactome. High-confidence and moderate-confidence predicted datasets were considered as new candidate proteins not found in the positive labels used for training models (see Supplementary Table 4.2 to obtain the complete list of predicted proteins including the predicted proteins categories).

Predicted dataset	Dataset 1	Dataset 2
Positive class (predicted by Unweighted and Weighted approaches)	672 (3.901%)	115 (0.668%)
Positive class (unweighted)	18 (0.104%)	10 (0.058%)
Positive class (weighted)	38 (0.221%)	8 (0.046%)

Predicted dataset	Dataset 1	Dataset 2
High confidence (predicted by unweighted and weighted approaches)	361 (2.095%)	331 (1.921%)
Moderate confidence (Unweighted)	1479 (8.585%)	1719 (9.978%)
Moderate confidence (Weighted)	1260 (7.314%)	1007 (5.845%)
Unlabeled	13400 (77.78%)	14038 (81.484%)

When the four scenarios were compared together, lists of unique proteins of 1106, 879, 1189, and 658 were unique for scenarios one to four respectively. Contrastingly, a list of 37 proteins that included the commonly predicted proteins for all scenarios was found (See Supplementary Table 4.4). This list of 37 proteins includes: three transcription factors, aluminum-related transporters, a cytokinin-related protein, peroxidases and 1-Aminocyclopropane-1-carboxylate synthase (ACC synthase). Which is involved in ethylene biosynthesis. Finally, 27 proteins were common for both datasets and were available in the four scenarios, as well as being present in the positive class. They were: LOC_Os01g01660, LOC_Os01g03530, LOC_Os01g69010, LOC_Os02g03900, LOC_Os02g18410, LOC_Os03g04060, LOC_Os03g13540, LOC_Os04g15840, LOC_Os04g45590, LOC_Os04g46810, LOC_Os04g46820, LOC_Os05g02310, LOC_Os05g04440, LOC_Os05g09440, LOC_Os05g41990, LOC_Os06g01920, LOC_Os06g48060, LOC_Os07g35860, LOC_Os08g33710, LOC_Os10g38670, LOC_Os10g40430, LOC_Os10g40480, LOC_Os10g40530, LOC_Os11g41840, LOC_Os11g42430, LOC_Os12g01760, and LOC_Os12g06660.

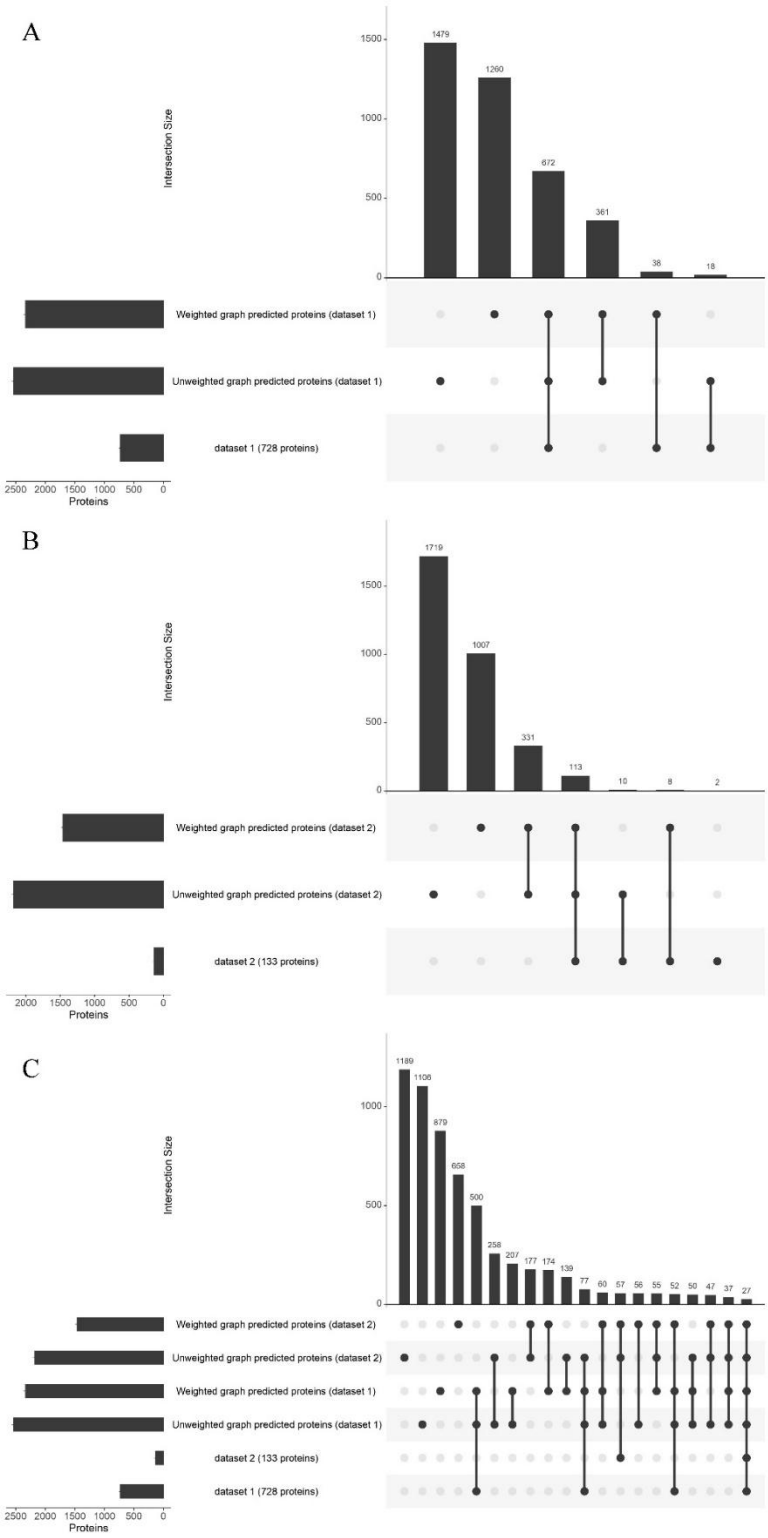


Figure 4.3. Upset plots with the positive class datasets and predicted proteins found per weighted or unweighted interactome approach A) Dataset 1: GOCompare positive class, B) Dataset 2: Differentially expressed genes ($|LFC \geq 1|$) C) Top 20 intersections for positive class datasets and

predicted proteins for the average of seven models for unweighted graph and added transcriptomics data to the interactome.

Functional enrichment analyses were performed on the predicted datasets as described in Table 4.3 For dataset 1, there were seven enriched biological processes (GO:BP) related to cell wall and mannan metabolic and catabolic processes,14 GO:BP related to carbohydrate, chitin, and cell wall metabolism and catabolism. Meanwhile, 31 GO:BP related to the positive classes (Aluminum-associated genes and genes derived from transcriptomic data and used for model training) were related to a wider spectrum of categories, such as cellular detoxification, hydrogen peroxide production, and central metabolism respiration (Supplementary Table 4.3).

For molecular function (GO:MF) enrichments, hydrolase function was enriched for high-confidence predicted proteins. A total of eight molecular functions linked to hydrolase activities were detected: hydrolase activity, hydrolyzing O-glycosyl compounds, beta-galactosidase activity, chitinase activity, beta-D-fucosidase activity, glucosidase activity, chitin binding, beta-glucosidase activity and alpha-L-arabinofuranosidase activity. Finally, six molecular functions related to ion binding, oxidoreductase activity, and antioxidant activity were identified.

Regarding enriched metabolic pathways for dataset 1, a list of 14 metabolic pathways were enriched for the positive class: glycolysis I (from glucose 6-phosphate), L-glutamine biosynthesis III, superpathway of glyoxylate cycle and fatty acid degradation, glycolysis IV (plant cytosol), superoxide radical degradation, ascorbate glutathione cycle, Rubisco shunt, superpathway of gibberellin GA_{12} biosynthesis, glyoxylate cycle, UDP-α-D-xylose biosynthesis, suberin monomers biosynthesis, pyruvate fermentation to ethanol II, ammonia assimilation cycle I, and fatty acid β-oxidation II (peroxisome). Three metabolic pathways were enriched for the moderate confidence results with transcriptomic integration: chitin degradation II (Vibrio), starch biosynthesis and (1,4)-β-D-xylan degradation. Similarly, for the moderate confidence derived dataset, but without transcriptomic data consideration, three pathways were enriched: soybean saponin I biosynthesis, very long chain fatty acid biosynthesis I, and very long chain fatty acid biosynthesis II.

For dataset 2, seven metabolic pathways were enriched when moderate confidence results with transcriptomic integration were considered: starch biosynthesis, very long chain fatty acid biosynthesis I, cutin biosynthesis, jasmonoyl-L-isoleucine inactivation, poly-hydroxy fatty acid biosynthesis, vernolate biosynthesis I and vernolate biosynthesis II). Additionally, one metabolic pathway was enriched when the medium-confidence dataset without transcriptomic data was considered: diterpene phytoalexins precursor biosynthesis (Supplementary Table 4.4).

Given that some datasets had no enriched metabolic pathways and with the purpose of obtaining insights about metabolic pathways similarities per the predicted dataset, a heatmap was implemented to explore the composition and frequency of metabolic pathways available per gene (Supplementary Figure 4.6). The heatmap revealed two clusters of predicted datasets. The first cluster was composed of most of the medium and high confidence of the dataset 1 and the positive class used for training models. This cluster had less diverse metabolic pathways including gluconeogenesis I, very long chain fatty acid biosynthesis I, threonine and lysine, cutin and fatty acid biosynthesis, and xyloglucan. In contrast, cluster two was mostly composed of medium confidence for datasets 1 and 2 and had more enriched metabolic pathways than cluster one. Moreover, the most frequent pathways for both clusters were: superpathway of cytosolic glycolysis (plants), pyruvate dehydrogenase and TCA cycle, superpathway of anaerobic sucrose degradation, glycolysis I (from glucose 6−phosphate), glycolysis IV (plant cytosol), Rubisco shunt, glycolysis II (from fructose 6−phosphate), L−ascorbate biosynthesis I (L−galactose pathway), GDP−mannose biosynthesis and gluconeogenesis I, and very long chain fatty acid biosynthesis I and II.

Finally, when degree is used as a topological measurement and analyzed, boxplots showed median differences for unlabeled and positive classes for training models in dataset 1 and 2 (p-value < 0.05) (Supplementary Figure 4.7A and 4.7B). In general, the high-confidence proteins and moderate confidence predicted proteins using edge weights had lower degree values in comparison with the positive class and unlabeled data for both datasets. In contrast, the betweenness centrality values had no median differences between then unlabeled and positive classes for dataset 2. Nevertheless, median differences were observed for both datasets (Supplementary Figure 4.7C and 4.7D). Similar to the degree inferences, lower betweenness centrality values were generally observed for the predicted proteins. Thus, it was suggested that weights could affect the prediction of candidate proteins and a wide range of topological values for the predicted datasets.

4.6. Discussion

The results presented here demonstrate the implementation of a new workflow based on complementary information that provides new insights into the aluminum tolerance mechanisms used by rice. The predicted proteins obtained herein offer a starting point for the description of new proteins and metabolic pathways linked to abiotic stress responses in rice, which cannot be determined using separate approaches. The current pipeline is the first attempt to implement combined strategies of machine learning, interactomics, and transcriptomics to generate a holistic perspective for predicting novel elements that are potentially related to abiotic stress in rice.

4.6.1. Considerations about the protein candidates prediction workflow

The strength of the described approach relies on four key features: (i) the use of the most recently published rice interactome, (ii) a set of genes and proteins extracted from literature related to aluminum tolerance (iii) a dataset of genes obtained from *in-house* transcriptomic experiments published and integrated with the interactome for testing its influence over predictions, and (iv), the combination of several machine learning strategies to obtain replicable results.

The use of the RicenetPPIs interactome has proven its usefulness by describing the interactions with a biotic stressor in rice, such as *Xanthomonas oryzae* (Mishra et al., 2022). RicenetPPIs was based on previous rice interactomes that demonstrated their usefulness in mapping agronomic traits, abiotic stress, virus-plant interactions, or Cd phytotoxicity in rice (Brizard et al., 2006; Seo et al., 2011; Zhu et al., 2011; Cantu et al., 2013; Bari et al., 2021; Liu et al., 2017a; Muthuramalingam et al., 2021). This study is the first to date, which includes the combination of graph embedding approaches and the use of a PULearning strategy to obtain predictions in rice and extend the use of interactome information to identify potential proteins related to particular features, which is valuable by (i) using conserved interactions among proteins found through species represented in the interactome used, (ii) proteins used as training data are supported in the literature, and (iii) it is highly reproducible to use an approach that uses less computational power in comparison with the algorithms based on communities detection or neighbors searching, and (iii) the edge weight embeddings as implemented in *Node2vec+* have demonstrated increased the classification accuracy tasks to regular graph embedding (Liu et al., 2023)

The second topic that ensures the strength of this analysis relies on the use of proteins and genes previously described in the literature to be related to the aluminum response in rice, which demonstrated shared functional similarities and some divergences in central metabolism between

Oryza sativa and *Arabidopsis thaliana* (Sosa et al., 2023). Thus, it is expected that the results described as products of dataset 1 are sufficiently diverse and can be promising for testing in future steps.

The results obtained after the analysis of both datasets were divergent given that only 37 common predicted proteins for the four prediction scenarios were obtained. This lack of common proteins can be attributed to the heterogeneity of the sources selected to compile the protein list used as the positive class in the first dataset. This list contains sources from proteomics, transcriptomics, GWAS, QTL, and SSR-related techniques, which can provide a more complex scenario than the one considered as the second dataset, comprising only transcriptional information and reflecting a lower number of proteins involved in the aluminum tolerance long response.

4.6.2. Superior Performance of Averaging Ensembles in Protein Classification

A four topic related to the strength of the predicted proteins presented here is that the results of the pipeline created using the single averaging ensemble show a good performance (AUC > 0.7), which is achieved by combining two different approaches: (i) PULearning, which leads to overfitting problems because classifiers are sensitive to size differences between unlabeled and positive labels and unlabeled data propagate noises that affect its performance (Yang et al., 2019; Li et al., 2022) and (ii) AdaSampling that use a weighted sampling to reduce the risk of selecting mislabeled data and creating generalizable models (Yang et al., 2019). Thus, the results of the workflow intended to obtain a balance between both approaches are justified by overfitting avoidance (Zhou, 2021). In addition, it is demonstrated that the single average ensemble results for PULearn and AdaEnsemble have a better performance than single classifiers, which can enhance the predictability of the approaches tried here.

4.6.3. Possible caveats of the protein prediction workflow

It is important to note that the workflow was highly dependent on the dataset used for training the models, and this was observed in the low number of overlapping proteins predicted (37 new predicted proteins) between datasets 1 and 2. Similarly, 10.8% and 1% for dataset 1 and 2, respectively, were common for unweighted and weighted edges, indicating that weighted and unweighted results obtained different predicted proteins.

In this regard, it is evident that the integration of transcriptomics data provides new information and alters the predictions because the predictions made by only edge weights are 32% and 40%

of the new predictions for dataset 1 and 2 respectively. This difference in predicted proteins could be caused because the transcriptomic data is highly specialized by tissue (Timmons et al., 2015). This bias could also be caused by the heterogeneous nature of the genes in dataset 1 which did not include the aluminum exposure to ten days used for dataset 2 and transcriptomics integration. Nevertheless, coherent results were observed in terms of identifying new candidate proteins based on protein interactions. For instance, the predicted proteins obtained via edge weights were more similar to the training dataset (see Figure 4.2) suggesting that the inclusion of transcriptomic data in the dataset 1 which is independent of the transcriptomic dataset improved the probability values and corroborated the pertinence of the prediction scenarios including experimental data.

The influence of transcriptomic integration was observed in the gene set enrichment analysis, in which the proteins predicted by the single averaging for both datasets possessed gene ontology terms enriched mostly related to cell wall functionalities, reactive oxygen species, and central metabolism, which are notable and coherent with the idea of the core of genes activated in rice as a response to abiotic stresses (Cohen and Leach, 2019), and plant stress response-related hub proteins enriched in protein-protein interaction networks: transcription factor hubs, kinase and phosphatase hubs, ubiquitin system associated hubs, chaperone hubs, redox signaling hubs and functionally unclear hubs (Vandereyken et al., 2018).

However, the topological measurement results with low betweenness values do not support the idea that aluminum-related proteins rely on bottlenecks as proteins with a high betweenness centrality (Yu et al., 2007) but in a bulk of diverse metabolically functions that act together to provide tolerance and the survival of rice under the stressor conditions. Furthermore, the similarities found among the positive class for dataset 1 and the medium confidence predicted proteins corroborate this fact of a high diversity in metabolic functions taken under several conditions for rice cultivars. Hence, the transcriptomics integration and the test with the dataset 2 are coherent with a scan of candidate proteins for more stringent conditions

4.6.4. High confidence proteins description is coherent with a conserved adaptation strategy using phytohormones

It is important to highlight that the list of 37 proteins contains important proteins involved in rice response stress such as: (i) a peroxidase (LOC_Os07g44590), peroxidases are present as part of the antioxidant strategy to handle the reactive oxygen species excess created by the aluminum stress (Hossain et al., 2012; Das and Roychoudhury, 2014; Mhamdi and Van Breusegem, 2018). (ii) the presence of the 1-aminocyclopropane-1-carboxylic acid synthase 1 (ACC1)

(LOC_Os03g51740) to be active under aluminum stress as an active plant response in rice (Keunen et al., 2016; Daspute et al., 2017). The presence of ACC1 suggests an ethylene role that has been widely described in plants when aluminum enters root cells, affecting root growth by modulating auxin biosynthesis, and which has been widely described to be part of plant development (Růžička et al., 2007; Keunen et al., 2016; Ahammed et al., 2020; Jingguang et al., 2020). (iii) The presence of the cytokinin oxidase (Os04g0523500), known as CKX8, reveals a potential role for cytokinin-related responses in the rice plant. CKX8 is related to the grain size, number, and panicle architecture in plants (Wu et al., 2024). In addition, cytokinin biosynthesis in *A. thaliana* inhibits root growth and the prediction of CKX8 suggests a similar mechanism in rice. However, little is known about the role of cytokinin in aluminum stress tolerance in plants currently (Yang et al., 2017; Ranjan et al., 2021a). (iv) The presence of three transcription factors belonging to B3, bHLH, and MYB protein families reveal the role of regulatory elements in the rice response given that the overexpression of some bHLH, MYB, and transcription factor families increases the tolerance of rice to other abiotic stress, such as salinity stress, and induces abscisic acid biosynthesis in response to abiotic stress in *Arabidopsis thaliana* as well (Abe et al., 2003; Tang et al., 2019). (v) The presence of the protein LOC_Os03g50410, which is related to triacylglycerol degradation provides evidence of detoxification of lipid intermediates produced by membrane lipid remodeling under abiotic stress (Lu et al., 2020), while LOC_Os09g32025, known as TKPR1, is related to sporopollenin, suggesting an active role of changes in membrane composition given that sporopollenin provides reinforcement of cell walls and protects pollen from extreme conditions (Grienenberger and Quilichini, 2021). (vi) The presence of LOC_Os07g44350 (OsNPSN13), an SNARE protein (soluble N-ethylmaleimide-sensitive factor attachment protein receptor), is related to an active role in regulating vesicle trafficking, which is a response against abiotic stress in rice, and it is hypothesized that SNAREs proteins allow root growth by regulating auxin transport (Bao et al., 2008; Luo et al., 2022; Zhu et al., 2023). Thus, the results obtained are consistent with a conserved response in plants where phytohormone presence and changes in the membrane are important as an evolutionary adaptative strategy to handle aluminum stress.

4.7. Conclusions and future perspectives

The workflow obtained herein combined with the integration of transcriptomic data provides a new framework for predicting proteins using physical or in silico predicted interactions between proteins as a new perspective to understand holistically complex phenomena, such as the abiotic stress caused by aluminum ions exposure. The introduction of unweighted and weighted interactomes embed can provide refinement for the classification for aluminum responsive proteins that cannot be elucidated using a unique type of omics technology. The two lists of predicted proteins indicate two possible scenarios to be tested experimentally: (i) a general list of new proteins filtered by two approaches (unweighted and weighted interactomes) for an extensive list of proteins and genes previously described. (ii) A list of proteins that can be tested for prolongated exposure to aluminum for two cultivars in rice. (iii) candidate protein lists suggest a differential response over time. This was observed in different proteins predicted for a general response in rice and a particular response in ten days of exposure to aluminum stress. In addition, these proteins will be as part of a genome-scale metabolic reconstruction. In the next chapter, to provide a complementary point of view that the use of this integrated approach cannot be observed, and it is how the combination of the metabolic reactions network acts together to provide the tolerance and what of the metabolic pathways are more relevant to provide new information for breeding purposes in rice.

Chapter 5

Ethylene and Auxin in response to aluminum stress affect amino acids and fatty acid metabolism in rice root meristematic cells

5.1. Link to the previous chapter

The previous chapter presented a list of possible candidate proteins that were not reported in the literature using omics-related techniques. The previous chapter reflects the complex landscape of the aluminum response by the rice plant where diverse metabolic pathways are involved, and a highly diverse response includes metabolic elements. Nevertheless, the metabolic response in the presence of aluminotoxic conditions is merely known in current literature, and there is a lack of information about how aluminum effectively disturbs the most affected zone in the plant, as is the root meristem.

Most of the current knowledge about the metabolic effects of stress comes from diverse species such as wheat, coffee, and *Arabidopsis thaliana* among other species. Even if some aspects of stress response are conserved during evolution and can be studied from a network perspective, interactomics cannot visualize stress effectively affects the cell environment and what is metabolically activated. Nevertheless, the understanding of auxin and ethylene as targets affected by the aluminum confers the opportunity to respond to how the aluminum metabolically affects the rice root meristem.

Thus, the other two research chapters wanted to offer a description of aluminum stress at the biological function level (how functionally similar is rice in comparison to a different studied species?) and find new elements that omics techniques that cannot identify themselves (e.g. what other possible proteins are possibly involved in the aluminum response in rice?). In comparison, this chapter aims to provide a holistic modeling approach to observe the consequences of stress on the plant at the metabolically level to obtain insights into how aluminum ions affect rice plants directly as is described in root growth and contribute to the specific objective three: To elucidate the aluminum effect over the metabolic response of rice, implementing a metabolic network reconstruction and analysis to generate a mathematical model that relates aluminum stress with the phenotypic response of rice plants.

5.2. Abstract

Aluminum stress mainly affects roots, causing economic losses in the rice crop. This is caused by the alteration of auxin and ethylene (two well-studied phytohormones) production in the rice root meristem by the metal, which leads to metabolic changes in the plant. Nevertheless, the metal upregulates the transcription of genes involved in the auxin and ethylene biosynthesis pathways, indirectly triggering a metabolic response that was mathematically modeled to obtain the possible metabolic effects of different concentration scenarios of aluminum on rice root growth via the implementation of a systems biology holistic approach. This approach is consistent with three stages: (i) A gene regulatory network of auxin and ethylene crosstalk including possible sensing and signaling by a receptor-like protein kinase (RLK), and Mitogen-activated protein kinases (MAPK) respectively (ii) A genome-scale metabolic reconstruction including a flux balance analysis (FBA) and (iii) a multistage FBA to dynamically model the influence of the aluminum in the phytohormones crosstalk and its metabolic effects on a rice root meristem cell through the cell cycle duration using 12 aluminum concentrations. The results of this blend of approaches reveal possible ethylene-auxin peaks production at 200 and 500 μM of aluminum, which is saturated at high aluminum concentrations (>1000 μM). A differential response of root meristem with a possible biomass production with aluminum concentration greater than 1 μM, where most of the reactions are involved in fatty acid biosynthesis and glutamate, and proline production for unique reactions with a core of 517 reactions common among aluminum concentrations. On the other hand, the multistage FBA approach considers three possible hormetic scenarios (100, 200, 500 μM), where the reaction diversity enhances biomass production but reduces the lifetime of the cell. Moreover, the results suggest that the aluminum response in the root meristem requires the presence of auxin polar transport to allow root survival and that the stage III model can be a visualization of a cell death model caused by different aluminum concentrations that reflect the metal effects on rice. On the other hand, the results indicate that aluminum stress is more complex than thought in rice, and the model potentially has caveats that do not reflect aluminum stress in real conditions. Consequently, the holistic approach was able to obtain results described widely in the literature, which can be improved by the inclusion of metabolomic information or auxin polar transport. Further experimental validation is recommended, and the exploration of possible gene regulatory networks can provide new insights into stress as a complement of the results of this chapter.

5.3. Introduction

For years, most of the aluminum stress research in rice has focused on understanding the exclusion mechanisms that plants use to obtain tolerance. For instance, a wide description of transporters that in combination with oxalate or citrate chelate aluminum ions keep out from the affected rice tissues (Kochian et al., 2015; Bojórquez-Quintal et al., 2017; Gallo-Franco et al., 2020). Nevertheless, the main agronomical effect of this stress is located in the roots where the most known physiological response is the affectation on auxin- ethylene crosstalk (Kochian et al., 2015; Kopittke, 2016; Qin et al., 2019; Chen et al., 2022; Liu et al., 2022).

This crosstalk works as follows: Aluminum ion upregulates the transcripts of methionine adenosyltransferase synthase (SAMS), 1-aminocyclopropane-1-1-carboxylate synthases (ACS), and oxidase (ACO) genes, promoting the release of ethylene as a fast burst response and auxin biosynthesis (Huang et al., 2022). Ethylene promotes the auxin biosynthesis by upregulating the TAA1, YUC5, YUC8, and YUC9 proteins (Zhou et al., 2022) which, allows the auxin accumulation and its transport to other plant organs via the polar transport via auxin transporter 1 (AUX1) and PIN-FORMED2 (PIN2) proteins to finally inhibit root growth (Liu et al., 2022). Thus, crosstalk is considered a signal integration from multiple inputs that affects a common biological output of phytohormones and multiple chemical compounds biosynthesis produced by plant metabolism, which in the case of aluminum in rice occurs mainly in the transition zone or meristematic zone of roots, which are the most sensitive parts of the plant to aluminum and inhibits cell growth and, therefore its elongation (Vert and Chory, 2011; Jingguang et al., 2020; Liu et al., 2022)

Rice is a model of study for abiotic stress. Plant stress responses are complex due to the presence of multiple interactions such as phytohormones such as ethylene, auxin, or abscisic acid (ABA), transcription factors, reactive oxygen species, and protein kinases, among others (Vandereyken et al., 2018; Huang et al., 2022). Many aspects of how aluminum affects rice metabolism are not well understood. For instance, the metabolic pathways that are affected when a plant is exposed to aluminum or how aluminum is sensed by the rice plant require the use of mathematical models such as those provided in systems biology that require expensive experimental information.

The systems biology perspective is helpful for abiotic stress. Stress responses in plants are complex, and tools such as genome-scale metabolic reconstruction or deterministic models can explain how a system works using an integrated perspective of biological information and mathematics (Cramer et al., 2011; Hillmer, 2015). These mathematical models have helped to model climate change impact and abiotic stress features until light influence in the plant (Poolman et al., 2013; Lakshmanan et al., 2015; Liu et al., 2017a; Shaw and Cheung, 2021). These tools

work without the use of literature information or even better, including experimental information, to obtain the research hypothesis to be tested further and improve plant breeding (Hong et al., 2019). Hence, the use of mathematical models in which ethylene and auxin crosstalk is functionally represented and genome-scale metabolic reconstruction together is intended in this chapter to obtain the possible metabolic effects of different concentration scenarios of aluminum on rice root growth using a systems biology holistic approach of three stages based on the fact that ethylene and auxin crosstalk is the most well-documented molecular mechanism affected by aluminum stress in plants (Sun et al., 2007; Daspute et al., 2017; Zhu et al., 2019) as well as the molecular machinery involved in the ethylene and auxin biosynthesis has been documented to be upregulated in rice under aluminum exposition (Yang et al., 2013).

5.4. Methodology

To determine the possible influence of aluminum entrance into rice root meristem cells, three stages were used: (i) construction of a deterministic model of the influence of aluminum in the gene regulatory network for the expression of auxin and ethylene in *O. sativa*, which are documented to be highly affected by aluminum stress and cause root growth inhibition (Daspute et al., 2017), thus, the purpose of this model was to provide flux reaction values to be used as constraints in a genome scale metabolic reconstruction, (ii) Refinement of the *O. sativa* genome-scale metabolic network, including auxin and ethylene biosynthetic reactions, using the flux balance analysis (FBA) approximation used to explore the possible metabolic reactions affected by the aluminum influence in meristematic rice cells ; and (iii) integration of these two network models into a unique model using the multistage FBA approximation to infer the influence of aluminum on the expression of different metabolic phenotypes related to auxin and ethylene biosynthesis (Figure 5.1 is a schematic representation of the approach used here).

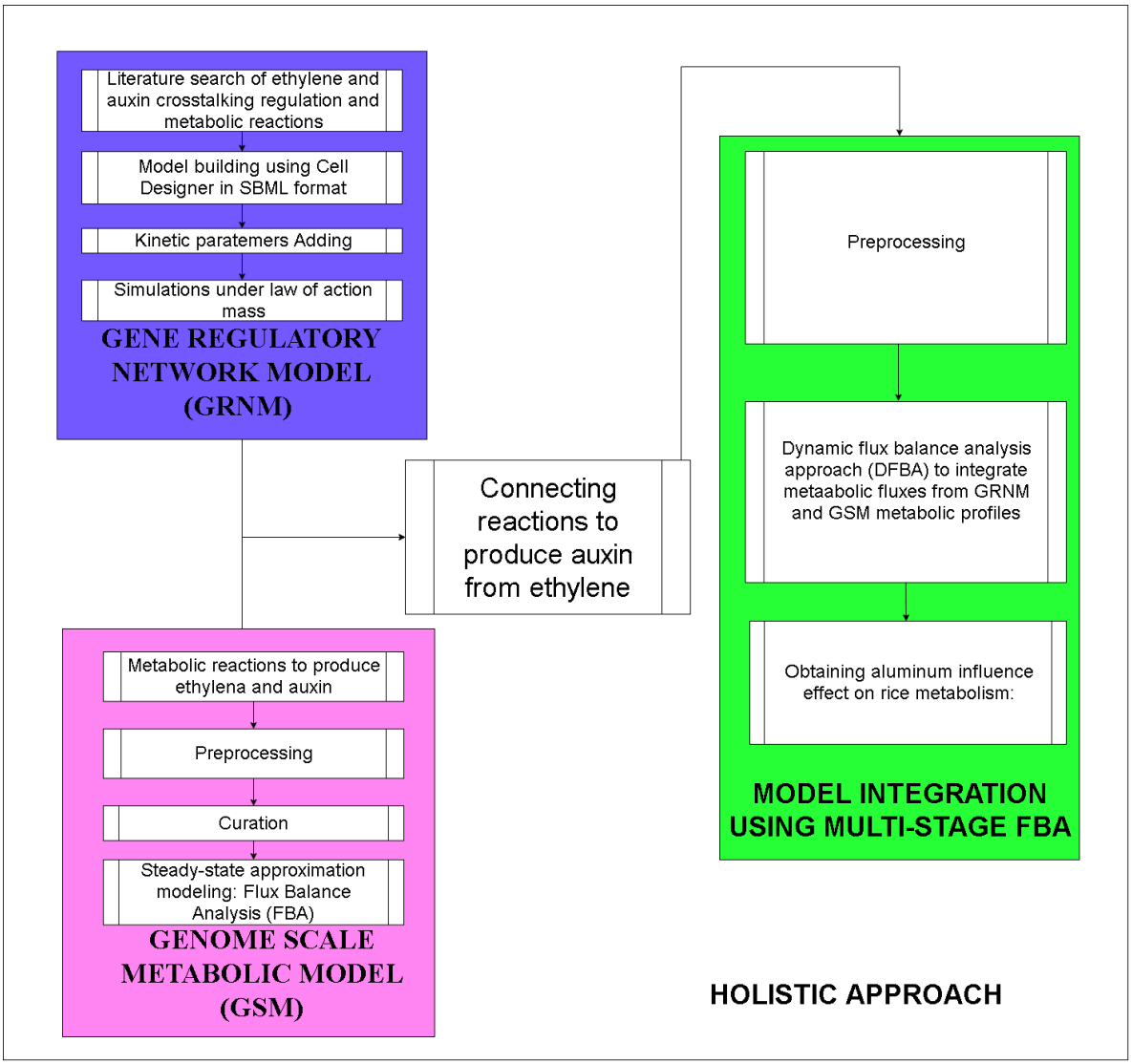


Figure 5.1. Schematic representation of the approach used in the chapter to model the metabolic influence of aluminum in rice root growth, (blue) gene network regulatory model, (violet) Genome-scale metabolic model, and (green) data integration into a dynamic flux balance approach using the shared Reaction fluxes obtained in the GRNM in the GSM as constraints to obtain the multistage FBA of the holistic approach.

5.4.1. Stage I: Deterministic model: Construction and Modeling of the Gene Regulatory Network to infer the aluminum influence in the auxin and ethylene biosynthesis in *O. sativa*

A deterministic model was designed to describe the effect of ethylene and auxin on root elongation when aluminum interacts with an Receptor-like kinases (RLK) receptor and triggers a response in the meristematic root zone which is one of the most affected tissues by aluminum toxicity in rice that finishes in the root growth inhibition (Liu et al., 2022). This effect was modeled during the

lapse of the 17-hour cell cycle and reflects the direct consequence of aluminum stress on the rice phenotype. To achieve this, a literature search was conducted to explore the crosstalk mechanisms between auxin and ethylene in plants, given that this crosstalk is the most well-described mechanism that controls root growth under stress (Jingguang et al., 2020). Moreover, the mechanisms of aluminum transport into the cells and polar auxin transport outside the cell were also included. This was done by considering their effects on genes and proteins related to aluminum stress, only focusing on the local response of meristem under aluminum stress without external auxin or ethylene stimuli. Therefore, this deterministic model aims to investigate the direct effects of aluminum on rice by (i) exploring the influence of aluminum on the transcription rate of genes involved in the auxin tryptophan-dependent and ethylene biosynthesis pathways and their regulatory mechanisms affected by aluminum sensing in meristematic cells.

The Gene Regulatory Network model to infer the aluminum influence in the auxin and ethylene biosynthesis in *O. sativa* was built using the standard System Biology Markup Language (SBML) format. Cell Designer v4.4.2 software (Funahashi et al., 2008) was used to model the biological processes to integrate auxin biosynthesis, ethylene biosynthesis pathways (Figure 5.2), and aluminum stress response mechanisms into a regulatory network by applying Systems Biology Graphical Notation (SBGN) and creating a mathematical representation of the biological model with a MathML layer.

To construct the deterministic model, elementary kinetics, based on the law of mass action for all chemical species, was assumed for all reactions in the system to generate a system of differential equations [ODEs]. For example, the differential equations for mRNAs and proteins in the transcription and translation processes are represented in Equations 5.1 and 5.2:

$$\frac{dYmRNA}{dt} = k_a[Ygene] - k_b[YmRNA] - k_c[YmRNA] \quad \textbf{(Equation 5.1)}$$

$$\frac{dYProtein}{dt} = k_1[YmRNA] - k_2[YProtein] - k_3[YProtein] \quad \textbf{(Equation 5.2)}$$

where Y represents chemical species, $\mathrm{k_a}$, the transcription rate constant, k_b, and k_1 translation rate constants whilst k_c, and k_3 the degradation rate constants, and k_2, a consumption rate of each reaction in a system (Clavijo-Buriticá, Arévalo-Ferro, et al., 2023b).

A parameter search was performed by thoroughly reviewing dedicated biological databases, including but not limited to Biomodels (Le Novère, 2006), BRENDA (Schomburg, 2004), and Sabio-RK (Wittig et al., 2012). Furthermore, we supplemented these searches by exploring the relevant scientific literature. Simulations were performed in the Cell Designer simulation platform

using the SBML ODE solver library [SOSlib] (Funahashi et al., 2008), which solves the rigidity problem of the ODEs and initiates numerical integration using the backward differentiation formula (BDF) or the Adams-Moulton (AM) method to calculate $x(t)$ for a series of time points (Machné et al., 2006). This model is available at https://data.mendeley.com/datasets/4w4pkcfm7y/4.

