DNA Methylation Patterns Associated with Aluminum Tolerance in Cultivated and Wild Rice Species

Jenny Johana Gallo Franco

Advisor: Mauricio Alberto Quimbaya
Co-Advisor: Thaura Ghneim Herrera
Assessor: Fabian Tobar Tosse

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

Cali, Valle del Cauca, Colombia
May 2023

Dedicated to my parents, my brother and my boyfriend
for all their love and unconditional support in every step
of this challenging adventure

Acknowledgements

This doctoral dissertation is the result of the collaborative work of a wonderful team of people who have been by my side in one way or another.

First, I would like to express my sincere gratitude to Professor Mauricio Quimbaya, who believed in me from the very beginning and became not only my academic guide but also a mentor for my personal development. Thank you, Professor Mauricio, for your unwavering support, guidance, motivation, and inspiration. It was a great honor to have you as my director.

I would also like to thank Professor Fabian Tobar, who has taught me a lot and has always been there for me whenever I needed guidance, motivation, or simply a piece of advice. I learned a lot from you, thank you.

I want to thank Professor Thaura Ghneim whose support from the beginning has provided me with the academic and personal tools necessary to navigate my path and become better every day. Thank you for being a woman of inspiration for me.

I would also like to express my immense gratitude to the Universidad Javeriana Cali and the Ómicas 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 my doctoral training. I extend my thanks to all the members of the Ómicas program, especially Leidi Rojas, Janeth Rodríguez, Mabren Camayo and Alejandra Cataño for their support in making this project possible. I want to acknowledge the doctoral program in engineering and applied sciences under the leadership of Professor Andres Jaramillo for the continuous support provided throughout my academic journey, and to Andrea Ramirez for guiding me every step of the way towards the completion of this goal.

I am grateful to the plant physiology group at Universidad Icesi for their experimental and emotional support, with special appreciation to Juliana Chaura and Vannesa Reyes for their teachings. Furthermore, I am immensely grateful to Dr. Frank Johannes for welcoming me into his research group. It has been an honor for me to work and learn from you and all your team, including Rashmi Hazarika, Ioanna Kakoulidou, Robert Piecyk, Luca Schlegel and Zhilin Zhang. I am left with great learning, many stories and experiences and new friends.

I thank my friends and doctoral colleagues, Miguel Romero, Camila Riccio, Christian Sossa, Nicolas Lopez, Carolina Saavedra, Gustavo Lara, David Jimenez and Carolina Clavijo for their unwavering support and all that each of you taught me. Thank you for being incredible friends in this journey, creating unforgettable memories and experiences. Likewise, thanks to my friends and colleagues at iOmicas for being part of this process in various ways and

enriching my journey in every conversation or discussion. I also want to thank my students Isabella Zuluaga, Ana Maria Restrepo, Sebastián Zapata and Juan Pablo Gutierrez who actively contributed to this process and taught me so much.

Last, but certainly not least, I want to extend my deepest thanks my parents Jenny Franco and Jhon Jairo Gallo, and my brother, Jhon Alexander Gallo, for their love, for always believing in me, for being my unconditional support and my inspiration to accomplish every goal in my life. To my life partner, Mauricio Peñuela, thanks for his love, unconditional support and for being always there to encourage me and accompany me in the most challenging moments.

Contents

List of abbreviations

Al: Aluminum

DEG: Differentially Expressed Gene

DMG-DEGs: Differentially methylated and Differentially Expressed Gene

DMG: Differentially Methylated Gene

DMR: Differentially Methylated Region

gbM: Genebody Methylated

HM: Heavy Metals

mCs: Methylated Cytosines

NDM-DEGs: Non-Differentially Methylated but Differentially Expressed Genes

OA: Organic Acids

PCC: Pearson Correlation Coefficient

ROS: Reactive Oxygen Species

teM: TE-like Methylated

TEs: Transposable Elements

UM: Unmethylated

WGBS: Whole Genome Bisulfite Sequencing

Abstract

Plants are sessile organisms that constantly face stressful biotic and abiotic conditions, which they must overcome to survive. To combat these challenges, plants have developed diverse strategies involving genetic, molecular, and physiological approaches. Recently, the role of epigenetics in plant stress response has gained significant attention. Epigenetic modifications, including alterations to DNA or histone proteins, can affect gene expression in a potentially stable and heritable manner without affecting DNA composition. Although there have been several studies investigating the role of epigenetics in plant stress response, significant knowledge gaps still exist, with most studies focusing on a limited number of species or specific stress conditions. This doctoral dissertation studies the role of DNA methylation, one of the most studied epigenetic mechanisms, in relation to aluminum stress tolerance levels in rice plants. The research addresses the proposed problem through several approaches: (i) Examination of pre-existing epigenetic marks in tolerant and susceptible genotypes of *Oryza sativa* and *Oryza glumaepatula,* as well as DNA methylation changes generated in response to Al exposure (ii) Evaluation of differential expression of aluminum stress response genes; and (ii) correlation of the gene expression patterns with changes in DNA methylation levels in response to stress. Presentation of evidence highlighting the significance of epigenetics as a key regulatory factor in aluminum stress response, along with a group of potentially epigenetically regulated genes involved in aluminum tolerance. These findings lay the bases for a better understanding of epigenetics as a crucial determinant in plant response to abiotic stresses, while also raising new questions and challenges for future research in the scientific field.

Resumen

Las plantas son organismos sésiles que se enfrentan constantemente a condiciones bióticas y abióticas estresantes que deben combatir para poder sobrevivir. Para ello, han desarrollado diversas estrategias que involucran enfoques genéticos, moleculares y fisiológicos. Recientemente, ha surgido un gran interés en el estudio del papel de la epigenética en la respuesta de las plantas a las condiciones de estrés. Las modificaciones epigenéticas, que incluyen alteraciones en el ADN o proteínas histonas, pueden afectar la expresión de los genes de manera potencialmente estable y heredable, sin afectar la composición del ADN. Aunque existen diversos estudios del papel de la epigenética en la respuesta a estrés por parte de las plantas, aún existen grandes vacíos alrededor del tema con la mayoría de los estudios enfocados en algunas especies o en condiciones de estrés específicas. Esta disertación doctoral estudia el papel de la metilación del ADN, uno de los mecanismos epigenéticos más estudiados, en relación con los niveles de tolerancia a estrés por aluminio en plantas de arroz. Esta investigación aborda el problema propuesto a través de varios enfoques: (i) evaluación de marcas epigenéticas preexistentes en genotipos tolerantes y susceptibles de *Oryza sativa* y *Oryza glumaepatula*, así como cambios en la metilación del ADN generados en respuesta a la exposición al Al (ii) Evaluación de la expresión diferencial de genes de respuesta al estrés por aluminio; y (ii) correlación de los patrones de expresión génica con los cambios en los niveles de metilación del ADN en respuesta al estrés. Los resultados de este estudio representan bases fundamentales para el entendimiento de la epigenética como un factor clave en la respuesta de las plantas a estrés por condiciones abióticas. A partir de estos resultados también se plantean nuevo interrogantes y retos para futuras investigaciones en el campo científico.

List of Figures

List of Tables

Chapter 1

Introduction

Rice is considered the main staple food for more than half of the world's population and the second most important cereal crop in the world. The *Oryza* genus comprises 24 species, but most of the world's rice falls into the *Oryza Sativa* (L.) classification. This species, along with *Oryza glaberrima* (Steudel) are the two cultivated species of rice (Pegoraro et al., 2018). In addition to its importance in human nutrition, rice has been considered a model organism due to its distinction as the second plant and the first cultivated species to have its genome sequenced. The sequencing of the rice genome had a great impact on rice genetics and plant breeding research, especially because it could be used as a model for other cereal crops with larger genomes, such as maize and wheat (Jackson, 2016).

The constant growth of the human population implies an increase in the demand for rice to supply nutritional needs. Currently, rice crops have a cultivated area of approximately 167 million hectares, with an increment of 5.55 million hectares between 2010 and 2017 period (Food and Agriculture Organization of the United Nations, 2020; Figure 1.1A). However, there is still a need to increase rice production by 50% by 2050 to feed the growing population (Lin, et al., 2019). Among the challenges that must be overcome in rice productivity is the impact of biotic and abiotic stresses that cause ~30% and ~50% yield losses worldwide respectively (Fahad et al., 2019). So far, different breeding strategies have been applied to increase the tolerance of rice crops to adverse conditions. One effective strategy to enhance rice crop tolerance is the study of wild genotypes which possess greater adaptability to adverse conditions and a diverse gene pool that can be transferred to cultivated varieties. Among the valuable genetic resources for rice breeding, wild species within the *Oryza* genus play a crucial role. Particularly,

Oryza glumaepatula, the only diploid (2n= 24) species native to South America, offers significant potential as a source of genetic diversity for rice breeding programs in the region, given its adaptation to local edaphoclimatic conditions (Fuchs et al., 2016).

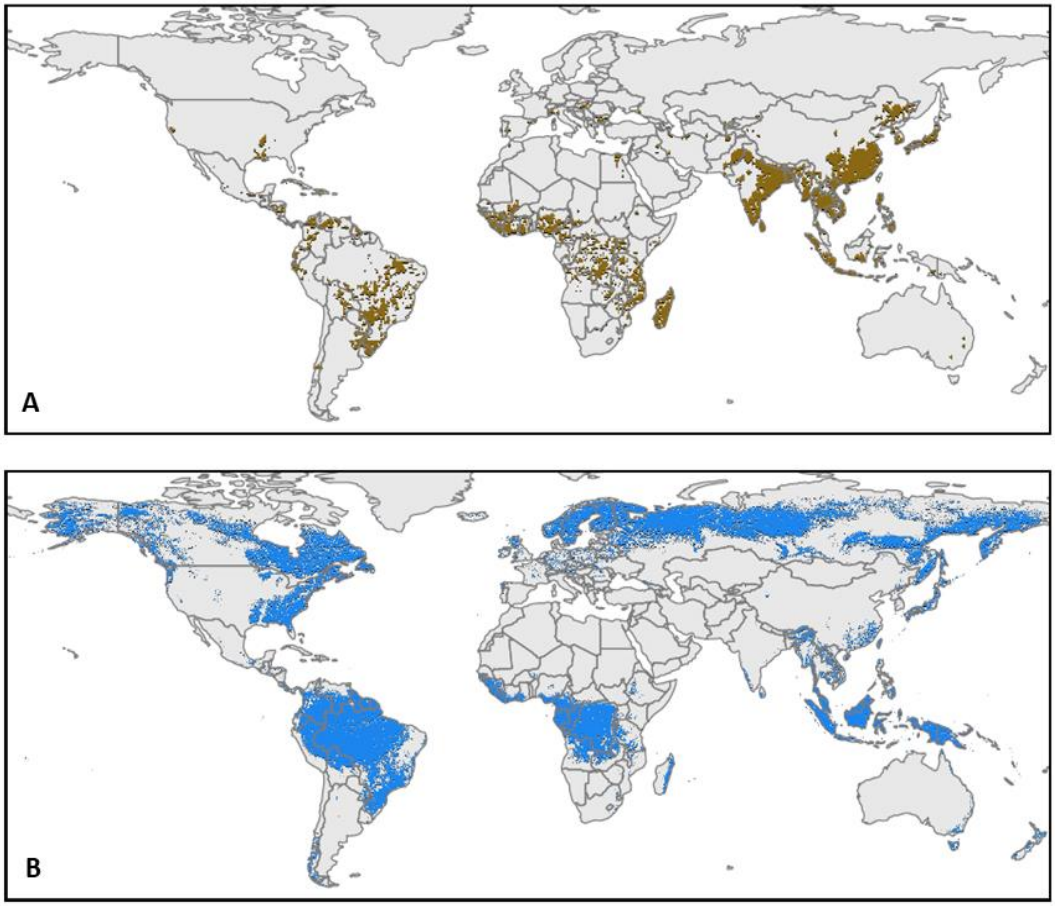


Figure 1.1. Worldwide distribution of acidic soils and rice crop areas (1 km^2 resolution). **A.** Worldwide rice crop area coverage (pixel probability > 0) (N. D. Jackson et al., 2019). **B.** Areas with a weighted averaged soil pH (0-30 cm) less than or equal to 5.5 (acidic soils) using data extracted from Soil grids (Hengl et al., 2017).

Among the most important abiotic stresses affecting rice production are drought, salinity, temperature, and heavy metals (HM). Although there have been substantial advances in increasing the resistance of current varieties to stressful conditions, there is still a very wide range of stresses affecting agricultural production (Zhu, 2016).

HMs are elements with densities above 5g/cm^3 that belong to the Earth's crust natural components. High concentrations of HMs can generate cytotoxic, genotoxic, and

mutagenic effects in living organisms. Under physiological conditions, HMs can be divided into two groups: (i). Essential elements that are necessary for plant growth being structural blocks in proteins with an enzymatic function, such as iron (Fe), manganese (Mn), zinc (Zn), magnesium (Mg), molybdenum (Mo), and copper (Cu), and (ii). Non-essential elements like cadmium (Cd), chromium (Cr), lead (Pb), aluminum (Al), and selenium (Se). While essential elements are necessary for plants in small amounts, high concentrations of both types of elements can lead to inhibition of plant growth and development (Rascio & Navari-Izzo, 2011).

HMs exert a significant impact for plants on acid soils, caused by the excess of cationic species such as magnesium (Mg^{2+}), calcium (Ca^{2+}), phosphorus (P), sodium (Na^{+}) and aluminum (Al^{3+}) which in turn, affect plant physiological responses leading to crop yield losses for breeders and farmers (Fryzova et al., 2017; Samac & Tesfaye, 2003). Acid soils represent nearly 30% of worldwide arable land, with 13% of staple crops cultivated in these areas. These types of soils classified as ultisols or oxisols are characterized by a pH lower than 5.5 (Bojórquez-Quintal et al., 2017; Rahman et al., 2018). In South America, acid soils constitute 33.3% of the land, with Brazil having the largest area (70.8%), followed by Colombia (8.3%), Peru (6.9%), Venezuela (6.4%), and Bolivia (4.9%) (Figure 1.1B) (Cochrane, 1979). In the specific case of rice, the prevalence of acid soils poses a significant challenge due to the substantial overlap between acid soil distribution and rice cultivation areas (Figure 1.1).