Given that the gene regulatory network model modeled the dynamics of aluminum sensing, which is still discussed until the biosynthesis of ethylene and auxin, the following assumptions were considered: (i) aluminum is recognized by possible receptor-like protein kinases (RLK) to initiate a secondary messenger cascade which leads to kinases activation (Poot-Poot and Hernandez-Sotomayor, 2011) (ii) kinases activation can lead to transcription factor activation. Thus, it is assumed that TF-X is a general transcription factor that can be each of the transcription factors found activated in response to stress conditions such as WRKY, MYB, MYB-related, NAC, AP-2, and bZIP, which have been described widely in plants (Pathak et al., 2013); (iii) Integration of the ethylene biosynthesis pathway involved in the conversion of methionine to S-adenosyl methionine, followed by its conversion to Aminocyclopropane carboxylate and ultimately its conversion to ethylene, catalyzed by the S-adenosyl-L-methionine synthetase (SAMS), 1-aminocyclopropane-1-carboxylic acid synthase (ACS), and 1-aminocyclopropane-1-carboxylic acid oxidase (ACO) enzymes, respectively. (iv) A similar approach was employed for the conversion of L-tryptophan into indole-3-pyruvate and indole-3-acetate (auxin), catalyzed by TAR2/TAA1 and YUC enzymes, respectively. (v) A constant presence of L-tryptophan and L-methionine in the simulations, (vi) ACS and ACO genes are considered to be always active during simulations, (vii) an alternative metabolic pathway from indole-3-butyric acid (IBA) to indole-3-acetic acid (auxin or IAA) was considered because it can reinforce the auxin response in roots (Chhun et al., 2004; Frick and Strader, 2018), (viii) ethylene exit was modeled due to the ethylene is one of the earliest responses of rice to stresses (Huang et al., 2022) (ix) Auxin was considered to be stored in the cell in a general inactive form (e.g. methyl-IAA, IBA conjugates, etc)(Korasick et al., 2013) and allowed to convert to its auxin active form, (x) auxin and ethylene were allowed to exit the cell via auxin transporters and diffusion respectively (Vandenbussche et al., 2012; Frick and Strader, 2018). (xi) Aluminum is needed to allow the production of ethylene and auxin in the gene regulatory network deterministic model, and (xii) only two compartments (cytosol and nucleus) were considered in the model, given that the reaction transport kinetic parameters for cell organelles are unknown.

The initial conditions were configured as follows: most of the genes were set with an initial concentration of 1.0 μM, for the ethylene biosynthesis enzymes SAMS gene was initialized at 10 μM to provide a sufficient amount of possible SAMS transcripts translated in SAMS enzyme, and

initiate methionine biosynthesis. ACO and ACS genes were initialized at 0.1 μM and kept constant during simulations to provide sufficient precursors for ethylene biosynthesis. Moreover, metabolic reaction precursors for ethylene and auxin biosynthesis: L-methionine and L-Tryptophan were initialized with an initial concentration of 1 μM and constant concentration during simulations under the same assumption used in, and ACS genes concentrations, nevertheless, the suggested concentrations of amino acids in rice coleoptiles tissue grown for 48 hours is around 3.91 μM g^{-1} fresh weight (Kato-Noguchi and Ohashi, 2006). Mitogen-activated protein kinases (MAPK), Mitogen-activated protein kinase kinase (MAPKK), and MAP kinase kinase kinase (MAPKKK) precursors were initialized with an initial concentration of 0.01 μM and constant concentration during simulations to provide the required precursors for the MAPK pathway normal working. As a complementary route for auxin biosynthesis, IBA-CoA, was initialized also, with an initial concentration of 0.1 μM for IBA-CoA enzyme to reflect that IBA is available in low levels in plants and stimulates lateral root elongation at 0.1 μM of IBA in rice (Chhun et al., 2004; Damodaran and Strader, 2019), and TF-X was initialized with a value of 2.0 μM, which represents average concentrations used for NAC, bZIP, and AP2 transcription factors used in Pathak et al., (2013). Finally, the decision to allow constant concentrations of chemical species was made to provide sufficient precursor concentrations for ethylene and auxin biosynthesis and signal transduction during the simulations.

5.4.1.1. Aluminum Simulation Scenario Conditions

5.4.1.1.1. Model sensitivity

To evaluate the model's usability to produce auxin and ethylene, a sensitivity analysis was conducted by modifying the aluminum ion exposure of the model using 14 aluminum concentrations from very low (0.1 μM) to very high concentrations (50000 μM). Usually, the aluminum concentrations evaluated in the literature for rice varieties or rice landraces range from,0 μM to 500 μM or even 1290 μM of aluminum (Zhang et al., 2016; Moreno-Alvarado et al., 2017; Bhattacharjee et al., 2023); however 540 μM of $AlCl_3$ has been suggested as the optimal concentration for testing aluminum tolerance in rice (Famoso et al., 2010). Finally, the aluminum concentrations (Table 5.1) were plotted together with the ethylene and auxin production of species and reaction fluxes (Table 5.1) to observe the model sensitivity.

Table 5.1. Aluminum concentrations used to test the sensitivity of the deterministic gene regulatory network to ethylene and auxin production under different aluminum stress conditions. The asterisks represent the aluminum concentration scenarios used for further analyses.

Aluminum concentration scenario Id	Aluminum concentration scenarios (μM)
Al1	0.1
Al2	0.5
Al3	1
***Al4**	20
***Al5**	50
***Al6**	100
***Al7**	200
***Al8**	500
***Al9**	1000
***Al10**	2000
***Al11**	5000
***Al12**	10000
Al13	20000
Al14	50000

After testing the model's sensitivity to the aluminum response, simulations were performed to obtain the steady state of the model. This steady state was achieved at 700 time intervals which was considered equal to the time of the cell cycle of meristem cells in plants (≈17 hours = 1020 minutes) (Hayashi et al., 2013) (Supplementary Figure 5.1). Subsequently, to evaluate the model's response to aluminum, the range of aluminum concentrations from 0.1 μM to 10000 μM (IDs Al1 to Al12, Table 5.1) was chosen for the rest of the analyses (Table 5.1). This aluminum concentration range was chosen because it includes from no-toxic to aluminotoxic conditions for sensitive rice varieties (100 μM of aluminum) (Goh and Lee, 1999; Zhang et al., 2016; Awasthi et al., 2019), an optimal aluminum tolerance point (500 μM) (Famoso et al., 2010), and excess of aluminum ions.

5.4.2. Stage II: Refinement and updating of a rice metabolic network including auxin and ethylene production

A genome-scale metabolic reconstruction was refined and updated by adding information to the previous metabolic model iOS2164 (2164 unique genes, 2283 reactions, and 1,999 metabolites) which covers central metabolism and rice metabolic subsystems (Lakshmanan et al., 2015) that was used as a template. To complement the photoperiod original model used, metabolic reactions associated with auxin-dependent tryptophan pathway represented in the indole-3-acetate

biosynthesis pathway (https://pmn.plantcyc.org/ORYZA/NEW-IMAGE?type=PATHWAY&object=PWYDQC-4), and ethylene production represented in the ethylene biosynthesis I pathway, available at https://pmn.plantcyc.org/ORYZA/NEW-IMAGE?type=PATHWAY&object=ETHYL-PWY (Table 5. **2**) were included to the reconstruction (Figure 5.2). Auxin production via tryptamine was not considered, given that tryptophan-dependent biosynthesis is the main pathway to produce auxin in rice (Zhou et al., 2022). Also, ethylene diffusion was considered as a reversible reaction.

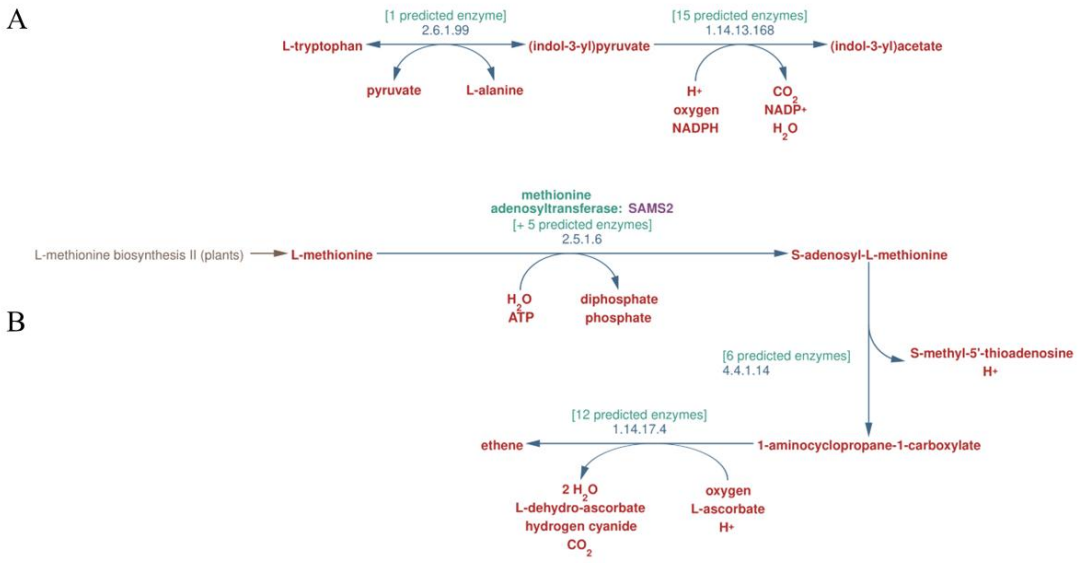


Figure 5.2. Chemical reactions used in the gene regulation network deterministic model (Stage I) and added to the genome-scale metabolic reconstruction (Stage II). Red represents the metabolic compounds; green represents the predicted enzymes acting according to the Oryzacyc database. These graphs were modified from the Oryzacyc database *(Hawkins et al., 2021)*. A) indole-3-acetate biosynthesis pathway (https://pmn.plantcyc.org/ORYZA/NEW-IMAGE?type=PATHWAY&object=PWYDQC-4) B) Ethylene biosynthesis I pathway, available at https://pmn.plantcyc.org/ORYZA/NEW-IMAGE?type=PATHWAY&object=ETHYL-PWY.

To update the iOS2614 model, directionality, Gibbs Free Energy, and metabolic subsystems for each reaction were added from (i) OryzaCyc (Hawkins et al., 2021), (ii) MetaCyc (Caspi et al., 2020) databases and (iii) Transport reactions were obtained from TransportDB 2.0 (Elbourne et al., 2017). Reactions related to ethylene and auxin are listed in Table 5. 2, and a list of the total reactions added is available at https://data.mendeley.com/datasets/3d3x4rdb78/4.

Table 5. 2. Metabolic reactions shared by the aluminum influence in the gene regulatory network model and genome-scale metabolic reconstruction. Values in parenthesis are the reaction ID present in the iOS2218 model available at https://data.mendeley.com/datasets/3d3x4rdb78/4.

Gene regulatory network model /metabolic model role	Enzyme	Reaction	EC number
S-Adenosyl-L-methionine production (R220)	Methionine adenosyltransferase	atp[c] + met-L[c] + h2o[c] -> pi[c] + ppi[c] + amet[c]	2.5.1.6
1-Aminocyclopropane-1-carboxylate production (R416)	1-aminocyclopropane-1-carboxylate synthase	amet[c] --> 1acpc[c] + 5mta[c] + h[c]	4.4.1.14
Ethylene transport (R2384)		ethylene[e] <=> ethylene[c]	
Ethylene production (R417)	aminocyclopropanecarboxylate oxidase	ascb-L[c] + 1acpc[c] + o2[c] + h[c] --> 2 h2o[c] + dhdascb[c] + co2[c] + cyan[c] + ethylene[c]	1.14.17.4
Indolepyruvate production (R2452)	L-tryptophan—pyruvate aminotransferase	trp-L[c] + pyr[c] <=> ala-L[c] + indpyr[c]	2.6.1.99
Indole-3-acetate production (R451)	indole-3-pyruvate monooxygenase	indpyr[c] + nadph[c] + o2[c] + h[c] --> ind3ac[c] + co2[c] + nadp[c] + h2o[c]	1.14.13.168
Indole-3-acetate transport (R2522)		ind3ac[c] --> ind3ac[e]	

5.4.2.1. Metabolic network curation

Curing a metabolic process refers to making the network mathematically feasible by removing pathologies such as metabolites that are not produced or consumed by any reaction in the model, eliminating thermodynamically infeasible cycles, and validating the production of metabolites in the biomass reaction (Clavijo-Buriticá , et al., 2023b). The straw biomass reaction was chosen as the objective function to maximize in the iOS2164 metabolic model (Lakshmanan et al., 2015). This biomass reaction considers biomass growth in rice straw without the production of auxin and ethylene phytohormones. Consequently, as the next step in the genome-scale metabolic reconstruction for rice, a model curation after including the phytohormones (auxin and ethylene) in the iOS2164 model reactions and objective function to ensure that the phytohormones-related metabolic pathways were active in the metabolic model since phytohormones are not part of the plant's central metabolism and allows to obtain the metabolic influence of aluminum in rice root cells in subsequent analysis steps.

The following steps were performed during the curation process:

(i) **Checking biomass precursors:** A biomass precursors check was performed to detect missing biomass precursor metabolites that were not available in the metabolic network. In this case, the biomass reaction was obtained from the straw biomass reaction used in Lakshmanan et al. (2015), which was originally used to simulate different light metabolic signatures in rice. This step was performed using the Senger engineering reverse methodology (Senger and Papoutsakis, 2008).

(ii) **Activation of ethylene and indole-3-acetate biosynthesis I (auxin) reactions:** Indole-3-acetate reactions from Oryzacyc with IDs INDOLE-3-ACETALDEHYDE-OXIDASE-RXN, and TRYPTOPHAN-AMINOTRANSFERASE-RXN were blocked because: (1) the adding of these two reactions into the reconstruction did not obtain auxin flux values. (2) With the blocking of the two reactions was ensured to only used the tryptophan dependent pathway known as indole-3-acetate biosynthesis I (BioCyc ID: PWYDQC-4) https://pmn.plantcyc.org/ORYZA/NEW-IMAGE?type=PATHWAY&object=PWYDQC-4 that is the main auxin pathway described in literature to produce the phytohormone (Korasick et al., 2013). Also, the ethylene biosynthesis I (BioCyc ID: ETHYL-PWY) https://pmn.plantcyc.org/ORYZA/NEW-IMAGE?type=PATHWAY&object=ETHYL-PWY (Figure 5.2) and both compounds were added to the biomass reaction to be used in further steps.

To obtain accurate auxin and ethylene values for inclusion in the objective function, a sensitivity analysis was performed to identify the stochiometric coefficients which maximize the production of the auxin and ethylene in meristematic cells. For this purpose, a screening of ethylene and auxin stoichiometric coefficient values from 0.02 to 1. The ranges from 0.02 to 1 were chosen to ensure that small flux values of ethylene and auxin were available in the model. These ranges from 0.02 to 1 represent 121 possible combinations of both metabolites to turn on the auxin and ethylene metabolic reactions in the objective function. As optimum values in the objective function for both metabolites, the following criteria were considered to adjust the auxin and ethylene in the objective function: (i) the auxin and ethylene values were obtained when the objective function had its highest flux value, (ii) auxin and ethylene biosynthesis connecting reactions with the deterministic model must have available flux values indicating that the corresponding metabolic pathways associated with ethylene and auxin biosynthesis were activated.

(iii) **Detecting and fixing root no-consumption and no-production metabolites:** The GapFind procedure (Satish Kumar et al., 2007; Saadat et al., 2022) was used to identify root no-consumption and no-production metabolite pathologies in the metabolic network. A manual Gap-fill procedure was performed to reduce the number of metabolite pathologies taking into consideration the following criteria: (a) ubcellular locations for all reactions were obtained using OryzaCyc and cropPAL (Hooper et al., 2020) , based on gene IDs linked to each reaction, (b) the metabolic pathways in which shared compounds were found in the iOS2164 model, and (c) the inclusion of reactions from the literature, when available.

The manual gap-filling procedure involved the following steps: (a) detect the pathway where the no production or consumption metabolites were identified, (b) determining the minimum number of reactions required to fill the gap and produce or consume the metabolite within the same compartment and pathway, (c) when required reactions occurred in different compartments, adding transport reactions supported by literature or data from TransportDB 2.0, and (d) for metabolites not found to be consumed in other reactions based on databases or literature, adding demand reactions.

(iv) **Search and solution of thermodynamically infeasible cycles (TICs) in the network**: Thermodynamically infeasible cycles (TICs) were removed from the metabolic network using Murabito and Schellenberg methodologies (Murabito et al., 2009; Schellenberger et al., 2011). Flux Variability Analyses (FVA) were performed to identify fluxes lesser than -900 and greater than 900. Furthermore, the null space was calculated for reactions with thermodynamically infeasible loops, and null spaces were calculated using MATLAB R2021A until the TICs were eliminated. Only null space values with one decimal point were considered for this step using when it was possible; thus, the lower and upper values were modified according to it or blocked if needed. All scripts used for the preprocessing and curation are available at https://github.com/ccsosa/Rice-aluminum-stress-GENRE-and-Deterministic-model, and the metabolic reconstruction file is available at https://data.mendeley.com/datasets/3d3x4rdb78/4.

5.4.2.2. Steady-state approximation modeling: Flux Balance Analysis

A Flux Balance Analysis (FBA) was performed by maximizing the biomass reaction as an objective function using a linear programming to solve the optimization problem (Equation 5.1)

implementing the CPLEX v.12.6.0.0 solver in the GAMS software (General Algebraic Modeling System: https://www.gams.com) (Equation 5.3).

$$Maximize\ z = v_j(biomass)$$

Subject to: **(Equation 5.3)**

$$\sum_{j=1}^{J} S_{ij} * v_j = 0, \qquad \forall i=1,\ldots,I$$

$$LB_j \leq v_j \leq UB_j\ , \qquad \forall j=1,\ldots,J$$

$$v_{o2} = K_1$$

$$v_{H2O} = K_2$$

$$v_{photon} = K_3$$

where z represents the objective function that denotes the biomass rate obtained as the weighted sum of reactions synthesizing biomass precursor, S_{ij}, represents the stoichiometric matrix including a metabolite i in a reaction j; v_j is the flux value of the reaction j. Otherwise, LB_j, and UBj are vectors of lower and upper bounds allowed for the reaction j. For irreversible reactions, LBj, and UBj values were 0 and 1000 and for reversible reactions, the LB_j, and UBj values were -1000 and 1000 (Maranas and Zomorrodi, 2016b).

To reflect similar environmental conditions found in paddy fields where rice is cropped, the fluxes for water, oxygen, and photon uptake reactions were limited to 10 mmol $gDW^{-1}h^{-1,}$ which are represented by K_1, K_2, K_3 respectively. These limitations were determined based on the availability of the compounds in the soil and the possible photon uptake of rice plant to perform underwater photosynthesis and limited the uptake of water and oxygen, as observed in paddy field rice during daylight (Winkel et al., 2013).

5.4.3. Stage III: Multi-Stage FBA Approximation: integration of the Gene Regulation Network and Genome-Scale Metabolic Network models following the methodology proposed by Clavijo-Buriticá et al. in 2023

The approach used here was proposed by Clavijo-Buriticá et al. (2023b) and was originally used to model the synthesis of virulence factor pyoverdin (PVD) regulated by the quorum-sensing (QS) phenomenon in *Pseudomonas aeruginosa.* QS regulation involves five regulatory proteins known as the *Pseudomonas* Quinolone Signal (PQS) system, which is initiated by an extracellular signal molecule, PQS (E-PQS). The methodological strategy combined two classes of models with different scales. The first model is a gene regulatory network model that controls PVD synthesis including different concentrations of E-PQS, while the second model (CCBM1146 model) implies

the construction of a metabolic network using flux balance analysis (FBA) to produce PVD. Finally, the two models were integrated on a dynamic flux balance analysis (DFBA) applied to the CCBM1146 model using the nine reaction flux values shared by both models and used as constraints to obtain profiles of biomass and affect the metabolic profiles of *P. aeruginosa* when PVD is produced (Clavijo-Buriticá, et al., 2023b).

Consequently, the Multi-Stage FBA Approximation presented here involves (i) the construction of a gene regulatory network for the production of auxin and ethylene, starting with the sensing of different concentrations of aluminum (Table 5.1) by an RLK receptor (this is explained in section 5.4.1. (ii) An updated genome-scale metabolic reconstruction named iOS2218, based on the iOS2164 metabolic model, is presented in Section 5.4.2. (iii) The Multistage FBA approximation consists of (i) obtaining a list of connecting reactions between both models, as presented in Table 5.3. (ii) Transform the flux reaction values for the connecting reactions from $\mu\text{mol}\ s^{-1}L^{-1}$ to $\text{mmol}\ \text{gDW}^{-1}h^{-1}$ to use the same units; (iii) create simulation scenarios of reactions (SSR) (Table 5.3) using combinations of the connecting reactions to observe which of these scenarios produces the most biomass; (iv) use the aluminum concentration scenarios described in Table 5.1 to use the dynamic flux balance analysis to obtain biomass profiles under several aluminum concentrations, and (v) use a clustering algorithm to obtain metabolic fluxes profiles. The steps are explained in detail as follows:

5.4.3.1. Simulation of aluminum influence over rice meristem cell under dynamic flux balance analysis (DFBA)

The stationary state of the gene regulation network model was achieved at 700 intervals, representing 17 hours of the cell cycle (Supplementary Figure 5.1). Given that the gene regulatory network model reaction fluxes were expressed in μmol $s^{-1}L^{-1}$, they were adjusted to mmol $gDW^{-1}h^{-1}$ using the following Equation (Equation 5.4):

(**Equation 5.4**)

$$\mu\text{mol}\ s^{-1}L^{-1} * \frac{1\ mmol}{1000\ \mu\text{mol}} * \frac{3600\ s}{1\ h} * \frac{1L}{7.6\ gDCW} = 0.473684211\ \text{mmol}\ \text{gDW}^{-1}h^{-1}$$

where the cell density used (gDCW) was derived from Brown and Beevers, (1987), which represents a rice cell culture containing Murashige and Skoog salts, 0.4 mg/1 thiamine HCI, 1.0 mg/l myoinositol, 2.0 mg/1 2,4-D, 5.0 g/l yeast extract (Sigma), 20.0 g/1 sucrose, and 8.0 g/l agar at eight days of establishment in a reference culture media ≈ (380 mg DCW in 50 mL of reference medium).

5.4.3.2. Simulation scenarios modeling for dynamic flux balance analysis

Under the multistage FBA approximation, both network models, the gene regulatory network deterministic model (stage I) and the genome-scale metabolic reconstruction (stage II) were combined into a dynamic flux balance analysis (DFBA) under a dynamic optimization-based approach (DOA) (stage III) (Clavijo-Buriticá, et al., 2023b; Mahadevan et al., 2002). It was run for 17 hours of the cell cycle of a rice meristematic cell) (Hayashi et al., 2013).

A total of seven reactions shared between the gene regulatory model (stage I) and the genome-scale metabolic model (stage II) were used for the multistage FBA approach (Table 5.2), and a total of six simulation scenarios of reactions (SSR) were designed to observe the effect of aluminum influence on rice metabolism considering possible combinations of reactions involved in ethylene and auxin biosynthesis and transport (Table 5.3).

The six simulation scenarios represent possible reactions in which aluminum influences rice metabolism. The first scenario (SSR1, Table 5.3) represents all the possible reactions involved in ethylene and auxin biosynthesis and extracellular transport, including the S-Adenosyl-L-methionine production, in which aluminum does not influence the gene regulatory network model. In the second scenario (SSR2, Table 5.3), aluminum influenced the direct reactions, creating the analyzed phytohormones. The third scenario (SSR3, Table 5.3) contained ethylene production and transport reactions, except for the S-Adenosyl-L-methionine and indole pyruvate production. This scenario SSR2 represents a scenario where the initial reactions for ethylene and auxin biosynthesis are not affected by aluminum in the root meristem. For SSR4 (SSR4, Table 5.3), the reactions of S-Adenosyl-L-methionine production, ethylene production, and transport, as well as indole pyruvate production and auxin transport, are present and represent the effect on the initial reaction of ethylene production and transport of ethylene with indole pyruvate production and auxin transport. Scenario five (SSR5, Table 5.3) corresponds to ethylene production, transport, and two auxin biosynthesis pathway reactions. Finally, in Scenario Six (SSR6, Table 5.3), the 1-Aminocyclopropane-1-carboxylate production, ethylene production, and the auxin biosynthesis pathway reactions were considered. None of the scenarios excluded the production of ethylene or auxin given that aluminum alleviation uses phytohormone crosstalk (Ranjan et al., 2021a) (Table 5.3).

Table 5.3. Simulation scenarios of reactions (SSR) intended in the multistage FBA approach.

Gene regulatory network model /metabolic model role	Scenario 1 (SSR1)	Scenario 2 (SSR2)	Scenario 3 (SSR3)	Scenario 4 (SSR4)	Scenario 5 (SSR5)	Scenario 6 (SSR6)
S-Adenosyl-L-methionine production (R220)	X			X		
1-Aminocyclopropane-1-carboxylate production (R416)	X		X			X
Ethylene production (R417)	X	X	X	X	X	X
Ethylene transport (R2384)	X		X	X	X	
Indolepyruvate production (R2452)	X			X	X	X
Indole-3-acetic acid production (R2451)	X	X	X		X	X
Auxin transport (R2522)	X		X	X		

For the multistep FBA approach, Equation 5.5 which represents a steady step FBA (stage II) was modified as follows to include the flux values obtained in stage I for the diverse simulation scenarios (Table 5.3):

$$Maximize\ z = v_j(biomass)$$

Subject to: (**Equation 5.5)**

$$\sum_{j=1}^{J} S_{ij} * v_j = 0, \qquad \forall i=1,\ldots,I$$

$$LB_j \leq v_j \leq UB_j, \qquad \forall j=1,\ldots,J$$

$$v_{o2} = K_1$$

$$v_{H2O} = K_2$$

$$v_{photon} = K_3$$

$$v_j = v_{\hat{j}}, \qquad \forall \hat{j}, \epsilon \text{ SS Reactions}$$

$$\forall t\ \epsilon\ [t_0, t_f]$$

where z, S_{ij}, v_j is the LB_j, and UB_j are defined in Equation 5.3, $\hat{j}$, represents the common reactions between the gene regulatory network deterministic model and the genome-scale metabolic reconstruction (Clavijo-Buriticá et al., 2023a) according to each simulation scenario (SSR)

represented in Table 5.3, and t_0, t_f are the initial and final times respectively (0 and 1020 minutes ≈ 17 hours of cell cycle) (Hayashi et al., 2013).

The optimization problem provided for the multistage FBA (Equation 5.5) was solved as a linear programming (LP) problem using the CPLEX v.12.6.0.0 solver in GAMS (https://www.gams.com/latest/docs/S_CPLEX.html). The simulations were run to obtain (i) the biomass flux profile of rice meristem cells under different aluminum stress ranges represented in Table 5.1, (ii) the metabolic flux profile under the influence of aluminum represented in the biosynthesis and extracellular transport of ethylene and rice over rice cell metabolism, and time in six simulation scenarios reactions (SSR) representing combinations of the seven common reactions between models (Table 5.3). Finally, the SSRs with biomass flux values were considered for further analysis.

5.4.3.3. Metabolic fluxes profiles obtention

After the obtention of the reaction fluxes from the multistage FBA approach for the 12 aluminum concentration conditions (Table 5.1), a K-means clustering was performed considering as variables the fluxes of the 2565 reactions of the metabolic model, the combination of the 500-integration steps and the aluminum concentrations scenarios (IDs Al1 – Al12, Table 5.1) as individuals (e.g., Integration step 1- 50 μM of aluminum). A Yeo-Johnson power transformation (Yeo and Johnson, 2000) was performed on the reaction fluxes with a variance greater than 0 to reduce noise and consider only reactions with changes in reaction flux values in the analysis. The optimal number of clusters was determined using the *silhouette coefficient*, and the reaction fluxes were plotted as heat maps per aluminum selected concentration scenarios. Upset plots were used to observe the most frequently found metabolic subsystems. Variable transformation, number of optimal clusters, K-means clustering, and heatmaps were performed in the caret v6.0-94, factoextra v1.0.7, stats v4.1.3 R package, in pheatmap v1.0.12, and upset plots were plotted using UpsetR 1.4.0 R packages. All codes, results and supplementary information is available at https://github.com/ccsosa/Rice-aluminum-stress-GENRE-and-Deterministic-model.

5.5. Results

5.5.1. Gene regulatory network deterministic model description

A full graphical representation of the gene network regulatory network to produce ethylene and auxin under aluminum stress is presented in Figure 5.3.

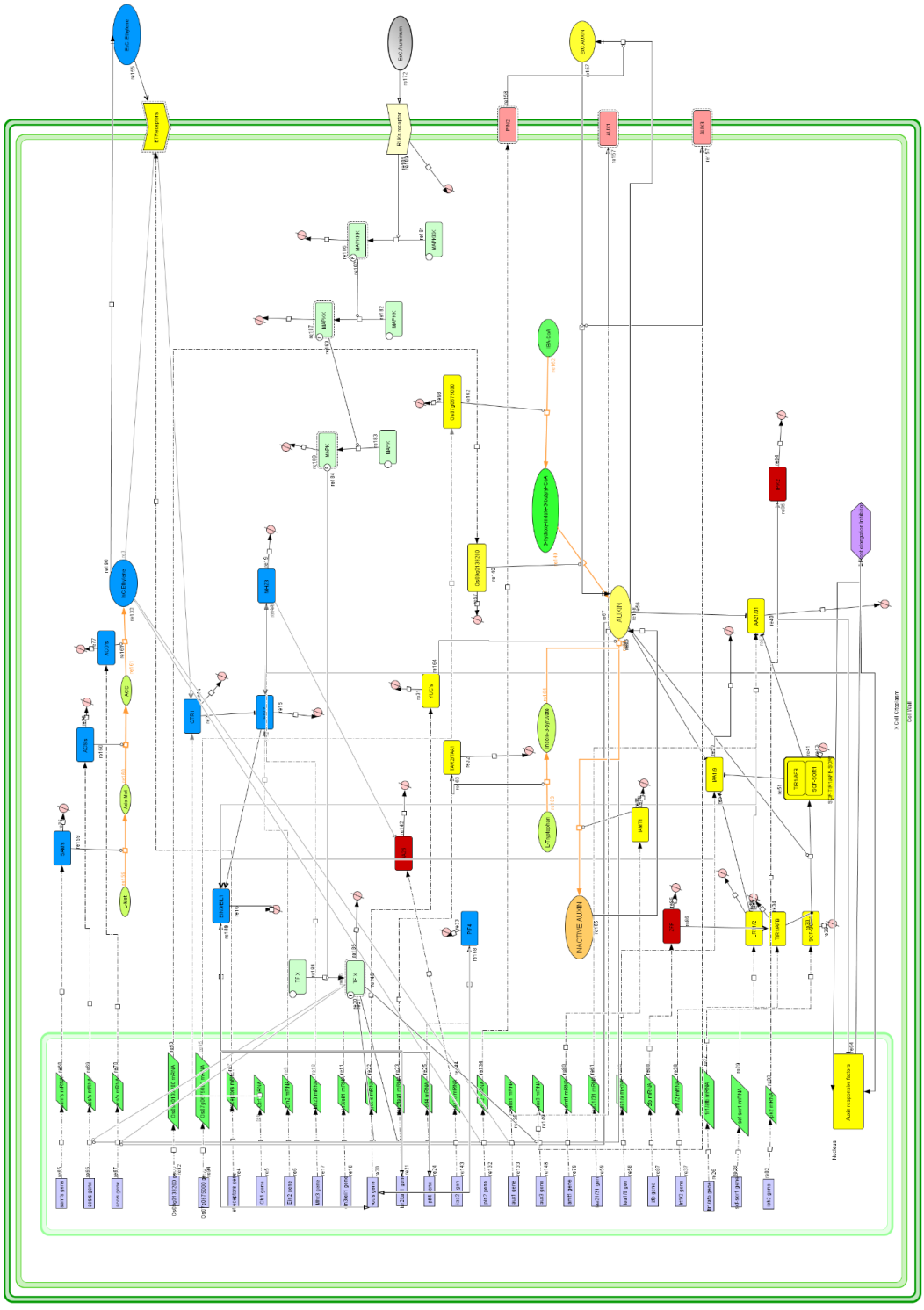


Figure 5.3. Schematic representation of the deterministic gene regulatory network model used to obtain flux values for ethylene and auxin biosynthesis. The two compartments were described in the model (nucleus and cytosol). Genes are represented on the left side of the figure in purple rectangles, whereas transcripts are represented as green rhomboids in the nuclear compartment.

The proteins are represented as rectangular shapes in the cytosol. Blue and yellow rectangles represent proteins involved in ethylene and auxin biosynthesis. Extracellular aluminum (green ellipse) is perceived in the RLK receptor (yellow shape in the cell membrane); after sensing, MAPK cascade signaling is activated, phosphorylating MAPKKK, MAPKK, and MAPK (the light green rectangular shapes in the cytosol) to activate a potential transcription factor X (TF X), which activates ACS, ACO, YUC, and TAR2/TAA1. Simultaneously, SAMS is represented by a blue color in the cytosol. AUX1 and AUX3 transcription is responsible for ethylene and auxin biosynthesis reactions (light green ellipses represent chemical species involved in hormone biosynthesis). Red rectangles (ZFP and IPK2) represent transcription factors linked to auxin. Yellow rectangles (IAMT1, IAA1/9, LRT1/2, TIR1/AFB, SCF/SOR1, and IAA21/31) are responsible for auxin regulation, which activates auxin-responsive factors that inhibit root elongation. Pink rectangles (AUX1, AUX3, and PIN2) represent auxin transport, including polar transport. Finally, pink rectangles represent protein degradation.

The gene regulatory network model obtained from stage I achieved a stationary state at 500 integration steps in 700-time intervals representing the 17 hours of duration of a cell cycle in a root meristem cycle in rice. The gene regulatory network model consisted of 127 species: 25 genes, 25 mRNA, 35 proteins, one protein complex, 22 degraded proteins, and 13 molecules (Supplementary Table 5.1). These chemical species interacted in 117 reactions including 25 transcription and 25 translation reactions, a unique heterodimer association, 12 negative influences, six positive influences, 40 transition states, three transport reactions, and five reduced physical stimulations were also presented.

5.5.1.1. Ethylene maximum production in rice meristem cells occurs at 500 μM of aluminum concentration

The results obtained from the gene regulatory network deterministic model suggested that ethylene production occurred for all aluminum scenarios (IDs Al1 – Al12, Table 5.1.). This production is presented in Table 5.1). Moreover, ethylene production is limited to almost no ethylene production in aluminum concentration scenarios less than 20 μM (IDs Al1 - Al3, Table 5.1), and ethylene is considerably higher in concentration and faster in production when aluminum concentrations are greater than or equal to 20 μM aluminum concentration (IDs Al4 – Al12, Table 5.1).

Notably, ethylene production peaked in a shorter time when the RLK receptor sensed high aluminum concentrations (IDs Al6 – Al12, Table 5.1). For instance, in the 50 μM aluminum concentration scenario (ID Al5, Table 5.1), the ethylene production started to increase at almost one hour and a half and decreased severely at almost four hours of the meristem cell cycle (Figure 5.4A). Furthermore, ethylene production reached its highest peak at the 500 μM aluminum

concentration scenario (ID Al8, Table 5.1) which started to increase from 46 minutes of the meristem cell cycle, producing a production peak at almost one hour and decreasing to no ethylene production in two hours of the meristem cell cycle. Additionally, for aluminum scenarios greater than 500 µM (IDs Al9-Al12, Table 5.1), the ethylene production curves decreased faster than those found in other aluminum scenarios. Ethylene production started from 30 minutes and decreased at two hours of the meristem cell cycle for aluminum concentration scenarios of 1000 µM, 2000 µM, 5000 µM, and 10000 µM, (Al9 - Al12, Table 5.1) (Figure 5.4A).

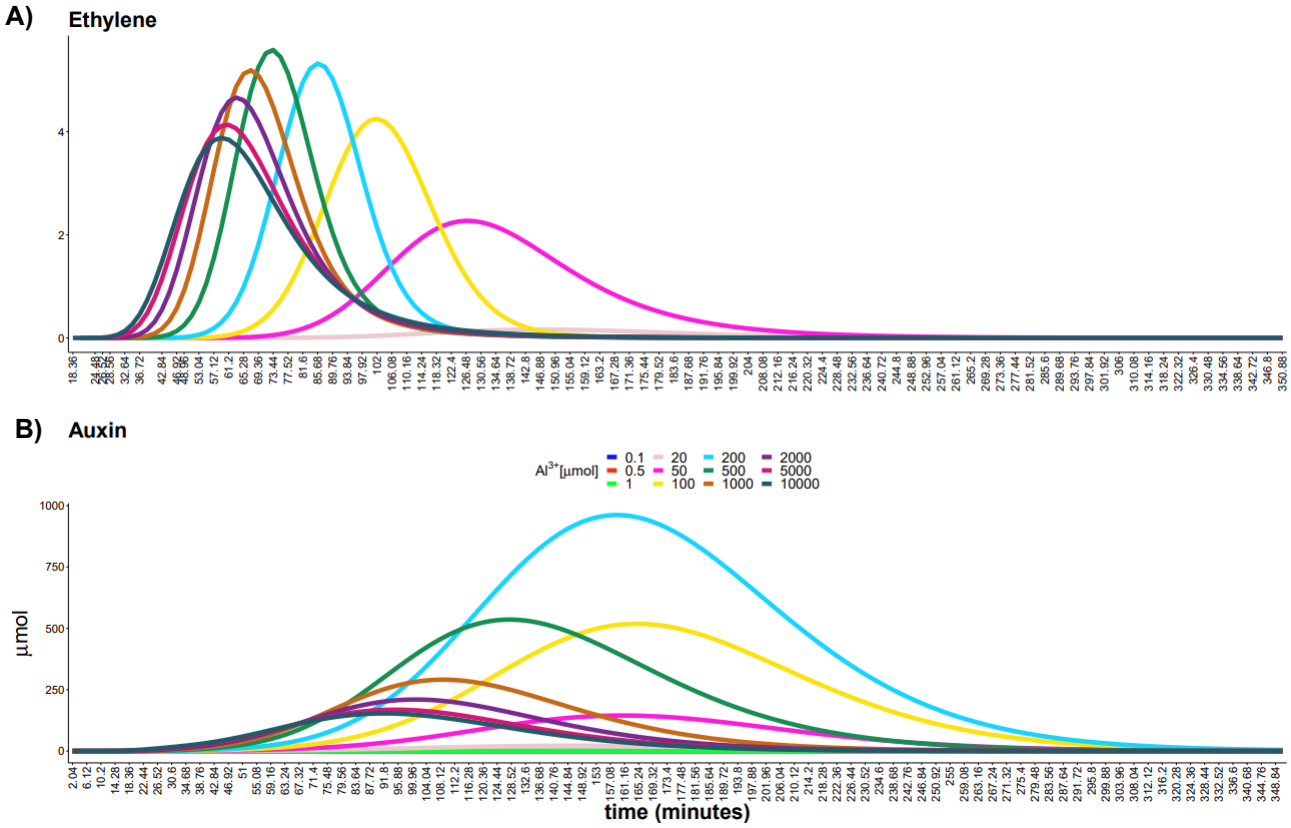


Figure 5.4. Phytohormone concentrations in micromoles for 12 aluminum concentration simulation scenarios for a time of six hours of the meristematic cell cycle obtained by the gene regulatory network model. A) Ethylene, B) Auxin. The colors represent the aluminum concentration simulation scenarios (Table 5.1).

5.5.1.2. Auxin production in rice meristem cells reaches its maximum production at 200 µM of aluminum

The gene regulatory network model was used to determine the auxin production in a rice root meristematic cell. This auxin production achieved higher concentrations than the ethylene production observed for the aluminum concentration scenarios simulated of 200 – 1000 µM (IDs Al7 – Al9, Table 5.1). In addition, a similar trend of production peaks for auxin and ethylene was observed (Figure 5.4A and 5.4 B), where there was a high peak of production at aluminum concentration scenarios of 200 – 1000 µM (IDs All7- Al9, Table 5.1) and auxin production decreased when high aluminum concentration scenarios such as greater than 2000 µM (All11, and A12, Table 5.1). However, for the difference observed in ethylene, the auxin maximum production peak was reached at 200 µM aluminum concentration scenario (ID Al7, Table 5.1) (Figure 5.4 and 5.4B).

It is important to highlight that auxin production started to increase after an almost one-hour cell cycle and increased until the maximum peak of production occurred at around three hours and finally reached almost no auxin production at five hours of cell cycle for the aluminum concentration scenarios lower than 500 µM (IDs Al1 - Al7, Table 5.1). In addition, for the aluminum concentration scenarios greater than 200 µM (IDs Al8 – Al12, Table 5.1), the increase of auxin production occurred quickly (after 30 minutes of the start of the cell cycle), and auxin production finished at approximately five hours of the cell cycle. Consequently, the auxin was produced slower than the ethylene production which agrees with Růžička et al., (2007). In addition, auxin production peaks for aluminum concentration scenarios occurred at aluminum concentrations greater than 20 µM (Table 5.1) when ethylene production decreased to almost zero (Figure 5.4A and 4B).

The flux reaction distributions obtained for the three reactions shared between the models for the ethylene biosynthesis pathway in the gene regulatory network model were as follows:

(i) The reaction mediated by S-methionine adenosyltransferase synthase (SAMS) had its maximum peak of S-adenosyl-L-methionine production early and decreased at two hours of the cell cycle (Supplementary Figure 5.2). It remained without changes in fluxes regardless of the aluminum concentration simulation scenarios applied because this reaction has no influence from the aluminum ion entering to the simulated meristem cell.

(ii) The ACS-mediated reaction reached its maximum reaction flux peak of 1-aminocyclopropane-1-carboxylate production at one hour for the 500 µM aluminum

concentration scenario (ID Al8, Table 5.1). Nevertheless, in the highest aluminum concentration scenarios (> 500 μM), (IDs Al9 - Al12, Table 5.1), fast decays in the production of 1-aminocyclopropane-1-carboxylate were observed (Supplementary Figure 5.2B).

(iii) The ACO-mediated reaction that synthetized ethylene followed a similar behavior as the ACS-mediated reaction; however, its maximum flux production was observed for the 500 μM aluminum concentration scenario (ID Al8, Table 5.1) and the flux values were lower than observed in the ACS-mediated reaction (Supplementary Figure 5.2C).

(iv) For ethylene transport, the same behavior was observed at 500 μM aluminum concentration simulation scenario (ID Al8, Table 5.1) as observed in the ACO and ACS mediated reactions, but also, the flux values were higher than those of the ethylene reaction. Thus, it is suggested that ethylene production is diffused rapidly to the extracellular environment in response to aluminum stimulation of the RLK receptor (Figure 5.3).

As observed in auxin production, the two reactions involved in the auxin biosynthesis pathway and transport to the extracellular environment obtained their maximum values in 200 μM aluminum concentration scenario (ID Al7, Table 5.1), and the reactions reached their maximum production peaks when ethylene production decreased. Both the tryptophan to indol-pyruvate and the indol-pyruvate to auxin reactions (Table 5. 2) followed a similar trend, with a maximum production peak at 2.6 hours of the cell cycle. On the other hand, auxin transport achieved its maximum peak at three hours, which was substantially lower than ethylene transport. Therefore, it is suggested that auxin biosynthesis reactions occur longer than ethylene reactions, and auxin accumulation occurs in meristem cells (Růžička et al., 2007).

5.5.2. Genome-scale metabolic reconstruction in steady state description (Stage II)

A total of 121 reactions, including auxin and indole butyric acid (IBA) metabolic pathways, transport, and gap-filling reactions, were added to the previous iOS2164 model (this data is available at https://data.mendeley.com/datasets/3d3x4rdb78/4). Consequently, a total of 2565 reactions, 2043 metabolites, 2218 genes, and eight compartments in the metabolic reconstruction were available (Table 5. ***4***).

Table 5. 4. Summary table of the genome-scale metabolic reconstruction (iOS2218) and its comparison with the iOS2164 model proposed by Lakshmanan et al., (2015).

		Model
Reactions and components	**iOS2164**	**iOS2218**
Genes	2164	2218
Reactions	2283	2565
Metabolites	1999	2043

Regarding possible pathologies of the metabolic network, 203 metabolites (9.94%) are in the metabolic network. (Supplementary Table 5.3). No thermodynamically infeasible cycles (TIC) were found in the metabolic network after applying the Murabito and Schellenberg methodologies (Murabito et al., 2009; Schellenberger et al., 2011). Nevertheless, after applying this methodology, 48 reactions were close either to the lower (<-900 mmol $gDCW^{-1}hr^{-1}$) or upper (>900 mmol $gDCW^{-1}hr^{-1}$) bound limits, mostly related to exchange reactions, photon uptake, and transport.

5.5.3. Multistage flux balance approach simulation reveals a possible impact of high aluminum concentrations to start sooner the biomass production in the rice root meristem

After screening of the six simulation scenarios of reactions (SSR) (Table 5. ***2***), only scenario two (SSR2) yielded results for biomass production after fixing the values of reactions R416 and R2451 (Table 5. ***2***, and 5.3), which are the main reactions that produce ethylene and auxin, respectively (Supplementary Table 5.4). Thus, only the SSR2 and the aluminum concentration scenarios Al4 – Al12 (Table 5.1) were considered for further steps, suggesting a possible direct influence of aluminum on ethylene and auxin biosynthesis, but not on their intermediate reactions.

Biomass production was present in nine of the twelve aluminum concentration scenarios (from 20 µM to 10000 µMm; IDs Al4 – Al12, Table 5.1). The response of biomass production curves followed a similar trend as observed in the flux values of the aminocyclopropanecarboxylate oxidase reaction flux values under higher aluminum concentrations scenarios, ethylene production was low and the peak of production occurred in 500 µM aluminum concentration scenario (ID Al8, Table 5.1). Smooth curves of biomass production were observed for the low aluminum concentration scenarios. But also, for the aluminum concentration scenarios of 100 µM, 200 µM, and 500 µM (IDs Al6 – Al8, Table 5.1), the biomass showed a strong increase in

comparison with lower aluminum concentration scenarios. Hence, the biomass decreased until stable production was achieved and decayed strongly until almost no biomass was produced (Figure 5.5C).

When the aluminum concentration increased, biomass production started earlier. This start ranged from one hour (20 μM Aluminum concentration scenario, Al4, c) to 22 minutes (>2000 μM Aluminum concentration scenario; AlI11 and Al12, Table 5.1). This suggests that the higher the aluminum concentration, the sooner biomass production starts (Figure 5.5C). However, the duration of biomass production under the aluminum concentration scenarios of less than 100 μM (IDs Al4-Al5, Table 5.1) was greater than five hours. At aluminum concentrations greater than 50 μM, it increased from four and a half to five hours (IDs Al6 – Al12, Table 5.1). Hence, it suggested that aluminum stimulus greater than 50μM (IDs Al6 – Al12, Table 5.1) affects the cell cycle of rice root meristems.

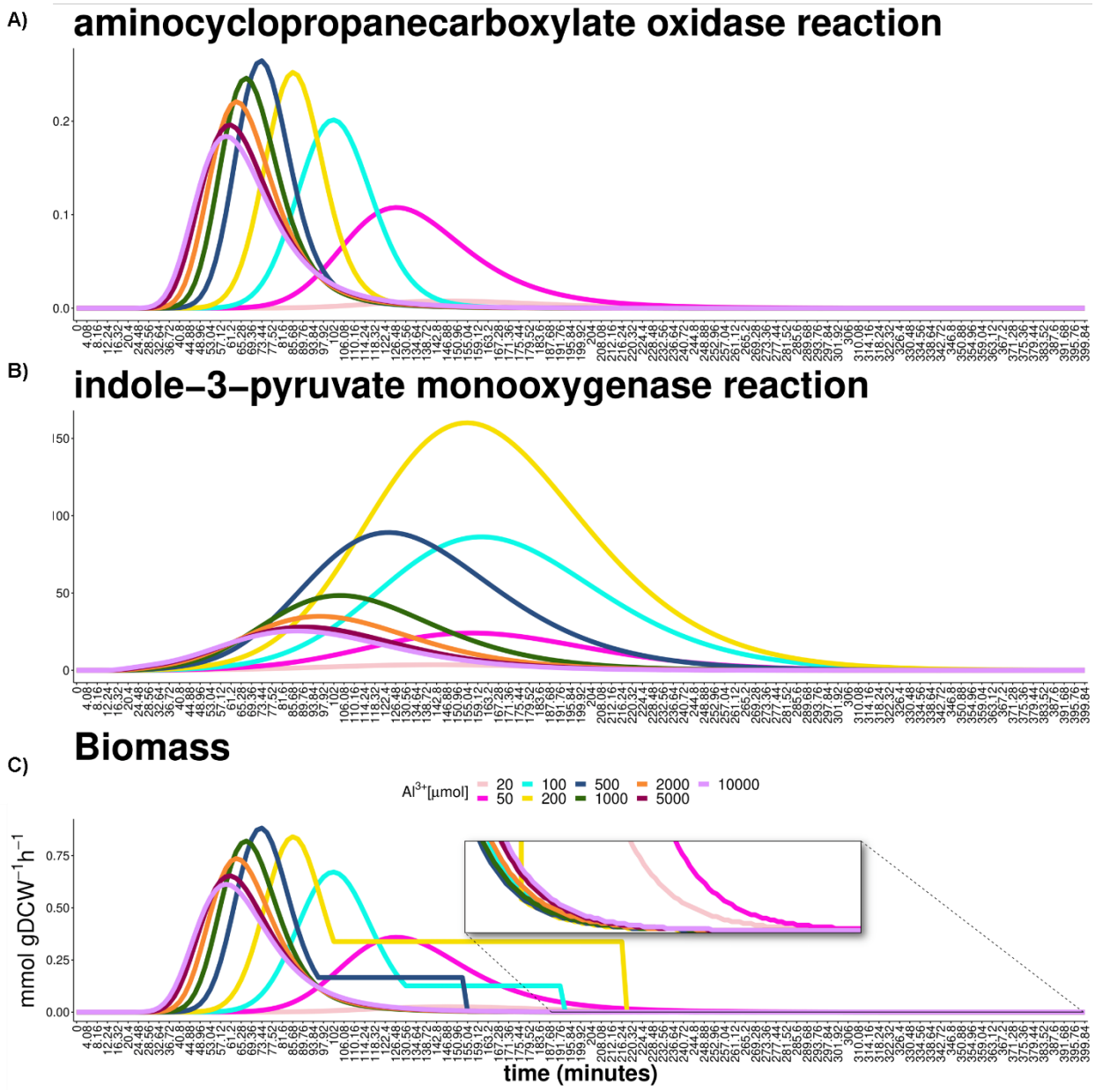


Figure 5.5. Simulated ethylene, auxin obtained from the gene network regulation model (stage I), and biomass production for nine aluminum concentration scenarios in 5.43 hours of rice meristem cell cycle when biomass is produced using the Multi-Stage FBA Approximation (stage III). Each color represents the aluminum concentration scenarios from 20 μM to 10000 μM (Table 5.1). A. aminocyclopropanecarboxylate oxidase reaction B. indole−3−pyruvate monooxygenase reaction C. biomass function. The box highlights flux values for biomass for intervals between three and seven hours of the cell cycle and flux values less than 0.0005 mmol $gDCW^{-1}hr^{-1}$.

5.5.3.1. A differential metabolic response with changes in reaction fluxes is suggested for medium aluminum concentration scenarios

To observe the metabolic similarities among the aluminum concentration scenarios using the reaction and flux values, non-metric multi-dimensional scaling (NMDS) was performed for all 12 aluminum concentration scenarios (Table 5.1) using the results of the multistep FBA approach. The NMDS showed that the aluminum concentration scenario of 20 μM (ID Al3, Table 5.1) had more similar reaction flux values over time, which indicates that there is a similar behavior of reaction flux values, even when the biomass increases (Figure 5.5C). On the other hand, for aluminum scenarios with 100 and 200 μM (IDs Al6, and Al7, Table 5.1), the changes in reaction flux values were more evident, as they were observed as outlier points in the graphs that corresponded to the biomass decay. In contrast to the behavior observed previously at higher aluminum concentrations (IDs Al10 - Al12, Table 5.1), the reaction flux values were similar among them. This pattern, similar to that observed in the scenario 20 μM of aluminum (ID Al3, Table 5.1), supports the hypothesis that exposure to medium concentrations promotes the activation of reactions in rice root, and consequently increases biomass production values.

5.5.3.2. Metabolic profile clustering reveals two main metabolic profiles and a possible role of folate and purine metabolism in the adaptation of rice to aluminum stress

The k-means clustering analysis revealed two metabolic profile clusters. A total of 181 reactions were found exclusively for cluster one. Most of these reactions involve Fatty acid biosynthesis, Primary Cell Wall Metabolism (Fructose), Pyrimidine metabolism, and Transport (plastidic) subsystems. In contrast, a total of eight reactions were exclusive to cluster two: Glucose exchange, glucose transport via diffusion (extracellular to cytosol), Cytidine deaminase, GTP:cytidine 5-phosphotransferase (plastidic), ATP:cytidine 5-phosphotransferase (plastidic), Tetrahydrofolate transport diffusion, (mitochondrial), 5,10-Methylenetetrahydrofolate uptake carrier (mitochondria), and Malate dehydrogenase (oxaloacetate decarboxylating) (NADP+) (plastidic).

In general, for the aluminum concentration scenarios from 100 to 500 μM (IDs Al6 – Al8, Table 5.1), the cell cycle, which comprised four and a half to five hours, belonged exclusively to cluster one. On the other hand, the cell cycle from 0 minutes to approximately 35 minutes in the aluminum concentration scenarios, and the metabolic response between one and a half and two hours or when a biomass production extension occurred were exclusively part of cluster two (Supplementary Figure 5.3).

5.5.3.3. Aluminum stress promotes fatty acid, and amino acid biosynthesis as a metabolic response

Hierarchical clustering using the presence of a reaction was performed to observe the similarity in the rice meristem metabolic response among the aluminum concentration scenarios (Table 5.1). The results showed that the 20 µM aluminum concentration scenario (IDs Al4, Table 5.1) was different from the other aluminum scenarios. The aluminum concentration scenarios of 50, 100, and 200 µM (IDs Al5 – Al7, Table 5.1) created a separate cluster linked to the cluster of the scenarios of 50, 1000, 2000, 5000, and 10000 µM (IDs Al5, Al9 – Al12, Table 5.1). Thus, the results suggested that there was a variable number of reactions during the simulations and a distinctive cluster of medium aluminum concentrations (Table 5.1) as observed in the biomass profile results (Figure 5.6).

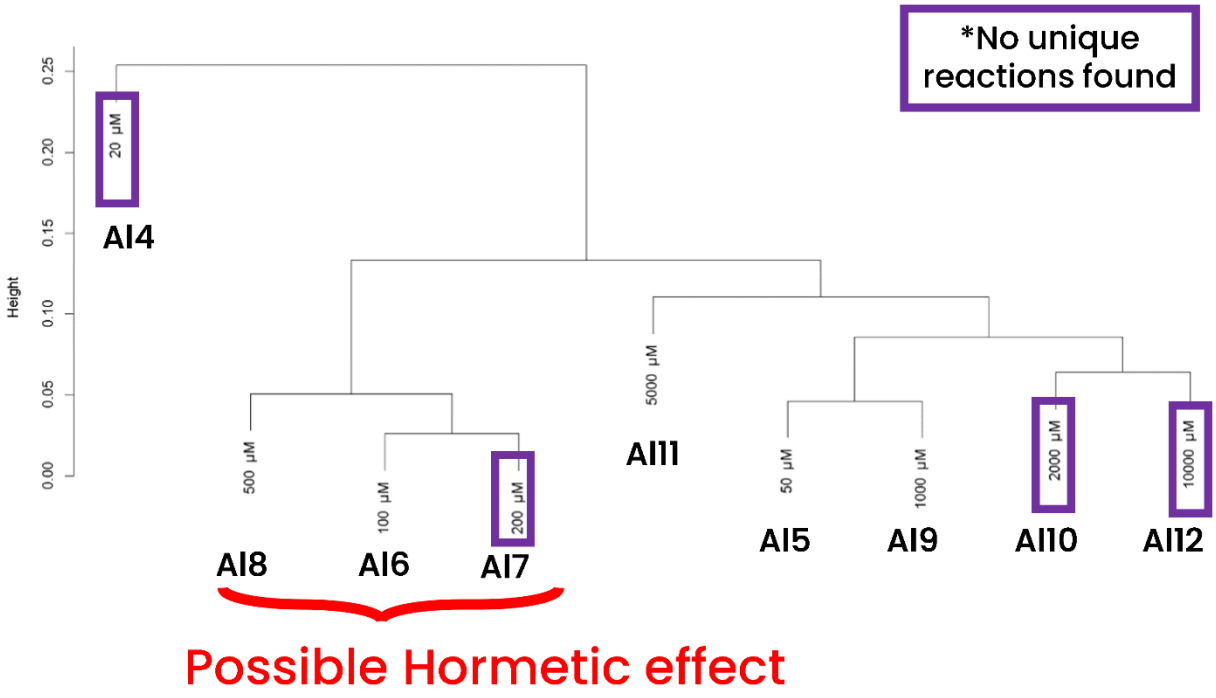


Figure 5.6. Dendrogram of the similarity among aluminum concentration scenarios in µM using the presence or absence of reactions with flux values. Aluminum concentrations without unique reactions are labeled as blue.

Unique reactions with flux values were obtained from the multistage FBA approach to observe the differential metabolic response, for the following aluminum concentration scenarios: 20, 200, 2000, and 10000 µM (Table 5.1). An absence of unique reactions active per aluminum scenario was observed, which indicates a large number of shared reactions for the aluminum concentration

scenarios. In contrast, the reactions of 1-pyrroline-5-carboxylate dehydrogenase (NADP) (mitochondrial) (R245) and Cystathionine beta-lyase (R754) were exclusive to the 50 μM aluminum concentration scenario (ID Al5, Table 5.1). In the 100 μM concentration scenario (ID Al6, Table 5.1), the glucose-6-phosphate 1-epimerase (cytosolic) (R786) reaction was exclusively observed. Additionally, six reactions were exclusive to the 500 μM aluminum scenario (ID Al8, Table 5.1): Glutamate 5-kinase in the plastid (R240), Pyrroline-5-carboxylate reductase (NADPH) in the plastid (R753), Glutamate-5-semialdehyde dehydrogenase in the plastid (R902), Proline biosynthesis, spontaneous reaction in the plastid (R905), Dicarboxylate/tricarboxylate carrier in mitochondria (R2070), and Amino acid transporter (pro-L) in the plastid (R2137). Additionally, For the 1000 μM aluminum (ID Al9, Table 5.1), the reactions of GTP:cytidine 5-phosphotransferase in the plastid (R173) and ATP:cytidine 5-phosphotransferase in the plastid (R187) were unique. Finally, a list of 24 reactions related to fatty acid biosynthesis was found for the 5000 μM aluminum concentration scenario (ID Al11, Table 5.1) (Supplementary Table 5.4).

From a different perspective, the intersection of reactions and aluminum scenarios showed a core of 517 reactions (20.15 % of the total reaction in the metabolic reconstruction) for the nine aluminum scenarios with biomass production (IDs Al4 – Al12, Table 5.1) (Supplementary Table 5.5). Furthermore, 88 reactions were observed at aluminum concentration scenarios greater than 20 μM (Table 5.1). A list of 23 reactions was exclusive to 100, 200, and 500 μM aluminum (medium concentration scenarios) (IDs Al5 -Al7, Table 5.1). Additionally, a list of seven reactions was commonly found for scenarios greater than 50 μM aluminum (Table 5.1), including Lysine decarboxylase, Serine O-acetyltransferase, Hexokinase (glc-B), cellobiose beta-glucosidase, cellulase, Amino acid transporter (lys-L), and Cadaverine sink (Figure 5.7).

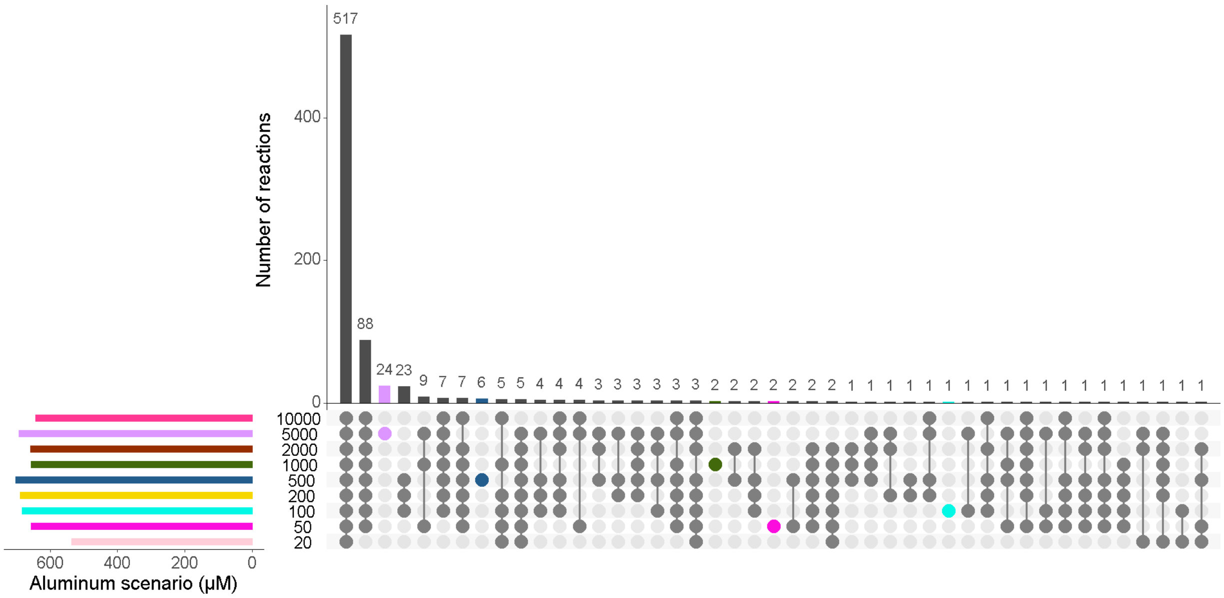


Figure 5.7. Upset plot of reaction with flux values available shared during the meristem cell cycle and nine aluminum concentration simulation scenarios (20, 50, 100, 200,500,1000, 2000, 5000, and 10000 µM) (Table 5.1). Each point represents an intersection among aluminum concentration scenarios where reactions are shared (several points) or unique reactions (a unique colored point). Colored bars represent each of the aluminum concentration scenarios. The colored points represent unique reactions to that scenario.

5.5.3.4. A possible active role of the purine and fatty acid-related metabolic subsystems was present regardless of aluminum concentrations simulated

A total of 71 subsystems were found for all 12 aluminum concentration scenarios (Table 5.1), and 43 subsystems were shared in all aluminum concentration scenarios. For the aluminum concentration scenarios, a total of 57 metabolic subsystems were shared; for the aluminum concentration scenarios > 20 µM (IDs Al5 – Al12, Table 5.1), a list of nine subsystems was found: Nicotinate and Nicotinamide metabolism, Other Vitamins Metabolism (Thiamine), Terpenoid biosynthesis (MEP Pathway, MVA Pathway), Terpenoid biosynthesis (MEP Pathway), Phytohormones biosynthesis (Zeatin), Sphingolipid metabolism, Aminoacyl-tRNA biosynthesis, Aroma Compound Biosynthesis, and Purine Nucleotide Biosynthesis. For the 5000 µM scenario (ID Al11, Table 5.1), the Secondary Cell Wall Metabolism (Cuticular Wax), and Fatty acid biosynthesis (Very long chain) were exclusive. Additionally, the Fatty acid biosynthesis (desaturation), and Lysine metabolism subsystems were found exclusively for the 100, 200, and 500 µM aluminum concentration scenarios (IDs Al6 – Al8, Table 5.1) (Figure 5.8).

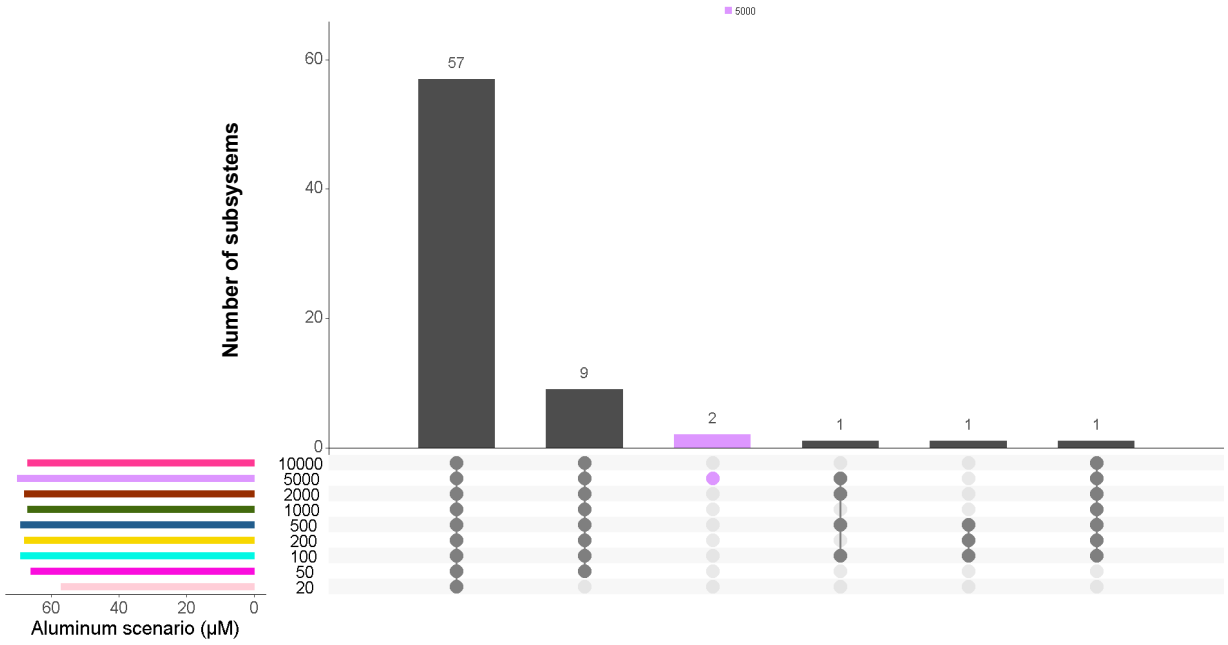


Figure 5.8. Upset plot of metabolic subsystems available shared during meristem cell cycle and aluminum concentration simulation scenarios with biomass production (20, 50, 100, 200, 500,1000, 2000, 5000, and 10000 μM) (Table 5.1). Each point represents an intersection aluminum concentration scenario. Colored bars represent each of the scenarios. Colores points represent unique subsystems for a given aluminum concentration scenario.

5.5.3.5. Central, amino acid and carbon-nitrogen ring metabolism would be altered under aluminum stress conditions in rice meristem cells

A total of four aluminum concentration scenarios (50, 200, 500, and 1000 μM representing aluminum concentrations of tolerance to high aluminum and aluminotoxic conditions found in different rice varieties) (Table 5.1) were plotted as heatmaps for reaction flux values and the time integration steps to observe changes in reaction flux values. The criteria to report relevant reactions whose flux values fluctuated the standard deviations that were greater than the third quantile (upper quantile) per scenario were considered. For the four aluminum concentration scenarios (IDs Al5, Al7, Al8, and Al9, Table 5.1), the reactions associated with the following subsystems had greater reaction flux variability: Glycolysis/Gluconeogenesis; Transport (Plastidic); Transport (Mitochondrial); Transport (Extracellular); Exchange Reactions; Pyrimidine metabolism; Nitrogen metabolism; Transport (Vacuolar); Starch and sucrose metabolism (Sucrose); Calvin cycle; Pentose Phosphate Pathway; Purine metabolism; Photorespiration;

Glycine, serine and threonine metabolism; Folates metabolism (Supplementary Figures 5.4 to 5.7).

Specifically, within the top 10 reactions with more variability during the simulation intervals for the selected aluminum concentration scenarios (IDs Al5, Al7, Al8, and Al9, Table 5.1), carbonic anhydrase cytosolic (R999), Dicarboxylate/tricarboxylate carrier (mal:cit) mitochondrial (R2063), and H2O transport from plastid (R2159) were the reactions with the highest variability for the four aluminum concentration scenarios selected. Dihydroxyacetone phosphate transport via triose-phosphate translocator from plastidic (R2150), and Triosephosphate isomerase cytosolic (R611) for the 50, 200, and 1000 μM aluminum concentration scenarios (IDs Al5, Al7, and Al9, Table 5.1). Dicarboxylate/tricarboxylate carrier (akg:cit), and mitochondria (R2066) were available at aluminum concentration scenarios of 200, 500, and 1000 μM (IDs Al7, Al8, and Al9, Table 5.1). Also, bicarbonate transport in the plastid (R2193) and carbonic anhydrase in the plastid (R998) were found at aluminum concentration scenarios of 500 and 1000 μM (IDs Al8, and Al9, Table 5.1). The CO_2 plastidic transport (R2160) reaction was observed for 200 and 500 μM aluminum concentrations (IDs Al7, and Al8, Table 5.1), while Glyceraldehyde 3-phosphate transport via triose-phosphate translocator (from plastid) (R2147) was observed at 50 and 1000 μM aluminum concentrations (IDs Al5, and Al9, Table 5.1). For the 50 μM aluminum concentration scenario (ID Al5, Table 5.1): Glucose-6-phosphate isomerase (g6p-B) plastidic (R2532), Glucose-6-phosphate 1-epimerase plastidic (R787), and Glucose-6-phosphate isomerase (g6p-A) plastidic (R530). For the 200 μM aluminum concentration scenario (ID Al7, Table 5.1), the following reactions were found in the third upper quantile: the L-aspartate transport via diffusion (extracellular to cytosol) (R111), and 3-Phospho-D-glycerate transport via the triose phosphate translocator (from plastidic) (R2151). Finally, L-Asparagine exchange (R23), L-asparagine transport via diffusion (extracellular to cytosol) (R100), and Asparaginase (R366) (Supplementary Figures 5.8 - 5.11).

5.6. Discussion

Understanding abiotic stress by plant systems biology approaches can help to elucidate molecular mechanisms such as phytohormones crosstalk, early downregulation of energy metabolism, observation of energy conservation mechanisms, and metabolic changes in response to abiotic stresses (Cramer et al., 2011). In this regard, it is clear that the aluminum stress behavior modeled here is a complex phenomenon that includes the regulation of genes responsible for ethylene and auxin biosynthesis as well as their crosstalk, and how the production of these phytohormones leads to metabolic changes that are modeled for the first time for a staple crop such as rice

integrating both the aluminum effect on ethylene and auxin biosynthesis pathways, as well as the metabolic machinery responsible for the production of these phytohormones.

The main findings are as follows: (i) The crosstalk dynamics of ethylene and auxin biosynthesis provide a fast mechanism for ethylene, and when its production decreases, auxin biosynthesis reaches its maximum peak as a response. (ii) The results obtained by the multistage FBA approach suggest that models under the modeled conditions provide a multistage FBA approach in which cell death under aluminum stress is described. (iii) The biomass response under different aluminum concentration scenarios suggested that scenarios greater than 50 μM and less than 1000 μM (IDs Al6 – Al 8, Table 5.1) would be hormetic effects of aluminum because under these aluminum concentrations, the biomass production is higher than under other aluminum concentrations scenarios (Figure 5.5), but the biomass production time decreases in comparison with aluminum scenarios <100μM (IDs Al4, and Al5, Table 5.1). (iv) Aluminum concentrations scenarios >500 μM biomass profiles (IDs Al9 – Al12, Table 5.1). suggests saturation of biomass production when the meristem cell cannot survive optimally. (v) In aluminum concentration scenarios from 100 to 1000 μM (IDs Al6 - Al9, Table 5.1), the variability of reactions involved is greater than the scenarios of <50 μM and >2000 μM (Table 5.1), and it suggests that hormetic effects enhance the appearance of different metabolic reactions than found in other aluminum concentrations. (vi) Two metabolic clusters (Supplementary Figures 5.3 – 5.11) suggest that the rice meristem biomass profile under aluminum stress depends on the reactions associated with fatty acids and the primary cell wall for the first moment of biomass production, which corresponds to cluster one. In contrast, cluster two refers to the maximum biomass peak and the primary metabolism involved. (vii) Fatty acid metabolism, methionine and cysteine, proline, glutamate, and transport reactions in plastids, as well as carbon-nitrogen rings such as pyrimidine and purine-involved reactions, would be affected when the plant receives the aluminum stimulus in the RLK receptor. (viii) The combination of models suggests that possible aluminum ion sensing in rice has not yet been sufficiently studied, and more complex biological mechanisms involved in metabolism changes under aluminum stress are not well understood.

5.6.1. A possible role of MAPK sensing aluminum in rice

The aluminum ion-sensing mechanism in rice must be better understood (Liu et al., 2014). Therefore, this chapter assumes that a general MAPK transduction pathway is used in plants, where aluminum interacts with a receptor and initiates a second messenger cascade (Poot-Poot and Hernandez-Sotomayor, 2011). Nevertheless, genome-scale metabolic reconstruction does

not include aluminum, but behaves as a heavy metal in plants by competing for membrane binding with other elements, such as calcium or magnesium, and affects cell wall components instead of being part of plant metabolism properly (Panda et al., 2009; Morkunas et al., 2018; Gallo-Franco et al., 2020). From this perspective, the use of MAPK as signal transduction, as described by Pathak et al. (2013), was justified by its function as a bridge from aluminum sensing, which triggers the transcription rate of auxin and ethylene crosstalk and provides a proxy for metabolic responses in rice given that the aluminum sensing mechanisms are still not clear, and it is believed that it includes transporters, calcium, or reactive oxygen species (ROS) (Liu et al., 2014).