Al toxicity on acid soils has been reported as one of the major factors limiting crop production, and becoming worse due to current fertilization practices, pasture management, and climate change (Kochian et al., 2015; Zheng, 2010). Normally, Al is found in the form of aluminosilicates and Al oxides, which do not affect organisms. However, when Al hydrolyzes water molecules, it forms Al hydroxide, which is toxic to living organisms. The toxic effect of different forms of Al (speciation) on plant growth decreases in the following order: Al, $Al(OH)^{2+}$, $Al(OH)_2^{+}$ and $Al(OH)_4^{-}$. The most common form of Al found in acidic soils is trivalent aluminum (Al^{3+}), which has the most significant impact on plant growth. Conversely, Al that is precipitated or chelated with organic compounds is not toxic to plants (Bojórquez-Quintal et al., 2017). When plants are exposed to Al^{3+} ions, damage to root apices occurs within a matter of minutes, with the distal zone being particularly affected, especially the last 2 to 3 mm of the root tips. The Al^{3+} ions enter the cell causing a decrease and/or cessation of cell growth by inhibiting the development of the cytoskeleton. Additionally, these molecules bind to DNA molecules, affecting their compaction and inhibiting DNA replication. Consequently, radical growth is slowed, and various types of serious cellular disorders are generated, including oxidative stress, low nutrient assimilation, and reduced water uptake (Kochian et al., 2004).

There are various methods for partially amending acid soils, such as applying calcareous compounds or mixing the soil with leaf litter. However, these practices only correct the pH of a thin layer of the surface soil and are limited in time, offering only partial solutions. In addition, soil amendment can increase production costs, making it impractical for producers in the most affected regions. Additionally, these methods can lead to soil salinization, creating yet another abiotic stress condition (Whitten et al., 2000). A more promising solution involves searching for plants that have adaptations to acid soils and are tolerant of the predominant stress conditions. These plants can be used in genetic improvement programs by introgression of genes associated with tolerance using molecular markers or incorporating the same genes through transgenic techniques. From a different perspective, it is crucial to comprehend the underlying of tolerance exhibited by tolerant varieties or species. This understanding can guide breeding programs through various strategies, including genetic editing, among others.

Staple food crops such as maize, wheat, sorghum, and rice have been extensively studied to unravel the functional mechanism involved in Al tolerance (Famoso et al., 2010). Significant progress has been made in the past decade, shedding light on the genomic and transcriptomic mechanisms underlying heavy metal (HM) tolerance. Rice, in particular, has served as a valuable model due to its notable tolerance to Al toxicity (Famoso et al., 2010; Mustafa & Komatsu, 2016). It has been proposed that rice has a complex response to Al stress, involving a wide range of strategies and a diversity of genes (Famoso et al., 2011; Zhang, et al., 2019). However, despite these advancements, several gaps in the understanding persist.

In addition, epigenetics has become a key actor of plant stress response, exerting a strong influence on gene expression and phenotype. However, there is a lack of studies investigating epigenetic mechanisms involved in rice Al tolerance (Chang et al., 2020; Sudan et al., 2018). Understanding the epigenetic factors underlying rice's ability to tolerate Al stress can provide invaluable insights into the regulatory mechanisms at play and unlock new avenues for enhancing Al tolerance in rice breeding programs. Likewise, there exist an urgent need for a comprehensive understanding of the link between genetic and epigenetic factors influencing rice tolerance to Al stress, which is critical for paving the way towards future improvements of rice varieties.

Therefore, this doctoral thesis studies the role of epigenetics in rice response to Al exposure through several approaches: (a) a comprehensive review of the current state of research on epigenetic involvement in response to heavy metal stress; (b) Examination of pre-existing epigenetic marks in tolerant and susceptible genotypes of *O. sativa* and *O. glumaepatula*, potentially associated with their tolerance levels prior to stress exposure; (c) Evaluation of differential expression of Al stress response genes after a long stress exposure time. No transcriptional studies on long-term Al-stress in rice plants has been

carried out up to the time of this research; and (d) correlation of the gene expression patterns with changes in DNA methylation levels in response to stress, to evaluate the possible regulatory role of epigenetics in such differential gene expression. The outcomes of this dissertation are as follows: (i) proposal of variety-specific methylome profiles, along with a set of genes involved in Al tolerance that possess pre-established epigenetic marks in different rice genotypes based on their tolerance level; (ii) Characterization of the transcriptomic profile of tolerant and susceptible rice genotypes under prolonged Al toxicity stress conditions; (iii) Exploration of the relationship between transcriptomic profiles and epigenetic marks; and (iv) presentation of evidence highlighting the significance of epigenetics as a key regulatory factor in Al stress response, along with a group of potentially epigenetically regulated genes involved in Al tolerance.

1.1. Research Hypothesis

H.1. DNA methylation patterns are related to Al tolerance levels in contrasting rice species and genotypes. Additionally, specific methylation patterns are regulating transcriptional responses associated with aluminum stress.

1.2. General Objective

To elucidate relevant genetic elements and biological mechanisms associated with physiological responses to aluminum stress, via epigenomic and transcriptional analyses of cultivated and wild rice, to complement the current knowledge, derived from a genetic perspective, of abiotic stress responses in rice.

1.3. Specific Objectives

SO.1. To characterize genome-wide DNA methylation patterns of commercial (*Oryza sativa*) and wild (*Oryza glumaepatula*) rice genotypes with contrasting response to aluminum stress grown under control conditions by whole genome bisulfite sequencing to identify the pre-established marks associated with Al-tolerance levels.

SO.2. To characterize the transcriptional response of Al-tolerant and susceptible genotypes of commercial (*Oryza sativa*) and wild (*Oryza glumaepatula*) rice plants under Al exposure using RNA sequencing techniques to propose a functional model of the plant response to aluminum stress.

SO.3. To identify genome-wide DNA methylation changes in Al-tolerant and susceptible genotypes of commercial (*Oryza sativa*) and wild (*Oryza glumaepatula*) rice plants in response to Al stress condition using whole-genome bisulfite sequencing.

SO.4. To evaluate the correlation between DNA methylation and the transcriptional response of aluminum tolerant and susceptible genotypes under aluminum stress conditions to determine genes or regions potentially regulated by epigenetic mechanism involved in the contrasting levels of Al tolerance.

SO.5. To evaluate the relationship between DNA methylation and chromosomal recombination through genome wide comparisons to identify the potential role of methylation as a feature for prediction of recombination rates.

1.4. Document Outline

The reminder of this dissertation is organized as follows:

1. **Chapter 2** explores and analyzes the existing scientific literature on epigenetics as an important factor that could regulate HM stress responses. In addition, the relationship between epigenetic and genetic elements related to HM tolerance is revised, with a special focus on Al tolerance in rice.

2. **Chapter 3** presents the analysis of predetermined DNA methylation marks in Al-tolerant with respect to the susceptible genotypes of *Oryza sativa* and *Oryza glumaepatula* grown under control conditions.

3. **Chapter 4** presents the transcriptome analysis of Al-tolerant and susceptible genotypes of *Oryza sativa* and *Oryza glumaepatula* under Al-stress conditions.

4. **Chapter 5** presents the methylome variations of Al-tolerant and susceptible genotypes of *Oryza sativa* and *Oryza glumaepatula* under Al-stress conditions, as well as the role of DNA methylation on gene regulation.

5. **Chapter 6** explores the relationship between DNA methylation and chromosomal recombination rates.

A graphical overview of the research questions addressed in each chapter of this dissertation is presented in Figure 1.2, along with the principal methods employed to obtain the results. Likewise, this figure shows the linearity as well as the relationship of the different chapters developed in this doctoral thesis.

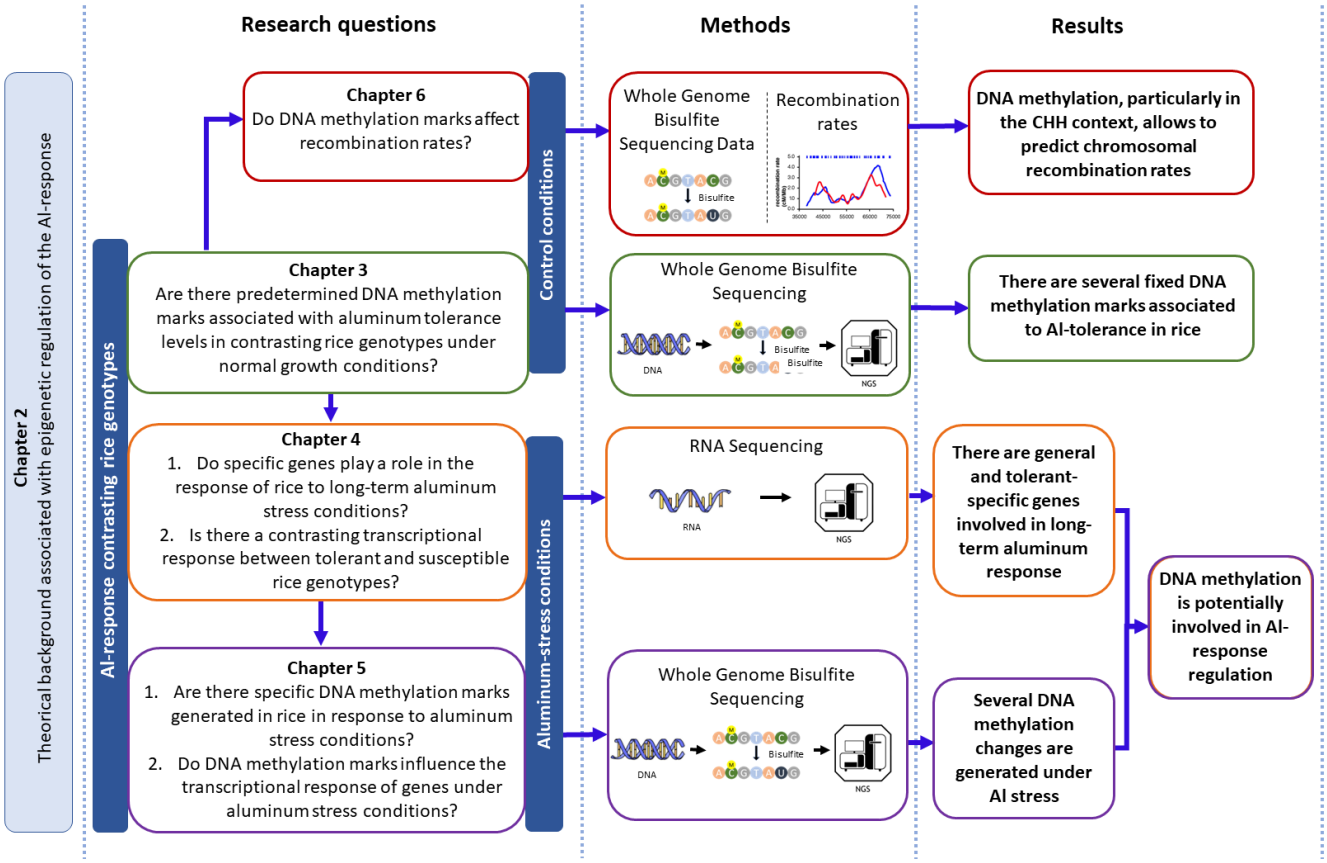


Figure 1.2. Graphical summary of the outline of this doctoral thesis. Three main sections are highlighted from left to right: research questions addressed in each chapter, methods employed, and key findings obtained.

Chapter 2

Epigenetic Control of Plant Response to Heavy Metal Stress: A New Window to Explain Aluminum Tolerance

This chapter was previously published as:

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

Chapter Summary

High concentrations of heavy metal (HM) ions impact agronomic staple crop production in acid soils (pH less than or equal to 5.5) due to their cytotoxic, genotoxic, and mutagenic effects. Among cytotoxic ions, the trivalent aluminum cation (Al^{3+}) formed by solubilization of aluminum (Al) into acid soils, is one of the most abundant and toxic elements under acidic conditions. In recent years, several studies have elucidated the different signal transduction pathways involved in HM responses, identifying complementary genetic mechanisms conferring tolerance to plants. Although epigenetics has become more relevant in abiotic stress studies, epigenetic mechanisms underlying plant responses to HM stress remain poorly understood. This chapter describes the main epigenetic mechanisms related to crop responses during stress conditions, specifically, the molecular evidence showing how epigenetics is related to plant adaptation responses to HM ions. Likewise, the pivotal relationship between epigenetic and genetic factors associated with HM tolerance was analyzed. Finally, using rice as a study case, a general analysis over previously whole-genome bisulfite-seq published data was performed. Specific genes related to Al tolerance, measured in contrasting tolerant and susceptible rice varieties, exhibited differences in DNA methylation frequency. The differential DNA methylation patterns could be associated with epigenetic regulation of rice responses to Al stress, highlighting the major role of epigenetics over specific abiotic stress responses.

Keywords: *Abiotic stress, aluminum tolerance, epigenetic response, heavy metals, rice.*

2.1 Genetic Mechanisms Underlying Aluminum Tolerance

Plants have evolved different strategies to cope with HMs, diverging according to distinct factors as the plant species or the HMs exposure time and concentrations (Horst et al., 2010). These strategies fall into two general mechanisms: (i) An exclusion mechanism, where plants exudate organic compounds to the rhizosphere to chelate HM ions, transforming them into nontoxic compounds, and avoiding their chemical intake through root cells; and (ii) A detoxification mechanism, where plants allow the entrance of HM ions for internal detoxification and sequestration (Figure 2.1; Kochian et al., 2015). Among the heavy metals, aluminum (Al) stress is widely considered as one of the primary limiting factors in plant crop production. As the most abundant metal on Earth and the third most prevalent element, following oxygen and silicon, aluminum constitutes approximately 8.1% of the Earth's crust by weight. Table 2.1 summarizes the different exclusion and tolerance mechanisms reported so far in plants.

Rice have been a model species for studying Al tolerance, as it is one of the plants with the highest tolerance to this element (Famoso et al., 2010, 2011). Rice has a complex response against Al stress, involving a wide range of strategies and a diversity of genes (Magalhaes et al., 2007). These genes are potentially involved in the exclusion of Al^{3+} ions through OA efflux; for instance, the MATE transporters *OsFRDL2* and *OsFRDL4*, has shown a role in OA transport (Delhaize et al., 2012; Famoso et al., 2010; Yokosho et al., 2014). Other rice Al responses include the modification of the cell wall properties (Che et al., 2018; Kochian et al., 2015), and Al^{3+} ions uptake and subsequent sequestration/translocation into the vacuole by different Al transporters like bacterial-type ABC and Nramp Al transporters (Huang et al., 2009; Liu et al., 2014; Xia et al., 2010). Other genetic elements associated with Al tolerance include genes encoding transcription factors as *ART1*, *ASR1* and *ASR5* (Arenhart et al., 2016; Che et al., 2016; Yamaji et al., 2009). The upregulation of specific genes as OsMGT1, a magnesium transporter, is also linked to high Al tolerance (Chen et al., 2012). More recently, Zhang et al., (2019) reported 69 potential candidate genes related to Al tolerance, identified in a collection of 150 rice landraces using a combined GWAS-transcriptomic approach. Complementarily, several QTLs associated with Al tolerance have been identified in rice using different inter and intra-specific mapping populations (Famoso et al., 2011; Ma et al., 2002; Nguyen et al., 2003; Wu et al., 2000; Xue et al., 2006, 2007; Zhang et al., 2019). Famoso et al. (2011) reported 48 QTLs located on chromosomes 1, 3, 9, and 12. The QTLs were generated based on mapping populations exposed to Al stress, using relative root growth as the experimental phenotypic readout. The major QTL was found on chromosome 12, explaining 19% of the phenotypic response. Findings reported in above mentioned studies support the hypothesis that Al tolerance in rice involves multiple genes, genomic regions, and mechanisms.