The results presented here show that signal transduced using the MAPK system is consistent with those found in rice. MAPKs are activated in rice, where the plants exposed to heavy metals such as cadmium, and copper, had similar response to the aluminum stress (Gallo-Franco et al., 2020; Majeed et al., 2023). The mitogen-activated protein kinases (MAPKs) are present in diverse abiotic stress responses (Vandereyken et al., 2018). In plants, MAPKs regulate auxin distribution via hydrogen peroxide under cadmium and zinc stress (Jagodzik et al., 2018). For instance, 16 MAPKs differentially expressed are affected by the exposition to aluminum in wheat (Liu et al., 2017b); MAPKs also are part of phytohormone response as abscisic acid, jasmonate, and Salicylic Acid in rice (Singh and Jwa, 2013). The presence of MAPK reinforces the idea of a similar mechanism in rice because some genes, such as OsMPK1, OsMPK5, and OsMPK12, are upregulated by ethylene (Singh and Jwa, 2013), whereas OsMPK3 and OsMPK4 are activated by arsenic stress in the indica rice subspecies. Furthermore, OSMPK expression increased until three hours of exposure and decreased after 9 hours of exposure to arsenic (Rao et al., 2011), the tolerance mechanism of which is similar to that of aluminum in plants. Furthermore, 75 MAPK-related proteins are involved in rice response to biotic, abiotic stresses and phytohormones signal transduction (Chen et al., 2021b). Alternatively, the results provided in the previous chapter include some MAPK-related proteins. Aluminum stress in rice causes a range of changes in root cells, such as changes in the structural properties of the cell wall, carbohydrates, and pectins, which are found to be in concordance with heavy metal stress responses in plants (Gallo-Franco et al., 2020; Jingguang et al., 2020). Thus, even though the MAPK system and aluminum stress response have not been well studied in rice yet, it is well supported that MAPKs respond similarly to aluminum and other heavy metals to trigger a fast response against the presence of aluminum stress.

5.6.2. A dynamic approach predicting cell death under aluminum stress?

A topic taken into consideration here is that both models (the gene regulatory network model and the metabolic network model) simulate local auxin and ethylene biosynthesis for a meristematic cell instead of a global response of the plant itself, given that (i) the metabolic dynamics take place in a meristematic cell at first glance when aluminum enters to roots without more participant elements in the aluminum response by plants under field conditions. Some examples of these participants are the alleviation by manganese, ammonium, or phosphorus that impedes the aluminum uptake by the plant (Bose et al., 2011; Zhao et al., 2013; Wang et al., 2015) and reactive oxygen species, which are part of stresses (Vandereyken et al., 2018). Both models used here to model the aluminum metabolic effect on rice roots capture a local response without supporting auxin provenance from leaves or shoots via polar transport, since simulation scenario two (SSR2) (Table 5. *2*) was the only one that obtained biomass profile results. This means that the multistage FBA approach only simulates the direct impact of aluminum on root growth mediated by auxin and ethylene crosstalk.

Highlighting this fact, in *Arabidopsis thaliana,* auxin local biosynthesis generates auxin gradients for plant development. However, polar transport from shoots to roots and the interaction of both systems, as well as root meristem auxin synthesis and metabolism, are required to maintain meristem in the plant to survive (Azizi et al., 2015; Brumos et al., 2018). Hence, the biomass profile results could be explained by the lack of this auxin polar transport mechanism in the blend of approaches (Figure 5.5). This fact led to fast cell death in the presence of aluminum as a stressor and a local response of a meristem cell without the external support of auxin in rice. In other words, without the external support of auxin or ethylene, the models provide an idea of a cell death effect by the impact of aluminum response without other elements helping for tolerance to aluminum due that among these elements, organic acid release, ammonium, pectin, magnesium alleviation, or phenylpropanoid biosynthesis, among other possible mechanisms are involved in tolerance (Bose et al., 2011; Zhao et al., 2013; Kochian et al., 2015; Wu et al., 2018; Wang et al., 2023a). This reinforces the idea found in *Arabidopsis thaliana* and other plants that auxin polar transport is important as part of the response of rice to aluminum stress effects. However, more information is required to enhance the aluminum response in rice roots behavior modeling by incorporating this information.

5.6.3. Ethylene and auxin production are part of the biggest game in rice root growth and root elongation in rice under aluminum stress conditions

Aluminum stress results using proteomics data show that under 50 μM aluminum, a decrease in hydrogen peroxide, an important role in putrescine inhibiting ethylene, and an increase in pectin methylation removal of aluminum in the cell wall, oxidative stress, and reduced energy synthesis occur in rice (Chunquan et al., 2021). To date, the auxin-ethylene crosstalk is the most known molecular mechanism affected by aluminum stress in rice (Liu et al., 2022). This auxin-ethylene crosstalk has a dual effect by stimulating and inhibiting root growth and elongation (Van De Poel et al., 2015). Ethylene triggers reactive oxygen species, hydrogen peroxide, and glutathione production which are key regulators of plant response to metal stress (Keunen et al., 2016). The results obtained in this work widely agree with Xie et al. (2022) in the sense that L-proline is associated with ethylene synthesis.

Nevertheless, auxin transport reduces the aluminum concentration in the cell wall, contributing to root elongation under aluminum phytotoxicity (Kopittke, 2016). Additionally, MAPK can activate other phytohormones such as jasmonic acid, salicylic acid, cytokinins, or ethylene which can be activated under biotic stress and which negatively control meristem activity (Azizi et al., 2015; Morkunas et al., 2018) that are not integrated into the genome-scale metabolic reconstruction as well as the five auxin biosynthesis pathways described, and only one is tryptamine-dependent (Azizi et al., 2015); however, only the main auxin biosynthesis pathway was considered to reduce the complexity of auxin biosynthesis in rice during the analysis. As a result, some aspects of aluminum stress, such as auxin polar transport or exclusion mechanisms, can alter the multistep FBA approach results if they are considered.

For instance, in addition to auxin polar transport, ethylene regulates the auxin gradient that moves towards the root apical meristem from leaves and accumulates auxin in the elongation zone and polar transport to inhibit root elongation in rice, altering auxin metabolism (Růžička et al., 2007; Vandenbussche et al., 2012). This behavior is partially captured in the gene regulatory network model by the rapid ethylene burst followed by high auxin production. However, the modeled behavior does not consider the whole complexity of the aluminum stress response. For instance, in plants, ethylene biosynthesis can be influenced by cytokinins and other hormones, such as abscisic acid (Ahammed et al., 2020). Nevertheless, the ethylene response modeled here followed a similar trend as suggested in *Arabidopsis thaliana* and *Lottus japonicus,* where the maximum production peak was observed after 30 minutes of exposure to 50 μM of aluminum (ID Al5, Table 5.1) and after six hours stabilized its production and upregulated ACO, and ACS genes

transcription (see Supplementary Figure 5.2) (Sun et al., 2007, 2010). Thus, even with the no considered complexity, the results of all three stages help to identify possible metabolic mechanisms of aluminum tolerance in rice, including purine catabolism which is a mechanism against abiotic stress in plants (Kaur et al., 2023) and other possible mechanisms that are not easily found in wet lab experiments.

In the case of exclusion mechanisms, rice releases organic acids such as malate, oxalate, or citrate to chelate the aluminum ion, converting it to a non-hazardous molecule for plant survival via transporters as the main detoxification mechanism in rice (Yokosho et al., 2011; Yan et al., 2022). These mechanisms do not eliminate the range of affectations provoked by the aluminum stress that can disrupt the balance of physiological reactions in plants. Aluminum leads to oxidative stress, inhibition of root cell growth, and agronomical losses in crop yield; however exclusion mechanisms reduce this oxidative stress (Kochian et al., 2015; Wang et al., 2022). Plant stress responses are complex and include primary sensory mechanisms, and signaling that translates the stimuli into a physiological response (Lamers et al., 2020), which is not completely observed in the holistic approach considered due to a possible lack of information known in rice, for instance, if ROS work as sensing or if a totally different mechanism is still not discovered.

5.6.4. Metabolic response to aluminum stress considers a diverse array of reactions associated with fatty acids, and amino acids metabolism

5.6.4.1. Hormesis was found for three aluminum concentration scenarios (100, 200, and 500 µM) via multistep FBA

It is important to highlight that aluminum inhibits rice root growth in a concentration-dependent manner in plants. For instance, after 30 minutes of exposure to aluminum, root growth inhibition started in rice cv. Youngnam (Goh and Lee, 1999). This root growth behavior agrees with the reported Indian rice varieties exposed to aluminum where the metal concentrations increase and the relative root length decreases (Awasthi et al., 2019). In this regard, the multistep FBA approach matches the behavior reported in the literature, given that biomass production starts to increase at a similar time. Furthermore, the biomass profile observed started to decline between an hour and two hours of exposure to aluminum, which can be explained by the decrease in ethylene concentrations in the meristem cell and the increase in auxin concentrations that finishes in root growth (Ranjan et al., 2021a). Thus, the behavior observed in Figure 5.5 could simulate the trend of auxin accumulation causing shortening of the root elongation zone, as reported in rice (Jingguang et al., 2020).

Aluminum can enhance root growth, chlorophyll content, and nutrients (phosphorus, iron, manganese, zinc, or copper) uptake by plants, such as tea or *Tabebuia chrysantha,* at low metal concentrations (Ofoe et al., 2023). Certainly, the multistep FBA approach would agree with this behavior at 100, 200, and 500 μM aluminum concentration scenarios (considered hormetic scenarios) by activating diverse metabolic reactions and probably improving sugar concentrations and micronutrients associated with photosynthesis as suggested by Moreno-Alvarado et al. (2017) as a benefit of including aluminum inside the cell and NAC transcription factor regulation. Hence, in the scenarios mentioned above, the hormesis effect was considered in the results obtained because there were changes in the number of reactions. The results of the NMDS showed that at these concentration scenarios, the metabolic profiles differed notoriously from the other aluminum scenarios modeled, and the biomass flux values were significantly greater than other aluminum concentration scenarios even if the biomass production time was the earliest and occured in less time.

As a complement to the hormetic effects, two series of reactions were promoted for the hormetic reactions: (i) fatty acids of 16 and 18 carbons (C16 and C18) and (ii) L- fuculose-related reactions. The very long fatty acid metabolism is involved in the form of fatty acyl-CoA synthases for C16 and C18 fatty acids, respectively. They are part of the responses to pathogens and hypoxia response in *A. thaliana,* and part of membrane lipids, energetic processes, and barriers against stresses (Batsale et al., 2021; Zhao et al., 2021). On the other hand, GDP-D-mannose dehydratase, GDP-4-oxo-L-fucose reductase, L-fuculose isomerase, L-fuculose aldolase, and L-fuculokinase mediated reactions produce L-fuculose which participates in plant organ development (Gonçalves et al., 2017). As a result, the possible activation of these reactions could promote the highest biomass production in a short period compared to the other aluminum concentration scenarios at sense aluminum in the cell wall and promote organogenesis.

5.6.4.2. Scenarios greater or equal than 50 μM of aluminum are affected by lysine, serine biosynthesis, and energy obtention-related reactions

For the aluminum concentration scenarios greater or equal than 50 μM (Al6 – Al12, Table 5.1), the presence of reactions such as those catalyzed by serine O-acetyltransferase is consistent with an enhanced detoxification complex similar to other heavy metal stress responses in plants. In this sense, this enzyme regulates sulfate and selenium uptake to promote glutathione production under arsenic stress (Sun et al., 2021). In addition, the presence of lysine catabolism to produce cadaverine in the model agrees with the idea of amino acid metabolism possibly involved in aluminum stress, given that lysine and serine are connected with a wide abiotic stress

tolerance in plants (Kishor et al., 2020). On the other hand, the cadaverine is a precursor of polyamines and S-adenosyl methionine from methionine catabolism conjugates to create more compounds and, as a response to stress, helps to synthesize reactive oxygen species and putrescine to control root growth (Jancewicz et al., 2016). Additionally, the presence of a cellobiose beta-glucosidase mediated reaction can be connected to a cell-wall oligosaccharide recycling as a strategy to produce glucose (Opassiri et al., 2007), and the presence of a cytosolic hexokinase has been suggested as a strategy for energy generation under anaerobic glycolysis in rice (Lakshmanan et al., 2013).

5.6.4.3. Unique reactions per aluminum concentration suggest an active role of proline, glutamate, and cell wall components

In the case of unique reactions, active roles of proline and glutamate were evident. For instance, proline accumulation and lipid peroxidation occur in maize under aluminum stress (Giannakoula et al., 2008). Similarly , proline accumulation increases under drought stress in rice (Dien et al., 2019). Proline is metabolized via glutamate, which in turn metabolizes glutamine, arginine, or proline (Forde and Lea, 2007). Furthermore, glutamate provides a direct link between nitrogen and carbon metabolism, inhibiting primary root growth, and is actively involved in signal transduction under stress conditions in rice (Kan et al., 2017). Specifically, aluminum affects glutamate metabolism in *Sorghum bicolor* and rice by scavenging oxidative stress scavenged in ascorbate-reduced glutathione and enhancing amino acid metabolism improved in tolerant varieties such as Azucena (Struys-Ponsar et al., 2000; Sharma et al., 2006; Ma et al., 2012; Xie et al., 2022).

The active roles of proline and glutamate are as follows:

(i) For the 50 µM aluminum concentration scenario, a reaction mediated by 1-Pyrroline-5-carboxylate dehydrogenase (P5CDH) was observed. P5CDH is a crucial enzyme in the conversion of proline to glutamate (Yuan et al., 2023), whereas Cystathionine beta-lyase is implicated in methionine biosynthesis, and its mutation in *A. thaliana* affects auxin response, PLETHORA 1 and 2 (PLT1 and PLT2 genes) affecting root stem maintenance (Liu et al., 2019). This effect has been observed in rice. In this species, at least 10 PLT genes have been described, which are upregulated by auxin and ABA treatment conditions (Li and Xue, 2011). This result supports the idea of a differential response for aluminum sensing at this aluminum concentration to complement the amino acid metabolism described above.

(ii) For the 500 μM aluminum concentration scenario (ID Al5, Table 5.1), the presence of Pyrroline-5-carboxylate reductase is associated with proline accumulation under drought stress in *Vitis yeshanensis,* transgenic *A. thaliana*, and rice (Antunes et al., 2019; Chen et al., 2021a). In addition, the presence of Dicarboxylate or tricarboxylate carrier-related reactions suggests an active role in amino acid transport due to they are described in nitrogen assimilation in rice leaves and are critical for the provision of amino acids such as proline or glutamate and ATP among other substrates for mitochondrial respiration (Huang et al., 2009c). Consequently, with at least 50 members of MC-likes genes in rice (Taylor et al., 2010), the presence of the carriers' reactions is coherent with an energetic mechanism alteration due to the aluminum presence in rice (Sosa et al., 2023) and observed in the heatmaps provided for the 50, 200, 500, and 1000 μM aluminum concentration scenarios (IDs Al5, Al7, Al8, Al9, Table 5.1) in the presence of glucose metabolism (see Supplementary Figures 5.4 - 5.11).

(iii) For the 1000 μM aluminum concentration scenario (ID Al9, Table 5.1), cytidine 5-phosphotransferase suggests a role for cytidine monophosphate (CMP) in water deficit in rice (Ray et al., 2011) and could be similar to a possible role of CMP biosynthesis in altering the cell wall in strawberry (Antunes et al., 2019). The effect of cell wall components could be more intense for the 5000 μM aluminum scenario (ID Al11, Table 5.1) because of the active role of fatty acid biosynthesis in cell wall compound components. In this regard, Very long fatty acid synthesis related to 24 and 26 monounsaturated carbon fatty acids (C24, C26) were found during the simulations. Very long-chain fatty acids (VLCFA) biosynthesis is affected by abiotic stresses such as heavy metals, and there is an increase in C24 and C26 in the presence of stresses such as cold, light, or osmosis. In addition, VLCFA is part of the cell wall and supports lateral root development in plants, such as *A. thaliana* (Batsale et al., 2021). Thus, the presence of VLCFA reactions is in accordance with the results obtained in the multistep FBA approach, where at higher aluminum concentrations, VLCFA would be affected, given that aluminum exposure can increase peroxidation and membrane lipid breakdown in tolerant rice cultivars (Peixoto et al., 2001). Therefore, the presence of this very long fatty acid is a product of the metabolic response to handle the aluminum stress and reinforces the usefulness of the combined approaches used in this chapter to gain insights into the possible effects of abiotic stress.

5.7. Conclusions and Future Perspectives

The results of the holistic workflow used in this chapter in three stages are important to obtain insights into the metabolic response of rice root meristem by including a possible aluminum sensing proposed using RLK receptor and MAPK signaling to finally obtain a physiological response in the form of a metabolic response visualized in the multistep FBA approach, which suggests central role of amino acid metabolism such as proline and fatty acid biosynthesis in the rice plant as a complement to the production of the most well-described physiological response caused by aluminum stress in the form of auxin-ethylene crosstalk that regulates root growth and its inhibition. For future steps, the following topics are considered: (i) the inclusion of metabolomic information to improve model accuracy is suggested, given that some key metabolites such as putrescine or other phytohormones such as cytokinins which are actively found in root meristem development (Kopittke, 2016). (ii) There is a need for in vitro tests at several time intervals and different aluminum concentrations to corroborate the results obtained here to provide feedback and corroborate the role of the proline, glutamate, and other metabolites obtained in the metabolomics research results obtained by Chunquan et al., (2021) or Xie et al., (2022). This is due to the possibility of differential metabolic responses at different aluminum concentrations observed in the biomass profiles obtained in this chapter. (iii) It is important to test the proteins predicted in aluminum responses in the previous chapter to determine whether a more complex plant response hypothesis is valid. (iv) The stress response in a plant is complex. It includes elements such as signaling and proteins that function as the main connectors in a network, including transcription factors (Vandereyken et al., 2018). Thus, it is important to understand the regulatory elements that control the abiotic stress response and will allow future approaches to modulate aluminum stress in rice.

In complement to the above observations and based on the comprehensive study of aluminum tolerance *in Oryza sativa* (rice) under aluminum stress, several evolutionary predictions can be made regarding how the species adapts to such abiotic stress. These predictions revolve around the complex interplay of phytohormones, metabolic pathways, and stress response mechanisms:

Phytohormone Crosstalk Adaptation:

- Ethylene and Auxin Biosynthesis Dynamics: Under aluminum stress, ethylene and auxin biosynthesis exhibit a rapid crosstalk mechanism where ethylene production quickly peaks and then decreases, leading to a subsequent peak in auxin biosynthesis. This mechanism

likely evolved to provide a quick and efficient response to aluminum stress, optimizing root growth and adaptation.

- Enhanced Auxin Transport Mechanisms: Given the importance of auxin in root development and aluminum stress response, rice varieties may evolve enhanced auxin transport systems to maintain root meristem function and growth under stress.

Metabolic Pathway Evolution:

- Fatty Acid and Primary Wall Biosynthesis: The adaptation includes a focus on fatty acid metabolism and primary wall biosynthesis during the initial phases of biomass production under aluminum stress. This suggests that rice has evolved mechanisms to reinforce cell walls and maintain cell integrity when faced with aluminum toxicity.
- Adaptations in Amino Acid Metabolism: The critical roles of methionine, cysteine, proline, and glutamate metabolism under aluminum stress indicate evolutionary adaptations that enhance stress tolerance. These amino acids are involved in stress response pathways and are crucial for maintaining cellular homeostasis.

Hormesis and Biomass Production:

- Hormetic Response to Aluminum: The observation of hormetic effects at aluminum concentrations between 50 μM and 1000 μM suggests that rice may have evolved to exploit low to moderate levels of aluminum for enhanced growth, optimizing biomass production within this range.
- Saturation of Biomass Production: At aluminum concentrations above 500 μM, a saturation point in biomass production indicates an evolutionary threshold beyond which rice cannot efficiently tolerate aluminum, leading to possible adaptations that prioritize survival overgrowth.

Signal Transduction Pathways:

- MAPK Signaling in Aluminum Sensing: The potential role of MAPK (Mitogen-Activated Protein Kinase) pathways in aluminum sensing and response suggests that rice may have evolved complex signaling networks that integrate aluminum stress signals to modulate gene expression and metabolic responses effectively.

- ROS and Calcium Signaling: The involvement of reactive oxygen species (ROS) and calcium signaling in aluminum stress response indicates evolutionary adaptations that utilize these molecules as secondary messengers to trigger defensive and adaptive responses in rice.

Cell Death and Survival Mechanisms:

- Local Cell Death Dynamics: The prediction of cell death models under aluminum stress highlights an evolutionary trade-off where rice may sacrifice certain root cells to protect the overall root system. This local cell death could be a strategy to prevent the spread of aluminum toxicity.
- Auxin Polar Transport and Root Maintenance: The necessity of auxin polar transport from shoots to roots for maintaining root meristem and overall plant survival under aluminum stress suggests that rice has evolved mechanisms to ensure continuous auxin supply to roots, critical for sustaining growth and development under stress conditions.

Exclusion and Detoxification Mechanisms:

- Organic Acid Exudation: Evolutionary adaptations likely include the exudation of organic acids such as malate, oxalate, and citrate to chelate aluminum ions, reducing their toxicity and preventing their detrimental effects on root growth.
- Cell Wall Modifications: Changes in cell wall composition, including pectin methylation, are predicted to be evolutionary strategies that rice employs to reduce aluminum uptake and mitigate its effects on cell wall integrity.

Future Research Directions:

- Metabolomic Studies: Further metabolomic research is needed to identify key metabolites involved in aluminum stress response and validate the predicted roles of proline, glutamate, and other amino acids.
- Genetic and Proteomic Analysis: In vitro experiments and genetic studies should focus on identifying and validating genes and proteins involved in aluminum tolerance, particularly those related to MAPK signaling and phytohormone biosynthesis.

- Integrative Modeling Approaches: Developing more comprehensive models that integrate metabolomic, proteomic, and transcriptomic data will enhance the understanding of aluminum stress responses and support the breeding of aluminum-tolerant rice varieties.

Chapter 6

General Discussion

New insights into the physiological and molecular responses of rice to abiotic stress, with a focus on aluminum tolerance, are presented in this document. A systematic approach was structured by: (i) providing a functional perspective associated with the biological processes that are triggered under toxic conditions through the implementation of comparative functional genomics method between *Arabidopsis thaliana* and rice, (ii) implementing machine learning strategies over a protein-protein interaction network to identify new proteins involved in aluminum-associated stress responses, and (iii) integrating a gene regulation network deterministic model and a genome-scale metabolic reconstruction to predict the possible metabolic responses of rice to different aluminum concentrations.

Functional enrichment analysis is a cornerstone of bioinformatics, as it makes it possible to identify functional information by using a gene list as a source. Nevertheless, quantitative procedures to compare GO terms among gene lists and species are unavailable. In the present study, a computational procedure implemented in R was established to infer functional information derived from comparative strategies. GOCompare provides a framework for functional comparative genomics, starting from comparable lists of GO terms. The combination of the gene lists and the biological and molecular functions obtained from functional enrichment together with the power of GOCompare software allowed to detect that reactive oxygen species responses and exclusion mechanisms play a pivotal role in facilitating the tolerance and adaptation of rice and *A. thaliana* to aluminum stress. These mechanisms are widely found in plant species such as *Camellia sinensis,* some *Brassicaceae* species*,* and *Glycine max,* among others, as primary responses to aluminum (Bojórquez-Quintal et al., 2017). Also, the described mechanisms are part of a well-defined plant stress response that diverges in rice and involves central metabolism induction in response to aluminum stress (Famoso et al., 2010; Kochian et al., 2015; Jingguang et al., 2020). Additionally, GOCompare was able to identify molecular functions that are highly prevalent in the aluminum response in both rice and *A. thaliana*. This also supports the hypothesis of possible downregulation of photosynthesis during stresses (Cohen and Leach, 2019). Hence, GOCompare shows its accuracy in providing a functional comparative genomics approach, which can provide functional insights that are not easy to obtain in a visual comparison or separate enriched biological or molecular functions (Figure 6.1).

Regarding the interaction among proteins, a considerable amount of genes have been described as part of the molecular response of rice to aluminum, including an activated core of genes whose translated proteins work as hubs that interact with each to generate a physiological response during abiotic and biotic stresses (Vandereyken et al., 2018; Cohen and Leach, 2019; Sosa, 2022; Sosa et al., 2023).

The major challenge in identifying these elements is that in traditional approaches, only one source of information is considered. In the present work, complementary knowledge was integrated: (i) a complete rice interactome was considered, (ii) a list of genes from an extensive bibliomics search, complemented with quantitative experimental information derived from RNA-seq experiments, was included in the analyses, and (iii) several classification algorithms based only on positive labels were integrated in the methodological approach.

The most recent advances in metabolomics in rice under aluminum stress show a complex landscape of metabolites from fatty acids and amino acids to central metabolism compounds (Xie et al., 2022; Wang et al., 2023a). Previously, separated omics approaches have provided different components of aluminum response mechanisms that are conserved in other plants, such as ROS scavenging or central metabolism deprivation, which are present in rice (Sosa et al., 2023). According to the results obtained by the machine learning approach, a description of other possible molecular elements not yet described in the aluminum response in rice is given, among these elements: (i) 1-aminocyclopropane-1-carboxylic acid synthase 1 (ACC1), (ii) a cytokinin oxidase (CKX8), (iii) tetraketide aplha-pyrone reductase (TKPR1), and (iv) Qb-SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) protein (OsNPSN13), suggesting the presence of the mechanisms of ethylene, cytokinin biosynthesis, sporopollenin precursor biosynthesis, and vesicle trafficking regulation in auxin transport as possible new complementary metabolic reactions to obtain physiological mechanisms affecting rice root growth (Figure 6.1).

The systemic approach that integrates omics information, genome-scale reconstruction and gene regulatory network deterministic models provides a good starting point to describe the metabolic response to aluminum in rice. The results obtained using this approach describe metabolic behaviors widely observed in plants in the form of common metabolic responses to different abiotic stresses. For example, from a metabolomics point of view, common abiotic stresses, such as flooding, drought, cold, heat, and salinity limit photosynthesis and energy metabolism, affecting Calvin-Benson, glycolysis, tricarboxylic acid cycle (TCA), and amino acid (Witt et al., 2012; Xu and Fu, 2022). In addition, the results provided evidence of possible hormetic responses at medium aluminum concentrations (100, 200, and 500 μM of aluminum). The generated model

also described the metabolic response in the root meristem, suggesting a possible role of auxin in polar transport to allow the survival of root cells. Interestingly, dynamic biomass profiles showed meristem cell death (Figure 6.1).

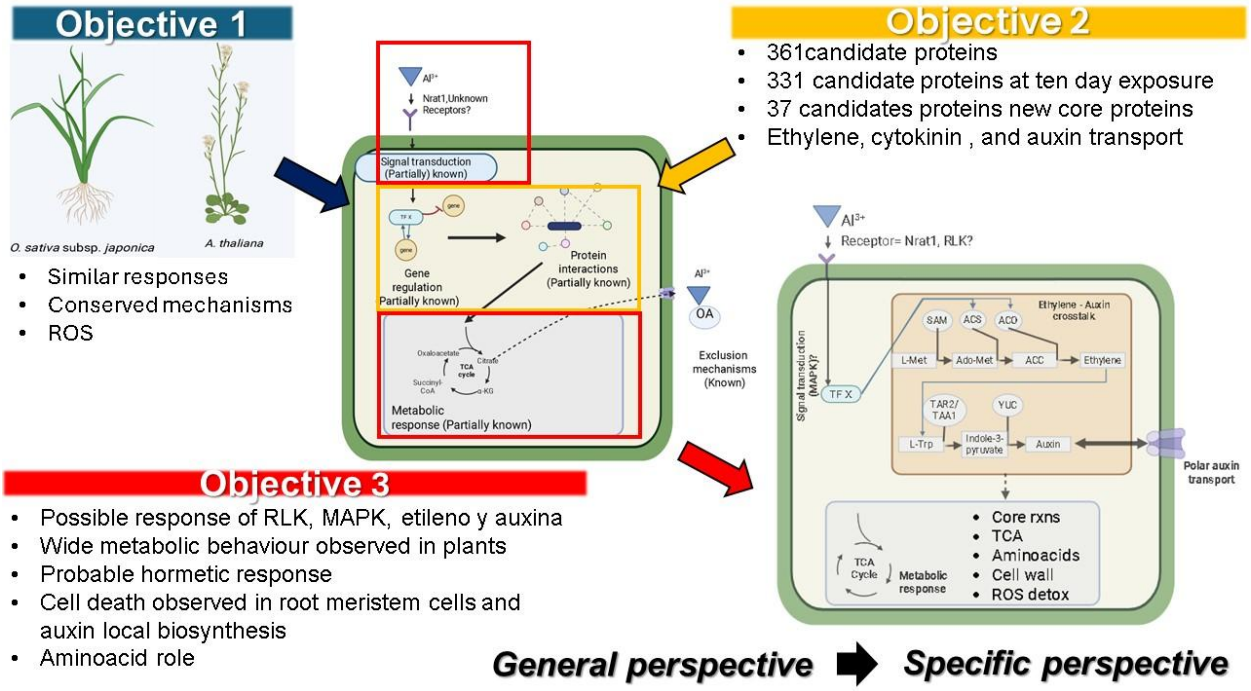


Figure 6.1. Schematic representation of the contributions of the thesis work per objective. Each objective's contribution is color-coded (objectives 1 - 3 are shown in blue, yellow, and red, respectively) and summarized in bullet points. The central figure highlights the main knowledge gaps addressed, with the contributions from each objective indicated. Black arrows illustrate the progression from a general perspective (left) to a more specific perspective (right). This specific perspective is focused on the contribution of the holistic multistep approach of objective 3 which focuses on the aluminum influence on ethylene and auxin crosstalk.

As main conclusions of this doctoral thesis:

- Overall, this research has contributed to extending the knowledge directly associated with responses to rice abiotic stresses. The identification of well-conserved molecular mechanisms, such as oxidative stress-related compound metabolism, was confirmed using three approaches implemented here.
- One of the main metabolic responses to aluminum stress is the detoxification of reactive oxygen species, which is in concordance with literature.

- The metabolic dynamic response in the root meristem depends on aluminum concentration, and hormetic effects are suggested by the generated multistage FBA approach.
- The integration of data from genes and proteins from transcriptomic results from experiments of stress Vs. non stress provides a feasible way to validate and obtain prospective genes that are not retained with RNA-Seq strategies alone.
- Even if auxin-ethylene crosstalk has been described as the most well-described physiological mechanism affecting rice root growth, the inclusion of complementary information from the literature can provide research hypotheses to confirm with wet-lab experiments.
- The present work represents the first model that integrates the sensing and transductional mechanisms for aluminum and the corresponding physiological response in rice plants.
- The results of this thesis represent an important advance in stress physiology and corroborate that computational approaches can accurately represent plant stress responses to abiotic stresses. Similarly, systemic approaches can enhance the understanding of stress physiology for better plant breeding programs.

6.1. Future perspectives

This doctoral work pointed out that, from complementary research perspectives (bioinformatics, machine learning, mathematical modeling), it is possible to obtain novel and useful insights associated with abiotic stresses such as aluminum and its effects in plants. To this end, the present research: (i) generated a new bioinformatics approach for functional comparative genomics in the structure of an R package transferable to different species; (ii) established a new bioinformatics pipeline that integrates the use of an interactome, transcriptional data, and the implementation of two series of classification algorithms in Python and R programming languages available in GitHub repositories; (iii) new candidate protein lists obtained from a network biology that in combination with a multistep FBA perspective suggest a strong metabolic response in rice related to energy metabolism including an active role of proline, and reactive oxygen species compound detoxification mechanisms; (iv) the combination of a gene regulatory network deterministic model and a genome-scale model, manually curated, which confirms the usability of plant systems biology approaches to study abiotic stress by allowing the study of perturbations of meristematic cells with different aluminum concentrations.

In complement to the outputs described, this document summarizes the obtention of an integrative approach that could be applied to other plant species, particularly because the generated approach could be tested in other cereals.

The novel knowledge obtained in rice is valuable given that:

(i) It can be useful for plant breeding programs for rice by adding potential target genes for experimental validation, enhancing the tolerance of agronomically important varieties to aluminum stress conditions in acidic soils.

(ii) The holistic approach generated here can be transferred to other staple crops, such as cereals, as a starting point to study specific physiological responses to a particular stress from an integrative perspective.

(iii) The gene network deterministic model can predict the dynamic response of molecular mechanisms, such as the auxin-ethylene crosstalk, representing a powerful tool to simulate complex biological systems and use the generated results to improve experimental designs in wet labs.

(iv) The results provide key insights into the possible reactions that have the potential to be altered by rice metabolism as a consequence of the aluminum response.

(v) The results obtained can be tested under experimental conditions to determine the role of the new proteins obtained by machine learning procedures and to evaluate their possible impact on important crop varieties.

(vi) The findings provide a different point of view about aluminum stress response, given that they combine results from functional, interacting elements, and a metabolic perspective. Nevertheless, experiments focused on regulatory elements, including a gene regulatory network, are needed to enhance the understanding of rice responses to aluminum conditions.

(vii) In the long term, a wet lab perspective is needed to evaluate rice materials under different aluminum concentrations in specific time frames to obtain the transcriptome and metabolome response of rice to validate the results obtained by the holistic approach intended here.

In summary, the research results obtained need to be expanded and supported by experimental validation to have a direct impact on plant stress responses in rice. Nevertheless, the results obtained have the potential to improve plant breeding strategies and allow better computational tools to advance decision-making procedures that contribute to Sustainable Development Goal Two (end hunger, achieve food security and improve nutrition and promote sustainable agriculture).

6.2. Future perspectives from evolutionary biology

From an evolutionary biology perspective, the insights presented in the document regarding rice's physiological and molecular responses to aluminum stress reveal several key points:

Conserved Mechanisms Across Species:

- The identification of conserved mechanisms such as reactive oxygen species (ROS) detoxification suggests that these stress responses are evolutionarily conserved across multiple plant species, including rice and *Arabidopsis thaliana*. This conservation indicates that these mechanisms likely offer a significant adaptive advantage in coping with abiotic stresses like aluminum toxicity.

Evolutionary Divergence in Stress Responses:

- The observed differences in regulatory mechanisms between rice and *A. thaliana* highlight the evolutionary divergence that has occurred in response to specific environmental pressures. For example, rice appears to have unique regulatory elements and central metabolic responses that differ from those in *A. thaliana*, pointing to species-specific adaptations that have evolved over time.

Protein Interactions and Evolutionary Adaptation:

- The interactome analysis shows how protein interactions in response to aluminum stress can uncover new proteins involved in these responses. This suggests an evolutionary adaptation where rice might have developed specialized proteins and pathways to enhance its tolerance to aluminum, which could be a result of selective pressures in environments with high aluminum content.

Auxin-Ethylene Crosstalk and Hormonal Evolution:

- The auxin-ethylene crosstalk model proposed indicates a complex hormonal regulation mechanism that might have evolved to fine-tune the stress response in rice. This crosstalk could represent an adaptive mechanism to modulate growth and development under stress conditions, which may have evolved differently in rice compared to other species.

Potential for Adaptive Evolution:

- The identification of new candidate proteins through machine learning and bioinformatics approaches provides potential targets for studying adaptive evolution in rice. These targets could help elucidate how rice has evolved to manage aluminum stress at a molecular level and how these adaptations can be leveraged for crop improvement.

Role of Phytohormones in Stress Response:

- The involvement of phytohormones such as auxin, ethylene, and potentially others like abscisic acid, melatonin, and jasmonic acid in aluminum tolerance suggests an intricate evolutionary adaptation where hormonal regulation plays a critical role in stress responses. Understanding these hormonal interactions can reveal how rice and other plants have evolved complex regulatory networks to cope with abiotic stresses.

Impact on Plant Breeding and Evolutionary Studies:

- The holistic approach integrating bioinformatics, machine learning, and systems biology provides a framework for studying evolutionary responses to abiotic stress. This approach can be extended to other crops, aiding in the development of stress-tolerant varieties through an understanding of evolutionary adaptations at the molecular and metabolic levels.

In summary, from an evolutionary biology perspective, the document provides valuable insights into how rice and other plants have evolved complex mechanisms to cope with aluminum stress. These adaptations involve conserved and divergent regulatory pathways, intricate hormonal crosstalk, and specialized protein interactions, highlighting the dynamic evolutionary processes that shape plant responses to abiotic stress.

References

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscisic Acid Signaling. *Plant Cell* 15, 63–78. doi: 10.1105/tpc.006130

Ahammed, G. J., Gantait, S., Mitra, M., Yang, Y., and Li, X. (2020). Role of ethylene crosstalk in seed germination and early seedling development: A review. *Plant Physiol. Biochem.* 151, 124–131. doi: 10.1016/j.plaphy.2020.03.016

Aittokallio, T. (2006). Graph-based methods for analysing networks in cell biology. *Brief. Bioinform.* 7, 243–255. doi: 10.1093/bib/bbl022

Alexa, A., Rahnenfuhrer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22, 1600–1607. doi: 10.1093/bioinformatics/btl140

Alföldi, J., and Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Res.* 23, 1063–1068. doi: 10.1101/gr.157503.113

Altenhoff, A. M., Studer, R. A., Robinson-Rechavi, M., and Dessimoz, C. (2012). Resolving the Ortholog Conjecture: Orthologs Tend to Be Weakly, but Significantly, More Similar in Function than Paralogs. *PLoS Comput. Biol.* 8, e1002514. doi: 10.1371/journal.pcbi.1002514

Antoniewicz, M. R. (2021). A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications. *Metab. Eng.* 63, 2–12. doi: 10.1016/j.ymben.2020.11.002

Antunes, A. C., Acunha, T. D. S., Perin, E. C., Rombaldi, C. V., Galli, V., and Chaves, F. C. (2019). Untargeted metabolomics of strawberry (Fragaria x ananassa 'Camarosa') fruit from plants grown under osmotic stress conditions. *J. Sci. Food Agric.* 99, 6973–6980. doi: 10.1002/jsfa.9986

Arbelaez, J. D., Maron, L. G., Jobe, T. O., Piñeros, M. A., Famoso, A. N., Rebelo, A. R., et al. (2017). ALUMINUM RESISTANCE TRANSCRIPTION FACTOR 1 (ART1) contributes to natural variation in aluminum resistance in diverse genetic backgrounds of rice (O. sativa). *Plant Direct* 1. doi: 10.1002/pld3.14

Arenhart, R. A., Bai, Y., Valter de Oliveira, L. F., Bucker Neto, L., Schunemann, M., Maraschin, F. dos S., et al. (2014). New Insights into Aluminum Tolerance in Rice: The ASR5 Protein Binds the STAR1 Promoter and Other Aluminum-Responsive Genes. *Mol. Plant* 7, 709–721. doi: 10.1093/mp/sst160

Arenhart, R. A., De Lima, J. C., Pedron, M., Carvalho, F. E. L., Da Silveira, J. A. G., Rosa, S. B., et al. (2013). Involvement of *ASR* genes in aluminium tolerance mechanisms in rice: ASR genes and aluminium tolerance in rice. *Plant Cell Environ.* 36, 52–67. doi: 10.1111/j.1365-3040.2012.02553.x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Aslam, S., Gul, N., Mir, M. A., Asgher, Mohd., Al-Sulami, N., Abulfaraj, A. A., et al. (2021). Role of Jasmonates, Calcium, and Glutathione in Plants to Combat Abiotic Stresses Through Precise Signaling Cascade. *Front. Plant Sci.* 12, 668029. doi: 10.3389/fpls.2021.668029

Ata, S. K., Ou-Yang, L., Fang, Y., Kwoh, C.-K., Wu, M., and Li, X.-L. (2018). Integrating node embeddings and biological annotations for genes to predict disease-gene associations. *BMC Syst. Biol.* 12, 138. doi: 10.1186/s12918-018-0662-y

Awasthi, J. P., Saha, B., Panigrahi, J., Yanase, E., Koyama, H., and Panda, S. K. (2019). Redox balance, metabolic fingerprint and physiological characterization in contrasting North East Indian rice for Aluminum stress tolerance. *Sci. Rep.* 9, 8681. doi: 10.1038/s41598-019-45158-3

Azizi, P., Rafii, M. Y., Maziah, M., Abdullah, S. N. A., Hanafi, M. M., Latif, M. A., et al. (2015). Understanding the shoot apical meristem regulation: A study of the phytohormones, auxin and cytokinin, in rice. *Mech. Dev.* 135, 1–15. doi: 10.1016/j.mod.2014.11.001

Bandumula, N. (2018). Rice Production in Asia: Key to Global Food Security. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* 88, 1323–1328. doi: 10.1007/s40011-017-0867-7

Bandurska, H. (2022). Drought Stress Responses: Coping Strategy and Resistance. *Plants* 11, 922. doi: 10.3390/plants11070922

Bao, Y.-M., Wang, J.-F., Huang, J., and Zhang, H.-S. (2008). Cloning and characterization of three genes encoding Qb-SNARE proteins in rice. *Mol. Genet. Genomics* 279, 291–301. doi: 10.1007/s00438-007-0313-2

Barabási, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Bari, M. A., El-Shehawi, A. M., Elseehy, M. M., Naheen, N. N., Rahman, M. M., and Kabir, A. H. (2021). Molecular characterization and bioinformatics analysis of transporter genes associated with Cd-induced phytotoxicity in rice (Oryza sativa L.). *Plant Physiol. Biochem.* 167, 438–448. doi: 10.1016/j.plaphy.2021.08.024

Basit, F., Liu, J., An, J., Chen, M., He, C., Zhu, X., et al. (2022). Seed priming with brassinosteroids alleviates aluminum toxicity in rice via improving antioxidant defense system and suppressing aluminum uptake. *Environ. Sci. Pollut. Res.* 29, 10183–10197. doi: 10.1007/s11356-021-16209-y

Batsale, M., Bahammou, D., Fouillen, L., Mongrand, S., Joubès, J., and Domergue, F. (2021). Biosynthesis and Functions of Very-Long-Chain Fatty Acids in the Responses of Plants to Abiotic and Biotic Stresses. *Cells* 10, 1284. doi: 10.3390/cells10061284

Ben Rejeb, I., Pastor, V., and Mauch-Mani, B. (2014). Plant responses to simultaneous biotic and abiotic stress: Molecular mechanisms. *Plants* 3, 458–475. doi: 10.3390/plants3040458

Bhattacharjee, B., Ali, A., Tuteja, N., Gill, S., and Pattanayak, A. (2023). Identification and expression pattern of aluminium-responsive genes in roots of rice genotype with reference to Al-sensitivity. *Sci. Rep.* 13, 12184. doi: 10.1038/s41598-023-39238-8

Bojórquez-Quintal, E., Escalante-Magaña, C., Echevarría-Machado, I., and Martínez-Estévez, M. (2017). Aluminum, a friend or foe of higher plants in acid soils. *Front. Plant Sci.* 8, 1–18. doi: 10.3389/fpls.2017.01767

Bose, J., Babourina, O., and Rengel, Z. (2011). Role of magnesium in alleviation of aluminium toxicity in plants. *J. Exp. Bot.* 62, 2251–2264. doi: 10.1093/jxb/erq456

Brizard, J. P., Carapito, C., Delalande, F., Van Dorsselaer, A., and Brugidou, C. (2006). Proteome Analysis of Plant-Virus Interactome. *Mol. Cell. Proteomics* 5, 2279–2297. doi: 10.1074/mcp.M600173-MCP200

Brown, D. J., and Beevers, H. (1987). Growth and respiration of rice (Oryza sativa L.) Cells in suspension culture. *Plant Cell Tissue Organ Cult.* 10, 175–186. doi: 10.1007/BF00037302

Brumos, J., Robles, L. M., Yun, J., Vu, T. C., Jackson, S., Alonso, J. M., et al. (2018). Local Auxin Biosynthesis Is a Key Regulator of Plant Development. *Dev. Cell* 47, 306-318.e5. doi: 10.1016/j.devcel.2018.09.022

Brunner, I., and Sperisen, C. (2013). Aluminum exclusion and aluminum tolerance in woody plants. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00172

Cai, Z., Mao, X., Li, S., and Wei, L. (2006). Genome comparison using Gene Ontology (GO) with statistical testing. *BMC Bioinformatics* 7, 374. doi: 10.1186/1471-2105-7-374

Campbell, M. T., Du, Q., Liu, K., Sharma, S., Zhang, C., and Walia, H. (2020). Characterization of the transcriptional divergence between the subspecies of cultivated rice (Oryza sativa). *BMC Genomics* 21, 394. doi: 10.1186/s12864-020-06786-6

Cantu, D., Yang, B., Ruan, R., Li, K., Menzo, V., Fu, D., et al. (2013). Comparative analysis of protein-protein interactions in the defense response of rice and wheat. *BMC Genomics* 14, 166. doi: 10.1186/1471-2164-14-166

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., et al. (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics* 25, 288–289. doi: 10.1093/bioinformatics/btn615

Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., et al. (2020). The MetaCyc database of metabolic pathways and enzymes-a 2019 update. *Nucleic Acids Res.* 48, D455–D453. doi: 10.1093/nar/gkz862

CGIAR (2017). CGIAR Research Program RICE contributions to the United Nations Sustainable Development Goals. Available at: https://ricecrp.org/wp-content/uploads/2017/03/RICE-and-SDGs.pdf

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nat. Clim. Change* 4, 287–291. doi: 10.1038/nclimate2153

Chatterjee, A., and Kundu, S. (2015). Revisiting the chlorophyll biosynthesis pathway using genome scale metabolic model of Oryza sativa japonica. *Sci. Rep.* 5. doi: 10.1038/srep14975

Chen, C., Cui, X., Zhang, P., Wang, Z., and Zhang, J. (2021a). Expression of the pyrroline-5-carboxylate reductase (P5CR) gene from the wild grapevine Vitis yeshanensis promotes drought resistance in transgenic Arabidopsis. *Plant Physiol. Biochem.* 168, 188–201. doi: 10.1016/j.plaphy.2021.10.004

Chen, J., Wang, L., and Yuan, M. (2021b). Update on the Roles of Rice MAPK Cascades. *Int. J. Mol. Sci.* 22, 1679. doi: 10.3390/ijms22041679

Chen, W., Tang, L., Wang, J., Zhu, H., Jin, J., Yang, J., et al. (2022). Research Advances in the Mutual Mechanisms Regulating Response of Plant Roots to Phosphate Deficiency and Aluminum Toxicity. *Int. J. Mol. Sci.* 23, 1137. doi: 10.3390/ijms23031137

Cheng, L., Kim, K.-W., and Park, Y.-J. (2019). Evidence for selection events during domestication by extensive mitochondrial genome analysis between japonica and indica in cultivated rice. *Sci. Rep.* 9, 10846. doi: 10.1038/s41598-019-47318-x

Chhun, T., Taketa, S., Tsurumi, S., and Ichii, M. (2004). IBA. *Plant Growth Regul.* 43, 135–143. doi: 10.1023/B:GROW.0000040120.37448.53

Chunquan, Z., Wenjun, H., Xiaochuang, C., Lianfeng, Z., Yali, K., Qianyu, J., et al. (2021). Physiological and Proteomic Analyses Reveal Effects of Putrescine-Alleviated Aluminum Toxicity in Rice Roots. *Rice Sci.* 28, 579–593. doi: 10.1016/j.rsci.2021.03.002

Clavijo-Buriticá, D. C., Arévalo-Ferro, C., and González Barrios, A. F. (2023a). A Holistic Approach from Systems Biology Reveals the Direct Influence of the Quorum-Sensing Phenomenon on Pseudomonas aeruginosa Metabolism to Pyoverdine Biosynthesis. *Metabolites* 13, 659. doi: 10.3390/metabo13050659

Clavijo-Buriticá, D. C., Sosa, C. C., Heredia, R. C., Mosquera, A. J., Álvarez, A., Medina, J., et al. (2023b). Use of Arabidopsis thaliana as a model to understand specific carcinogenic events: Comparison of the molecular machinery associated with cancer-hallmarks in plants and humans. *Heliyon* 9. doi: 10.1016/j.heliyon.2023.e15367

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., et al. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theor. Appl. Genet.* 132, 627–645. doi: 10.1007/s00122-019-03317-0

Cohen, S. P., and Leach, J. E. (2019). Abiotic and biotic stresses induce a core transcriptome response in rice. *Sci. Rep.* 9, 1–11. doi: 10.1038/s41598-019-42731-8

Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M., and Shinozaki, K. (2011). Effects of abiotic stress on plants: a systems biology perspective.pdf (application/pdf Object). *BMC Plant Biol.* 11, 1–14.

Currie, G., Burgess, N., White, L., Lockett, A., Gladman, J., and Waring, J. (2014). *A qualitative study of the knowledge-brokering role of middle-level managers in service innovation: managing the translation gap in patient safety for older persons' care*. Southampton (UK). doi: 10.3310/hsdr02320

Cusick, M. E., Klitgord, N., Vidal, M., and Hill, D. E. (2005). Interactome: gateway into systems biology. *Hum. Mol. Genet.* 14, R171–R181. doi: 10.1093/hmg/ddi335

Damodaran, S., and Strader, L. C. (2019). Indole 3-Butyric Acid Metabolism and Transport in Arabidopsis thaliana. *Front. Plant Sci.* 10, 851. doi: 10.3389/fpls.2019.00851

Das, K., and Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* 2, 1–13. doi: 10.3389/fenvs.2014.00053

Das, S., and Maity, A. (2022). "Utility of Network Biology Approaches to Understand the Aluminum Stress Responses in Soybean," in *Soybean Improvement: Physiological, Molecular and Genetic Perspectives*, eds. S. H. Wani, N. ul R. Sofi, M. A. Bhat, and F. Lin (Cham: Springer International Publishing), 109–124. doi: 10.1007/978-3-031-12232-3_5

Daspute, A. A., Sadhukhan, A., Tokizawa, M., Kobayashi, Y., Panda, S. K., and Koyama, H. (2017). Transcriptional Regulation of Aluminum-Tolerance Genes in Higher Plants: Clarifying the Underlying Molecular Mechanisms. *Front. Plant Sci.* 8, 1358. doi: 10.3389/fpls.2017.01358

De Diego, I. M., Redondo, A. R., Fernández, R. R., Navarro, J., and Moguerza, J. M. (2022). General Performance Score for classification problems. *Appl. Intell.* 52, 12049–12063. doi: 10.1007/s10489-021-03041-7

de Oliveira Dal'Molin, C. G., Quek, L.-E., Palfreyman, R. W., Brumbley, S. M., and Nielsen, L. K. (2010). AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis. *Plant Physiol.* 152, 579–589. doi: 10.1104/pp.109.148817

Depuydt, T., and Vandepoele, K. (2021). Multi-omics network-based functional annotation of unknown Arabidopsis genes. *Plant J.* 108, 1193–1212. doi: https://doi.org/10.1111/tpj.15507

Dharmawardhana, P., Ren, L., Amarasinghe, V., Monaco, M., Thomason, J., Ravenscroft, D., et al. (2013). A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. *Rice* 6, 1–15. doi: 10.1186/1939-8433-6-1

Dien, D. C., Thu, T. T. P., Moe, K., and Yamakawa, T. (2019). Proline and carbohydrate metabolism in rice varieties (Oryza sativa L.) under various drought and recovery conditions. *Plant Physiol. Rep.* 24, 376–387. doi: 10.1007/s40502-019-00462-y

Ding, X., Richter, T., Chen, M., Fujii, H., Seo, Y. S., Xie, M., et al. (2009). A Rice Kinase-Protein Interaction Map. *Plant Physiol.* 149, 1478–1492. doi: 10.1104/pp.108.128298

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., et al. (2017). Sequencing and de novo assembly of a near complete indica rice genome. *Nat. Commun.* 8, 1–12. doi: 10.1038/ncomms15324

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48. doi: 10.1186/1471-2105-10-48

Elbourne, L. D. H., Tetu, S. G., Hassan, K. A., and Paulsen, I. T. (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* 45, D320–D324. doi: 10.1093/nar/gkw1068

Elmsallati, A., Msalati, A., and Kalita, J. (2018). Index-Based Network Aligner of Protein-Protein Interaction Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15, 330–336. doi: 10.1109/TCBB.2016.2613098

Escobar-Sepúlveda, H. F., Trejo-Téllez, L. I., García-Morales, S., and Gómez-Merino, F. C. (2017). Expression patterns and promoter analyses of aluminum-responsive NAC genes suggest a possible growth regulation of rice mediated by aluminum, hormones and NAC transcription factors. *PLOS ONE* 12, e0186084. doi: 10.1371/journal.pone.0186084

Fàbregas, N., and Fernie, A. R. (2021). The interface of central metabolism with hormone signaling in plants. *Curr. Biol.* 31, R1535–R1548. doi: 10.1016/j.cub.2021.09.070

Fahad, S., Bajwa, A. A., Nazir, U., Anjum, S. A., Farooq, A., Zohaib, A., et al. (2017). Crop production under drought and heat stress: Plant responses and management options. *Front. Plant Sci.* 8, 1–16. doi: 10.3389/fpls.2017.01147

Famoso, A. N., Clark, R. T., Shaff, J. E., Craft, E., McCouch, S. R., and Kochian, L. V. (2010). Development of a Novel Aluminum Tolerance Phenotyping Platform Used for Comparisons of Cereal Aluminum Tolerance and Investigations into Rice Aluminum Tolerance Mechanisms. *Plant Physiol.* 153, 1678–1691. doi: 10.1104/pp.110.156794

Famoso, A. N., Zhao, K., Clark, R. T., Tung, C.-W., Wright, M. H., Bustamante, C., et al. (2011). Genetic Architecture of Aluminum Tolerance in Rice (Oryza sativa) Determined through Genome-Wide Association Analysis and QTL Mapping. *PLoS Genet.* 7, e1002221. doi: 10.1371/journal.pgen.1002221

Fernandes, A., and Vinga, S. (2016). Improving Protein Expression Prediction Using Extra Features and Ensemble Averaging. *PLOS ONE* 11, e0150369. doi: 10.1371/journal.pone.0150369

Fisberg, R. M., Leme, A. C. B., Previdelli, Á., de Mello, A. V., Martinez, A. G., Hermes Sales, C., et al. (2021). Contribution of food groups to energy, grams and nutrients-to-limit: the Latin American Study of Nutrition and Health/Estudio Latino Americano de Nutrición y Salud (ELANS). *Public Health Nutr.* 24, 2424–2436. doi: 10.1017/S136898002100152X

Fondi, M., and Liò, P. (2015). "Genome-Scale Metabolic Network Reconstruction," in *Bacterial Pangenomics*, eds. A. Mengoni, M. Galardini, and M. Fondi (New York, NY: Springer New York), 233–256. doi: 10.1007/978-1-4939-1720-4_15

Frick, E. M., and Strader, L. C. (2018). Roles for IBA-derived auxin in plant development. *J. Exp. Bot.* 69, 169–177. doi: 10.1093/jxb/erx298

Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K., et al. (2006). Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr. Opin. Plant Biol.* 9, 436–442. doi: 10.1016/j.pbi.2006.05.014

Fukuda, T., Saito, A., Wasaki, J., Shinano, T., and Osaki, M. (2007). Metabolic alterations proposed by proteome in rice roots grown under low P and high Al concentration under low pH. *Plant Sci.* 172, 1157–1165. doi: 10.1016/j.plantsci.2007.02.020

Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., and Kitano, H. (2008). CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. *Proc. IEEE* 96, 1254–1265. doi: 10.1109/JPROC.2008.925458

Gallo-Franco, J. J., Sosa, C. C., Ghneim-Herrera, T., and Quimbaya, M. (2020). Epigenetic Control of Plant Response to Heavy Metal Stress: A New View on Aluminum Tolerance. *Front. Plant Sci.* 11, 2025–2025. doi: 10.3389/fpls.2020.602625

Gallo-Franco, J. J., Zuluaga-Yusti, I., Restrepo-García, A. M., Sosa, C. C., Zapata-Balanta, S., Gutiérrez-Marín, J. P., et al. (2023). Transcriptional analysis in four wild and cultivated rice genotypes identifies aluminum-induced genes. *Plant Stress* 10, 100247. doi: 10.1016/j.stress.2023.100247

Gao, J. P., Chao, D. Y., and Lin, H. X. (2008). Toward understanding molecular mechanisms of abiotic stress responses in rice. *Rice* 1, 36–51. doi: 10.1007/s12284-008-9006-7

Gao, J., Yan, S., Yu, H., Zhan, M., Guan, K., Wang, Y., et al. (2019). Sweet sorghum (*Sorghum bicolor* L.) SbSTOP1 activates the transcription of a *β* -1,3-glucanase gene to reduce callose deposition under Al toxicity: A novel pathway for Al tolerance in plants. *Biosci. Biotechnol. Biochem.* 83, 446–455. doi: 10.1080/09168451.2018.1540290

Garay-Arroyo, A., De La Paz Sánchez, M., García-Ponce, B., Azpeitia, E., and Álvarez-Buylla, E. R. (2012). Hormone symphony during root growth and development. *Dev. Dyn.* 241, 1867–1885. doi: 10.1002/dvdy.23878

Gaudet, P., and Dessimoz, C. (2017). "Gene Ontology: Pitfalls, Biases, and Remedies," in *The Gene Ontology Handbook*, eds. C. Dessimoz and N. Škunca (New York, NY: Springer New York), 189–205. doi: 10.1007/978-1-4939-3743-1_14

Ge, S. X., Jung, D., Jung, D., and Yao, R. (2020). ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628–2629. doi: 10.1093/bioinformatics/btz931

The Gene Ontology Consortium (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338. doi: 10.1093/nar/gky1055

Giannakoula, A., Moustakas, M., Mylona, P., Papadakis, I., and Yupsanis, T. (2008). Aluminum tolerance in maize is correlated with increased levels of mineral nutrients, carbohydrates and proline, and decreased levels of lipid peroxidation and Al accumulation. *J. Plant Physiol.* 165, 385–396. doi: 10.1016/j.jplph.2007.01.014

Goff, S. A. (2002). A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. japonica). *Science* 296, 92–100. doi: 10.1126/science.1068275

Goh, C.-H., and Lee, Y. (1999). Aluminum uptake and aluminum-induced rapid root growth inhibition of rice seedlings. *J. Plant Biol.* 42, 151–158. doi: 10.1007/BF03031024

Gomez, J. A., Höffner, K., and Barton, P. I. (2014). DFBAlab: A fast and reliable MATLAB code for dynamic flux balance analysis. *BMC Bioinformatics* 15, 1–10. doi: 10.1186/s12859-014-0409-8

Gonçalves, B., Maugarny-Calès, A., Adroher, B., Cortizo, M., Borrega, N., Blein, T., et al. (2017). GDP-L-fucose is required for boundary definition in plants. *J. Exp. Bot.* 68, 5801–5811. doi: 10.1093/jxb/erx402

Grafahrend-Belau, E., Junker, A., Eschenröder, A., Müller, J., Schreiber, F., and Junker, B. H. (2013). Multiscale metabolic modeling: Dynamic flux balance analysis on a whole-plant scale. *Plant Physiol.* 163, 637–647. doi: 10.1104/pp.113.224006

Grienenberger, E., and Quilichini, T. D. (2021). The Toughest Material in the Plant Kingdom: An Update on Sporopollenin. *Front. Plant Sci.* 12, 703864. doi: 10.3389/fpls.2021.703864

Grover, A., and Leskovec, J. (2016). node2vec., in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York, NY, USA: ACM), 855–864. doi: 10.1145/2939672.2939754

Hansen, D. L., Shneiderman, B., Smith, M. A., and Himelboim, I. (2020). “Calculating and visualizing network metrics,” in *Analyzing Social Media Networks with NodeXL*, (Elsevier), 79–94. doi: 10.1016/B978-0-12-817756-3.00006-6

Hawe, J. S., Theis, F. J., and Heinig, M. (2019). Inferring Interaction Networks From Multi-Omics Data. *Front. Genet.* 10, 535. doi: 10.3389/fgene.2019.00535

Hawkins, C., Ginzburg, D., Zhao, K., Dwyer, W., Xue, B., Xu, A., et al. (2021). Plant Metabolic Network 15: A resource of genome-wide metabolism databases for 126 plants and algae. *J. Integr. Plant Biol.* 63, 1888–1905. doi: 10.1111/jipb.13163

Hayashi, K., Hasegawa, J., and Matsunaga, S. (2013). The boundary of the meristematic and elongation zones in roots: endoreduplication precedes rapid cell expansion. *Sci. Rep.* 3, 2723. doi: 10.1038/srep02723

Hillmer, R. A. (2015). Systems Biology for Biologists. *PLOS Pathog.* 11, e1004786. doi: 10.1371/journal.ppat.1004786

Hong, W.-J., Kim, Y.-J., Chandran, A. K. N., and Jung, K.-H. (2019). Infrastructures of systems biology that facilitate functional genomic study in rice. *Rice* 12, 15. doi: 10.1186/s12284-019-0276-z

Hooper, C. M., Castleden, I. R., Aryamanesh, N., Black, K., Grasso, S. V., and Millar, A. H. (2020). CropPAL for discovering divergence in protein subcellular location in crops to support strategies for molecular crop breeding. *Plant J.* 104, 812–827. doi: 10.1111/tpj.14961

Hossain, M. A., Piyatida, P., da Silva, J. A. T., and Fujita, M. (2012). Molecular Mechanism of Heavy Metal Toxicity and Tolerance in Plants: Central Role of Glutathione in Detoxification of Reactive Oxygen Species and Methylglyoxal and in Heavy Metal Chelation. *J. Bot.* 2012, 1–37. doi: 10.1155/2012/872875

Huang, B., Liu, Y., Xue, X., and Chang, L. (2002). Comparison of aluminium tolerance in the brassicas and related species. *Plant Breed.* 121, 360–362. doi: 10.1046/j.1439-0523.2002.00723.x

Huang, C. F., Yamaji, N., Mitani, N., Yano, M., Nagamura, Y., and Ma, J. F. (2009a). A Bacterial-Type ABC Transporter Is Involved in Aluminum Tolerance in Rice. *Plant Cell* 21, 655–667. doi: 10.1105/tpc.108.064543

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., et al. (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 35, W169–W175. doi: 10.1093/nar/gkm415

Huang, G., Kilic, A., Karady, M., Zhang, J., Mehra, P., Song, X., et al. (2022). Ethylene inhibits rice root elongation in compacted soil via ABA- and auxin-mediated mechanisms. *Proc. Natl. Acad. Sci.* 119, e2201072119. doi: 10.1073/pnas.2201072119

Huang, S., Taylor, N. L., Narsai, R., Eubel, H., Whelan, J., and Millar, A. H. (2009c). Experimental Analysis of the Rice Mitochondrial Proteome, Its Biogenesis, and Heterogeneity. *Plant Physiol.* 149, 719–734. doi: 10.1104/pp.108.131300

Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., and Mortensen, D. A. (2017). Agriculture in 2050: Recalibrating targets for sustainable intensification. *BioScience* 67, 386–391. doi: 10.1093/biosci/bix010

Ideker, T., and Sharan, R. (2008). Protein networks in disease. *Genome Res.* 18, 644–652. doi: 10.1101/gr.071852.107

ImmunoMind Team (2019). immunarch: An R Package for Painless Bioinformatics Analysis of T-Cell and B-Cell Immune Repertoires. doi: 10.5281/zenodo.3367200

Itoh, T., Tanaka, T., Barrero, R. A., Yamasaki, C., Fujii, Y., Hilton, P. B., et al. (2007). Curated genome annotation of Oryza sativa ssp. japonica and comparative genome analysis with Arabidopsis thaliana. *Genome Res.* 17, 175–183. doi: 10.1101/gr.5509507

Jackson, S. A. (2016). Rice: The First Crop Genome. *Rice* 9, 14. doi: 10.1186/s12284-016-0087-4

Jagodzik, P., Tajdel-Zielinska, M., Ciesla, A., Marczak, M., and Ludwikow, A. (2018). Mitogen-Activated Protein Kinase Cascades in Plant Hormone Signaling. *Front. Plant Sci.* 9, 1387. doi: 10.3389/fpls.2018.01387

Jancewicz, A. L., Gibbs, N. M., and Masson, P. H. (2016). Cadaverine's Functional Role in Plant Development and Environmental Response. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00870

Jantzen, S. G., Sutherland, B. J., Minkley, D. R., and Koop, B. F. (2011). GO Trimming: Systematically reducing redundancy in large Gene Ontology datasets. *BMC Res. Notes* 4, 267. doi: 10.1186/1756-0500-4-267

Jia, L., Xie, L., Lao, S., Zhu, Q.-H., and Fan, L. (2021). Rice bioinformatics in the genomic era: Status and perspectives. *Crop J.* 9, 609–621. doi: 10.1016/j.cj.2021.03.003

Jingguang, C., Qi, L., Baiquan, Z., Longbiao, G., and Guoyou, Y. (2020). Progress on Molecular Mechanism of Aluminum Resistance in Rice. *Rice Sci.* 27, 454–467. doi: 10.1016/j.rsci.2020.09.003

Kan, C.-C., Chung, T.-Y., Wu, H.-Y., Juo, Y.-A., and Hsieh, M.-H. (2017). Exogenous glutamate rapidly induces the expression of genes involved in metabolism and defense responses in rice roots. *BMC Genomics* 18, 186. doi: 10.1186/s12864-017-3588-7

Kar, D., Pradhan, A. A., and Datta, S. (2021). The role of solute transporters in aluminum toxicity and tolerance. *Physiol. Plant.* 171, 638–652. doi: 10.1111/ppl.13214

Kato-Noguchi, H., and Ohashi, C. (2006). Effects of Anoxia on Amino Acid Levels in Rice Coleoptiles. *Plant Prod. Sci.* 9, 383–387. doi: 10.1626/pps.9.383

Kaur, R., Chandra, J., Varghese, B., and Keshavkant, S. (2023). Allantoin: A Potential Compound for the Mitigation of Adverse Effects of Abiotic Stresses in Plants. *Plants* 12, 3059. doi: 10.3390/plants12173059

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., Mccombie, W. R., Ouyang, S., et al. (2013). Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. *Rice* 6, 3–10. doi: 10.1186/1939-8433-6-4

Keunen, E., Schellingen, K., Vangronsveld, J., and Cuypers, A. (2016). Ethylene and Metal Stress: Small Molecule, Big Impact. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00023

Kim, Y., Chung, Y. S., Lee, E., Tripathi, P., Heo, S., and Kim, K. H. (2020). Root response to drought stress in rice (Oryza sativa L.). *Int. J. Mol. Sci.* 21, 12–14. doi: 10.3390/ijms21041513

Kishor, P. B. K., Suravajhala, R., Rajasheker, G., Marka, N., Shridhar, K. K., Dhulala, D., et al. (2020). Lysine, Lysine-Rich, Serine, and Serine-Rich Proteins: Link Between Metabolism, Development, and Abiotic Stress Tolerance and the Role of ncRNAs in Their Regulation. *Front. Plant Sci.* 11, 546213. doi: 10.3389/fpls.2020.546213

Kitomi, Y., Inahashi, H., Takehisa, H., Sato, Y., and Inukai, Y. (2012). OsIAA13-mediated auxin signaling is involved in lateral root initiation in rice. *Plant Sci.* 190, 116–122. doi: 10.1016/j.plantsci.2012.04.005

Klamt, S., Haus, U.-U., and Theis, F. (2009). Hypergraphs and Cellular Networks. *PLoS Comput. Biol.* 5, e1000385. doi: 10.1371/journal.pcbi.1000385

Kobayashi, Y., and Weigel, D. (2007). Move on up, it's time for change—mobile signals controlling photoperiod-dependent flowering. *Genes Dev.* 21, 2371–2384. doi: 10.1101/gad.1589007

Kochian, L. V., Piñeros, M. A., Liu, J., and Magalhaes, J. V. (2015). Plant Adaptation to Acid Soils: The Molecular Basis for Crop Aluminum Resistance. *Annu. Rev. Plant Biol.* 66, 571–598. doi: 10.1146/annurev-arplant-043014-114822

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler [version 2; peer review: 2 approved]. *F1000Research* 9. doi: 10.12688/f1000research.24956.2

Kopittke, P. M. (2016). Role of phytohormones in aluminium rhizotoxicity. *Plant Cell Environ.* 39, 2319–2328. doi: 10.1111/pce.12786

Korasick, D. A., Enders, T. A., and Strader, L. C. (2013). Auxin biosynthesis and storage forms. *J. Exp. Bot.* 64, 2541–2555. doi: 10.1093/jxb/ert080

Kumar, M., Kumar Patel, M., Kumar, N., Bajpai, A. B., and Siddique, K. H. M. (2021). Metabolomics and Molecular Approaches Reveal Drought Stress Tolerance in Plants. *Int. J. Mol. Sci.* 22, 9108. doi: 10.3390/ijms22179108

Kuriya, Y., and Araki, M. (2020). Dynamic Flux Balance Analysis to Evaluate the Strain Production Performance on Shikimic Acid Production in Escherichia coli. *Metabolites* 10, 198. doi: 10.3390/metabo10050198

Lakshmanan, M., Lim, S.-H., Mohanty, B., Kim, J. K., Ha, S.-H., and Lee, D.-Y. (2015). Unraveling the light-specific metabolic and regulatory signatures of rice through combined in silico modeling and multi-omics analysis. *Plant Physiol.*, pp.01379.2015. doi: 10.1104/pp.15.01379

Lakshmanan, M., Zhang, Z., Mohanty, B., Kwon, J., Choi, H., Nam, H. J., et al. (2013). Elucidating Rice Cell Metabolism under Flooding and Drought Stresses Using Flux-Based Modeling. *Plant Physiol.* 162, 2140–2150. doi: 10.1104/pp.113.220178

Lamers, J., Der Meer, T. V., and Testerink, C. (2020). How plants sense and respond to stressful environments. *Plant Physiol.* 182, 1624–1635. doi: 10.1104/PP.19.01464

Le Novère, N. (2006). Model storage, exchange and integration. *BMC Neurosci.* 7, S11. doi: 10.1186/1471-2202-7-S1-S11

Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52, W78–W82. doi: 10.1093/nar/gkae268

Levy, A., Guenoune-Gelbart, D., and Epel, B. L. (2007). β-1,3-Glucanases: Plasmodesmal Gate Keepers for Intercellular Communication. *Plant Signal. Behav.* 2, 404–407. doi: 10.4161/psb.2.5.4334

Lexer, C., and Fay, M. F. (2005). Adaptation to environmental stress: a rare or frequent driver of speciation? *J. Evol. Biol.* 18, 893–900. doi: 10.1111/j.1420-9101.2005.00901.x

Lexer, C., Welch, M. E., Raymond, O., and Rieseberg, L. H. (2003). The origin of ecological divergence in Helianthus paradoxus (Asteraceae): selection on transgressive characters in a novel hybrid habitat. *Evol. Int. J. Org. Evol.* 57, 1989–2000.