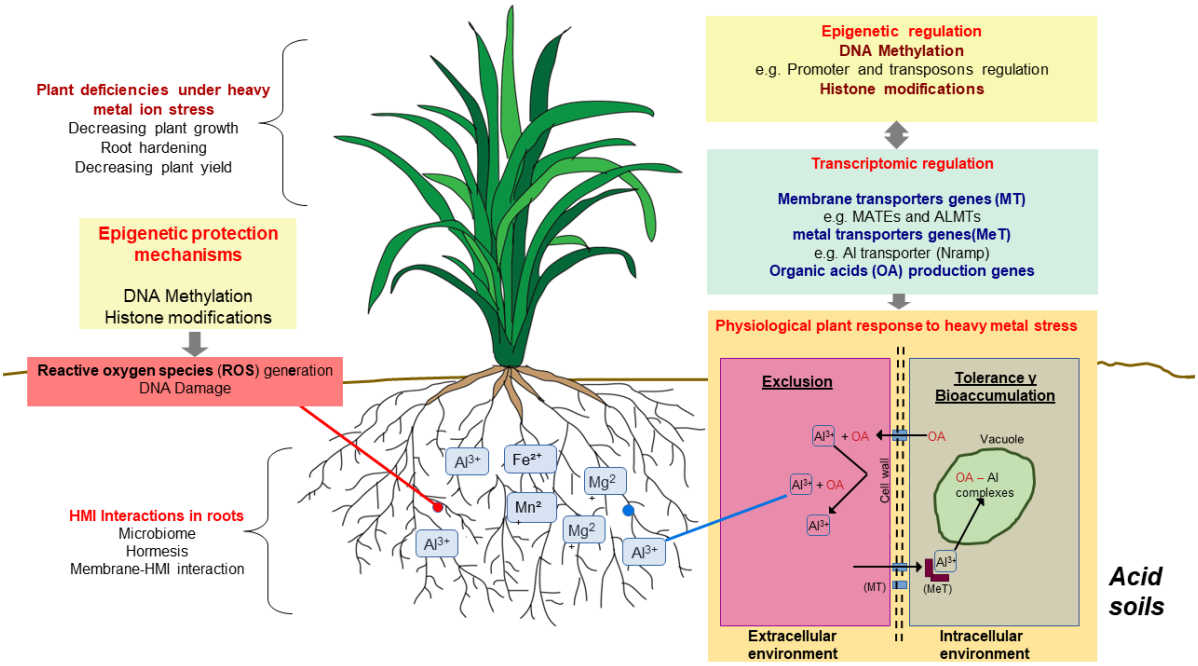


Figure 2.1. Schematic representation of physiological, genetic, transcriptional, and epigenetic mechanisms involved in plant responses to heavy metals (HM) exposure. Plant exposure to HMs induces different physiological deficiencies that could be countered by two principal tolerance mechanism shown at the bottom right of the figure: an exclusion mechanism, where the plant secretes organic acids (OAs) out of the root, avoiding the entrance of HM ions or, a detoxification mechanism and sometimes bioaccumulation, wherein plants internalize HM ions through membrane transport proteins such as ALMT or MATE carriers, and subsequently, HMs can be chelated by organic acids (OA) or translocated into the vacuoles through ABC carriers or aquaporins. The regulation of HM responsive genes has been related to epigenetic mechanisms as DNA methylation and histone modifications which can repress or activate gene expression through promoter or gene body methylation as well as avoiding transposon movement (top right). Another important epigenetic mechanism involved in the HM stress response is the hypermethylation along the genome to protect DNA from possible damages caused by metal subproducts.

The previous evidence relates both, genic elements, and specific genic mechanisms with the phenotypic response to cope with HMs stresses. Besides the genetic control that exists to regulate these responses, additional regulation layers might exist, being epigenetics a controlling mechanism of paramount importance to adapt to abiotic stresses, and specifically, to HMs restrictive conditions. The following sections will review the current evidence associating epigenetics with HMs stress responses. Giving its agronomic relevance, special attention is given to rice epigenetics as integrated strategies to cope with HMs and aluminum stresses.

Table 2.1. Summary of main exclusion and tolerance mechanisms reported in plants.

Species	Genes	Mechanism	Specific mechanism	Function	References
P. vulgaris, T. aestivum, S. bicolor, H. vulgare, Zea mays, snapbean, oat, rye, *Glicine max*, *Colocasia esculenta, Triticale* sp., *Helianthus annuus*	*ALMT, MATE, OSALMT4*	Exclusion	Organic acid exudation	Chelate (release of malate, citrate, or oxalate) located in the root apex	Kochian et al., 2004, 2015; Liu et al., 2018
Zea mays, *Cinnamomum camphora, Eucalyptus camaldulensis*		Exclusion	Phenolic compounds exudation	Release of other organic compounds (e.g. catechol, catechin, and quercetin), cenothein B and proanthocyanidin in roots	Kochian et al., 2015
Cucurbita pepo, wheat, tea	ATPases	Al detoxification	Changes in the Rhizosphere pH	pH rhizosphere changes to induce to Al detoxification mechanisms	Bojórquez-Quintal et al., 2017
Oryza sativa, Solanum tuberosum, Arabidopsis thaliana, Petunia inflata	*XTH, XET, XTH31*, pectin methylesterases, *OsFRDL4, STAR1, STAR2*, ABC transporters*, HMG2, HMG3, WAK1*	Al detoxification	Cell wall modification	Changes in the structural properties of cell wall such as reduction of wall plasticity/elasticity, carbohydrates, methylated pectins, and reduced pectin methylesterases; increased sterols biosynthesis; negativity of apoplast to enhance Al transport	Horst et al., 2010; Kochian et al., 2015; Morkunas et al., 2018; Schmohl et al., 2000; Wagatsuma et al., 2018
Arabidopsis thaliana, Oryza sativa,	*Nramp, OsNrat1, OsALS1*, aquaporine family, ABC, *ALMT, OsCDT3*	Al detoxification	Al transportation	Arrest Al from cell wall to root cell vacuole	Kochian et al., 2015; Arbelaez et al., 2017
Brassica napus, Nicotiana tabacum, wheat*, Arabidopsis thaliana, Zea mays*	*ALMT, MATE, SbMATE, TaALMT1, OsFRDL4*	Al detoxification	ALMT/MATE proteins Al transportation	Passive efflux of malate; carriers that mediate citrate efflux coupled to H+ influx	Kochian et al., 2015; Liu et al., 2014
Oryza sativa, Arabidopsis thaliana, Andropogon virginicus	*Nramp, OsALS1, Nrat1*	Al detoxification	Nramp proteins Al transportation	Specific transporter for aluminum ions (no divalent cations) transport from cell wall to vacuoles	Yokosho et al., 2011; Ezaki et al., 2013; Kochian et al., 2015
*Oryza sativa, Arabidopsis thalian*a	*OsSTAR1, OsSTAR2, AtALS3, OsALS1, AtALS1*	Al detoxification	ABC proteins Al transport	ATP-driven pumps (ABC transporters);	Huang et al., 2009; Delhaize et al, 2012; Kochian et al., 2015
Oryza sativa, Arabidopsis thaliana, Hydrangea macrophylla	Aquaporins such as *HmVALT*, *HmPALT1*	Al detoxification	Aquaporins transportation	Transport and store in shots	(Kochian et al., 2015; Negishi et al., 2012)

2.2 Epigenetic Mechanisms in Plants

Epigenetics refers to the study of heritable and stable changes in gene expression without DNA sequence modifications (Wu and Morris, 2001). Three epigenetic mechanisms have been described in gene expression regulation: (i) DNA methylation (modifications at genomic level), (ii) histone modifications (chromatin modifications) and (iii) Small RNA modifications (RNA directed DNA Methylation-RdDM pathway) (Chang et al., 2020; Sudan et al., 2018). Currently, DNA methylation is the most documented epigenetic modification, and it is recognized as a relatively stable, and inheriting transgenerational mark involved in a set of biological processes such as the activity of transposable elements, genomic imprinting, alternative splicing, and regulation of temporal and spatial gene expression (Ou et al., 2012; Zhang et al., 2006). Mammals and plants differ in their DNA methylation patterns. In plants, DNA methylation is more widespread and complex, and occurs in cytosine residues in the CG, CHG, and CHH sequence context (H can be A, C, or T), while in mammals it occurs mainly in a CG context (Bender, 2004; He et al., 2010). Studies on general DNA methylation profiles conducted on plants have shown that transposable elements and repetitive sequences are the most heavily methylated DNA regions in the rice genome. Overall, gene methylation occurs mainly in the CG context, while transposon methylation occurs in all three described contexts (He et al., 2010; Li et al., 2012; Yan et al., 2010).

The methylome in plants is mainly monitored and maintained during DNA replication and cell division by DNA methyltransferases. There are three major classes of DNA methyltransferases: DNA methyltransferases (METs), which are the main CG methylases in charge of maintaining CG methylation; the plant specific enzymes chromomethyltransferases (CMTs), that are known to mantain CHH and CHG methylation; and the domain rearranged methyltransferases (DRMs), that are involved in the maintenance and de novo methylation in all three contexts: CG, CHG and CHH. In contrast, DNA demethylation is performed by DNA glycosylases such as ROS1 (Repressor Of Silencing 1) and the DME (Demeter) enzyme (Lanciano & Mirouze, 2017).

2.3 Epigenetic Regulation of Plant Stress Response

Abiotic stresses can generate a diverse range of phenotypes in plants, which are a consequence of complex molecular, biochemical, and physiological changes. Plants responses and adaptation to these stress conditions vary in different ways and at various levels, including short term physiological responses such as metabolic and gene expression changes, and long-term responses such as genetic and epigenetic genome modifications (Turner, 2009). The mechanisms of signal transduction, as well as the

genetic variability underlying plants responses to stress, have been widely studied and, in many cases, successfully exploited by plant breeders to improve resistance to abiotic stress through traditional breeding or marker-assisted selection (Kantar et al., 2015; Zhu, 2016). Recently, epigenetic marks have gained attention as important factors of abiotic stress-related gene control (Kumar, 2018). For example, a stress signal can promote DNA methylation changes in the promoter regions of stress-responsive genes, thus modifying their expression pattern, generating histone conformational changes, and promoting transcriptional repression by preventing transcription factors binding to their target sites (Boyko et al., 2010; Ou et al., 2012; Ueda & Seki, 2020). Since methylation affects how genes are transcribed, it is hypothesized that DNA methylation is involved in the long-term transgenerational maintenance of transcriptional regulation patterns.

DNA methylation states can be complemented by additional mechanisms such as histone modifications (Mirouze & Paszkowski, 2011). These modifications refer to covalent changes that occur at the histone tail, including methylation, acetylation, phosphorylation, ubiquitination, sumoylation, glycosylation and ADP-ribosylation (Chang et al., 2020). These modifications are known to be capable of altering the chromatin structure and thus the accessibility of genes for transcription factors and other regulatory elements, which eventually, may influence gene transcription (Zhao & Shilatifard, 2019). Although considered a more dynamic and transitory mechanism because most changes that occur under stress conditions revert to their initial state quickly, histone modifications could play a role in the inheritance of certain stress-tolerant phenotypes (Pecinka & Mittelsten Scheid, 2012). For example, Kim et al., (2012) showed that H3K4me3 and H3K9ac histone modifications were abundant in several drought associated genes in *Arabidopsis thaliana* plants subjected to water-deficit regimes. When plants were irrigated, the H3K9ac modifications were rapidly eliminated, while H3K4me3 ones remained, indicating that the latter modification can be stably inherited through generations. Histone modification effects on gene regulation have also been reported for other stress conditions. Sokol et al., (2007) reported transient H3Ser-10 phosphorylation, H3 phosphoacetylation, and histone H4 acetylation under salinity and cold-stress related to the expression of stress-specific genes. Likewise, the trimethylation of H3K4 and acetylation of H3K9 in *A. thaliana* was generated by exposure to drought, ABA, and salt stress, causing stress-responsive genes expression (Kim et al., 2008).

Stress-induced epigenetic changes, especially DNA methylation, occur regularly in all plant species, reinforcing the importance of this mechanism for regulating plant responses to environmental changes; most of these changes are heritable and play an important role in plant adaptation (Feng et al., 2010). Genomic sequences whose changes in their methylation status are maintained over generations, without altering the acquired methylated pattern, are known as epialleles (Kalisz & Purugganan, 2004). There is evidence that epialleles can occur over stress-related genes, however, they can also be

present in genetic regions that are not directly related with the specific stress response, generating random changes across the genome. Moreover, both types of variations could be affected by natural selection according to the phenotypic effects they may cause (Verhoeven et al., 2010).

Transposons can also play a role in suppressing gene expression. This can occur due to the methylation state of a transposon located in or near a gene, which can directly affect the regulation of that gene through a methylation spread mechanism. Thus, transposon silencing through epigenetic marks contributes to the establishment of epigenetic variations affecting gene modulation in plants (Galindo-González et al., 2018; Saze & Kakutani, 2007).

Although the heritability of stress-induced methylation in plants remains poorly understood, some studies show that most of the induced variations are faithfully inherited to the offspring. For instance, Boyko et al., (2010) showed that *A. thaliana* plants exposed to salinity, cold, heat, and flooding, showed an overall increase in DNA methylation, associated with a higher stress tolerance in the progeny. In addition, Herman & Sultan, (2016) reported that in *Polygonum persicaria*, DNA methylation is involved in increasing offspring drought tolerance when parental plants are subjected to this stress. Some studies have even found epialleles with direct effects on economically important traits; for instance, heritable methylation changes induced in rice due to nitrogen deficiency (Kou et al., 2011), heavy metal toxicity (Ou et al., 2012), and drought (Zheng et al., 2017) have been described. This last study showed the conservation of several non-random methylation changes generated under drought conditions (>40%) through several generations. Zheng et al., (2017) also found that these epigenetic changes are related to stress responsive genes, and they seemed to influence rice long term adaptation to drought conditions. Thus, these studies support the potential role of epigenetic variation, and its inheritance across generations, as a relevant evolutionary process in crops. Similarly, they show that in rice, the mechanisms of epigenetic regulation of stress responses may be related to the type of stressor.

2.4 Epigenetic Mechanisms Involved in Heavy Metal Toxicity

A recent recurring question is whether there is a general pattern of DNA methylation related to HMs exposure in plants. Evidence from previous studies suggests that DNA methylation might play a role in the regulation of plant responses to HMs through at least two mechanisms (Aina et al., 2004; Arif et al., 2016; Choi & Sano, 2007; Greco et al., 2012). The first mechanism is related to a protective effect of methylation against HM-induced DNA damage through single-strand breaks or multi-copy transposition (Figure 2.1) (Bender, 1998). For example, Aina et al. (2004) compared methylation levels between clover (*Trifolium repens* L.), which is sensitive to Cr, Ni, and Cd, and hemp (*Cannabis sativa*

L.), which is partially tolerant to these HMs. The study found that in the absence of HM stress, the level of methylation of hemp roots was significantly higher than in clover. Similarly, Gullì et al., (2018) found that *Noccaea caerulescens* plants (a niquel (Ni) hyperaccumulator species) grown under high Ni concentrations were significantly hypermethylated at the genome level in comparison to *A. thaliana* Ni susceptible plants exposed to high Ni concentrations. These authors also showed that *MET1*, *DRM2*, and *HDA8* genes, which are involved in DNA methylation and histone modification, were differentially expressed between *N. caerulescens* and *A. thaliana*. Hypermethylation has also been reported to act as a defense mechanism to counteract radiation genotoxic effect as shown by (Kovalchuk et al., 2003); Volkova et al., (2018) who reported that pine trees plants (*Pinus silvestris*) adapted to survive high ionizing radiation, exhibited significantly hypermethylated loci compared to less adapted plants.

A second type of epigenetic response to HM stresses involves gene expression control (Figure 2.1). This regulation is not limited to the promoter region of genes but includes their coding regions (Choi & Sano, 2007). DNA methylation on gene promoters usually represses genetic transcription but, in some cases, it can also promote it (Zhang et al., 2006). In the meantime, exon/intron methylation occurs mainly in CG context and its function remains unclear. Gene body methylation has been related to transcriptional upregulation and has been suggested to protect genes from aberrant transcription caused by cryptic promoters (Feng et al., 2016; Zhang et al., 2006). The local acetylation of histones located near the promoter region of genes can induce transcriptional activation (Finnegan, 2001). Although there are no reports of specific histone modifications related to HM stresses in plants, some studies in animals have revealed a direct relation between HM exposition and histone modifications (Cheng et al., 2012).

Gene expression changes generated by HM exposure in rice have been described extensively in the literature and linked to variations in DNA methylation levels. For instance, Oono et al., (2016) showed a positive correlation between Cd dose-response in plants and the expression of genes coding for metal ion transporters where DNA methylation marks were detected. Similarly, using whole-genome bisulfite sequencing (WGBS), Feng et al., (2016) evaluated DNA methylation changes induced by specific Cd stress in rice plants (*Oryza sativa* ssp. japonica cv. Nipponbare). The authors found specific differentially methylated regions after Cd treatment, with patterns of methylation closely associated with transcriptional differences of stress response genes involved in metal transport, metabolic processes, and transcriptional regulation. Likewise, some studies have shown the heritability and stability of HM stress-induced methylation changes (Ou et al., 2012; Rahavi, 2011). For instance, in *A. thaliana*, improved tolerance to HMs has been observed in the progeny under the same stress experienced by parental plants (Ou et al., 2012). More recently, Cong et al., (2019) showed that specific methylation changes induced by HM stress, specifically methylation changes at the Tos17 retrotransposon,

displayed transgenerational inheritance through three generations. Therefore, the evidence suggests that epigenetic mechanisms contribute to HM stress adaptation through successive plant generations.

2.5 Epigenetic Mechanism Involved in Aluminum Toxicity

Al exposure can trigger DNA damage and cell death through a strong binding of Al ions to pectins and other structural components of the cell wall (Murali Achary & Panda, 2010). Although there are currently few studies that have explored the relationship between epigenetic regulation and Al tolerance (Table 2.2), current evidence suggests that Al tolerance might be conferred through DNA methylation as specific methylation changes frequently occur after Al exposure. For example, Bednarek et al., (2017) subjected five Al-tolerant and five non-tolerant triticale lines to Al exposure. Using methylation-sensitive amplification polymorphisms (MSAP), the study showed that Al exposition in both Al tolerant and non-tolerant plants induced demethylation.

These findings are consistent with other reports that describe the effects of HMs on methylation patterns (Aina et al., 2004; Feng et al., 2016; Ou et al., 2012). However, the opposite pattern has also been reported; for example, by using coupled restriction enzyme digestion and random amplification (CRED-RA) in corn (*Zea mays* cv. RX9292), Taspinar et al., (2018) established that exposure to Al induced mobilization of long terminal repeat retrotransposons (LTR) and triggered DNA hypermethylation as a protective response to the stress condition. Complementarily, Agnieszka, (2018) compared liquid chromatography (RP-HPLC), MSAP analysis and methylation amplified fragment length polymorphisms (metAFLP) to detect DNA methylation levels of triticale lines showing contrasting tolerance to Al treatments. After Al exposure, a reduction in DNA methylation across nontolerant lines was identified with the RP-HPLC technique, in contrast, increased methylation was seen in tolerant plants; this outcome was independent of the Al dose. When MSAP was used, increased demethylation was found in the roots of both non-tolerant and tolerant lines, with no differences between them. Finally, metAFLP results demonstrated no differences in DNA methylation under stress conditions, suggesting that only a portion of the genome responds to Al stress.

Hossein Pour et al., (2019) used CRED_RA in three wheat cultivars (*Triticum aestivum* cv. aymana79, Kılçıksız, and Bezostaja1) to evaluate genetic and epigenetic variations to different Al conditions (7.5 and 30mM). DNA hypermethylation was observed in wheat plants at higher Al concentration (30 mM) and hypomethylation at lower Al concentration (7.5 mM). These results suggest a gradual effect of Al on methylation, with concomitant cellular damages associated with increased Al toxicity. A methylation increase along the genome was concluded to confer a protective response in

the affected plants. Thus, the existing evidence points to a complex influence of DNA methylation on the response to Al-induced stress in a species-dependent manner.

Table 2.2. Summary of epigenetic studies related to Al stress responses in plants.

Plant	Variety	Epigenetic modification	Method	Reference
Nicotiana tabaccum	*Xan-thi nc*	DNA Methylation	HPLC, Direct bisulfite sequencing.	Choi & Sano, 2007
Sorghum bicolor	*inbred lines, YN336 and YN267*	DNA Methylation	MSAP	Kimatu et al., 2011
Zea Mays	*Kenyan tropical maize (KTM)*	DNA Methylation	MSAP	Kimatu et al., 2013
Arabidopsis thaliana	*Col-0 ecotype*	DNA Methylation, Histone modifications	Chromatin Immuno-Precipitation (ChIP), direct bisulfite sequencing.	Ezaki et al., 2016
Triticale inbred lines		DNA Methylation	MSAP	Bednarek et al., 2017
Zea mays	*cultivar RX9292*	DNA Methylation	CRED–RA	Taspinar et al., 2018
Triticale inbred lines		DNA Methylation	metAFLP, MSAP, HPLC	Agnieszka, 2018
Triticum aestivum	*Haymana 79, Kılçıksız, and Bezostaja 1.*	DNA Methylation	CRED-iPBS	Hossein Pour et al., 2019

2.6 DNA Methylation as a Regulatory Factor in Plant Responses to Aluminum Stress: Rice as a Study Case

Epigenetics has the potential to explain mechanistically, at least part of the molecular responses to different abiotic stresses, including HM toxicity (Figure 2.1). Although there are no studies related to the epigenetic regulation of Al tolerance in rice, it is hypothesized that epigenetic mechanisms, like DNA methylation, could play an important role as a regulatory factor in this response. Potentially, several of the genes mentioned in this chapter might be regulated through differential patterns of DNA methylation. As a first

approach on this dissertation to test this assumption, a brief analysis was conducted to quantify the methylation status of specific Al responsive genes in three different rice varieties (IR64, Nipponbare, and Pokkali) with contrasting responses to Al exposure.

For this exploratory evaluation, publicly available data from (Stroud et al., 2013) was analyzed for the Nipponbare cultivar (highly tolerant to Al toxicity) and from Garg et al. (2015) for IR64 and Pokkali varieties (both susceptible to Al toxicity). These published data are from plants grown under control conditions as the aim is to identify fixed epigenetic marks in pre-existing rice varieties before being subjected to stress conditions. To explore the possible role of methylated cytosines over gene expression, in a set of 250 genes associated with Al tolerance in rice (Arbelaez et al., 2017; Arenhart et al., 2014), the number of methylated cytosines (mCs) was calculated considering the different methylation contexts (counting was performed 1000 bps before and after the transcription initiation site). According to the reported experimental data, these 250 genes showed significant changes in expression after Al exposure (upregulated genes Log2FC ≤ 1, downregulated genes Log2FC ≥ 1) (Supplementary Table 2.1). As a result, a group of 72 genes was kept, representing 10% of genes with the highest counts for methylated cytosines (Supplementary Table 2.2). Additionally, to increase the probability that the effects over gene expression were caused by an epigenetic regulation solely, genes with copy number differences or SNP variations in the coding region were filtered out from this list, retaining for the analysis only single-copy genes identified from the rice genes paralogous list generated by Lin et al., (2008) and without SNPs variants identified from the database Rice SNP Seek database (Mansueto et al., 2017, https://snp-seek.irri.org/). After filtering by gene duplication and SNPs variants, 26 candidate genes were retained (Supplementary Figure 2.1 and Supplementary Table 2.3).

Among the three analyzed varieties, considering the different methylation contexts, and the localization of the methylated cytosines, Nipponbare exhibited more methylated sites than the other two varieties ($p \leq 0.01$ in an FDR analysis), while IR64 and Pokkali did not show differences in methylation (Figure 2.2). These results are interesting since Nipponbare has been extensively reported as a cultivar highly tolerant to Al (Famoso et al., 2010). These exploratory analyses show indications of the role of methylation in the response of plants to aluminum stress conditions. However, it is possible that this result reflects a higher level of genome-wide methylation of Nipponbare in general relative to the other two varieties analyzed.

At the top of the list, representing highly methylated genes (Table 2.3), some genes previously reported as important players in rice Al tolerance were found. For example, the Calmodulin binding protein (*Loc_Os09g13890*) is a calcium ion-binding molecule that regulates different cellular processes, and recently, the association of the Calmodulin signal transduction pathway to Al stress has been reported (Zhang et al., 2016). This study

showed that transgenic *Saccharomyces cerevisiae* strains transformed with the Calmodulin gene were more tolerant to Al toxicity, suggesting that the gene is a good candidate for improving Al tolerance in plants through transgenic approaches. Similarly, the analyses also showed the proteins *STAR1* (*Loc_Os06g48060*) and *ART1* (*Loc_Os12g07280*) as relevant in Al-related methylation. *STAR1* encodes a nucleotide-binding domain that associates with *STAR2*, which encodes a transmembrane domain, to form a bacterial-type ABC transporter required for Al detoxification in roots (Table 3.1; Huang et al., 2009). On the other hand, the ART1 zinc finger protein is a transcription factor that regulates around 31 genes, probably involved in Al detoxification at different cellular levels, including *STAR1* and *STAR2* genes (Yamaji et al., 2009). The results suggest that the methylation status of reported Al response genes, could play a role in Nipponbare's Al tolerance.

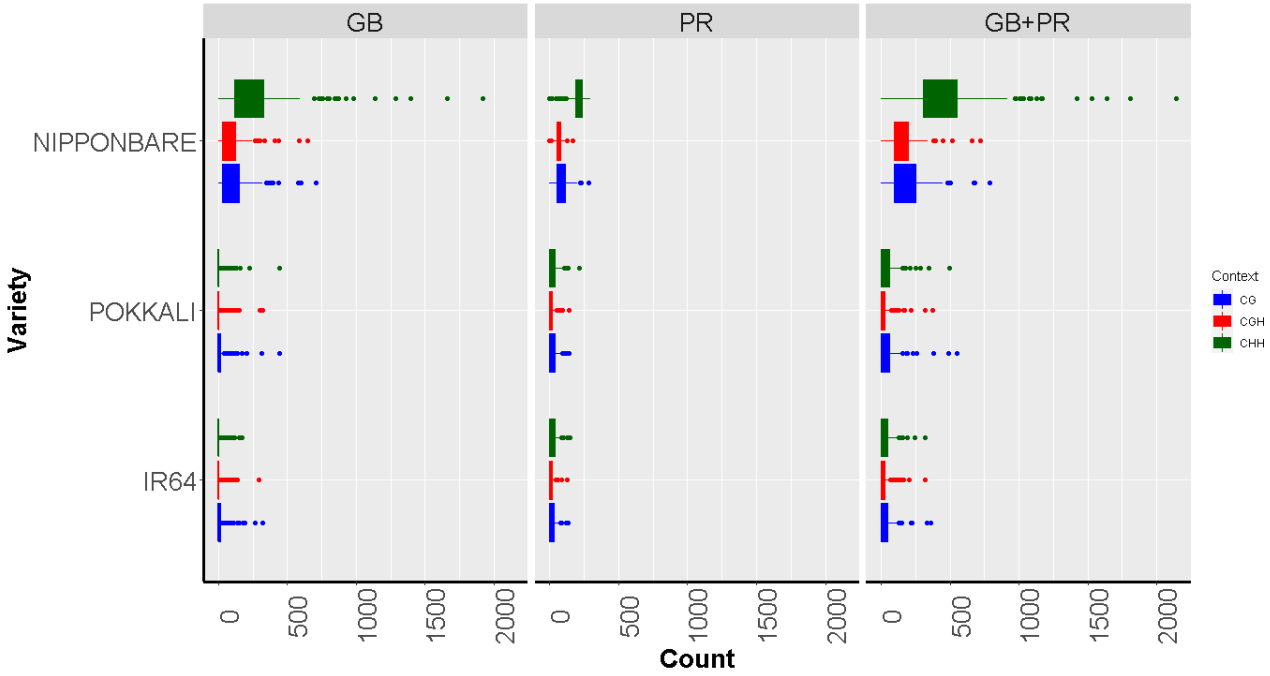


Figure 2.2. Boxplots showing methylated cytosine counts in three sequence contexts: CG (blue), CHG (red), and CHH (green) for three different rice varieties with contrast responses to aluminum exposure: Nipponbare (Tolerant), Pokkali, and IR64 (Susceptible). The results are discriminated according to the location of the epigenetic. mark, either inside the gene body region (GB), the promoter (PR), or both the promoter and inside the gene body region of analyzed genes (PR + GB).