Li, F., Dong, S., Leier, A., Han, M., Guo, X., Xu, J., et al. (2022). Positive-unlabeled learning in bioinformatics and computational biology: a brief review. *Brief. Bioinform.* 23, bbab461. doi: 10.1093/bib/bbab461

Li, J.-Y., Wang, J., and Zeigler, R. S. (2014). The 3,000 rice genomes project: new opportunities and challenges for future rice research. *GigaScience* 3, 8. doi: 10.1186/2047-217X-3-8

Li, K., Jiang, W., Hui, Y., Kong, M., Feng, L.-Y., Gao, L.-Z., et al. (2021). Gapless indica rice genome reveals synergistic contributions of active transposable elements and segmental duplications to rice genome evolution. *Mol. Plant* 14, 1745–1756. doi: 10.1016/j.molp.2021.06.017

Li, P., and Xue, H. (2011). Structural characterization and expression pattern analysis of the rice PLT gene family. *Acta Biochim. Biophys. Sin.* 43, 688–697. doi: 10.1093/abbs/gmr068

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, G., Yang, W., Zhang, X., Peng, T., Zou, Y., Zhang, T., et al. (2019). Cystathionine beta-lyase is crucial for embryo patterning and the maintenance of root stem cell niche in Arabidopsis. *Plant J.* 99, 536–555. doi: https://doi.org/10.1111/tpj.14343

Liu, H., Sachidanandam, R., and Stein, L. (2001). Comparative Genomics Between Rice and *Arabidopsis* Shows Scant Collinearity in Gene Order. *Genome Res.* 11, 2020–2026. doi: 10.1101/gr.194501

Liu, H., Zhu, R., Shu, K., Lv, W., Wang, S., and Wang, C. (2022). Aluminum stress signaling, response, and adaptive mechanisms in plants. *Plant Signal. Behav.* 17, 2057060. doi: 10.1080/15592324.2022.2057060

Liu, J., Piñeros, M. A., and Kochian, L. V. (2014). The role of aluminum sensing and signaling in plant aluminum resistance: Al signaling activates plant resistance responses. *J. Integr. Plant Biol.* 56, 221–230. doi: 10.1111/jipb.12162

Liu, L., Mei, Q., Yu, Z., Sun, T., Zhang, Z., and Chen, M. (2013). An integrative bioinformatics framework for genome-scale multiple level network reconstruction of rice. *J. Integr. Bioinforma.* 10, 223–223. doi: 10.2390/biecoll-jib-2013-223

Liu, R., Hirn, M., and Krishnan, A. (2023). Accurately modeling biased random walks on weighted networks using *node2vec+*. *Bioinformatics* 39, btad047. doi: 10.1093/bioinformatics/btad047

Liu, R., and Krishnan, A. (2021). PecanPy: a fast, efficient and parallelized Python implementation of node2vec. 37, 3377–3379. doi: 10.1093/bioinformatics/btab202

Liu, S., Liu, Y., Zhao, J., Cai, S., Qian, H., Zuo, K., et al. (2017a). A computational interactome for prioritizing genes associated with complex agronomic traits in rice (Oryza sativa). *Plant J.* 90, 177–188. doi: 10.1111/tpj.13475

Liu, X., Lin, Y., Liu, D., Wang, C., Zhao, Z., Cui, X., et al. (2017b). MAPK-mediated auxin signal transduction pathways regulate the malic acid secretion under aluminum stress in wheat (Triticum aestivum L.). *Sci. Rep.* 7, 1620. doi: 10.1038/s41598-017-01803-3

Love, M. I., Anders, S., and Huber, W. (2014). *DESeq2 package__Differential analysis of count data*. doi: 10.1101/002832

Lu, J., Xu, Y., Wang, J., Singer, S. D., and Chen, G. (2020). The Role of Triacylglycerol in Plant Stress Response. *Plants* 9, 472. doi: 10.3390/plants9040472

Lu, L., Chen, X., Tan, Q., Li, W., Sun, Y., Zhang, Z., et al. (2024a). Gibberellin-Mediated Sensitivity of Rice Roots to Aluminum Stress. *Plants* 13, 543. doi: 10.3390/plants13040543

Lu, L., Chen, X., Tan, Q., Li, W., Sun, Y., Zhang, Z., et al. (2024b). Gibberellin-Mediated Sensitivity of Rice Roots to Aluminum Stress. *Plants* 13, 543. doi: 10.3390/plants13040543

Luo, C., Shi, Y., and Xiang, Y. (2022). SNAREs Regulate Vesicle Trafficking During Root Growth and Development. *Front. Plant Sci.* 13, 853251. doi: 10.3389/fpls.2022.853251

Luo, R., Liao, S., Tao, G., Li, Y., Zeng, S., Li, Y., et al. (2006). Dynamic analysis of optimality in myocardial energy metabolism under normal and ischemic conditions. *Mol. Syst. Biol.* 2, 2006.0031. doi: 10.1038/msb4100071

Luo, R., Wei, H., Ye, L., Wang, K., Chen, F., Luo, L., et al. (2009a). Photosynthetic metabolism of C $_3$ plants shows highly cooperative regulation under changing environments: A systems biological analysis. *Proc. Natl. Acad. Sci.* 106, 847–852. doi: 10.1073/pnas.0810731105

Luo, R., Wei, H., Ye, L., Wang, K., Chen, F., Luo, L., et al. (2009b). Photosynthetic metabolism of C_3 plants shows highly cooperative regulation under changing environments: A systems biological analysis. *Proc. Natl. Acad. Sci.* 106, 847–852. doi: 10.1073/pnas.0810731105

Ma, B., Gao, L., Zhang, H., Cui, J., and Shen, Z. (2012). Aluminum-induced oxidative stress and changes in antioxidant defenses in the roots of rice varieties differing in Al tolerance. *Plant Cell Rep.* 31, 687–696. doi: 10.1007/s00299-011-1187-7

Ma, C., Zhang, H. H., and Wang, X. (2014a). Machine learning for Big Data analytics in plants. *Trends Plant Sci.* 19, 798–808. doi: 10.1016/j.tplants.2014.08.004

Ma, J. F., Chen, Z. C., and Shen, R. F. (2014b). Molecular mechanisms of Al tolerance in gramineous plants. *Plant Soil* 381, 1–12. doi: 10.1007/s11104-014-2073-1

Ma, S., and Bohnert, H. J. (2007). Integration of Arabidopsis thaliana stress-related transcript profiles, promoter structures, and cell-specific expression. *Genome Biol.* 8. doi: 10.1186/gb-2007-8-4-r49

Machné, R., Finney, A., Müller, S., Lu, J., Widder, S., and Flamm, C. (2006). The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* 22, 1406–1407. doi: 10.1093/bioinformatics/btl086

Madani, B., Mirshekari, A., and Imahori, Y. (2019). “Physiological Responses to Stress,” in *Postharvest Physiology and Biochemistry of Fruits and Vegetables*, (Elsevier), 405–423. doi: 10.1016/B978-0-12-813278-4.00020-8

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* 21, 3448–3449. doi: 10.1093/bioinformatics/bti551

Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth. *Biophys. J.* 83, 1331–1340.

Mahood, E. H., Kruse, L. H., and Moghe, G. D. (2020). Machine learning: A powerful tool for gene function prediction in plants. *Appl. Plant Sci.* 8, e11376. doi: 10.1002/aps3.11376

Majeed, Y., Zhu, X., Zhang, N., ul-Ain, N., Raza, A., Haider, F. U., et al. (2023). Harnessing the role of mitogen-activated protein kinases against abiotic stresses in plants. *Front. Plant Sci.* 14, 932923. doi: 10.3389/fpls.2023.932923

Maranas, C. D., and Zomorrodi, A. (2016a). “Flux Balance Analysis and LP problems,” in *Optimization methods in metabolic network*, (Wiley), 52–80.

Maranas, C. D., and Zomorrodi, A. R. (2016b). *Optimization methods in Metabolic Networks*., ed.John Wiley & Sons Inc. New Jersey.

Mathur, P., Tripathi, D. K., Baluška, F., and Mukherjee, S. (2022). Auxin-mediated molecular mechanisms of heavy metal and metalloid stress regulation in plants. *Environ. Exp. Bot.* 196, 104796. doi: 10.1016/j.envexpbot.2022.104796

Mazandu, G. K., Chimusa, E. R., and Mulder, N. J. (2016). Gene Ontology semantic similarity tools: survey on features and challenges for biological knowledge discovery. *Brief. Bioinform.*, bbw067. doi: 10.1093/bib/bbw067

McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci.* 106, 12273–12278. doi: 10.1073/pnas.0900992106

Meijer, R. J., and Goeman, J. J. (2016). Multiple Testing of Gene Sets from Gene Ontology: Possibilities and Pitfalls. *Brief. Bioinform.* 17, 808–818. doi: 10.1093/bib/bbv091

Meng, F., Xiang, D., Zhu, J., Li, Y., and Mao, C. (2019). Molecular Mechanisms of Root Development in Rice. *Rice* 12, 1. doi: 10.1186/s12284-018-0262-x

Mhamdi, A., and Van Breusegem, F. (2018). Reactive oxygen species in plant development. *Dev. Camb.* 145. doi: 10.1242/dev.164376

Mishra, B., Kumar, N., and Shahid Mukhtar, M. (2022). A rice protein interaction network reveals high centrality nodes and candidate pathogen effector targets. *Comput. Struct. Biotechnol. J.* 20, 2001–2012. doi: 10.1016/j.csbj.2022.04.027

Montgomery, S. H., and Mank, J. E. (2016). Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Mol. Ecol.* 25, 5114–5128. doi: 10.1111/mec.13824

Mordelet, F., and Vert, J.-P. (2014). A bagging SVM to learn from positive and unlabeled examples. *Partial. Supervised Learn. Pattern Recognit.* 37, 201–209. doi: 10.1016/j.patrec.2013.06.010

Moreno-Alvarado, M., García-Morales, S., Trejo-Téllez, L. I., Hidalgo-Contreras, J. V., and Gómez-Merino, F. C. (2017). Aluminum Enhances Growth and Sugar Concentration, Alters Macronutrient Status and Regulates the Expression of NAC Transcription Factors in Rice. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00073

Morkunas, I., Wozniak, A., Mai, V. C., Rucinska-Sobkowiak, R., and Jeandet, P. (2018). The role of heavy metals in plant response to biotic stress. *Molecules* 23, 1–30. doi: 10.3390/molecules23092320

Mosa, K. A., Ismail, A., and Helmy, M. (2017). "Introduction to Plant Stresses," in *SpringerBriefs in Systems Biology*, 1–19. doi: 10.1007/978-3-319-59379-1_1

Mulder, N. J., Akinola, R. O., Mazandu, G. K., and Rapanoel, H. (2014). Using biological networks to improve our understanding of infectious diseases. *Comput. Struct. Biotechnol. J.* 11, 1–10. doi: 10.1016/j.csbj.2014.08.006

Murabito, E., Simeonidis, E., Smallbone, K., and Swinton, J. (2009). Capturing the essence of a metabolic network: A flux balance analysis approach. *J. Theor. Biol.* 260, 445–452. doi: 10.1016/j.jtbi.2009.06.013

Muthuramalingam, P., Jeyasri, R., Selvaraj, A., Kalaiyarasi, D., Aruni, W., Pandian, S. T. K., et al. (2021). Global transcriptome analysis of novel stress associated protein (*SAP*) genes expression dynamism of combined abiotic stresses in *Oryza sativa* (L.). *J. Biomol. Struct. Dyn.* 39, 2106–2117. doi: 10.1080/07391102.2020.1747548

Nagayama, T., Nakamura, A., Yamaji, N., Satoh, S., Furukawa, J., and Iwai, H. (2019). Changes in the Distribution of Pectin in Root Border Cells Under Aluminum Stress. *Front. Plant Sci.* 10, 1216. doi: 10.3389/fpls.2019.01216

Nakamura, Y., Itoh, T., and Martin, W. (2007). Rate and Polarity of Gene Fusion and Fission in Oryza sativa and Arabidopsis thaliana. *Mol. Biol. Evol.* 24, 110–121. doi: 10.1093/molbev/msl138

Narusaka, Y., Narusaka, M., Seki, M., Umezawa, T., Ishida, J., Nakajima, M., et al. (2004). Crosstalk in the responses to abiotic and biotic stresses in Arabidopsis: Analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray. *Plant Mol. Biol.* 55, 327–342. doi: 10.1007/s11103-004-0685-1

Nelson, W., Zitnik, M., Wang, B., Leskovec, J., Goldenberg, A., and Sharan, R. (2019). To Embed or Not: Network Embedding as a Paradigm in Computational Biology. *Front. Genet.* 10, 381. doi: 10.3389/fgene.2019.00381

Netotea, S., Sundell, D., Street, N. R., and Hvidsten, T. R. (2014). ComPlEx: conservation and divergence of co-expression networks in A. thaliana, Populus and O. sativa. *BMC Genomics* 15, 106. doi: 10.1186/1471-2164-15-106

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Rev.* 45, 167–256.

Nykiel, M., Gietler, M., Fidler, J., Prabucka, B., Rybarczyk-Płońska, A., Graska, J., et al. (2022). Signal Transduction in Cereal Plants Struggling with Environmental Stresses: From Perception to Response. *Plants* 11, 1009. doi: 10.3390/plants11081009

O'Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using Genome-scale Models to Predict Biological Capabilities. *Cell* 161, 971–987. doi: 10.1016/j.cell.2015.05.019

Ofoe, R., Thomas, R. H., Asiedu, S. K., Wang-Pruski, G., Fofana, B., and Abbey, Lord (2023). Aluminum in plant: Benefits, toxicity and tolerance mechanisms. *Front. Plant Sci.* 13, 1085998. doi: 10.3389/fpls.2022.1085998

Opassiri, R., Pomthong, B., Akiyama, T., Nakphaichit, M., Onkoksoong, T., Ketudat Cairns, M., et al. (2007). A stress-induced rice (Oryza sativa L.) β-glucosidase represents a new subfamily of glycosyl hydrolase family 5 containing a fascin-like domain. *Biochem. J.* 408, 241–249. doi: 10.1042/BJ20070734

Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi: 10.1038/nbt.1614

Panda, S. K., Baluška, F., and Matsumoto, H. (2009). Aluminum stress signaling in plants. *Plant Signal. Behav.* 4, 592–597. doi: 10.4161/psb.4.7.8903

Pandey, P., Irulappan, V., Bagavathiannan, M. V., and Senthil-Kumar, M. (2017). Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Front. Plant Sci.* 8, 1–15. doi: 10.3389/fpls.2017.00537

Pandey, P., Srivastava, S., Pandey, A. K., and Dubey, R. S. (2020). "Abiotic-stress tolerance in plants-system biology approach," in *Plant Life Under Changing Environment*, (Elsevier), 577–609. doi: 10.1016/B978-0-12-818204-8.00025-4

Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V., and Jugulam, M. (2020). Role of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants* 9, 454. doi: 10.3390/antiox9050454

Pareek, A., Singh, A., Kumar, M., Kushwaha, H. R., Lynn, A. M., and Singla-Pareek, S. L. (2006). Whole-Genome Analysis of Oryza sativa Reveals Similar Architecture of Two-Component Signaling Machinery with Arabidopsis. *Plant Physiol.* 142, 380–397. doi: 10.1104/pp.106.086371

Pathak, R. K., Taj, G., Pandey, D., Arora, S., and Kumar, A. (2013). Modeling of the MAPK machinery activation in response to various abiotic and biotic stresses in plants by a system biology approach. *Bioinformation* 9, 443–449. doi: 10.6026/97320630009443

Pazhamala, L. T., Kudapa, H., Weckwerth, W., Millar, A. H., and Varshney, R. K. (2021). Systems biology for crop improvement. *Plant Genome* 14, e20098. doi: 10.1002/tpg2.20098

Peixoto, P. H. P., Cambraia, J., Sant'Anna, R., Mosquim, P. R., and Moreira, M. A. (2001). ALUMINUM EFFECTS ON FATTY ACID COMPOSITION AND LIPID PEROXIDATION OF A PURIFIED PLASMA MEMBRANE FRACTION OF ROOT APICES OF TWO SORGHUM CULTIVARS. *J. Plant Nutr.* 24, 1061–1070. doi: 10.1081/PLN-100103803

Per, T. S., Khan, N. A., Reddy, P. S., Masood, A., Hasanuzzaman, M., Khan, M. I. R., et al. (2017a). Approaches in modulating proline metabolism in plants for salt and drought stress tolerance: Phytohormones, mineral nutrients and transgenics. *Plant Physiol. Biochem.* 115, 126–140. doi: 10.1016/j.plaphy.2017.03.018

Per, T. S., Khan, N. A., Reddy, P. S., Masood, A., Hasanuzzaman, M., Khan, M. I. R., et al. (2017b). Approaches in modulating proline metabolism in plants for salt and drought stress tolerance: Phytohormones, mineral nutrients and transgenics. *Plant Physiol. Biochem.* 115, 126–140. doi: 10.1016/j.plaphy.2017.03.018

Pereira, A. (2016). Plant Abiotic Stress Challenges from the Changing Environment. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01123

Peres da Rocha Oliveiros Marciano, D., Toledo Ramos, F., Neiva Alvim, M., Ronaldo Magalhaes, J., and Giovanni Costa França, M. (2010). Nitric oxide reduces the stress effects of aluminum on the process of germination and early root growth of rice. *J. Plant Nutr. Soil Sci.* 173, 885–891. doi: https://doi.org/10.1002/jpln.200900312

Pesquita, Catia (2017). "Semantic Similarity in the Gene Ontology," in *The gene ontology handbook*, (New York, NY: Springer New York), 161–174. Available at: https://link.springer.com/protocol/10.1007/978-1-4939-3743-1_12

Poole, N., Donovan, J., and Erenstein, O. (2022). Continuing cereals research for sustainable health and well-being. *Int. J. Agric. Sustain.* 20, 693–704. doi: 10.1080/14735903.2021.1975437

Poolman, M. G., Kundu, S., Shaw, R., and Fell, D. A. (2013). Responses to Light Intensity in a Genome-Scale Model of Rice Metabolism. *Plant Physiol.* 162, 1060–1072. doi: 10.1104/pp.113.216762

Poolman, M. G., Kundu, S., Shaw, R., and Fell, D. A. (2014). Metabolic trade-offs between biomass synthesis and photosynthate export at different light intensities in a genomeâ€"scale metabolic model of rice. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00656

Poot-Poot, W., and Hernandez-Sotomayor, S. M. T. (2011). Aluminum stress and its role in the phospholipid signaling pathway in plants and possible biotechnological applications. *IUBMB Life* 63, 864–872. doi: https://doi.org/10.1002/iub.550

Poschenrieder, C., Busoms, S., and Barceló, J. (2019). How Plants Handle Trivalent (+3) Elements. *Int. J. Mol. Sci.* 20, 3984. doi: 10.3390/ijms20163984

Qin, H., He, L., and Huang, R. (2019). The Coordination of Ethylene and Other Hormones in Primary Root Development. *Front. Plant Sci.* 10, 874. doi: 10.3389/fpls.2019.00874

Rahman, M. A., Lee, S. H., Ji, H. C., Kabir, A. H., Jones, C. S., and Lee, K. W. (2018). Importance of mineral nutrition for mitigating aluminum toxicity in plants on acidic soils: Current status and opportunities. *Int. J. Mol. Sci.* 19. doi: 10.3390/ijms19103073

Rajewska, I., Talarek, M., and Bajguz, A. (2016). Brassinosteroids and Response of Plants to Heavy Metals Action. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00629

Ranjan, A., Sinha, R., Lal, S. K., Bishi, S. K., and Singh, A. K. (2021a). Phytohormone signalling and cross-talk to alleviate aluminium toxicity in plants. *Plant Cell Rep.* 40, 1331–1343. doi: 10.1007/s00299-021-02724-2

Ranjan, A., Sinha, R., Sharma, T. R., Pattanayak, A., and Singh, A. K. (2021b). Alleviating aluminum toxicity in plants: Implications of reactive oxygen species signaling and crosstalk with other signaling pathways. *Physiol. Plant.* 173, 1765–1784. doi: 10.1111/ppl.13382

Rao, K. P., Vani, G., Kumar, K., Wankhede, D. P., Misra, M., Gupta, M., et al. (2011). Arsenic stress activates MAP kinase in rice roots and leaves. *Arch. Biochem. Biophys.* 506, 73–82. doi: 10.1016/j.abb.2010.11.006

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi: 10.1093/nar/gkz369

Ray, S., Dansana, P. K., Giri, J., Deveshwar, P., Arora, R., Agarwal, P., et al. (2011). Modulation of transcription factor and metabolic pathway genes in response to water-deficit stress in rice. *Funct. Integr. Genomics* 11, 157–178. doi: 10.1007/s10142-010-0187-y

Rico-Chávez, A. K., Franco, J. A., Fernandez-Jaramillo, A. A., Contreras-Medina, L. M., Guevara-González, R. G., and Hernandez-Escobedo, Q. (2022). Machine Learning for Plant Stress Modeling: A Perspective towards Hormesis Management. *Plants* 11, 970. doi: 10.3390/plants11070970

Rigault, P., Boyle, B., Lepage, P., Cooke, J. E. K., Bousquet, J., and MacKay, J. J. (2011). A White Spruce Gene Catalog for Conifer Genome Analyses. *Plant Physiol.* 157, 14–28. doi: 10.1104/pp.111.179663

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., and Mittler, R. (2004). When Defense Pathways Collide. The Response of Arabidopsis to a Combination of Drought and Heat Stress. *Plant Physiol.* 134, 1683–1696. doi: 10.1104/pp.103.033431

Rosa-Téllez, S., Anoman, A. D., Flores-Tornero, M., Toujani, W., Alseek, S., Fernie, A. R., et al. (2018). Phosphoglycerate Kinases Are Co-Regulated to Adjust Metabolism and to Optimize Growth. *Plant Physiol.* 176, 1182–1198. doi: 10.1104/pp.17.01227

Růžička, K., Ljung, K., Vanneste, S., Podhorská, R., Beeckman, T., Friml, J., et al. (2007). Ethylene Regulates Root Growth through Effects on Auxin Biosynthesis and Transport-Dependent Auxin Distribution. *Plant Cell* 19, 2197–2212. doi: 10.1105/tpc.107.052126

Saadat, N. P., van Aalst, M., and Ebenhöh, O. (2022). Network Reconstruction and Modelling Made Reproducible with moped. *Metabolites* 12. doi: 10.3390/metabo12040275

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013). Rice annotation project database (RAP-DB): An integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54, 1–11. doi: 10.1093/pcp/pcs183

Salicrú, M., Ocaña, J., and Sánchez-Pla, A. (2011). Comparison of lists of genes based on functional profiles. *BMC Bioinformatics* 12, 401. doi: 10.1186/1471-2105-12-401

Santos, F., Boele, J., and Teusink, B. (2011). “A Practical Guide to Genome-Scale Metabolic Models and Their,” in *Methods in Systems Biology*, (Elsevier Inc.), 509–532. doi: 10.1016/B978-0-12-385118-5.00024-4

Satish Kumar, V., Dasika, M. S., and Maranas, C. D. (2007). Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* 8, 1–16. doi: 10.1186/1471-2105-8-212

Schellenberger, J., Lewis, N. E., and Palsson, B. Ø. (2011). Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models. *Biophys. J.* 100, 544–553. doi: 10.1016/j.bpj.2010.12.3707

Schläpfer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., et al. (2017). Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. *Plant Physiol.* 173, 2041–2059. doi: 10.1104/pp.16.01942

Schomburg, I. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* 32, 431D – 433. doi: 10.1093/nar/gkh081

Schoof, H., and Karlowski, W. M. (2003). Comparison of rice and Arabidopsis annotation. *Curr. Opin. Plant Biol.* 6, 106–112. doi: 10.1016/S1369-5266(03)00003-7

Scott, F., Wilson, P., Conejeros, R., and Vassiliadis, V. S. (2018). Simulation and optimization of dynamic flux balance analysis models using an interior point method reformulation. *Comput. Chem. Eng.* 119, 152–170. doi: 10.1016/j.compchemeng.2018.08.041

Senger, R. S., and Papoutsakis, E. T. (2008). Genome-scale model for Clostridium acetobutylicuM : Part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng.* 101, 1036–1052. doi: 10.1002/bit.22010

Seo, Y.-S., Chern, M., Bartley, L. E., Han, M., Jung, K.-H., Lee, I., et al. (2011). Towards Establishment of a Rice Stress Response Interactome. *PLoS Genet.* 7, e1002020. doi: 10.1371/journal.pgen.1002020

Shaik, R., and Ramakrishna, W. (2014). Machine Learning Approaches Distinguish Multiple Stress Conditions using Stress-Responsive Genes and Identify Candidate Genes for Broad Resistance in Rice. *PLANT Physiol.* 164, 481–495. doi: 10.1104/pp.113.225862

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sharan, A., Soni, P., Nongpiur, R. C., Singla-Pareek, S. L., and Pareek, A. (2017). Mapping the ‘Two-component system’ network in rice. *Sci. Rep.* 7, 9287. doi: 10.1038/s41598-017-08076-w

Sharma, A. D., Vasudeva, R., and Kaur, R. (2006). Expression of a boiling-stable protein (BsCyp) in response to heat shock, drought and ABA treatments in Sorghum bicolor. *Plant Growth Regul.* doi: 10.1007/s10725-006-0010-x

Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M., and Zheng, B. (2019). Response of Phenylpropanoid Pathway and the Role of Polyphenols in Plants under Abiotic Stress. *Molecules* 24, 2452. doi: 10.3390/molecules24132452

Shaw, R., and Cheung, C. Y. M. (2018). A Dynamic Multi-Tissue Flux Balance Model Captures Carbon and Nitrogen Metabolism and Optimal Resource Partitioning During Arabidopsis Growth. *Front. Plant Sci.* 9, 884. doi: 10.3389/fpls.2018.00884

Shaw, R., and Cheung, C. Y. M. (2021). Integration of crop growth model and constraint-based metabolic model predicts metabolic changes over rice plant development under water-limited stress. *Silico Plants* 3, diab020. doi: 10.1093/insilicoplants/diab020

Shen, F., Wu, X., Shi, L., Zhang, H., Chen, Y., Qi, X., et al. (2018). Transcriptomic and metabolic flux analyses reveal shift of metabolic patterns during rice grain development. *BMC Syst. Biol.* 12. doi: 10.1186/s12918-018-0574-x

Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø., and Ruppin, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* 26, 1003–1010. doi: 10.1038/nbt.1487

Shu, K., Zhou, W., Chen, F., Luo, X., and Yang, W. (2018). Abscisic Acid and Gibberellins Antagonistically Mediate Plant Development and Abiotic Stress Responses. *Front. Plant Sci.* 9, 416. doi: 10.3389/fpls.2018.00416

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* 21, 110–124. doi: 10.1016/j.tplants.2015.10.015

Singh, R., and Jwa, N.-S. (2013). The rice MAPKK–MAPK interactome: the biological significance of MAPK components in hormone signal transduction. *Plant Cell Rep.* 32, 923–931. doi: 10.1007/s00299-013-1437-y

Singh, R., Lee, M.-O., Lee, J.-E., Choi, J., Park, J. H., Kim, E. H., et al. (2012). Rice Mitogen-Activated Protein Kinase Interactome Analysis Using the Yeast Two-Hybrid System. *Plant Physiol.* 160, 477–487. doi: 10.1104/pp.112.200071

Smet, D., Opdebeeck, H., and Vandepoele, K. (2023). Predicting transcriptional responses to heat and drought stress from genomic features using a machine learning approach in rice. *Front. Plant Sci.* 14, 1212073. doi: 10.3389/fpls.2023.1212073

Sosa, C. C. (2022). Genes/proteins involved in aluminum tolerance in Oryza sativa subsp. japonica. doi: 10.17632/fzxjwrt53m.3

Sosa, C. C., Clavijo-Buriticá, D. C., García-Merchán, V. H., López-Rozo, N., Riccio-Rengifo, C., Diaz, M. V., et al. (2023). GOCompare: An R package to compare functional enrichment analysis between two species. *Genomics* 115, 110528. doi: 10.1016/j.ygeno.2022.110528

Struys-Ponsar, C., Guillard, O., and Van Den Bosch De Aguilar, Ph. (2000). Effects of Aluminum Exposure on Glutamate Metabolism: A Possible Explanation for Its Toxicity. *Exp. Neurol.* 163, 157–164. doi: 10.1006/exnr.2000.7355

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Sun, P., Tian, Q.-Y., Chen, J., and Zhang, W.-H. (2010). Aluminium-induced inhibition of root elongation in Arabidopsis is mediated by ethylene and auxin. *J. Exp. Bot.* 61, 347–356. doi: 10.1093/jxb/erp306

Sun, P., Tian, Q.-Y., Zhao, M.-G., Dai, X.-Y., Huang, J.-H., Li, L.-H., et al. (2007). Aluminum-Induced Ethylene Production is Associated with Inhibition of Root Elongation in Lotus japonicus L. *Plant Cell Physiol.* 48, 1229–1235. doi: 10.1093/pcp/pcm077

Sun, S.-K., Xu, X., Tang, Z., Tang, Z., Huang, X.-Y., Wirtz, M., et al. (2021). A molecular switch in sulfur metabolism to reduce arsenic and enrich selenium in rice grain. *Nat. Commun.* 12, 1392. doi: 10.1038/s41467-021-21282-5

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* 6, e21800. doi: 10.1371/journal.pone.0021800

Swindell, W. R. (2006). The association among gene expression responses to nine abiotic stress treatments in Arabidopsis thaliana. *Genetics* 174, 1811–1824. doi: 10.1534/genetics.106.061374

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Takehisa, H., Sato, Y., Igarashi, M., Abiko, T., Antonio, B. A., Kamatsuki, K., et al. (2012). Genome-wide transcriptome dissection of the rice root system: implications for developmental and physiological functions. *Plant J.* 69, 126–140. doi: https://doi.org/10.1111/j.1365-313X.2011.04777.x

Tang, Y., Bao, X., Zhi, Y., Wu, Q., Guo, Y., Yin, X., et al. (2019). Overexpression of a MYB Family Gene, OsMYB6, Increases Drought and Salinity Stress Tolerance in Transgenic Rice. *Front. Plant Sci.* 10, 168. doi: 10.3389/fpls.2019.00168

Taylor, N. L., Howell, K. A., Heazlewood, J. L., Tan, T. Y. W., Narsai, R., Huang, S., et al. (2010). Analysis of the Rice Mitochondrial Carrier Family Reveals Anaerobic Accumulation of a Basic Amino Acid Carrier Involved in Arginine Metabolism during Seed Germination. *Plant Physiol.* 154, 691–704. doi: 10.1104/pp.110.162214

Tello-Ruiz, M. K., Naithani, S., Gupta, P., Olson, A., Wei, S., Preece, J., et al. (2021). Gramene 2021: harnessing the power of comparative genomics and pathways for plant research. *Nucleic Acids Res.* 49, D1452–D1463. doi: 10.1093/nar/gkaa979

Thanmalagan, R. R., Roy, A., Jayaprakash, A., and Lakshmi, P. T. V. (2022). Comprehensive meta-analysis and machine learning approaches identified the role of novel drought specific genes in Oryza sativa. *Plant Gene* 32, 100382. doi: 10.1016/j.plgene.2022.100382

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408, 796–815. doi: 10.1038/35048692

Thiele, I., and Palsson, B. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121. doi: 10.1038/nprot.2009.203

Thomas, Paul, D. (2017). “The Gene Ontology and the Meaning of Biological Function,” in *The Gene Ontology Handbook*, (New York, NY: Springer New York). Available at: http://link.springer.com/10.1007/978-1-4939-3743-1

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: doi.org/10.1093/nar/gkx382

Timmons, J. A., Szkop, K. J., and Gallagher, I. J. (2015). Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biol.* 16, 186. doi: 10.1186/s13059-015-0761-7

Tipney, H., and Hunter, L. (2010). An introduction to effective use of enrichment analysis software. *Hum. Genomics* 4, 202. doi: 10.1186/1479-7364-4-3-202

Tippmann, S. (2015). Programming tools: Adventures with R. *Nature* 517, 109–110. doi: 10.1038/517109a

Tripathi, B., Parthasarathy, S., Sinha, H., Raman, K., and Ravindran, B. (2019). Adapting community detection algorithms for disease module identification in heterogeneous biological networks. *Front. Genet.* 10, 1–17. doi: 10.3389/fgene.2019.00164

Tyagi, W., Yumnam, J. S., Sen, D., and Rai, M. (2020). Root transcriptome reveals efficient cell signaling and energy conservation key to aluminum toxicity tolerance in acidic soil adapted rice genotype. *Sci. Rep.* 10, 4580. doi: 10.1038/s41598-020-61305-7

Uhrig, J. F. (2006). Protein interaction networks in plants. *Planta* 224, 771–781. doi: 10.1007/s00425-006-0260-x

Van De Poel, B., Smet, D., and Van Der Straeten, D. (2015). Ethylene and Hormonal Cross Talk in Vegetative Growth and Development. *Plant Physiol.* 169, 61–72. doi: 10.1104/pp.15.00724

Vandenbussche, F., Vaseva, I., Vissenberg, K., and Van Der Straeten, D. (2012). Ethylene in vegetative development: a tale with a riddle. *New Phytol.* 194, 895–909. doi: 10.1111/j.1469-8137.2012.04100.x

Vandereyken, K., Van Leene, J., De Coninck, B., and Cammue, B. P. A. (2018). Hub Protein Controversy: Taking a Closer Look at Plant Stress Response Hubs. *Front. Plant Sci.* 9, 694. doi: 10.3389/fpls.2018.00694

Varoquaux, N., Cole, B., Gao, C., Pierroz, G., Baker, C. R., Patel, D., et al. (2019). Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. *Proc. Natl. Acad. Sci.* 116, 27124–27132. doi: 10.1073/pnas.1907500116

Vemanna, R. S., Bakade, R., Bharti, P., Kumar, M. K. P., Sreeman, S. M., Senthil-Kumar, M., et al. (2019). Cross-talk signaling in rice during combined drought and bacterial blight stress. *Front. Plant Sci.* 10, 1–11. doi: 10.3389/fpls.2019.00193

Vert, G., and Chory, J. (2011). Crosstalk in Cellular Signaling: Background Noise or the Real Thing? *Dev. Cell* 21, 985–991. doi: 10.1016/j.devcel.2011.11.006

Waadt, R., Seller, C. A., Hsu, P.-K., Takahashi, Y., Munemasa, S., and Schroeder, J. I. (2022). Plant hormone regulation of abiotic stress responses. *Nat. Rev. Mol. Cell Biol.* 23, 680–694. doi: 10.1038/s41580-022-00479-6

Wagatsuma, T., Maejima, E., Watanabe, T., Toyomasu, T., Kuroda, M., Muranaka, T., et al. (2017). Dark conditions enhance aluminum tolerance in several rice cultivars via multiple modulations of membrane sterols. *J. Exp. Bot.* 69, 567–577. doi: 10.1093/jxb/erx414

Wang, C. Y., Shen, R. F., Wang, C., and Wang, W. (2013). Root protein profile changes induced by Al exposure in two rice cultivars differing in Al tolerance. *J. Proteomics* 78, 281–293. doi: 10.1016/j.jprot.2012.09.035

Wang, J., Su, C., Cui, Z., Huang, L., Gu, S., Jiang, S., et al. (2023a). Transcriptomics and metabolomics reveal tolerance new mechanism of rice roots to Al stress. *Front. Genet.* 13, 1063984. doi: 10.3389/fgene.2022.1063984

Wang, M., Qiao, J., Yu, C., Chen, H., Sun, C., Huang, L., et al. (2019). The auxin influx carrier, OsAUX3, regulates rice root development and responses to aluminium stress. *Plant Cell Environ.* 42, 1125–1138. doi: 10.1111/pce.13478

Wang, P., Wan, N., Horst, W. J., and Yang, Z.-B. (2023b). From stress to responses: aluminium-induced signalling in the root apex. *J. Exp. Bot.* 74, 1358–1371. doi: 10.1093/jxb/erac516

Wang, T., Hu, Y., Chen, H., Tan, J., Xu, H., Li, P., et al. (2022). Transcriptome Analysis of Response to Aluminum Stress in Pinus massoniana. *Forests* 13, 837. doi: 10.3390/f13060837

Wang, W., Zhao, X. Q., Hu, Z. M., Shao, J. F., Che, J., Chen, R. F., et al. (2015). Aluminium alleviates manganese toxicity to rice by decreasing root symplastic Mn uptake and reducing availability to shoots of Mn stored in roots. *Ann. Bot.* 116, 237–246. doi: 10.1093/aob/mcv090

Wang, Z. Q., Xu, X. Y., Gong, Q. Q., Xie, C., Fan, W., Yang, J. L., et al. (2014). Root proteome of rice studied by iTRAQ provides integrated insight into aluminum stress tolerance mechanisms in plants. *J. Proteomics* 98, 189–205. doi: 10.1016/j.jprot.2013.12.023

Wanichthanarak, K., Boonchai, C., Kojonna, T., Chadchawan, S., Sangwongchai, W., and Thitisaksakul, M. (2020). Deciphering rice metabolic flux reprograming under salinity stress via in silico metabolic modeling. *Comput. Struct. Biotechnol. J.* 18, 3555–3566. doi: 10.1016/j.csbj.2020.11.023

Wei, Y., Han, R., Xie, Y., Jiang, C., and Yu, Y. (2021). Recent Advances in Understanding Mechanisms of Plant Tolerance and Response to Aluminum Toxicity. *Sustainability* 13, 1782. doi: 10.3390/su13041782

Wen, J.-Q., Oono, K., and Imai, R. (2002). Two Novel Mitogen-Activated Protein Signaling Components, OsMEK1 and OsMAP1, Are Involved in a Moderate Low-Temperature Signaling Pathway in Rice. *Plant Physiol.* 129, 1880–1891. doi: 10.1104/pp.006072

Weng, J.-K., Ye, M., Li, B., and Noel, J. P. (2016). Co-evolution of Hormone Metabolism and Signaling Networks Expands Plant Adaptive Plasticity. *Cell* 166, 881–893. doi: 10.1016/j.cell.2016.06.027

Wijesooriya, K., Jadaan, S. A., Perera, K. L., Kaur, T., and Ziemann, M. (2022). Urgent need for consistent standards in functional enrichment analysis. *PLOS Comput. Biol.* 18, e1009935. doi: 10.1371/journal.pcbi.1009935

Wimalagunasekara, S. S., Weeraman, J. W. J. K., Tirimanne, S., and Fernando, P. C. (2023). Protein-protein interaction (PPI) network analysis reveals important hub proteins and sub-network modules for root development in rice (Oryza sativa). *J. Genet. Eng. Biotechnol.* 21, 69. doi: 10.1186/s43141-023-00515-8

Winkel, A., Colmer, T. D., Ismail, A. M., and Pedersen, O. (2013). Internal aeration of paddy field rice (Oryza sativa) during complete submergence – importance of light and floodwater. *New Phytol.* 197, 1193–1203. doi: 10.1111/nph.12048

Witt, S., Galicia, L., Lisec, J., Cairns, J., Tiessen, A., Araus, J. L., et al. (2012). Metabolic and Phenotypic Responses of Greenhouse-Grown Maize Hybrids to Experimentally Controlled Drought Stress. *Mol. Plant* 5, 401–417. doi: 10.1093/mp/ssr102

Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., et al. (2012). SABIO-RK--database for biochemical reaction kinetics. *Nucleic Acids Res.* 40, D790–D796. doi: 10.1093/nar/gkr1046

Wu, D., Shen, H., Yokawa, K., and Baluška, F. (2014). Alleviation of aluminium-induced cell rigidity by overexpression of OsPIN2 in rice roots. *J. Exp. Bot.* 65, 5305–5315. doi: 10.1093/jxb/eru292

Wu, L., Kobayashi, Y., Wasaki, J., and Koyama, H. (2018). Organic acid excretion from roots: a plant mechanism for enhancing phosphorus acquisition, enhancing aluminum tolerance, and recruiting beneficial rhizobacteria. *Soil Sci. Plant Nutr.* 64, 697–704. doi: 10.1080/00380768.2018.1537093

Wu, S., Kyaw, H., Tong, Z., Yang, Y., Wang, Z., Zhang, L., et al. (2024). A simple and efficient CRISPR/Cas9 system permits ultra-multiplex genome editing in plants. *Crop J.* 12, 569–582. doi: 10.1016/j.cj.2024.01.010

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2, 100141. doi: 10.1016/j.xinn.2021.100141

Xie, L., Li, H., Zhong, Z., Guo, J., Hu, G., Gao, Y., et al. (2022). Metabolome Analysis under Aluminum Toxicity between Aluminum-Tolerant and -Sensitive Rice (Oryza sativa L.). *Plants* 11, 1717. doi: 10.3390/plants11131717

Xu, J., Wang, X., and Guo, W. (2015). The cytochrome P450 superfamily: Key players in plant development and defense. *J. Integr. Agric.* 14, 1673–1686. doi: 10.1016/S2095-3119(14)60980-1

Xu, Y., and Fu, X. (2022). Reprogramming of Plant Central Metabolism in Response to Abiotic Stresses: A Metabolomics View. *Int. J. Mol. Sci.* 23, 5716. doi: 10.3390/ijms23105716

Yadav, P., Li, G., and Mulet, J. M. (2024). Editorial: Transcriptome & metabolic profiling: an insight into the abiotic stress response crosstalk in plants. *Front. Plant Sci.* 15, 1370817. doi: 10.3389/fpls.2024.1370817

Yan, L., Riaz, M., Liu, J., Yu, M., and Cuncang, J. (2022). The aluminum tolerance and detoxification mechanisms in plants; recent advances and prospects. *Crit. Rev. Environ. Sci. Technol.* 52, 1491–1527. doi: 10.1080/10643389.2020.1859306

Yang, L., Tian, D., Todd, C. D., Luo, Y., and Hu, X. (2013). Comparative Proteome Analyses Reveal that Nitric Oxide Is an Important Signal Molecule in the Response of Rice to Aluminum Toxicity. *J. Proteome Res.* 12, 1316–1330. doi: 10.1021/pr300971n

Yang, P., Ormerod, J. T., Liu, W., Ma, C., Zomaya, A. Y., and Yang, J. Y. H. (2019). AdaSampling for Positive-Unlabeled and Label Noise Learning With Bioinformatics Applications. *IEEE Trans. Cybern.* 49, 1932–1943. doi: 10.1109/TCYB.2018.2816984

Yang, Z., Liu, G., Liu, J., Zhang, B., Meng, W., Müller, B., et al. (2017). Synergistic action of auxin and cytokinin mediates aluminum-induced root growth inhibition in *Arabidopsis*. *EMBO Rep.* 18, 1213–1230. doi: 10.15252/embr.201643806

Yeo, I., and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika* 87, 954–959. doi: 10.1093/biomet/87.4.954

Yin, C.-C., Zhao, H., Ma, B., Chen, S.-Y., and Zhang, J.-S. (2017). Diverse Roles of Ethylene in Regulating Agronomic Traits in Rice. *Front. Plant Sci.* 8, 1676. doi: 10.3389/fpls.2017.01676

Ying, X. (2019). An Overview of Overfitting and its Solutions. *J. Phys. Conf. Ser.* 1168, 022022. doi: 10.1088/1742-6596/1168/2/022022

Yokosho, K., Yamaji, N., and Ma, J. F. (2011). An Al-inducible MATE gene is involved in external detoxification of Al in rice. *Plant J.* 68, 1061–1069. doi: 10.1111/j.1365-313X.2011.04757.x

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 26, 976–978. doi: 10.1093/bioinformatics/btq064

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* 16, 284–287. doi: 10.1089/omi.2011.0118

Yu, H., Kim, P. M., Sprecher, E., Trifonov, V., and Gerstein, M. (2007). The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. *PLOS Comput. Biol.* 3, e59. doi: 10.1371/journal.pcbi.0030059

Yu, J. (2002). A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. indica). *Science* 296, 79–92. doi: 10.1126/science.1068037

Yuan, Z., Wang, J., Qu, Q., Zhu, Z., Xu, M., Zhao, M., et al. (2023). Celastrol Combats Methicillin-Resistant *Staphylococcus aureus* by Targeting Δ ^{1} -Pyrroline-5-Carboxylate Dehydrogenase. *Adv. Sci.* 10, 2302459. doi: 10.1002/advs.202302459

Zhang, H., Zhu, J., Gong, Z., and Zhu, J.-K. (2021). Abiotic stress responses in plants. *Nat. Rev. Genet.* 23, 104–119. doi: 10.1038/s41576-021-00413-0

Zhang, P., Ding, Z., Zhong, Z., and Tong, H. (2019a). Transcriptomic Analysis for Indica and Japonica Rice Varieties under Aluminum Toxicity. *Int. J. Mol. Sci.* 20, 997. doi: 10.3390/ijms20040997

Zhang, P., Zhong, K., Tong, H., Shahid, M. Q., and Li, J. (2016). Association Mapping for Aluminum Tolerance in a Core Collection of Rice Landraces. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01415

Zhang, P., Zhong, K., Zhong, Z., and Tong, H. (2019b). Mining candidate gene for rice aluminum tolerance through genome wide association study and transcriptomic analysis. *BMC Plant Biol.* 19, 490. doi: 10.1186/s12870-019-2036-z

Zhang, Y., Lv, Y., Jahan, N., Chen, G., Ren, D., and Guo, L. (2018). Sensing of abiotic stress and ionic stress responses in plants. *Int. J. Mol. Sci.* 19, 1–16. doi: 10.3390/ijms19113298

Zhang, Y., Xu, J., Li, R., Ge, Y., Li, Y., and Li, R. (2023). Plants' Response to Abiotic Stress: Mechanisms and Strategies. *Int. J. Mol. Sci.* 24, 10915. doi: 10.3390/ijms241310915

Zhao, C., and Wang, Z. (2018). GOGO: An improved algorithm to measure the semantic similarity between gene ontology terms. *Sci. Rep.* 8, 15107. doi: 10.1038/s41598-018-33219-y

Zhao, H., Kosma, D. K., and Lü, S. (2021). Functional Role of Long-Chain Acyl-CoA Synthetases in Plant Development and Stress Responses. *Front. Plant Sci.* 12, 640996. doi: 10.3389/fpls.2021.640996

Zhao, M., Song, J., Wu, A., Hu, T., and Li, J. (2018). Mining Beneficial Genes for Aluminum Tolerance Within a Core Collection of Rice Landraces Through Genome-Wide Association Mapping With High Density SNPs From Specific-Locus Amplified Fragment Sequencing. *Front. Plant Sci.* 9, 1838. doi: 10.3389/fpls.2018.01838

Zhao, X. Q., Guo, S. W., Shinmachi, F., Sunairi, M., Noguchi, A., Hasegawa, I., et al. (2013). Aluminium tolerance in rice is antagonistic with nitrate preference and synergistic with ammonium preference. *Ann. Bot.* 111, 69–77. doi: 10.1093/aob/mcs234

Zhou, Y., Ma, B., Tao, J.-J., Yin, C.-C., Hu, Y., Huang, Y.-H., et al. (2022). Rice EIL1 interacts with OsIAAs to regulate auxin biosynthesis mediated by the tryptophan aminotransferase MHZ10/OsTAR2 during root ethylene responses. *Plant Cell* 34, 4366–4387. doi: 10.1093/plcell/koac250

Zhou, Y., Xiong, Q., Yin, C., Ma, B., Chen, S., and Zhang, J. (2020). Ethylene Biosynthesis, Signaling, and Crosstalk with Other Hormones in Rice. *Small Methods* 4, 1900278. doi: 10.1002/smtd.201900278

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10, 1523. doi: 10.1038/s41467-019-09234-6

Zhou, Z.-H. (2021). “Ensemble Learning,” in *Machine Learning*, ed. Z.-H. Zhou (Singapore: Springer Singapore), 181–210. doi: 10.1007/978-981-15-1967-3_8

Zhu, C. Q., Cao, X. C., Bai, Z. G., Zhu, L. F., Hu, W. J., Hu, A. Y., et al. (2019). Putrescine alleviates aluminum toxicity in rice (Oryza sativa) by reducing cell wall Al contents in an ethylene-dependent manner. *Physiol. Plant.* 167, 471–487. doi: 10.1111/ppl.12961

Zhu, P., Gu, H., Jiao, Y., Huang, D., and Chen, M. (2011). Computational Identification of Protein-Protein Interactions in Rice Based on the Predicted Rice Interactome Network. *Genomics Proteomics Bioinformatics* 9, 128–137. doi: 10.1016/S1672-0229(11)60016-8

Zhu, X. F., and Shen, R. F. (2023). Towards sustainable use of acidic soils: Deciphering aluminum-resistant mechanisms in plants. *Fundam. Res.*, S2667325823000857. doi: 10.1016/j.fmre.2023.03.004

Zhu, X., Yin, J., Guo, H., Wang, Y., and Ma, B. (2023). Vesicle trafficking in rice: too little is known. *Front. Plant Sci.* 14, 1263966. doi: 10.3389/fpls.2023.1263966

Zur, H., Ruppin, E., and Shlomi, T. (2010). iMAT: an integrative metabolic analysis tool. *Bioinformatics* 26, 3140–3142. doi: 10.1093/bioinformatics/btq602

List of publications

Review articles:

- Gallo-Franco, J.J., **Sosa, C.C**, Ghneim-Herrera, T. & Quimbaya, M.A. (2020) Epigenetic control of plant response to heavy metal stress: A new window to explain Aluminum tolerance. Frontiers in plant science. 11: 602625. https://doi.org/10.3389/fpls.2020.602625

Research articles:

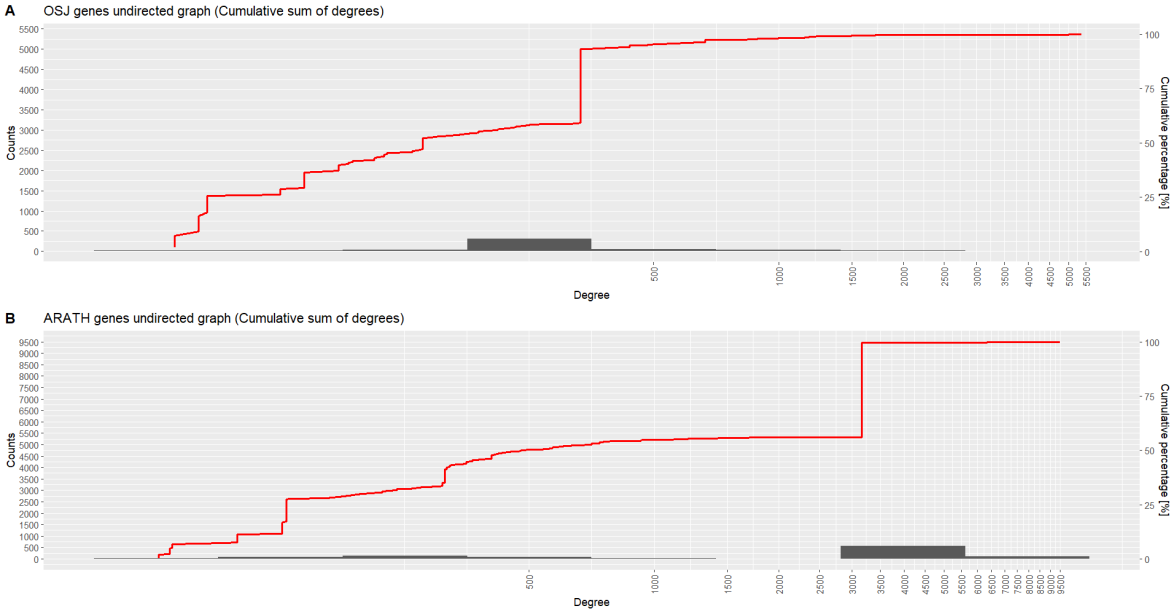
- **Sosa, Chrystian C.,** Diana Carolina Clavijo-Buriticá, Victor Hugo García-Merchán, Nicolas López-Rozo, Camila Riccio-Rengifo, Maria Victoria Diaz, David Arango Londoño, y Mauricio Alberto Quimbaya. «GOCompare: An R package to compare functional enrichment analysis between two species». *Genomics* 115, n.º 1 (2023): 110528. https://doi.org/10.1016/j.ygeno.2022.110528.
- Clavijo-Buriticá, Diana Carolina, **Chrystian C. Sosa**, Rafael Cárdenas Heredia, Arlen James Mosquera, Andrés Álvarez, Jan Medina, y Mauricio Quimbaya. «Use of Arabidopsis thaliana as a model to understand specific carcinogenic events: Comparison of the molecular machinery associated with cancer-hallmarks in plants and humans». *Heliyon*. https://doi.org/10.1016/j.heliyon.2023.e15367.
- Gallo-Franco, J.J., I. Zuluaga-Yusti, A.M. Restrepo-García, **C.C. Sosa**, S. Zapata-Balanta, J.P. Gutiérrez-Marín, T. Ghneim-Herrera, y M. Quimbaya. «Transcriptional Analysis in Four Wild and Cultivated Rice Genotypes Identifies Aluminum-Induced Genes». *Plant Stress* 10 (diciembre de 2023): 100247. https://doi.org/10.1016/j.stress.2023.100247.

Abstracts presented at scientific conferences:

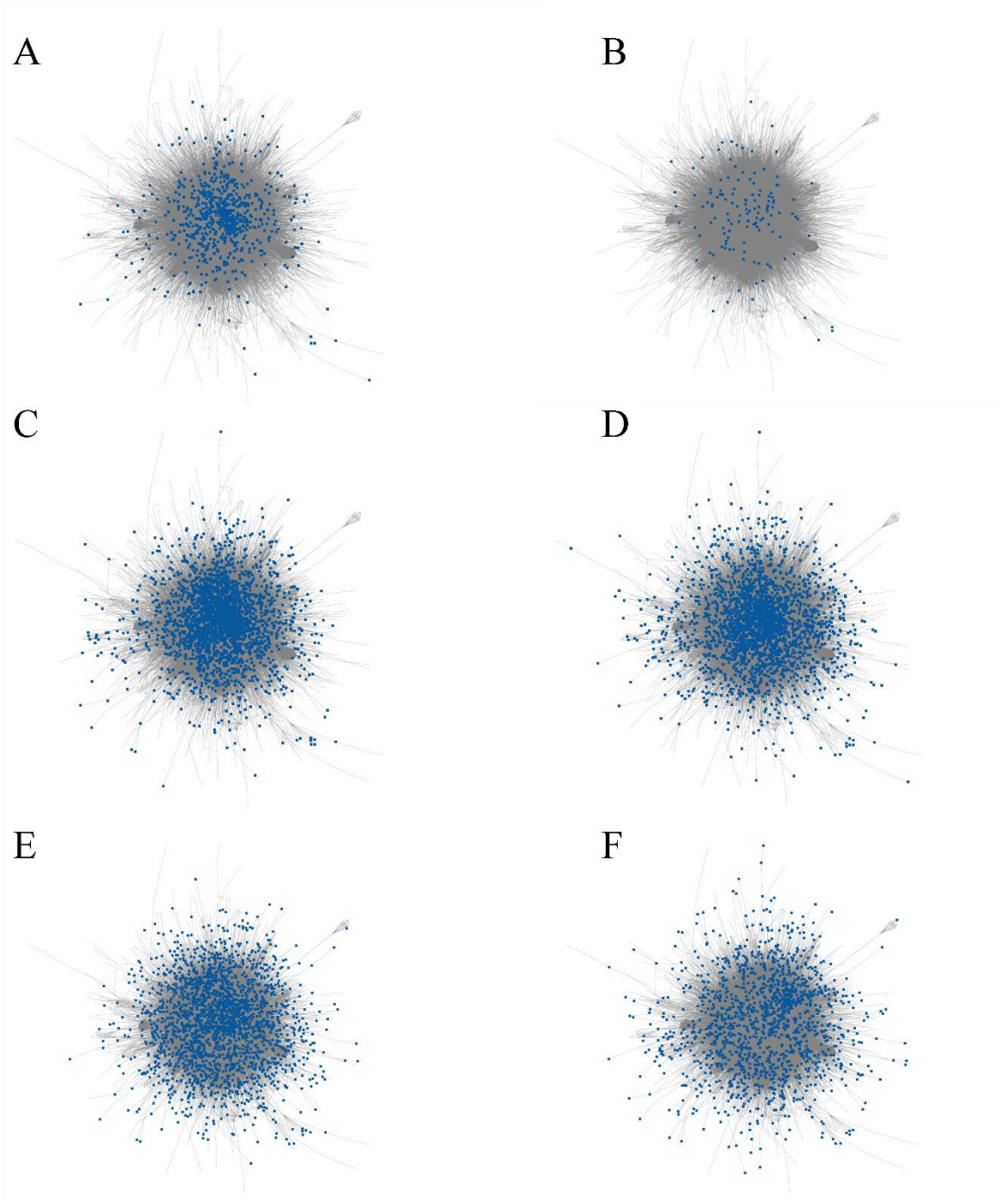
- Machine learning-guided PPI network analysis for aluminum-tolerant proteins prediction (2024). **Sosa C**; Clavijo-Buriticá D; García Merchán V; Quimbaya M. Tercera Conferencia Colombiana de Matemáticas Industriales MAPI 3. Bucaramanga, Santander, Colombia. https://scm.org.co/wp-content/uploads/2024/06/Libro-de-Resumenes_Mapi3_Version06.pdf
- *GOCOmpare: an R package to compare functional enrichment analysis results among gene lists and two species* (2022). **Sosa C**; Clavijo-Buriticá D; López-Rozo N; Riccio-Rengifo C, Diaz M, García Merchán V; Quimbaya M. X Bioinformatics and Genomics Symposium (BGS-X) Valencia, España. https://scb.iec.cat/wp-content/uploads/2022/12/BookOfAbstracts.pdf
- *GOCOmpare: an R package to compare functional enrichment analysis results among gene lists and two species* (2022). **Sosa C**; Clavijo-Buriticá D; López-Rozo N; Riccio-Rengifo C, Diaz M, García Merchán V; Quimbaya M. Sexto congreso colombiano de biología computacional y bioinformática. https://www.omicas.co/sites/default/files/2022-07/CCBCOLVI_2022_04_01_CCS.pdf
- *Cancer hallmarks orthologues in Arabidopsis thaliana suggests partial overlap and potentiality as a research model for the disease* (2020). **Sosa C**.; Quimbaya M.; Cárdenas-Heredia R. *Applied Bioinformatics in Life Sciences (3rd edition). February 12 - 14, 2020. Leuven, Belgium* https://www.omicas.co/sites/default/files/2022-07/VIB_2019_02_04.pdf
- Uso de *Arabidopsis thaliana* como modelo de estudio para entender y encontrar potenciales nuevos genes asociados a procesos carcinogénicos. (2022). **Sosa C**; Clavijo-Buriticá D; Mosquera A; Álvarez A; Medina J y Quimbaya M. Tercer Simposio de Bioinformática RSG-Colombia. https://www.omicas.co/sites/default/files/2022-07/P1_EVENTO%20CIENT_PARTICIPACION_Chrystian%20Sosa%20%281%29.pdf

Supplementary Figures

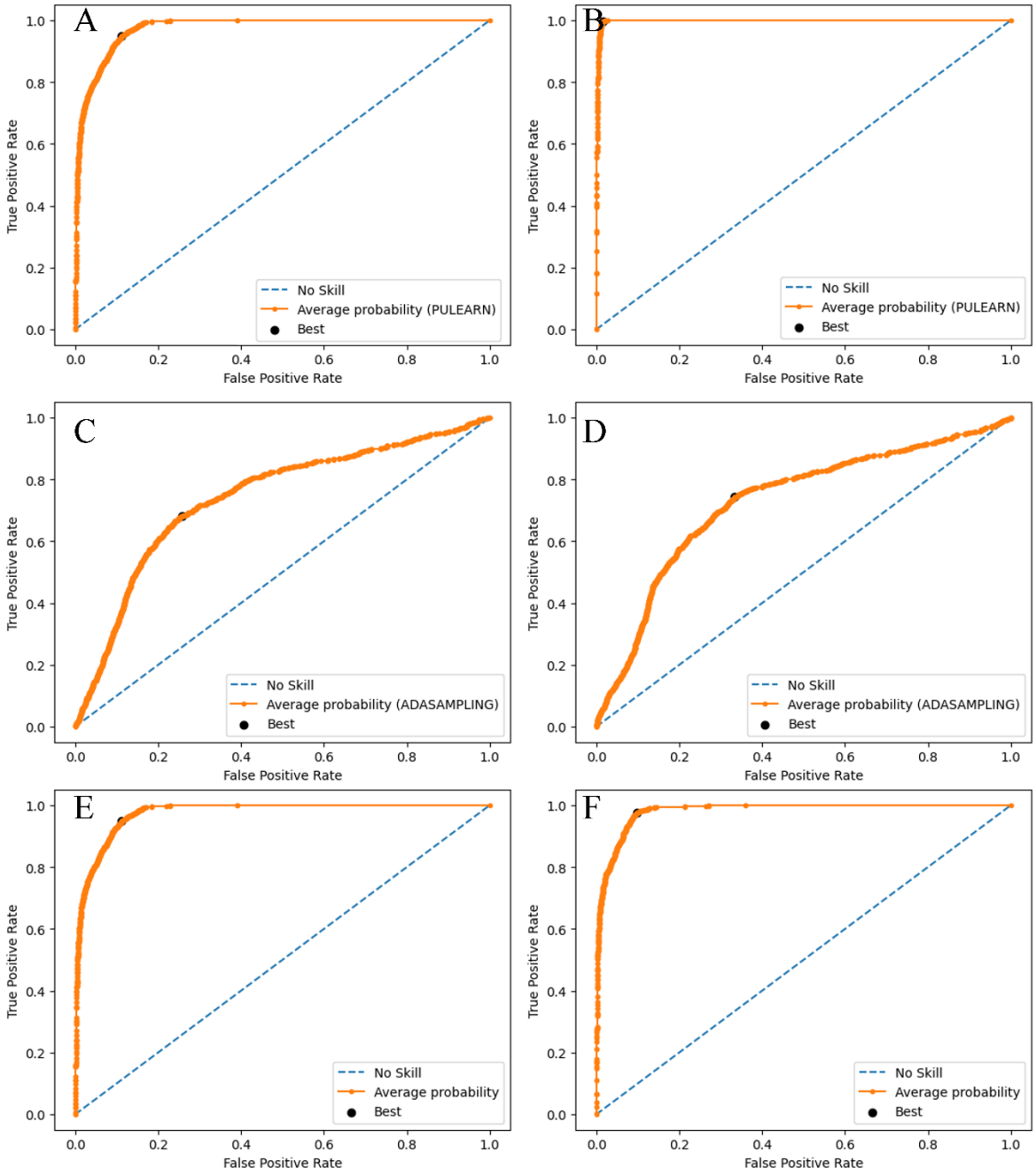
Chapter 3



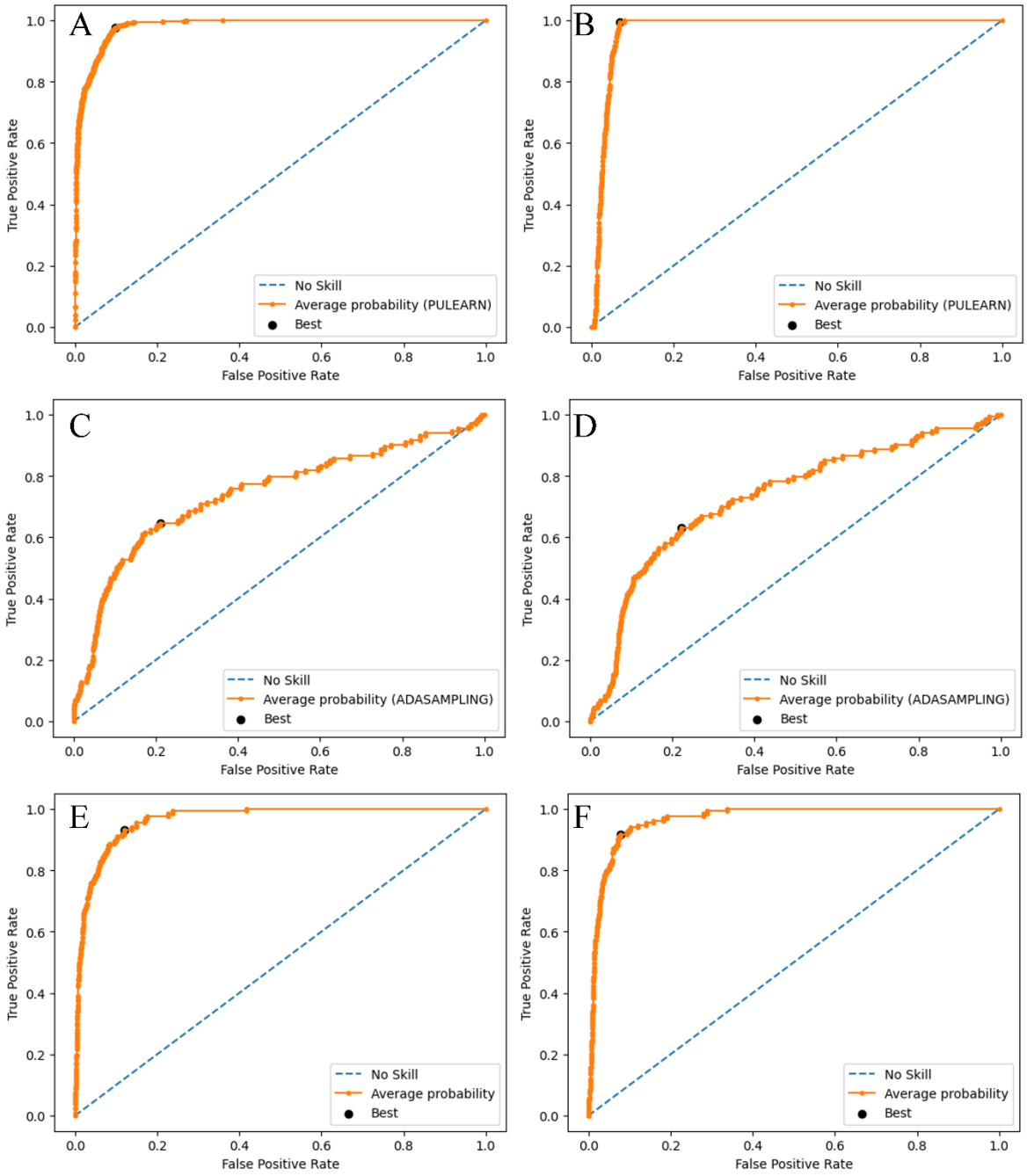
Supplementary Figure 3.1. Histograms of the cumulative sum of degrees for the gene networks. Red line represents the degree cumulative percentage (%). A) *Oryza sativa* subspecies *japonica* B...) *Arabidopsis thaliana*



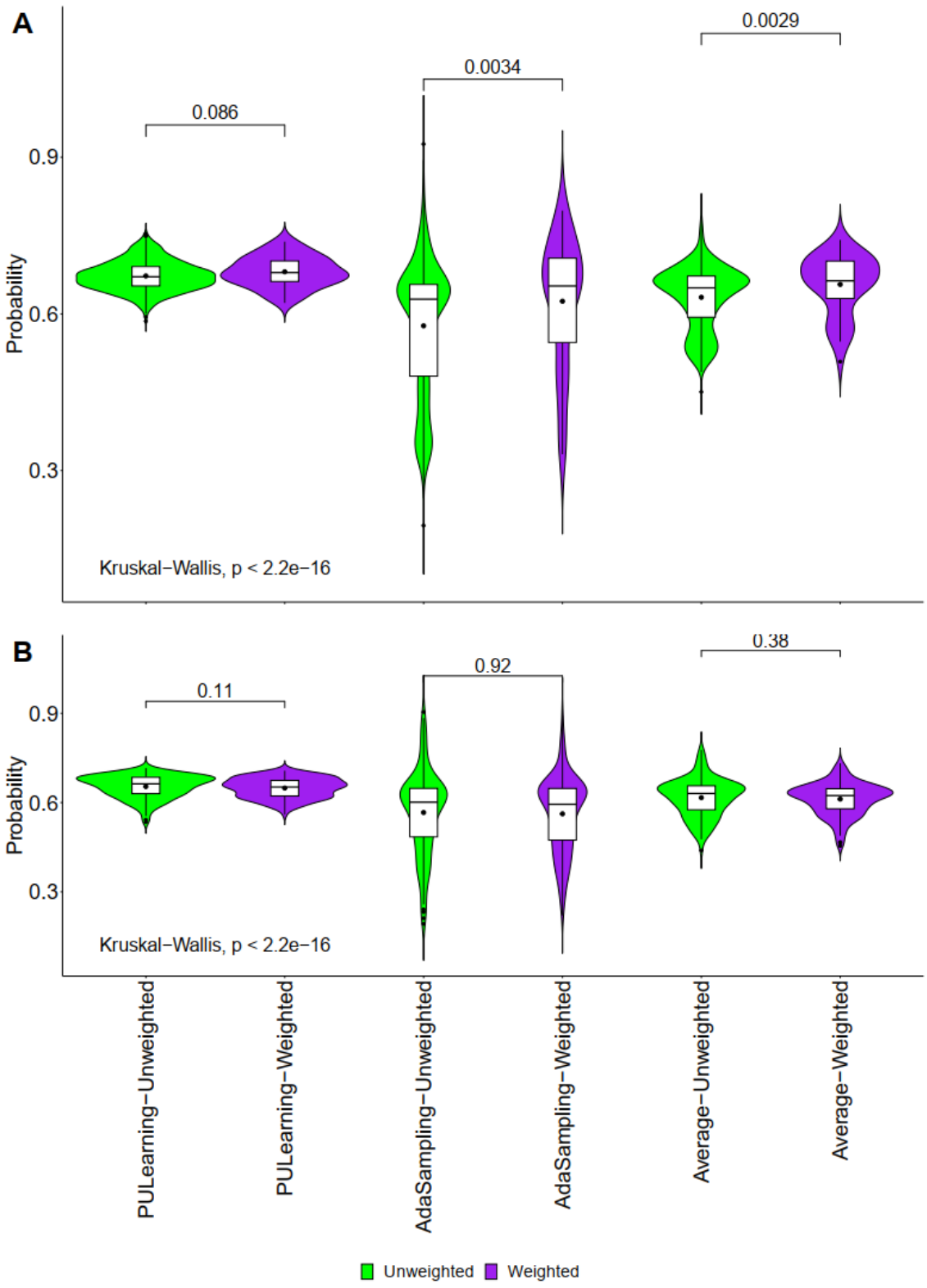
Supplementary Figure 4.1. RicePPInets interactome with positive class datasets and predicted proteins. Blue indicates positive class or predicted datasets and gray indicates unlabeled proteins. A.) Dataset 1 positive class dataset, B.) Dataset 2 positive class dataset C.) Unweighted graph embedding predictions for dataset 1 D.) Unweighted graph embedding predictions for dataset 2 E.) Graph embedding integrating transcriptomics data predictions for dataset 1 F.) Graph embedding integrating transcriptomics data predictions for dataset 2.



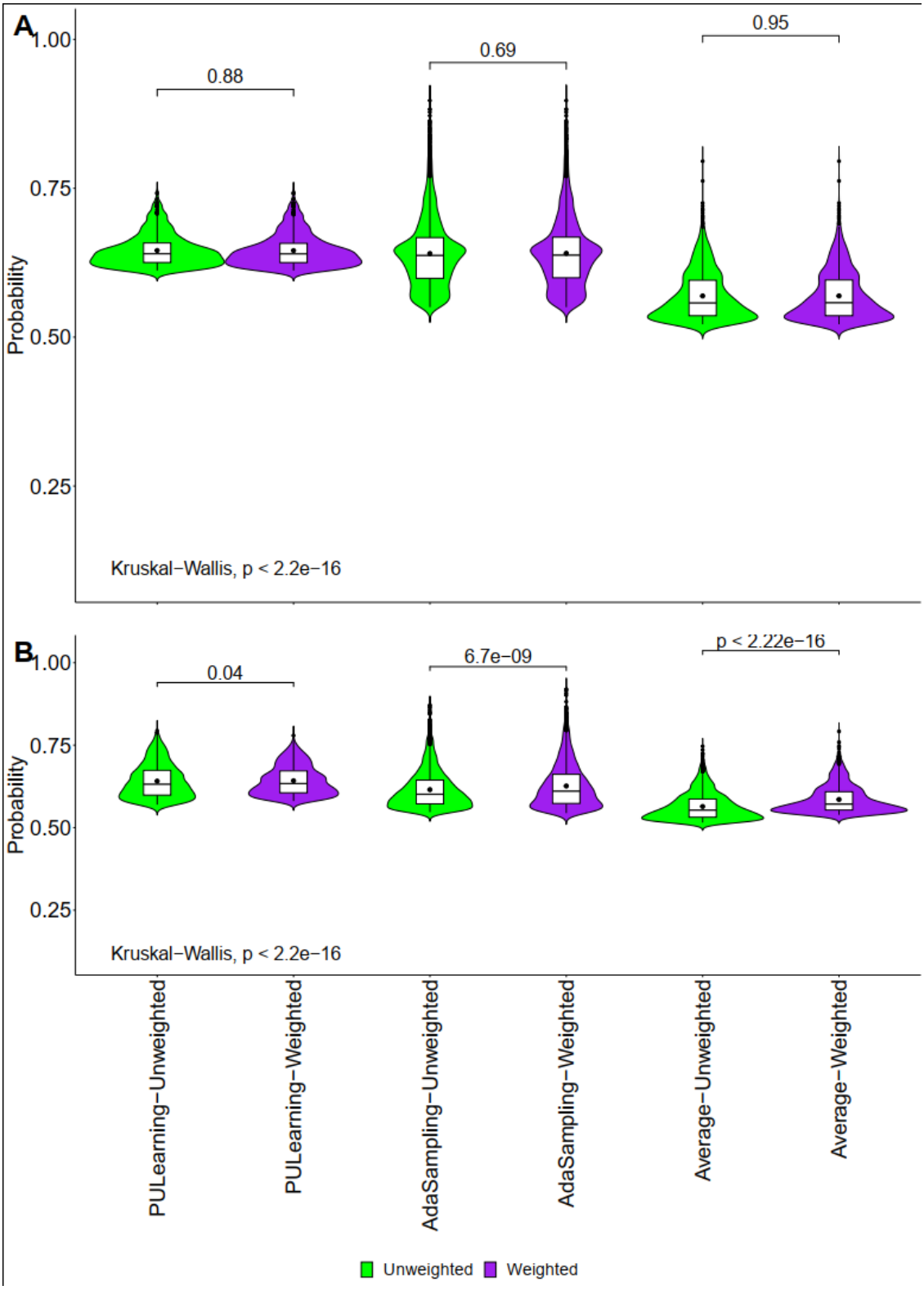
Supplementary Figure 4.2. ROC curves for single averaging ensembles for both positive class datasets using graph embedding without transcriptomics integration. The black point represents the probability threshold used. A.) PULearning ensemble for dataset 1 B.) PULearning ensemble for dataset 2 C.) AdaSampling ensemble for dataset 1 D.) AdaSampling ensemble for dataset 2 E.) Average of PULearning and AdaSampling ensembles for dataset 1 F.)Average of PULearning and AdaSampling ensembles for dataset 2.