Concluding Remarks

Current knowledge of HM and Al tolerance in plants has been extensively documented with a direct focus on the physiological, and biochemical effects of these molecules, and

their negative impacts on crop production. In rice, there is abundant information about genes and QTLs involved in Al tolerance in comparison with other staple cultivars such as Barley or even the model plant *A. thaliana*. Nevertheless, recently, epigenetic mechanisms have emerged as important factors in the response of plants to HM stresses. Based on existing knowledge, two epigenetic strategies for coping with Al stress in plants were identified: (i) epigenetic marks are used as a mechanism to protect plants from possible DNA damage caused by metal ions through random DNA methylation along the genome, and (ii) epigenetic changes are used for the regulation of transposon and stress-responsive genes (Figure 2.1).

The studies conducted so far provide evidence of the potential impact of epigenetics in plant response to HMs. However, there remains a significant knowledge gap regarding the epigenetic mechanisms involved in plant response to Al stress. Addressing this knowledge gap is the main focus of this doctoral dissertation, which aims to evaluate the role of DNA methylation in rice plants response to Al stress through various approaches. The experiments and analyses conducted in the subsequent chapters of this dissertation represent pioneering research in this field, offering compelling evidence for considering epigenetics as a crucial factor in regulating the response to heavy metal stress in rice.

Table 2.3. Top 10 of genes with the highest methylated cytosines count for three *O. sativa* varieties with different aluminum tolerance levels.

Gene (MSU ID)	Annotation	IR64	Nipponbare	Pokkali
Loc_Os12g32850	Cytochrome P450 71E1, putative	202	949	273
Loc_Os09g13890	Calmodulin binding protein, putative, expressed	202	1075	159
Loc_Os12g42860	Cysteine dioxygenase	161	937	219
Loc_Os03g11950	CRAL/TRIO domain containing protein, expressed	137	1059	156
Loc_Os06g48060	Protein STAR1	130	1155	175
Loc_Os05g51470	2-aminoethanethiol dioxygenase, putative, expressed	115	1053	143
Loc_Os12g07280	Zinc finger protein ART1	109	1024	99
Loc_Os12g06660	Actin-7, putative, expressed	99	990	121
Loc_Os04g33640	Glycosyl hydrolases family 17, putative, expressed	83	1357	94
Loc_Os09g37510	DUF292 domain containing protein, expressed	69	941	82

Chapter 3

Whole-Genome DNA Methylation Patterns of *Oryza sativa* and *Oryza glumaepatula* Associated with Aluminum Tolerance Under Control Conditions

This chapter was previously published as:

Gallo-Franco, J.J., Ghneim-Herrera, T., Tobar-Tosse, F., Romero, M., Chaura, J. & Quimbaya, M.A. (2022) Whole-genome DNA methylation patterns of *Oryza sativa* (L.) and *Oryza glumaepatula* (Steud) genotypes associated with aluminum response. Plant Direct: e430.

Chapter Summary

Epigenetic mechanisms in crops have emerged as a fundamental factor in plant adaptation and acclimation to biotic and abiotic stresses. Among described epigenetic mechanisms, DNA methylation has been defined as the most studied epigenetic modification involved in several developmental processes. In this chapter, single-base resolution methylome maps were analyzed for *Oryza sativa* and *Oryza glumaepatula* genotypes grown under control conditions. The results showed that overall, genome-wide methylation profile is mainly conserved between both species, nevertheless, there are several differentially methylated regions with species-specific methylation patterns. In addition, the association of identified DNA methylation marks in relation with Al-tolerance levels of studied genotypes was analyzed. Several differentially methylated regions (DMRs) and differentially methylated genes (DMGs) that are linked with Al tolerance were found. Some of these DMGs have been previously reported as differentially expressed under Al exposure in *O. sativa*. Complementarily a Transposable Elements (TEs) analysis revealed that specific aluminum related genes have associated-TEs potentially regulated by DNA methylation. Interestingly, the DMRs and DMGs between Al-tolerant and susceptible genotypes were different between *O. sativa* and *O. glumaepatula*, suggesting that methylation patterns related to Al responses are unique for each rice species.

Keywords: *Abiotic stress, aluminum, bisulfite sequencing, epigenetic, heavy metals, methylome, rice*

Connection to Previous Chapter

Chapter 2 provided an in-depth description showing evidence of epigenetics as a possible regulator of the response to Al stress in several economically important crops. In addition, an exploratory analysis of DNA methylation patterns in rice was also performed, which showed strong differences in DNA methylation between Al tolerant and susceptible genotypes, based on bisulfite sequencing data from the literature. To confirm the exploratory results from the previous chapter through new experimental and analytical evidence, in this chapter single-base resolution methylome maps were analyzed for genotypes of *Oryza sativa*, a cultivated species, and *Oryza glumaepatula*, a wild species, with contrasting levels of Al tolerance that were grown under control conditions. The findings provide new insights into the pre-established DNA methylation patterns in tolerant and susceptible rice genotypes under control conditions, and how they may be contributing to the regulation of rice plant responses to Al stress.

3.1 Introduction

Rice is an important crop as it represents the primary food source for over half of the world population. Similarly, it has been established as a biological model for monocots molecular and evolutionary research. As it was described in the previous chapter, numerous studies have shown differential methylation patterns that are related to specific developmental stages in rice (Liu et al., 2017; Stroud et al., 2013; Xing et al., 2015), and with contrasting responses to stress conditions (Feng et al., 2016; Hu et al., 2015; Ou et al., 2012; Schmidt et al., 2018). However, most of these analyses have been restricted to elite *Oryza sativa* (L.) (cultivated rice) crops, analyzing few stress conditions, mainly salinity and drought (Ferreira et al., 2019; Garg et al., 2015; Rajkumar et al., 2019; Wang et al., 2011; Zheng et al., 2013, 2017). Additionally, it has been shown that DNA methylation differences might explain gene expression variations between cultivated and wild species (Li et al., 2012). Nevertheless, the epigenomic divergence between wild and cultivated rice and how the model crop differs from its wild relative has been poorly studied.

The aluminum (Al) toxicity is one of the most limiting factors for plant growth and crop yield on acid soils (Rascio & Navari-Izzo, 2011; Zheng, 2010) and is particularly severe in the tropical and subtropical regions that represent most of the worldwide arable land affected by soil acidity. Among staple food crops, rice is the most aluminum (Al) tolerant cereal and there is genetic variation for Al-tolerance at the genus/species level, with several genotypes exhibiting highly contrasting tolerance responses (Famoso et al., 2010). Remarkably, greater genetic variability and higher Al-tolerance levels have been reported for wild rice species (Cao et al., 2011; Mao, 2003). The identification of mechanisms that are associated with Al responses have been studied from a genomic and transcriptomic perspective. However, to date, there are no studies on DNA methylation influence over Al-tolerance in rice, although there is some evidence pointing towards epigenetics as a regulatory factor in Al stress responses in other plant species as described in chapter 2.

To address the lack of information regarding epigenomic variation between cultivated and wild rice species, single-base resolution methylome maps were generated for four genotypes of the cultivated rice species *O. sativa*, and two genotypes of the wild specie *O. glumaepatula* (Steud) grown under control conditions. The results showed that the genome-wide methylation profiles are conserved between both species, but there are several differentially methylated regions (DMRs) with species-specific methylation patterns. Likewise, overall methylome profiles were not found to be shaped by the response to Al toxicity. However, there exist several DMRs between Al-tolerant and susceptible species, some of which correlate with differentially expressed genes under Al-stress conditions previously reported in the literature (Arbelaez et al., 2017; Arenhart

et al., 2014). The findings shed light on the methylation patterns associated with cultivated and wild rice species that will serve as a reference scaffold for future studies of rice genetics and epigenetics.

3.2 Materials and Methods

3.2.1 Plant Material and Genomic DNA extraction

This chapter presents the evaluation of the global methylation patterns for the *O. sativa* genotypes: Azucena (AZU), Nipponbare (NIP), IR64 and BGI9311 (BGI), and for the *O. glumaepatula* genotypes: OG131 and OG97 (Supplementary table 3.1). Seeds used in this study were given by the plant physiology laboratory from ICESI University, Cali, Colombia. Tolerance levels for *Oryza sativa* genotypes were defined according to literature data where AZU and NIP are the most tolerant varieties followed by IR64 which is considered intermediate tolerance and BGI has the lowest tolerance index and is considered a susceptible genotype (Famoso et al., 2010). These authors used the Relative Root growth (RRG) of the total root system to define the tolerance level of the different varieties. On the other hand, the plant physiology laboratory at Icesi University previously characterized the tolerance levels of *Oryza glumaepatula* genotypes using the Total Root System Length Index. Based on this characterization, OG131 was classified as a susceptible genotype, whereas OG97 was identified as tolerant. After sterilization and break of dormancy, seeds were subjected to dark conditions at 30°C for 4 days. Later, seeds were grown in a culture room at 30°C and 12:12 dark/light conditions for 10 days. Seedlings were transferred to a hydroponic medium with a Kimura B solution (pH 7) and Arnon micronutrients. Roots from three weeks-old seedlings were collected and stored at -80°C. Total genomic DNA was extracted from frozen root tissue by CTAB 2X protocol with modifications (Maropola et al., 2015). Genomic DNA quality was evaluated on agarose gels and DNA quantity was measured using a Nanodrop spectrophotometer (Thermo Scientific).

3.2.2 Whole-Genome Bisulfite Sequencing (WGBS)

Bisulfite-seq (BS-seq) libraries were made from genomic DNA isolated from *O. sativa* and *O. glumaepatula* seedlings roots. DNA from three independent seedlings was pooled as one sample, and two samples were sequenced per genotype. DNA was first fragmented by sonication from 100 to 300 base pairs (bp) in size, followed by end-blunting, dA addition at the 3' end, and ligation of adapters. Next, adaptor-ligated molecules of 200 to 300 bp were isolated by agarose gel electrophoresis and subjected to a treatment of sodium bisulfite conversion using the ZYMO EZ DNA Methylation-Gold kit

(ZYMO Research Corporation, Irvine, California, USA). Finally, the Polymerase chain reaction (PCR) enriched libraries were purified and subjected to high-throughput sequencing with Illumina HiSeq X ten platform to achieve an approximately 30X sequencing depth for each sample. BS-seq was performed by CD Genomics (CD Genomics Inc., Shirley, New York, USA).

3.2.3 Data Filtering, Read Alignment, and General Statistics

The FastQC tool was used to perform basic statistics on the quality of the raw reads. Then, sequencing adapters and low-quality data of the sequencing data were removed by Trimmomatic (version 0.36). Pre-processed reads were aligned to the Os-Nipponbare-Reference-IRGSP-1.0 genome downloaded from de Rice annotation project database (RAP-DB) (https://rapdb.dna.affrc.go.jp/download/irgsp1.html) using Bismark v.0.16.3 (Krueger & Andrews, 2011) and the Bowtie2 v.2.2.8 tool (Langmead & Salzberg, 2012) with default parameters. The same reference genome was used for both species since one of the objectives of this chapter is to make direct comparisons between the two species. Only the uniquely aligned reads were maintained and all the samples were de-duplicated using the Bismark deduplication module. The methylation calling data obtained from Bismark was used for further analysis. The methylation level of each cytosine was defined as the proportion of reads displaying mCs among all the reads covering the same cytosine position.

3.2.4 Analysis of DNA Methylation Landscape Between Rice Species

A comparative analysis among the rice genotypes methylome was made using the methylation level per cytosine position throughout the complete genome. The methylation level was also calculated for each sequence context (CG, CHG and CHH) inside gene and TE-body as well as their 2Kb upstream and 2Kb downstream regions. Each region was divided into 20 bins of equal size and the average methylation level for each bin was calculated and plotted. In addition, the number of unmethylated (uM), gene body methylated only in CG context (gbM), and TE-like methylated (teM) genes were determined using the method reported by (Kawakatsu et al., 2016).

3.2.5 Methylation Patterns in Relation to Genes and TEs

The rice genome was divided into 300Kb windows and for each one of them (i) the number of genes, Gypsy TEs and Mite TEs were computed; and (ii) the average methylation level per sequence context for each species. The Gypsy (Class I) and Mite TEs (Class II) were selected because they are the most common and diverse TEs families in rice (Song & Cao, 2017). The correlation among the characteristics computed per window was compare using a PCC. For the two families of TEs, Gypsy and Mite, the distance

between the TE and the closest gene was calculated. Based on this information, TEs were classified into two categories: close TEs as those located less than 2 Kbp to a gene, and distant TEs as those located more than 2Kb away from a gene. Finally, it was defined whether TEs were located inside the centromere region or in the chromosome arms. The coordinates used to define the area of influence of the centromere in each chromosome was 2Mb on each side of the window with the highest frequency of CentO (AA) sequences (Lee et al., 2005).

3.2.6 Identification of Differentially Methylated Regions (DMRs) Between Rice Species

A comparison between methylation level of the cytosines was performed using the Pearson Correlations Coefficients (PCC). Thus, the methylome of each species is compared to the others. The hierarchical clustering between all the comparisons was made using 1-PCC distance metric. Likewise, differentially methylated regions (DMRs) between rice species were identified using the tiling window approach with a window's size of 200bp and step size of 100 in the software Methylkit (v.1.16.1.). Only the methylated cytosines covered by ≥10 reads and windows with at least five cytosines each were considered for the analysis. A linear regression model was used to determine statistical significance (q-value) followed by a Sliding Linear Model (SLIM) correction. A window with methylation difference of 75%, 50% and 15% for CG, CHG and CHH respectively, and a q-value ≤ 0.01 as compared with the reference samples was considered as a DMR. Neighboring DMRs with a gap less than 100bp were merged. Pairwise comparisons were made between all the *O. sativa* genotypes against the *O. glumaepatula* genotypes to determine DMRs (*O. sativa* genotype was considered as the reference group). Finally, overlapping DMRs for all the pairwise comparisons were selected. The location of DMRs in the genome was defined with at least 1 bp overlapping between the DMR and a functional feature (Genebody, upstream -2Kb, downstream +2Kb and TEs) (Sun et al., 2019; Wang et al., 2018). Genes overlapping with DMRs in the functional or promoter region were defined as DMR-associated genes (DMGs).

Pairwise comparisons were also made between Al-tolerant and susceptible genotypes for *O. sativa* (NIP-BGI, AZU-BGI) and *O. glumaepatula* (OG97-OG131) using the same procedure described above. The susceptible genotype was considered as the reference group. Overlapping DMRs for the two comparisons made inside *O. sativa* were selected for further analyses.

DMR-associated genes identified among Al-tolerant and susceptible genotypes for *O. sati*va and *O. glumaepatula* were compared with a set of 250 genes previously associated to Al stress-response in rice (Arbelaez et al., 2017; Arenhart et al., 2014). According to the

reported experimental data, the 250 genes showed significant changes in expression after Al exposure (upregulated genes Log2FC≥1, downregulated genes Log2FC≤-1) (Arbelaez et al., 2017; Arenhart et al., 2014). This comparison allowed us to explore the association between expression and methylation patterns of Al responsive genes.

3.3 Results

3.3.1 Genome-Wide DNA Methylation Patterns in Cultivated and Wild Rice Genotypes

In this chapter, the methylome of four cultivated and two wild rice genotypes with contrasting responses to Al stress were generated and characterized, grown under control conditions. In this way, it was possible to analyze the methylation profiles associated with each of these species, and their role as a mechanism associated with Al tolerance. In total, 47 - 60 million and 55 - 63 million high-quality reads were obtained for *O. sativa* and *O. glumaepatula*, respectively, whose mapping statistics are presented in Supplementary Table 3.1. The genome coverage for all samples ranged between 33-44X, with an average depth per base between 15-19X. The cleaned reads for all the samples were aligned to the reference genome: Os-Nipponbare-Reference-IRGSP-1.0 from the Rice Annotation Project Database (RAP-DB). Only those sequences that mapped uniquely were considered, therefore, duplicated sequences were removed from all samples. The percentage of unique aligned reads ranged between 49 - 65% for *O. sativa* and 42 - 43% for *O. glumaepatula*. The bisulfite conversion rate for all the libraries was above 99.5%. The depth and quality of the sequencing were enough to ensure a high-quality genome-wide methylation analysis in all the samples.

The percentage of mCs varied from 10 to14% in *O. sativa* and from 8 to 15% in *O. glumaepatula* samples. The CG context had the highest methylation level for both rice species, i.e., number of reads showing mCs for all reads covering the same cytosine site, followed by CHG and CHH contexts. Because DNA methylation has a known role in TEs silencing and gene regulation, the methylation levels for genes and TEs was analyzed. This analysis was performed inside genes and TEs bodies, and their 2kb upstream and downstream regions (Figure 3.1A). The general methylation patterns for these genomic features were similar for both species, but the methylation levels in the OG131 genotype of *O. glumaepatula* tended to be lower with respect to the other samples. Consistent with other plant species, for genes and TEs, the methylation level was higher in the CG context, followed by CHG and CHH contexts (Niederhuth et al., 2016). Furthermore, TEs had a higher methylation level compared with genes for all sequence contexts. In addition, the methylation levels variation for all the genes and TEs was calculated (Figure 3.1B), as well

as the number of genebody methylated genes (gbM), TE-like methylated genes (teM), and unmethylated genes (uM) for all the genotypes (Figure 3.1C). Overall, no clear differences were observed between species for methylation patterns inside these genomics features.

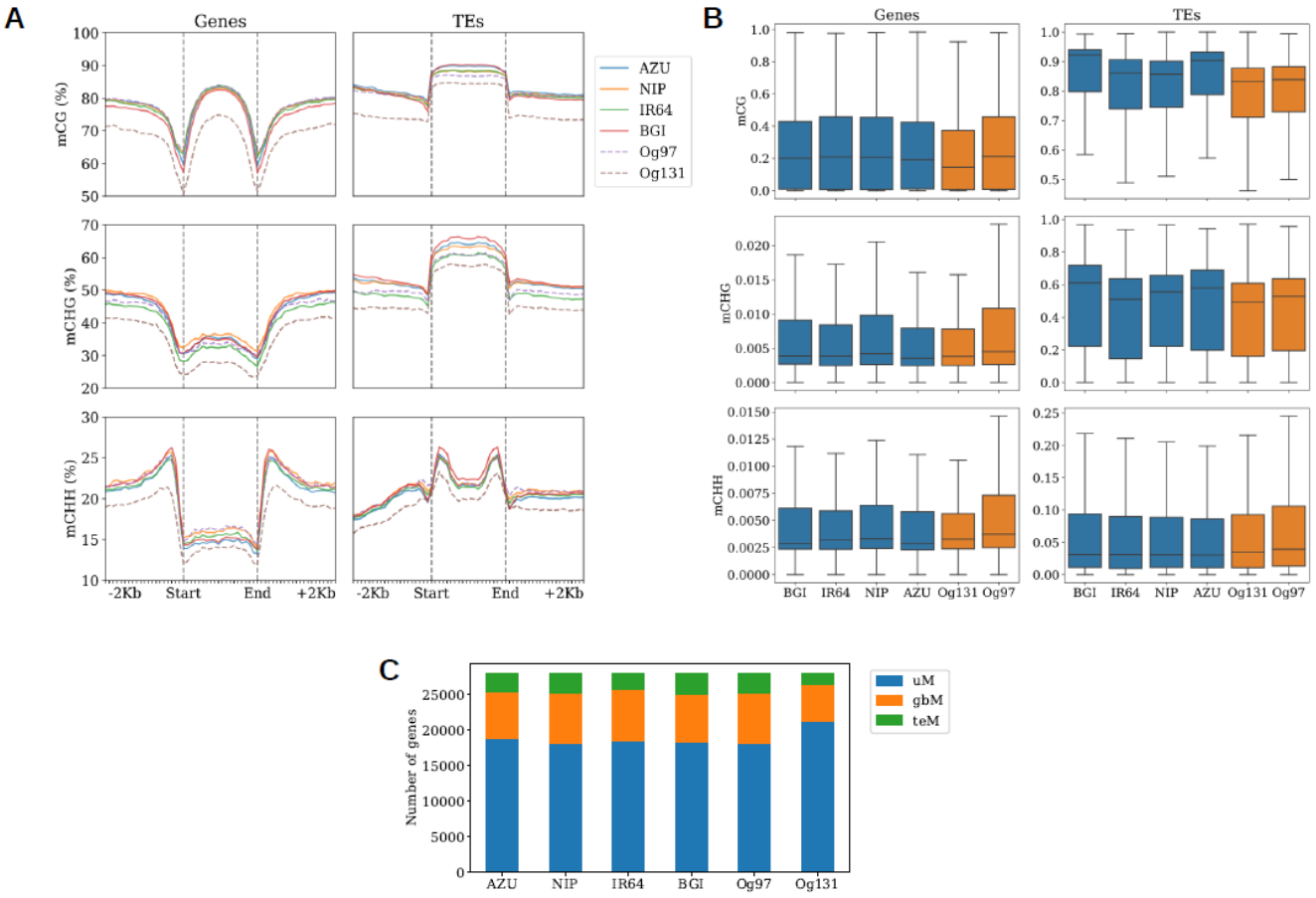


Figure 3.1. DNA methylation levels in *Oryza sativa* and *Oryza glumaepatula* genotypes. **A.** Average methylation levels for genes and TEs bodies, upstream (-2Kb) and downstream region (+2Kb). Each region was divided into 20 bins and the average methylation level was calculated for each bin. **B.** Boxplot showing methylation levels variation located in genes and TEs, for all the sequence context in rice genotypes. Blue bars: *O. sativa*; Orange bars: *O. glumaepatula* **C).** Number of unmethylated genes (uM), genebody methylated genes (gbM), and TE-like methylated genes (teM) in rice genotypes.

3.3.2 DNA Methylation Profiles of TEs in Rice

The rice genome was divided into windows of 300 Kb and for each one of them (i) the number of genes, Gypsy TEs, and Mite TEs; and (ii) the average methylation level per

sequence context for each species were computed. To understand the relationship between the DNA methylation patterns and the distribution of genes and TEs throughout the genome, the PCC among the computed characteristics per window was compare (Figure 3.2A). This approach showed that DNA methylation for CG and CHG context for both species correlated positively with Gypsy TEs density (PCC: 0.69 - 0.72, Figure 3.2A), which means that windows with a higher methylation level had a greater number of Gypsy TEs. In contrast, the methylation level correlated negatively with the number of genes and Mite TEs per window.

A

	Genes	Mite	Gypsy	Os_CG	Og_CG	Os_CHG	Og_CHG	Os_CHH	Og_CHH
Genes	1.00	0.51	-0.59	-0.85	-0.82	-0.81	-0.76	0.54	-0.16
Mite	0.51	1.00	-0.49	-0.60	-0.57	-0.64	-0.60	0.56	0.07
Gypsy	-0.59	-0.49	1.00	0.69	0.67	0.72	0.68	-0.40	0.21
Os_CG	-0.85	-0.60	0.69	1.00	0.95	0.92	0.85	-0.63	0.13
Og_CG	-0.82	-0.57	0.67	0.95	1.00	0.88	0.90	-0.59	0.20
Os_CHG	-0.81	-0.64	0.72	0.92	0.88	1.00	0.93	-0.51	0.26
Og_CHG	-0.76	-0.60	0.68	0.85	0.90	0.93	1.00	-0.46	0.34
Os_CHH	0.54	0.56	-0.40	-0.63	-0.59	-0.51	-0.46	1.00	0.38
Og_CHH	-0.16	0.07	0.21	0.13	0.20	0.26	0.34	0.38	1.00

B

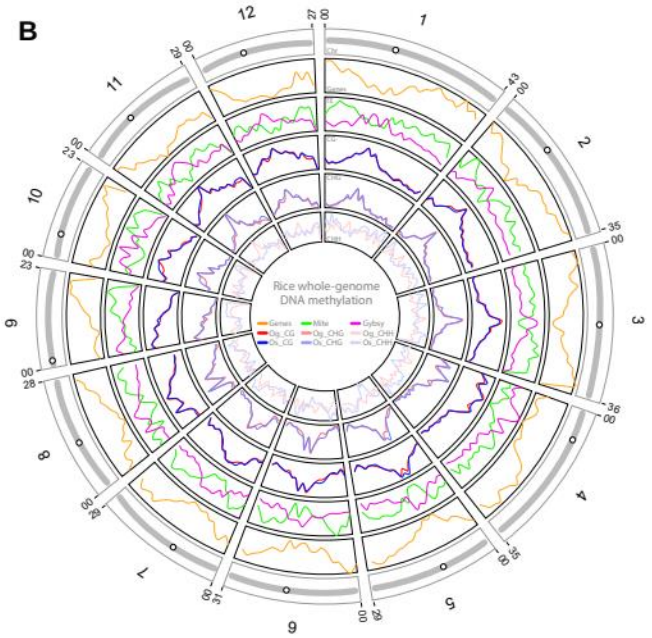


Figure 3.2. A. Pearson Correlation Coefficients (PCC) calculated comparing the methylation levels for *Oryza sativa* and *Oryza glumaepatula*, associated with the number of genes, Mite TEs, and Gypsy TEs along the genome. **B.** Circos plot showing in a genome-wide context, genes, Mite TEs and Gypsy TEs distribution in association with average methylation levels in the CG, CHG and CHH context for *O. sativa* and *O. glumaepatula*.

The calculated PCC between methylation level and genes density ranged from -0.76 to -0.85, while for methylation level and Mite density it varied from -0.57 to -0.64 (Figure 3.2 A, B). It is worth noting that the CHH context showed a different trend, mainly for *O. sativa* species. Thus, CHH-methylation showed a positive correlation with respect to genes and Mite densities, and a negative correlation with respect to the methylation in CG and CHG context, as well as Gypsy density (Figure 3.2A y B, Supplementary figure 3.1). Even though the rice methylation levels along the genome are differentially related to the Gypsy and Mite TEs density, there is not a clear association between the TEs methylation levels and their impact on the gene structure or regulation. In consequence, TE methylation

levels was explored with respect to genes distance, classifying them as close (<2Kb from the gene) and distant (>2Kb from the gene) elements.

Our results showed no significant correlations between TEs methylation levels and the distance to the closest gene, but the methylation patterns were clearly different among CG, CHG, and CHH contexts, especially for Gypsy TEs as shown in Figure 3.3A. In fact, for the CG context, the higher distances to a gene correlate with higher methylation levels, but it is an opposite trend in the CHH context, where higher distances to a gene correspond to lower methylation levels. Likewise, for Gypsy TEs there is a clear difference regarding methylation levels for close and distant TEs in all methylation contexts. For Mite TEs similar methylation levels were observed between these two categories (Figure 3.3B). In addition, the methylation levels distribution of TEs located in the centromeres or chromosome arms was explored, but clear differences between them were not found (Supplementary figure 3.2).

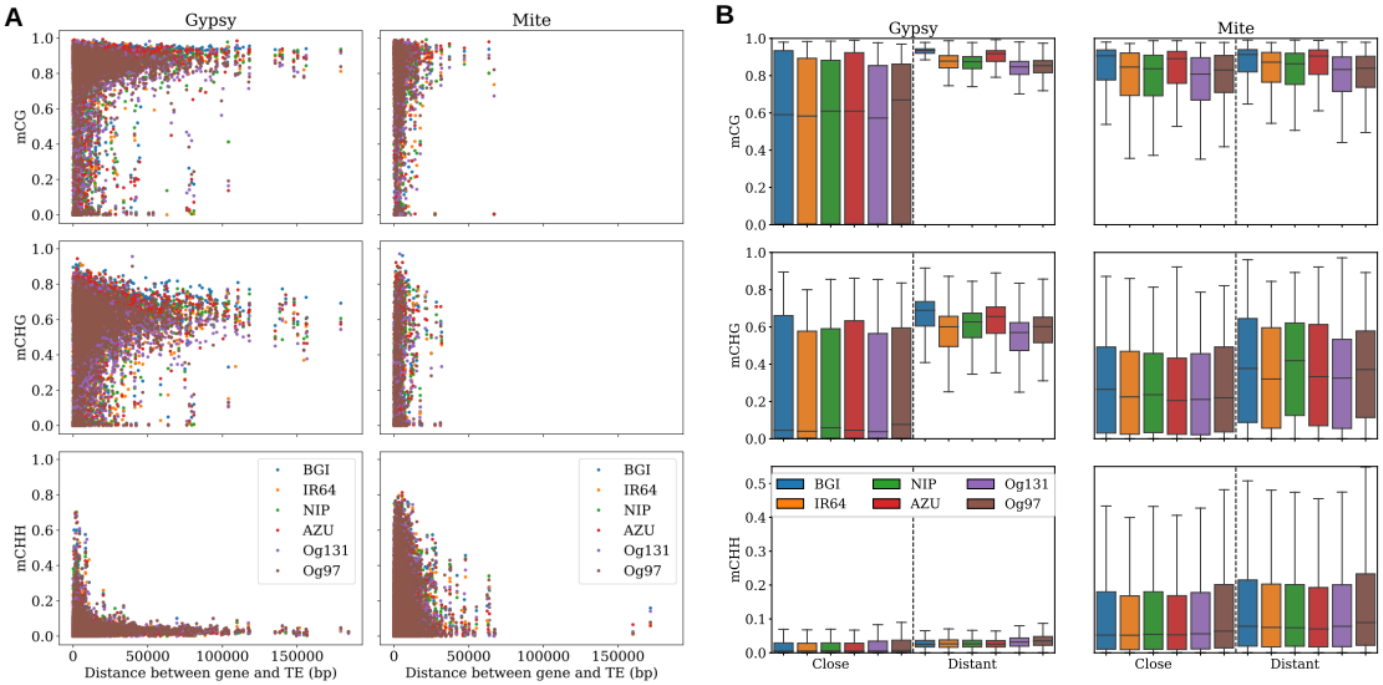


Figure 3.3. A. Gypsy and Mite TEs methylation levels in relation to the distance to the nearest gene for all the sequence contexts. B. Boxplot showing methylation level variation between Gypsy and Mite TEs close and distant from genes for all the sequence contexts.