Supplementary Figure 4.3. ROC curves for single averaging ensembles for both positive class datasets using graph embedding and transcriptomics data integration. The black point represent the probability threshold used. A.) PULearning ensemble for dataset 1 B.) PULearning ensemble for dataset 2 C.) AdaSampling ensemble for dataset 1 D.) AdaSampling ensemble for dataset 2 E.) Average of PULearning and AdaSampling ensembles for dataset 1 F.)Average of PULearning and AdaSampling ensembles for dataset 2.

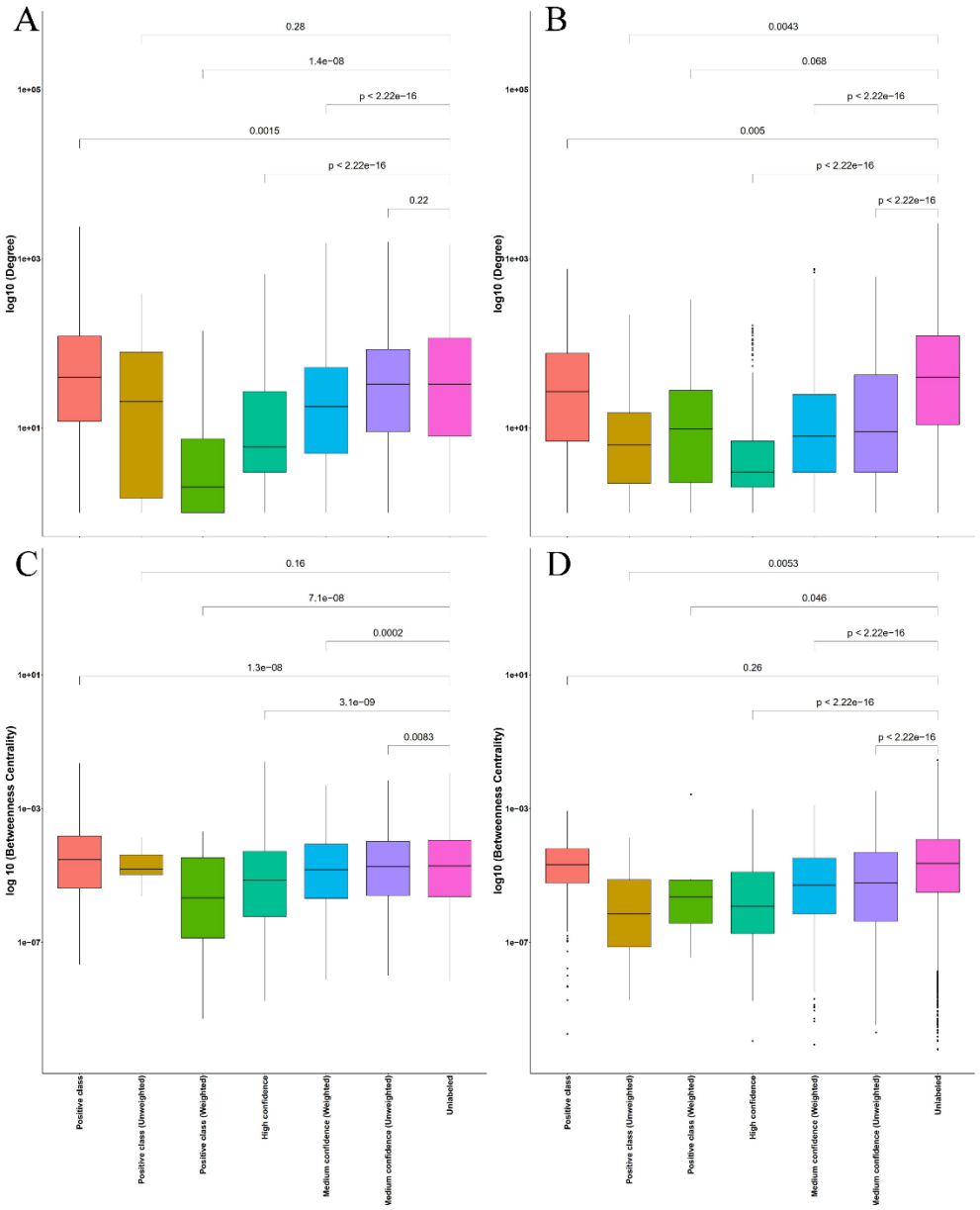


Supplementary Figure 4.4. Impact of transcriptomics integration on the prediction of the positive class used for training the models, Unweighted interactome (green), Weighted interactome using transcriptomics data (violet) A.) Dataset 1 B.) Dataset 2.



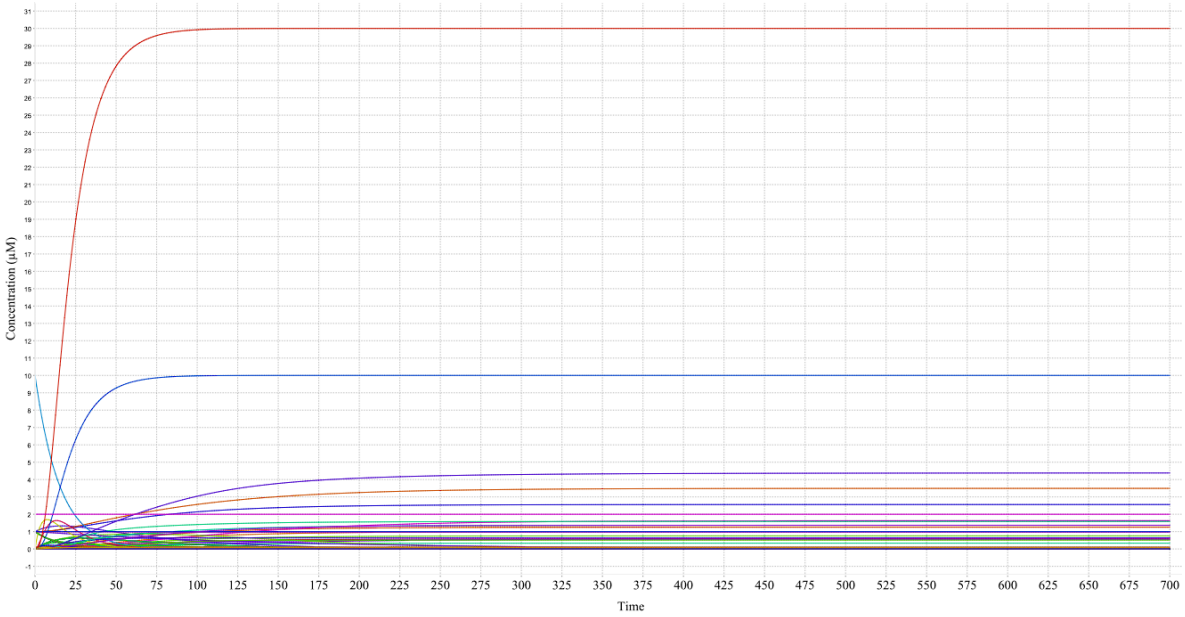
Supplementary Figure 4.5. Impact of transcriptomics integration on the prediction of new candidate proteins, Unweighted interactome (green), Weighted interactome using transcriptomics data (violet) A.) Dataset 1 B.) Dataset 2.

Supplementary Figure 4.6. Heatmap of metabolic pathways frequency per predicted dataset, in the x-axis are shown the predicted datasets categories, in the y-axis are displayed the metabolic pathways, in color red and blue are the predicted dataset categories for the dataset 1 and 2 respectively. This figure is available at https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER4/SUP_FIG6.pdf

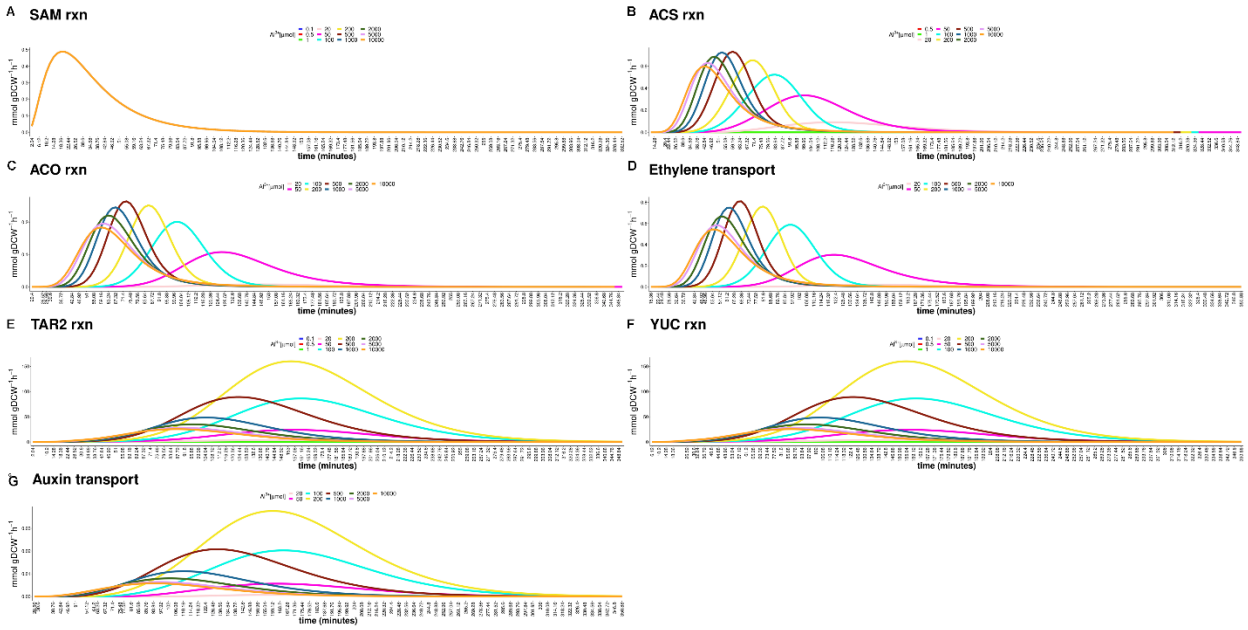


Supplementary Figure 4.7. Topological measurements of predicted proteins and positive class categories using the average of seven models. (Values in brakes represent *post hoc* p-values A) Degree of predicted proteins for dataset 1 B) Betweenness centrality of predicted proteins for dataset 1 C) Degree of predicted proteins for dataset 2 D) Betweenness centrality of predicted proteins for dataset 2.

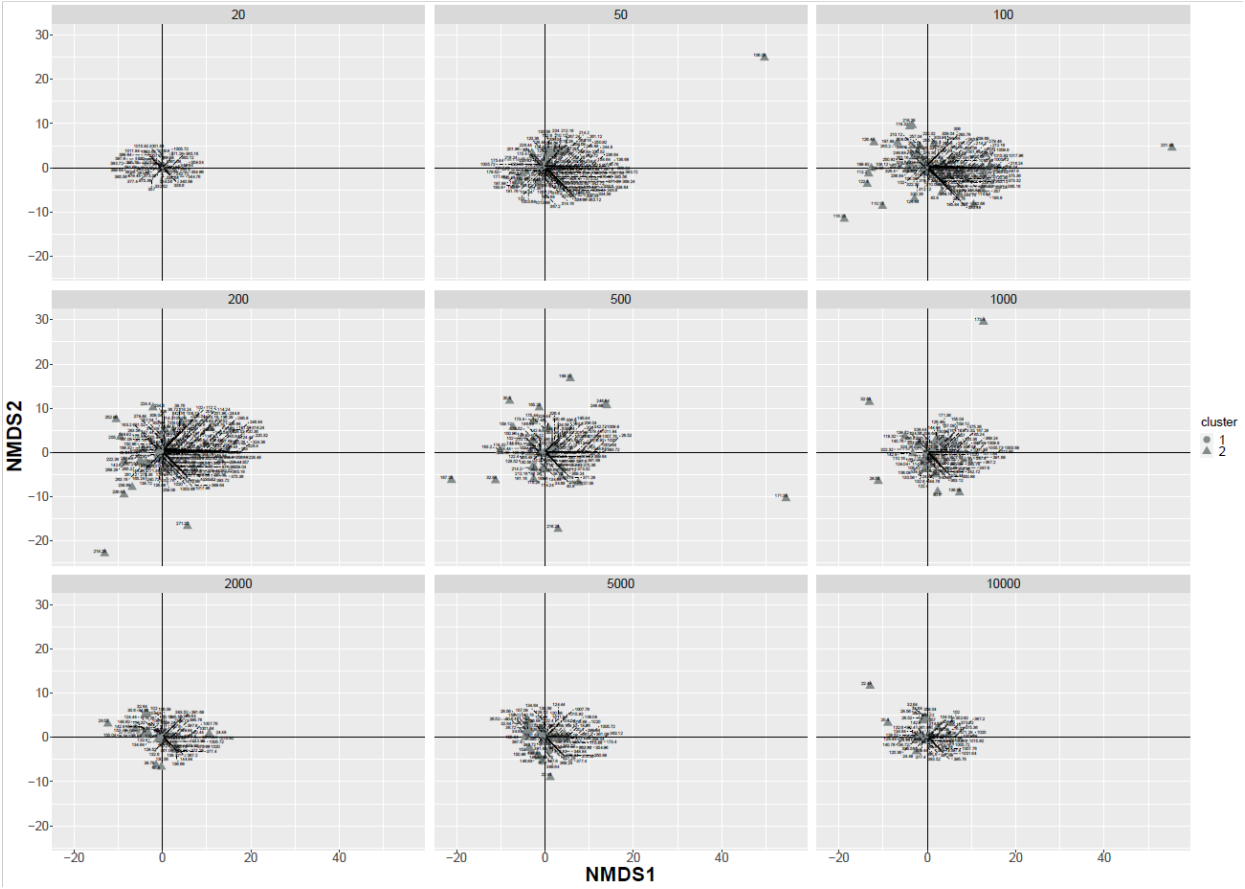
Chapter 5



Supplementary Figure 5.1. Simulation of the gene network regulatory network for ethylene and auxin production under aluminum stress. Steady State was achieved approximately in the interval 700.



Supplementary Figure 5.2. Flux distributions calculated from the gene regulatory network deterministic model for five reactions shared with the metabolic model. A to C represent ethylene biosynthesis pathway reactions mediated by methionine adenosyltransferase (SAMS), 1-aminocyclopropane-1-carboxylate synthase (ACS), and aminocyclopropanecarboxylate oxidase (ACO) respectively. D and E represent indole-3-acetate biosynthesis I reactions: L-tryptophan—pyruvate aminotransferase (TAR2) and indole-3-pyruvate monooxygenase (YUC) respectively. This figure is available at https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER5/Supplementary_figure2.pdf



Supplementary Figure 5.3. Non-metric multi-dimensional scaling of reaction flux values exhibiting simulation time intervals as individuals, grids show nine aluminum concentration scenarios from 20 μM to 10000 μM (Table 5.1) where biomass production was observed. In addition, shapes represent k-means clusters. Figures for each aluminum simulation scenario (Table 5.1) are available at https://github.com/ccsosa/Rice-aluminum-stress-GENRE-and-Deterministic-model/tree/main/Supplementary_information/Suppl_fig3_zoom_in.

- **Supplementary Figures 5.4 to 5.7**: Heatmaps for metabolic subsystem frequency (columns) activated for each time interval (rows) simulated under the multistage FBA approach for four aluminum concentration scenarios (Table 5.1). Clusters one and two are shown in blue and red respectively. **Supplementary Figure 5.4**: 50 μM (ID Al5, Table 5.1), **Supplementary Figure 5.5**: 200 μM (ID Al7, Table 5.1), **Supplementary Figure 5.6**: 500 μM (ID Al8, Table 5.1), **Supplementary Figure 5.7**: 10000 μM (ID Al12, Table 5.1). These figures are available at https://github.com/ccsosa/Supplementary-information/tree/main/CHAPTER5
- **Supplementary Figures 5.8 to 5.11**: Heatmaps for reaction flux values (rows) activated for each integration step(columns) simulated under the multistage FBA approach for four aluminum concentration scenarios. Clusters one and two are shown in blue and red respectively. The black color represents reactions available at the upper quartile of the reaction flux values standard deviation. **Supplementary Figure 5.8**: 50 μM (ID Al5, Table 5.1), **Supplementary Figure 5.9**: 200 μM (ID Al7, Table 5.1), **Supplementary Figure 5.10**: 500 μM (ID Al8, Table 5.1), **Supplementary Figure 5.11**: 10000 μM (ID Al12, Table

5.1). These figures are available at https://github.com/ccsosa/Supplementary-information/tree/main/CHAPTER5

Supplementary Tables

Chapter 2

Supplementary Table 2.1. Comparative table of some methods based on optimization used in Genome Scale Metabolic Reconstructions (GENREs).

Method	Inputs	Initial metabolite concentration information	Assumption	Purpose	Solver	ODE	Source
Flux Balance Analysis (FBA)	Genomic annotation, chemical reactions associated with genome annotation and literature, an objective function	No	Pseudo steady state: production and consumption of metabolites inside the cell are balanced	Calculate the flow of metabolites through a metabolic network	Lineal Programming (LP)	No	(Orth et al., 2010)
Dynamic Flux Balance Analysis (DFBA)	Genomic annotation, chemical reactions associated with genome annotation and literature, an objective function	Yes	Incorporates rate of change of flux constraints	Dynamic predictions of substrate, biomass, and product concentrations for growth in batch or fed-batch cultures	Non linear programming (NLP)	Yes	(Mahadevan et al., 2002)
Metabolic flux analysis (MFA)	Substrate uptake rate, oxygen uptake rate, growth rate and product secretion rates	Yes	Simplified metabolic network models must be used for this analysis	Metabolic fluxes. Similar to FBA	Least-squares regression	No	(Antoniewicz, 2021)

Method	Inputs	Initial metabolite concentration information	Assumption	Purpose	Solver	ODE	Source
Metabolic Adjustment Dynamic Flux Balance Analysis (MDFBA)	Genomic annotation, chemical reactions associated with genome annotation and literature, and objective function	Yes	Differences of extracellular concentration and intracellular concentrations	Minimization of the fluctuation of concentration vector X to be the optimal objective	Non linear programming (NLP)	Yes	(Luo et al., 2006, 2009b)
Integrative Metabolic Analysis Tool (iMAT)	Gene and/or protein expression data, and a GENRE	No	Gene expression threshold based in three categories: Low, Medium and Highly expressed genes	Flux activity states over all network reactions maximizing the number of reactions whose activity is consistent with Shlomi et al., (2008) method	Mixed-Integer Linear Programming (MILP)	No	(Shlomi et al., 2008; Zur et al., 2010)

Chapter 3

Supplementary Tables 3.1 to 3.6 content is described below, and they are available as spreadsheets in an Excel file available at:
https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER3/SUPPLEMENTARY_TABLE.xlsx

- **Supplementary Table 3.1.** Results for the frequency of molecular functions GO terms per functional category obtained by the *mostFrequentGOs* function for *Oryza sativa***.**

- **Supplementary Table .2.** Results for the frequency of molecular functions GO terms per functional category obtained by the *mostFrequentGOs* function for *Arabidopsis thaliana***.**

- **Supplementary Table 3.3.** Calculated node weights for the undirected-weighted graph obtained for *Oryza sativa* and *A. thaliana* using the *graphGOspecies* function The underlined numbers represent molecular functions, GO terms and node weights greater or equal to the upper quartile per each species respectively.

- **Supplementary Table 3.4.** Node weights calculated for the undirected-weighted graph obtained for *Oryza sativa* subsp**.** *japonica* and *A. thaliana* using the *graph_two_GOspecies* function The underlined numbers represent the node weights greater or equal to the upper quartile for molecular functions.

- **Supplementary Table 3.5.** Node description for the network done with *Oryza sativa* subsp. *japonica* genes as nodes and molecular functions as edges to identify potential main genes with high degree values which could be associated with aluminum tolerance. A Candidate gene list is provided in the column Candidate genes.

- **Supplementary Table 3.6.** Node description for the network done with *Arabidopsis thaliana* genes as nodes and molecular functions as edges to identify potential main genes with high degree values which could be associated with aluminum tolerance. A Candidate gene list is provided in the column Candidate genes.

Supplementary Table 3.7. General recommendations to perform analyses in GOCompare R package and functional enrichment analysis. A set of general recommendations is provided for GOCompare users in order to obtain reproducible and accurate results.

Step	Recommendations
Define gene lists (categories)	Define your gene lists regarding your experiments or hypothesis
Genes have orthologues	Use the best reciprocal hits approach when it is possible (It can be helpful for no model species)
in species 2?	Use standardized data such as ENSEMBL database via BiomaRt or PLAZA database to obtain orthologues if the species analyzed are well studied
Assign orthologues to the same lists as the original lists as a species one	Once defined your categories, assign the orthologues to the same lists (e.g. if gene A in species 1 is in the list 1, 2, and 3 and it has the orthologue A1 in species 2, assign gene A1 to list 1,2,3 for species 2.)
Functional enrichment analysis	Prepare an accurate gene background for analyses (Avoid using the whole genome information when it is possible)
	Select a functional enrichment analysis according to your analysis needs: Take into consideration:
	1.) If the user wants to perform analysis for specific GO terms levels, It is encouraged the use of tools such as topGO. Nevertheless, a method to obtain p values should be justified
	2.) Tools such as BinGO, ShinyGO, or gprofiler can get good results for well-annotated species but they cannot be accurate for nonmodel species
	Use multiple test correction procedures (e.g. FDR, gSCS) if the tool used allows their use
	Only use GO terms with p values or fixed p values < 0.05 to reduce noise data in your datasets

Chapter 4

Supplementary Tables 4.1 to 4.3 content is described below and they are available as spreadsheets in an Excel file available at: https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER4/SUPPL_TABLE1_3.xlsx.

Supplementary Table 4.1. Evaluation metrics values and probability thresholds calculated for single averaging ensembles. Data available at available at: https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER4/SUPPL_TABLE1_3.xlsx.

Supplementary Table 4.2. Predicted protein for single averaging ensembles for dataset 1 and dataset 2 including metabolic pathways extracted from OryzaCyc. Data available at available at: https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER4/SUPPL_TABLE1_3.xlsx.

Supplementary Table 4.3. Enriched gene ontology molecular functions and biological processes for predicted proteins:https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER4/SUPPL_TABLE1_3.xlsx.

Supplementary Table 4.4. List of 37 predicted proteins obtained by weighted and unweighted graphs with their respective name and gene family obtained by PLAZA 5.0.

MSU gene id	RAP-DB gene id	Gene name	Orthologous gene family
LOC_Os02g15750	Os02g0257300	Os02t0257300-00, STR12, SULFURTRANSFERASE 12, LOC_Os02g15750.1, OsTROL, Str12, TROL, OsStr12, Os-TROL, Q6ETQ7	Rhodanese-like domain containing protein., Sulfurtransferase 12
LOC_Os03g63760	Os03g0854600	Os03t0854600-01, XTH24, XYLOGLUCAN ENDOTRANSGLUCOSYDASE/HYDROLASE 24, LOC_Os03g63760.1, OsXTH24, Q84T83	Concanavalin A-like lectin/glucanase domain containing protein., xyloglucan endotransglucosylase/hydrolase, xyloglucan endotransglucosylase 24
LOC_Os02g03550	Os02g0127800	Os02t0127800-01, XTH26, XYLOGLUCAN ENDOTRANSGLUCOSYDASE/HYDROLASE 26, LOC_Os02g03550.1, OsXTH26, Q6Z2L4	Concanavalin A-like lectin/glucanase domain containing protein., xyloglucan endotransglucosylase/hydrolase, xyloglucan endotransglucosylase 26
LOC_Os11g34360	Os11g0546100	Os11t0546100-01, LOC_Os11g34360.1, OsGPCR, Q2R2X1	Transmembrane receptor, eukaryota domain containing protein.
LOC_Os10g39190	Os10g0537100	Os10t0537100-00, LOC_Os10g39190.1, Q8LNN8	Hypothetical conserved gene.
LOC_Os03g51740	Os03g0727600	Os03t0727600-01, OsACS1, ACC1*, Os-ACS1, ACS1, ACC1, ACC SYNTHASE 1, LOC_Os03g51740.1, Q10DK7	ACC synthase, Ethylene biosynthesis, 1-aminocyclopropane-1-carboxylic acid synthase 1

MSU gene id	RAP-DB gene id	Gene name	Orthologous gene family
LOC_Os11g47510	Os11g0701000	Os11t0701000-01, LOC_Os11g47510.1, Q53NL1	Class III chitinase homologue (OsChib3H-c).
LOC_Os03g30250	Os03g0416350	Os03t0416350-00, LOC_Os03g30250.1, A0A0P0VZK8	Hypothetical protein.
LOC_Os06g39230	Os06g0592400	Os06t0592400-01, ALDH2B2, ALDEHYDE DEHYDROGENASE 2B2, LOC_Os06g39230.1, OsALDH2B2, Q0DB86	Similar to Cytosolic aldehyde dehydrogenase RF2C., Aldehyde dehydrogenase 2B2, aldehyde dehydrogenase
LOC_Os10g31780	Os10g0456100	Os10t0456100-00, LOC_Os10g31780.1, Q7XDQ8	Similar to Brn1-like protein.
LOC_Os08g43070	Os08g0543700	Os08t0543700-00, LOC_Os08g43070.1, OsbHLH011, Q6ZBI4	Hypothetical conserved gene., basic helix-loop-helix protein 011
LOC_Os09g32025	Os09g0493500	Os09t0493500-02, OsDFR2, OsTKPR1, LOC_Os09g32020.1, LOC_Os09g32025.1, DFR2A, OS-DFR2, OsDFR2A, TKPR1, DFR2, Q0J0M1	Similar to H0404F02.2 protein., Tetraketide aplha-pyrone reductase, Anther cuticle development, Pollen wall formation
LOC_Os07g39980	Os07g0588800	Os07t0588800-01, LOC_Os07g39980.1, Q84Z13	Similar to predicted protein., Protein of unknown function DUF2419 domain containing protein.
LOC_Os01g68330	Os01g0911300	Os01t0911300-01, ABCB24, ABC TRANSPORTER B FAMILY MEMBER 24, LOC_Os01g68330.1, OsABCB24, Q0JGQ4	Similar to Transporter associated with antigen processing-like protein., ABC transporter superfamily ABCB subgroup member 24
LOC_Os02g44320	Os02g0662100	Os02t0662100-01, HYPRP2, HYBRID PROLINE- OR GLYCINE-RICH PROTEIN 2, LOC_Os02g44320.1, OsHyPRP02, HyP/GRP02, OsHyPRP2, HyPRP2, HyP/GRP2, Q6H6L8	Similar to Tfm5 protein., hybrid proline- or glycine-rich protein 2, Hybrid Proline-Rich Protein 2
LOC_Os06g30780	Os06g0503800	Os06t0503800-00, LOC_Os06g30780.1, A0A0P0WXE7	Hypothetical conserved gene.
LOC_Os07g44590	Os07g0639400	Os07t0639400-01, LOC_Os07g44590.1, prx108, Q8GVF7	Similar to Peroxidase 1., class III peroxidase 108
LOC_Os01g36460	Os01t0545100-01, LOC_Os01g36460.1, OsMYB86-L1, Q5JKA5	HOM05M000012	

MSU gene id	RAP-DB gene id	Gene name	Orthologous gene family
LOC_Os03g54100	Os03g0752300	Os03t0752300-01, TPKA, TPKa, OsTPKa, TWO-PORE K+ CHANNEL A, LOC_Os03g54100.1, LOC_Os03g54100.2, OsTPK1a, TPK1a, OsTpka, Q850M0	Two-pore K+ channel family protein, K+ homeostasis
LOC_Os09g36040	Os09g0530000	Os09t0530000-02, STR4, SULFURTRANSFERASE 4, LOC_Os09g36040.1, OsStr4, Q69NF7	Rhodanese-like domain containing protein., Sulfurtransferase 4
LOC_Os04g46490	Os04g0550800	Os04t0550800-01, 1, TIP5, TONOPLAST INTRINSIC PROTEIN 5, LOC_Os04g46490.1, OsTIP5, Q7XU31	Major intrinsic protein family protein., Probable aquaporin TIP5-1, Tonoplast intrinsic protein 5-1
LOC_Os06g45370	Os06g0664200	Os06t0664200-01, LOC_Os06g45370.1, Q655Y9	Inositol monophosphatase/Fructose-1, 6-bisphosphatase domain containing protein., Hypothetical gene., Fru-1, 6-bisphosphatase
LOC_Os04g44230	Os04g0523500	Os04t0523500-00, CKX8, CYTOKININ OXIDASE/DEHYDROGENASE 8, LOC_Os04g44230.1, OsCKX8, A3AVP1	Hypothetical conserved gene., cytokinin oxidase 8
LOC_Os11g16260	Os11g0264000	Os11t0264000-00, LAC22, LACCASE 22, LOC_Os11g16260.1, OsLAC22, Q53LU4	Cupredoxin domain containing protein., laccase 22
LOC_Os03g50410	Os03g0712000	Os03t0712000-00, LOC_Os03g50410.1, Q10E06	Similar to Lipase family protein.
LOC_Os05g50250	Os05g0578500	Os05t0578500-01, CCR23, CINNAMOYL-COA REDUCTASE 23, LOC_Os05g50250.1, OsCCR23, Q0DFP6	NAD(P)-binding domain containing protein., Cinnamoyl-CoA reductase 23
LOC_Os07g07440	Os07g0170000	Os07t0170000-00, LOC_Os07g07440.1, Q69LD5	Similar to Brn1-like protein.
LOC_Osp1g00770			
LOC_Os10g13670	Os10g0204100	Os10t0204100-01, LOC_Os10g13670.1, Q0IYJ5	Zinc finger, RING/FYVE/PHD-type domain containing protein.
LOC_Os04g38520	Os04g0458200	Os04t0458200-01, IIP4, ILA1 INTERACTING PROTEIN 4, LOC_Os04g38520.1, LOC_Os04g38525.1, A0A5S6RDA4	Uncanonical CCCH-tandem zinc-finger protein, Repression of secondary wall synthesis, Control of mechanical strength, Similar to cDNA clone:J033037C08, full insert sequence.
LOC_Osp1g00780			

MSU gene id	RAP-DB gene id	Gene name	Orthologous gene family
LOC_Os01g14050	Os01g0242600	Os01t0242600-01, NTMC2T6.1, N-TERMINAL-TM-C2 DOMAIN PROTEIN 6.1, LOC_Os01g14050.2, LOC_Os01g14050.1, OsNTMC2T6.1, Q5NA77	C2 calcium-dependent membrane targeting domain containing protein., NTMC2 family gene 6.1, N-terminal-TM-C2 domain protein 6.1, N-terminal trans- membrane and C2 domain protein 6.1, NTMC2 type 6.1 protein
LOC_Os04g44570	Os04g0527900	Os04t0527900-01, 2, TIP3, TONOPLAST INTRINSIC PROTEIN 3, LOC_Os04g44570.1, OsTIP3;2, OsEnS-70, TIP3-2, Q7XKI5	Similar to Tonoplast membrane integral protein ZmTIP3-2., Probable aquaporin TIP3-2, Tonoplast intrinsic protein 3-2, endosperm-specific gene 70
LOC_Os04g06860			
LOC_Os09g25610	Os09g0424200	Os09t0424200-01, LOC_Os09g25610.1, Q69QJ1	Glutamine amidotransferase class-I domain containing protein.
LOC_Os03g61280	Os03g0828500	Os03t0828500-01, LOC_Os03g61280.2, LOC_Os03g61280.1, Q10B67	Similar to (1-4)-beta-mannan endohydrolase precursor (EC 3.2.1.78).
LOC_Os07g44350	Os07g0637400	Os07t0637400-01, OsNPSN13, NPSN13, NOVEL PLANT SNARE 13, LOC_Os07g44350.1, OryzaNPSN13, Q8H5R6	Qb-SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) protein, Abiotic stress response, SNARE 13

Chapter 5

Supplementary Table 5.1. Gene regulatory deterministic model statistics including species, reactions, and their initial concentrations considered for simulation in Cell Designer software is available at https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER5/Supplementary_table1.xlsx.

Supplementary Table 5.2. Sensitivity analysis including ethylene and auxin production in the objective function before starting manual curation is available at https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER5/Supplementary_table2.xlsx.

Supplementary Table 5.3. Pathologies found in the Genome-scale metabolic reconstruction after curation is available at https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER5/Supplementary_table3.xlsx.

Supplementary Table 5.4. Unique reactions found for aluminum concentration scenarios that generated biomass

Aluminum Scenario (μM)	Reaction id	Description	Subsystem
50	R245	1-pyrroline-5-carboxylate dehydrogenase (NADP), mitochondrial	Arginine and proline metabolism
50	R754	Cystathionine beta-lyase	Cysteine and methionine metabolism
100	R786	glucose-6-phosphate 1-epimerase, cytosolic	Glycolysis/Gluconeogensis
500	R240	Glutamate 5-kinase, plastidic	Arginine and proline metabolism
500	R753	Pyrroline-5-carboxylate reductase (NADPH), plastidic	Arginine and proline metabolism
500	R902	Glutamate-5-semialdehyde dehydrogenase, plastidic	Arginine and proline metabolism
500	R905	Proline biosynthesis; spontaneous reaction, plastidic	Arginine and proline metabolism
500	R2070	Dicarboxylate/tricarboxylate carrier (oaa:akg), mitochondrial	Transport (Mitochondrial)
500	R2137	Amino acid transporter (pro-L), plastidic	Transport (Plastidic)
1000	R173	GTP:cytidine 5-phosphotransferase, plastidic	Pyrimidine metabolism
1000	R187	ATP:cytidine 5-phosphotransferase, plastidic	Pyrimidine metabolism
5000	R306	Citrate synthase, peroxisomal	Glyoxylate Cycle
5000	R307	Citrate synthase, mitochondrial	TCA Cycle
5000	R356	Malate synthase, peroxisomal	Glyoxylate Cycle
5000	R453	Malonyl-CoA decarboxylase	Alanine, aspartate and glutamate metabolism
5000	R1484	alcohol-forming fatty acyl-CoA reductase (n-C26:0CoA)	Secondary Cell Wall Metabolism (Cuticular Wax)
5000	R1528	3-oxoacyl-CoA synthase (n-C20:0)	Fatty acid biosynthesis (Very long chain)
5000	R1529	3-oxoacyl-CoA reductase (n-C20:0)	Fatty acid biosynthesis (Very long chain)
5000	R1530	3-hydroxyacyl-CoA dehydratase (n-C20:0)	Fatty acid biosynthesis (Very long chain)
5000	R1531	enoyl-CoA reductase (NADPH) (n-C20:0)	Fatty acid biosynthesis (Very long chain)
5000	R1532	3-oxoacyl-CoA synthase (n-C22:0)	Fatty acid biosynthesis (Very long chain)
5000	R1533	3-oxoacyl-CoA reductase (n-C22:0)	Fatty acid biosynthesis (Very long chain)
5000	R1534	3-hydroxyacyl-CoA dehydratase (n-C22:0)	Fatty acid biosynthesis (Very long chain)
5000	R1535	enoyl-CoA reductase (NADPH) (n-C22:0)	Fatty acid biosynthesis (Very long chain)
5000	R1536	3-oxoacyl-CoA synthase (n-C24:0)	Fatty acid biosynthesis (Very long chain)
5000	R1537	3-oxoacyl-CoA reductase (n-C24:0)	Fatty acid biosynthesis (Very long chain)
5000	R1538	3-hydroxyacyl-CoA dehydratase (n-C24:0)	Fatty acid biosynthesis (Very long chain)
5000	R1539	enoyl-CoA reductase (NADPH) (n-C24:0)	Fatty acid biosynthesis (Very long chain)

Aluminum Scenario (μM)	Reaction id	Description	Subsystem
5000	R1540	3-oxoacyl-CoA synthase (n-C26:0)	Fatty acid biosynthesis (Very long chain)
5000	R1541	3-oxoacyl-CoA reductase (n-C26:0)	Fatty acid biosynthesis (Very long chain)
5000	R1542	3-hydroxyacyl-CoA dehydratase (n-C26:0)	Fatty acid biosynthesis (Very long chain)
5000	R1543	enoyl-CoA reductase (NADPH) (n-C26:0)	Fatty acid biosynthesis (Very long chain)
5000	R2259	citrate transport, peroxisomal	Transport (Peroxisomal)
5000	R2316	Malonyl-CoA transport, ER	Transport (ER)
5000	R2358	Sink needed to allow Cerotyl alcohol to leave system	Demand

Supplementary Table 5.5. Core reactions found in the aluminum response in different aluminum concentration scenarios (Table 5.1). Table is available at https://github.com/ccsosa/Supplementary-information/blob/main/CHAPTER5/Supplementary_table5.xlsx.