3.3.3 Differential Methylation Patterns Between Cultivated and Wild Rice Genotypes.

Although methylation patterns were similar between rice species, a detailed comparison between their methylomes was carried out by performing a pairwise comparison between species per each sequence context. Here, the Pearson Correlation Coefficient (PCC) was used to contrast cytosine methylation levels. For this analysis, only the cytosines covered in all samples were considered. The hierarchical clustering of all samples using 1-PCC distance, showed clustering by species but not grouping by subspecies inside *O. sativa* for all sequence contexts (Supplementary figure 3.3). In fact, the most Al-tolerant (AZU) and the most Al-susceptible (BGI) genotypes were clustered together. Additionally, the comparisons considering only the cytosines inside genes and TEs for all the sequence contexts were repeated (Supplementary figure 3.4, Figure 3.4A). The results showed the same clustering profile as the one obtained using whole genome methylation data.

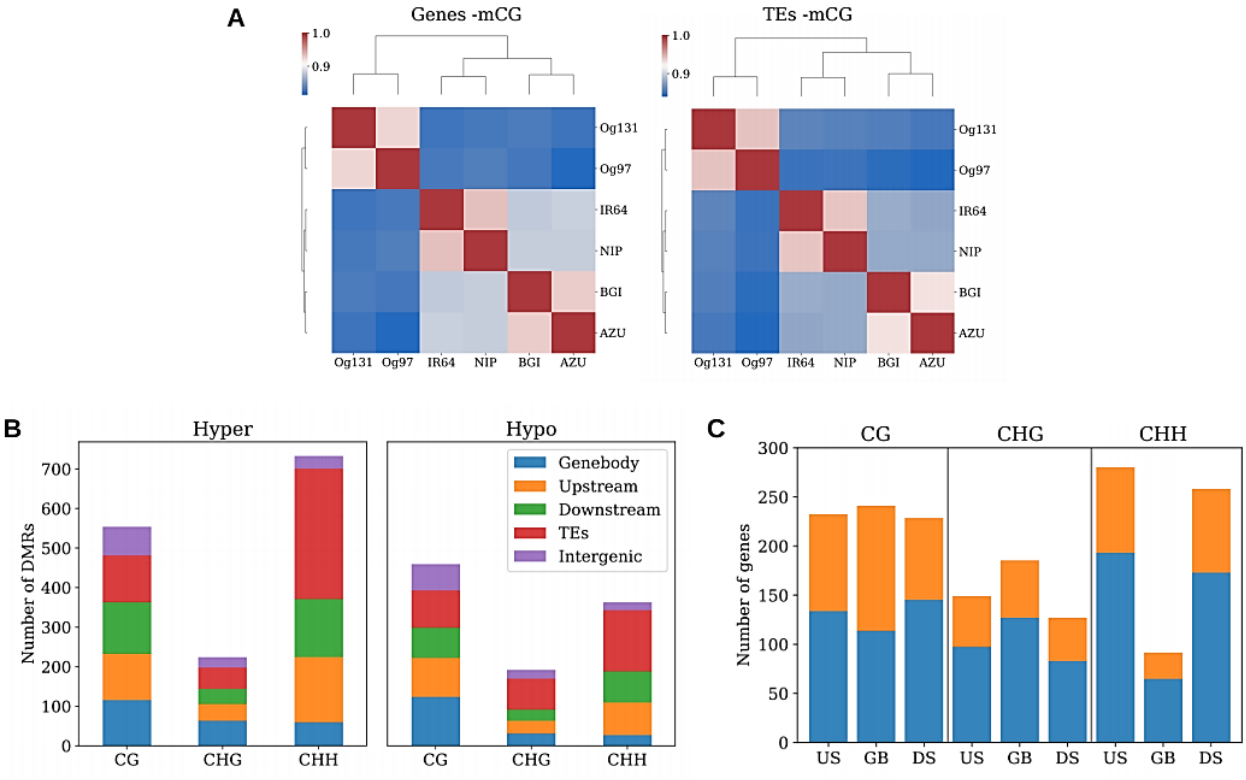


Figure 3.4. Differential methylation between wild and cultivated rice species. **A.** Clustering analysis of rice genotypes according to the methylation levels of mCs inside genes and TEs for CG context **B.** Number and location of hyper and hypomethylated DMRs between *Oryza sativa* and *Oryza glumaepatula* varieties for each sequence context. **C.** Number of DMR-associated genes found between *O. sativa* and *O. glumaepatula* species. Genes were considered as DMR-associated

genes if DMRs were located at gene-body (GB), upstream (US) (-2Kb), or downstream (DS) (+2Kbps) regions.

Since the rice species are separated according to the methylome correlation analysis (Figure 3.4A), it was decided to explore which regions differ between the analyzed species. Therefore, differentially methylated regions (DMRs) were calculated by a tiling window approach (200bp window size) between *O. sativa* and *O. glumaepatula*. *O. sativa* genotypes were used as reference for all the pairwise comparisons. In total, 1601 DMRs were found of which 60% were hypermethylated and 40% were hypomethylated in *O. glumaepatula* with respect to *O. sativa* (Figure 3.4B). Notably, hypermethylation was mainly found for the CHH context with most DMRs located in TEs, while hypomethylation was mainly for the CG context with most DMRs located in the genebody region. Additionally, for the identification of differentially methylated genes (DMGs), the genes overlapping with at least 1 bp of distance to a DMR on its annotated genebody, upstream (-2Kb) or downstream regions (+2 Kb) were tagged. A total of 1324 DMGs were reported, of which, 799 genes were hypermethylated and 547 were hypomethylated (Figure 3.4C).

3.3.4 Differential Methylation Patterns Associated with Al-Tolerance: A Comparison Between Wild and Cultivated Rice.

It is well known that the methylome could be related to stress tolerance phenotypes in plants, so it was decided to evaluate whether there are specific DMRs in relation to established Al tolerance in different rice genotypes. DMRs between Al-tolerant and susceptible genotypes were calculated within each rice species, using the susceptible genotypes as control for all the pairwise comparisons. For *O. sativa* the overlapping DMRs between NIP-BGI and AZU-BGI pairwise comparisons were selected for further analysis. The results showed 4633 DMRs for *O. sativa* of which 38% were hypermethylated and 62% were hypomethylated, with most hypomethylated DMRs in the CHH context (mainly inside TEs) (Figure 3.5A). Complementarily, 8048 DMRs were found for *O. glumaepatula* of which 72% were hypermethylated and 38% were hypomethylated (Figure 3.5B). It is worth noting that most DMRs were hypermethylated at the CHH context contrary to the DMRs found for *O. sativa*. Lastly, 3024 DMGs were identified for *O. sativa* and 5484 DMGs for *O. glumaepatula* (Figure 3.5C). These DMGs are considered as potentially regulated by DNA methylation patterns between Al-tolerant and susceptible genotypes within each rice species. Besides, obtained DMRs between species were compared to evaluate whether there are similarities in DNA methylation patterns associated with Al-tolerance. Here, 91 shared DMRs were found between these two species, 63% hypermethylated and 37% hypomethylated of which 58 were associated with 83 genes.

To gain more insights into the methylation profiles associated with Al tolerance in rice, the methylation patterns of 250 genes that were previously reported as differentially expressed under Al-exposure were analyzed (upregulated genes Log2FC ≥ 1 and downregulated genes Log2FC ≤ -1), reported by Arbelaez et al. (2017) and Arenhart et al. (2014). Here 21 DMGs were reported for *O. sativa* (Supplementary table 3.2) and 37 DMGs for *O. glumaepatula* (Supplementary table 3.3) that were associated with Al-responsive genes of which 11 were shared between both species (Table 3.1). Additionally, for each genotype, the average methylation level for the upstream, genebody, and downstream regions of these 250 genes were calculated (Supplementary figure 3.5). Then, the average methylation level between genotypes were compared. The results showed a similar clustering among genotypes for all sequence contexts as previously obtained using whole-genome methylation data (Figure 3.1D), where *O. sativa* and *O. glumaepatula* belong to different clusters, but inside *O. sativa* the most Al-tolerant (AZU) and the most Al-susceptible genotype (BGI) are grouped together (Supplementary figure 3.5). These results suggest that there are no differential methylation patterns for all these Al-responsive genes between Al-tolerant and susceptible species, grown under control conditions. But a few genes are being epigenetically regulated in relation to the Al-stress response.

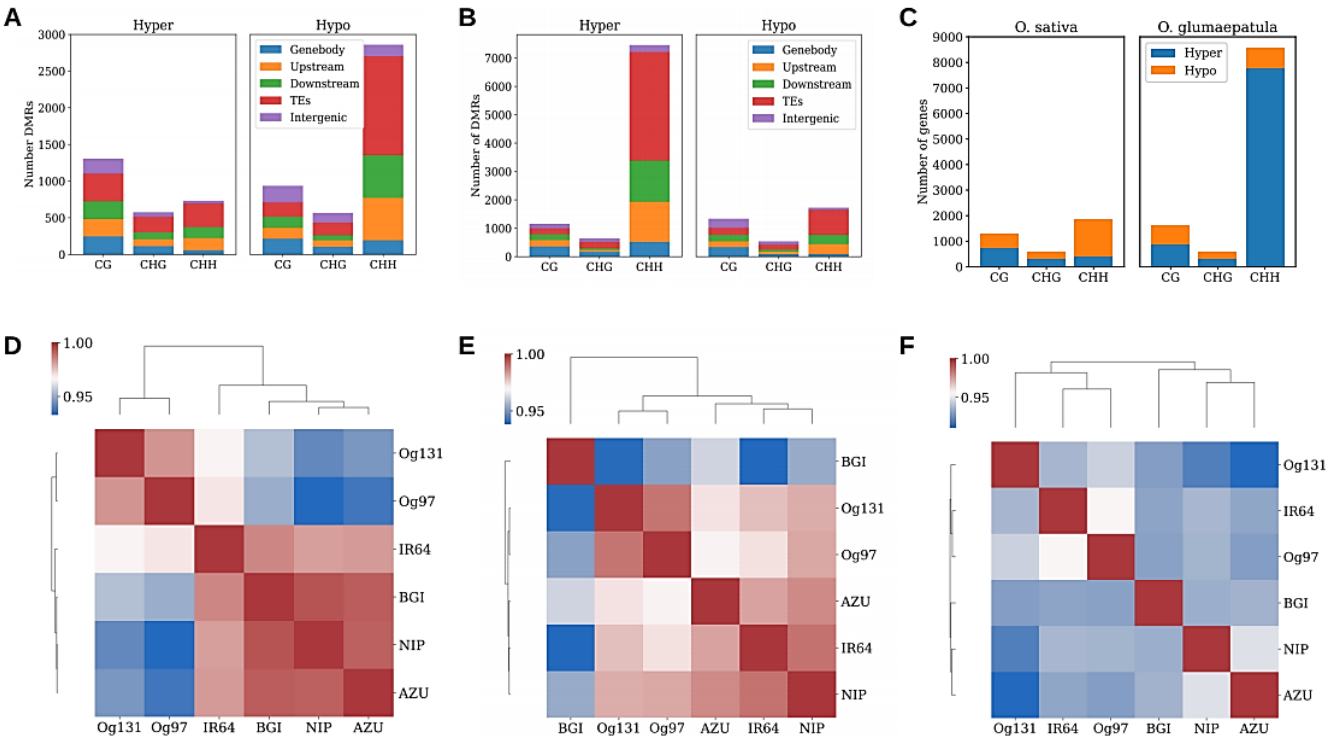


Figure 3.5. DNA methylation patterns associated to aluminum tolerance in wild and cultivated rice species. **A.** Hypo and hyper DMRs detected between Al- tolerant and susceptible genotypes for *Oryza sativa*. **B.** Hypo and hyper DMRs detected between Al- tolerant and susceptible genotypes for *Oryza glumaepatula*. **C.** Number of DMR-associated genes in *O. sativa* and *O. glumaepatula*

genotypes. **D).** Clustering analysis of rice genotypes built by using the average methylation levels of TEs close to Al-responsive genes for CG context. **E.** Clustering analysis of rice genotypes built by using the average methylation levels of TEs close to Al-responsive genes for CHG context. **F.** Clustering analysis of rice genotypes built by using the average methylation levels of TEs close to Al-responsive genes for CHH context.

Finally, the methylation status of 127 TEs close to Al-responsive genes was analyzed. 79 of these TEs were associated with a DMR in *O. glumaepatula* and 70 were associated with a DMR in *O. sativa*. In addition, implementing a PCC analysis, the methylation levels of TEs in proximity to Al-responsive genes were compared between varieties. Genotypes were clustered according to Al-tolerance level, at least, for the CG methylation context (Figure 3.5D, E and F). This result is especially interesting within *O. sativa* species where the two most tolerant genotypes (AZU and NIP) were grouped together while IR64 (intermediate genotype) and BGI (susceptible genotype) were clustered apart. These results suggest a possible role of TEs in the epigenetic regulation of Al-genes.

Table 3.1. DMR-associated genes found in both O. sativa and O. glumaepatula between aluminum-susceptible and tolerant genotypes. These DMR-associated genes were reported as differentially expressed under aluminum stress conditions by Arbelaez et al., 2017 and Arenhart et al., 2014.

	O. sativa			***O. glumaepatula***		
Gene ID	**Location**	**Status**	**Context**	**Location**	**Status**	**Context**
Os01g0949900	Upstream	Hypo	CHH	Upstream	Hyper	CHH
Os01g0639600	Upstream	Hypo	CHH	Upstream	Hyper	CHH
Os12g0210500	Upstream	Hypo	CHH	Genebody	Hypo	CHH
	Genebody	Hypo	CHH			
Os02g0186800	Upstream	Hypo	CHH	Upstream	Hyper	CHH
Os05g0472400	Upstream	Hypo	CHH	Genebody	Hypo	CHH
	Genebody	Hypo	CHH	Downstream	Hyper	CHH
Os08g0158200	Genebody	Hyper	CHG	Genebody	Hyper	CHG
Os07g0509800	Genebody	Hyper	CHH	Genebody	Hyper	CHH
Os01g0609300	Genebody	Hypo	CG	Genebody	Hypo	CG

				Upstream	Hyper	CHH
Os11g0134900	Genebody	Hypo	CHG	Genebody	Hypo	CHG
Os01g0597800	Downstream	Hyper	CG	Upstream	Hyper	CHH
Os10g0459300	Downstream	Hypo	CHH	Downstream	Hyper	CHH

3.4 Discussion

DNA methylation is the most studied epigenetic variation, having an important role in the regulation of many genomic features such as chromatin structure, transposon silencing, gene expression regulation and recombination. From this perspective, DNA methylation patterns are considered an evolutionary result of life histories in plant species (Niederhuth et al., 2016). However, most epigenetic variation studies have focused on model species.

In this chapter, DNA methylation patterns between *O. sativa,* a rice cultivated species, and *O. glumaepatula*, a South American endemic wild rice, were compared. Two genetically and ecologically different species (Stein et al., 2018). The results showed that, despite the genetic and ecological differences in the studied models, their general methylation profiles are very alike. Here, it is demonstrated that gene methylation showed the typical behavior reported previously for each context with a similar number of gbM, teM, and uM genes between species. Likewise, for all the methylation contexts, TEs showed a strong methylation level compared with genes. This result agrees with the important role of epigenetics in the TEs silencing, preventing their mobilization along the genome (Jones, 2012; Rabinowicz, 2003) and they are in accordance with the global genome-wide methylation patterns reported in several rice species and genotypes (Feng et al., 2016; Garg et al., 2015; Li et al., 2020; Li et al., 2012; Stroud et al., 2013; Zheng et al., 2017). In fact, previous studies have shown that general methylation trends are conserved in phylogenetically distant plant species (Niederhuth et al., 2016; Zemach et al., 2010).

It has been reported that the methylation status of TEs can regulate nearby genes expression, affecting plant responses to environmental conditions. In this regard, a novel finding from this study is that methylation levels throughout *O. sativa* and *O. glumaepatula* genomes were positively correlated with Gypsy TEs density. Conversely, these levels were negatively correlated with genes and Mite TEs density. Previous studies have reported the relationship that exists between methylation level, genes and TEs densities (Bhatia et al., 2018; Rajkumar et al., 2019; Zhang et al., 2006). For instance, Bhatia et al., (2018) evaluated the methylation levels of different organs of chickpea plants and reported the overlapping of low methylation levels with high gene density regions, while high methylation levels positively correlate with high TEs density. These reports agree with the high methylation level found in the TEs body region, possibly associated with TEs silencing

through methylation mechanisms (Jones, 2012; Rabinowicz, 2003). However, the mentioned studies reported a generalized trend of high methylation levels for all TEs. In contrast, the results show that methylation patterns vary according to the TEs family and the DNA methylation sequence context.

Our findings showed that Gypsy TEs have lower average methylation levels when they are located close to a gene, whereas for Mite TEs this trend is not clear. It has been previously reported that methylation levels of TEs can affect the transcriptional activity of neighboring genes (Hollister & Gaut, 2009; Zhang et al., 2015). For instance, Choi & Purugganan, (2018) showed that Class I TEs near to genes have reduced methylation levels in comparison with TEs located far from genetic regions. Likewise, they showed that Gypsy TEs located far from genes heavily methylate as a defensive mechanism against TEs transposition. Therefore, the clear methylation level differences found for Gypsy TEs close and distant to genes, in this study, suggests a possible interaction between TEs and surrounding genes regulation. Regarding methylation levels for Mite TEs, there is not a clear difference between mobile elements located close or distant from genes. One of the possible hypotheses for this behavior is that Class II TEs (DNA) such as the MITE group of transposons do not have an independent transposition mechanism, so the silencing of these elements through DNA methylation is probably not as necessary as it is for Class I transposons. In fact, these types of TEs are preferentially inserted near or within genes and have been observed to contribute to allelic diversity and act as transcriptional regulators. (Dubin et al., 2018). Likewise, Class II TEs might have strong regulation effects on gene expression associated with methylation spread (Choi & Purugganan, 2018). Given that Mite family is one of the most diverse TEs families in rice, with almost four times more copy numbers in comparison with other sequenced plant genomes (Song & Cao, 2017; Wang et al., 2014; Zhang et al., 2014), it would be interesting to further study, from a functional perspective, the potential relation between Mite TEs distribution and transcriptional control.

Another result derived from TEs methylation patterns was the high methylation levels of Gypsy and Mite TEs close to genes in the CHH context. The genome-wide methylation levels for CHH context in several plant species, including this study, are lower than CG and CHG methylation (Niederhuth et al., 2016). However, higher CHH methylation levels of TEs close to genes were found. This behavior has been previously related to the presence of mCHH islands, which are defined as short and highly methylated regions, typically found upstream and downstream genes (Martin et al., 2021). In fact, mCHH islands were reported by the first time associated with Mite TEs in rice plants (Zemach et al., 2010). Likewise, Martin et al. (2021) showed a high level of genes associated with CHH islands (78%) in rice plants compared with other *Poaceae* species. Although the function of this CHH island is not clear, it has been proposed that they modify chromatin states, to prevent the spread of epigenetic modifications into genes or vice versa (Gent et al., 2013).

So far, the general patterns of methylation shared between wild and cultivated rice genotypes have been analyzed. Samples clustering using whole-genome methylation as well as specific genes and TEs methylation patterns allowed us to group samples to the species level. Specific genome regions that differ in their methylation status between *O. sativa* and *O. glumaepatula* were identified, representing species-specific epigenetic marks. Then, DMGs found in this study were contrasted with genes, previously reported as differentially methylated when comparing wild and cultivated rice species (Li et al., 2012). 11 common genes differentially methylated in both studies were found: (*Os08g0236800, Os03g0170200, Os04g0460900, Os03g0760000, Os06g0608401, Os06g0690600, Os05g0525800, Os06g0499900, Os09g0434600, Os11g0275500, Os04g0277400*). Likewise, two DMGs (*Os07g0669500* and *Os08g0424500*) that have been previously reported as key domestication-related genes in rice were identified (Chen et al., 2021; Kovach et al., 2007). *Os07g0669500* (FRIZZI PANICLE-FZP) is a gene that increased the number of secondary branch and grains per panicle (Chen et al. 2021) and *Os08g0424500* (SCENTED KERNEL – SK2) gene has been related to the grain fragrance (Kovach et al., 2007). Although more studies are needed, these results provide evidence for the role of DNA methylation during the domestication process in rice plants.

Interestingly, the wild rice *O. glumaepatula* showed a high number of hypermethylated TEs with respect to *O. sativa*, specifically in the CHH context. Even though epigenetic variations can arise spontaneously in different organisms (Becker et al., 2011; Schmitz et al., 2011; van der Graaf et al., 2015), genetic background and environmental conditions are considered the most important factors that structure epigenetic patterns (Kawakatsu et al., 2016). Given that *O. sativa* and *O. glumaepatula* are phylogenetically distant and have been subjected to their own evolutionary histories the different methylomes must reflect adaptation processes (Stein et al., 2018). Nevertheless, the analyzed *O. sativa* genotypes were not clustered according to the subspecies division. It is possible that methylome differences at the sub-species level could be strongly affected by the environmental conditions in which they have evolved, or by stable epialleles segregation, generated by spontaneous DNA methylation (Zhang et al., 2018). The segregating epimutations have been reported, in several cases, as factors that contribute to inheritable variation, independently of DNA sequence changes (Cortijo et al., 2014; Richards, 2006).

Considering DNA methylation level in different rice genotypes, the aluminotoxic conditions in contrasting rice varieties could be associated with differential methylation patterns. When analyzing contrasting genotypes in both, *O. sativa* and *O. glumaepatula*, several DMRs were reported between tolerant and susceptible varieties, suggesting a possible role of DNA methylation in Al-response regulation. For *O. sativa* mainly hypomethylated regions in the CHH context were found while hypermethylated ones were found in the CHH context for *O. glumaepatula.* Likewise, most of the found DMRs between Al-tolerant and susceptible varieties were unique for each rice species, suggesting that epigenetic regulation mechanisms related to Al-tolerance are different between *O. sativa*

and *O. glumaepatula.* These variations might be associated with different adaptation processes in contrasting environments, for example, *O. glumaepatula* being a wild species, endemic from South America, has been permanently subjected to acid soils guiding an adaptation process to Al toxicity. Likewise, epimutation accumulation in plants is also a source of heritable epigenetic and phenotypic diversity in plants such as the generation of epialleles related to stress tolerance (van der Graaf et al., 2015).

Our experimental strategy depicted 91 shared DMRs and 83 common DMGs between *O. sativa* and *O. glumaepatula*. Additionally, DMGs that have been reported as differentially expressed under Al-stress conditions were found (Arbelaez et al., 2017; Arenhart et al., 2014). (Supplementary Table S1 and S2). Notably, 11 of these genes were differentially methylated between Al-tolerant and susceptible genotypes for both species (Table 1). These convergent features represent promising genetic elements that might structure a core machinery in the epigenetic regulation of Al-tolerance. Further work must elucidate their functional role in Al-tolerance by evaluating the transcriptional regulation of identified DMGs under stress conditions, helping to confirm the causal relationship between methylation and transcriptional repression or activation. The results suggest epigenetic mechanisms of aluminum tolerance based on a set of limited core genes rather than generalized regulation mechanisms exerting phenotypic effects through thousand genes.

Concluding Remarks

According to these results the methylation patterns associated with genes and TEs for both rice species are conserved. Interestingly, a positive correlation between the methylation level and the density of Gypsy TEs were found. But a negative correlation was found between the methylation level and the density of Mite TEs and genes. Likewise, for Gypsy TEs a clear difference between methylation levels according to their distance to the closest gene were found, while no differences for Mite TEs were observed. However, there exist several genomic regions with species-specific methylation patterns, reflecting the own evolutionary histories of *O. sativa* and *O. glumaepatula*. Complementarily, several regions potentially regulated through epigenetics that are related to Al-tolerance in rice were identified. Different DMRs and DMGs in *O. sativa* and *O. glumaepatula* were reported independently. Few of these reported DMGs between Al-tolerant and susceptible genotypes for both rice species were shared between them, suggesting different mechanism of Al-response between cultivated and wild rice species, nevertheless, these convergent features, represent promising genetic elements that might structure a core machinery in the epigenetic regulation of Al-tolerance. These findings represent a first approach in the understanding of the differential methylation patterns between wild and cultivated genotypes and suggest a participation of DNA methylation in the regulation of Al tolerance in rice.

Chapter 4

Transcriptional Analysis in Wild and Cultivated Rice Genotypes Reveals Core Genes and Mechanisms Associated with Aluminum Tolerance

This chapter is submitted as a research article in the journal *Plant Science* as:

Gallo-Franco, J.J., Zuluaga-Yusti, I., Restrepo-García, A.M., Zapata-Balanta, S., Gutiérrez, J.P., Sosa, C.C., Ghneim-Herrera, T., and Quimbaya, M. Transcriptional analysis in wild and cultivated rice genotypes identifies core genes and mechanisms associated with aluminum tolerance.

Chapter Summary

Differential gene expression analysis under stress conditions is a widely used strategy to identify relevant genes and functional mechanisms that shape plant tolerance phenotypes to abiotic stresses. In this study, RNA-Seq experiments were conducted to analyze the transcriptional responses after 10 days of aluminum (Al) exposure in rice genotypes belonging to both cultivated (*Oryza sativa*) and wild (*Oryza glumaepatula*) rice species, which exhibit contrasting responses to Al stress conditions. After comparing transcriptional responses under control and Al treatment conditions, differentially expressed genes (DEGs) were identified for each genotype. In addition, functional analyses carried out on identified DEGs revealed the coordinated action of both exclusion and detoxification mechanisms associated with Al responses in rice. A group of 28 genes that were differentially expressed in all analyzed species and genotypes were reported, pinpointing a core set of mechanisms that are triggered by Al exposure. Similarly, 14 DEGs were identified only in Al-tolerant genotypes, representing potential key elements to understanding the successful adaptive response to toxic conditions adopted by tolerant genotypes. Given that many of these identified genes have not been previously reported or characterized, they could be considered novel regulators linked with Al tolerance responses, and promising candidates for the generation of agronomically improved rice varieties.

Keywords: *Abiotic stress, Aluminum tolerance, Differential expression, RNA-seq exeriments, Transcriptome, Wild rice, Cultivated rice, Al-detoxification mechanisms*

Connection to Previous Chapter

Chapter 3 identified specific DNA methylation patterns associated with Al stress tolerance levels rice varieties. However, the question remains as to how these epigenetic marks affect transcriptional regulation under conditions of Al toxicity. To answer this question, the variation in the transcriptome as well as in the methylome of tolerant and susceptible rice varieties under Al stress conditions must be contrasted. As a first step to resolve this question, this chapter analyzes the transcriptional variation of two tolerant and two contrasting rice genotypes after 10 days of exposure to Al conditions. The results of this chapter will be the starting point to evaluate the effect of methylation on gene regulation that will be developed in the following chapter.

4.1 Introduction

One of the most used approaches to understand the plant response to any specific stress is the analysis of differential transcription of genes under the stress conditions. It has been proposed the existence of an universally stress-regulated pathways base on a network of core stress-responsive genes in rice under both biotic and abiotic stress conditions (Cohen & Leach, 2019). However, several studies have also propose the existence of a stress-specific group of responsive genes in different abiotic stresses such as drought, heavy metals, salinity and cold (Baldoni et al., 2021; do Amaral et al., 2016; Shen et al., 2014).

As described in previous chapters, one of the most critical stresses affecting rice production in the tropics and subtropics is Al toxicity in acid soils. Several approaches have been explored to understand the mechanism underlying the response of rice to these stressful conditions. Previous research has described several QTLs and genes involved in Al tolerance in rice using massive sequencing techniques (Kochian et al., 2015). Likewise, some transcriptome studies have been conducted in rice to decipher the regulatory networks underlying Al stress response (Arbelaez et al., 2017; Arenhart et al., 2014; Tyagi et al., 2020; Zhang et al., 2019). However, most of the genes reported in those studies reflect responses to Al^{3+} toxicity rather than mechanisms of tolerance. This is especially true when comparing genotypes with different levels of tolerance, because the sensitive genotype may experience more severe stress than will the tolerant genotype (Delhaize et al., 2012). Therefore, genetic mechanism of Al tolerance in rice is considered a complex trait that requires further studies to reveal the different mechanisms at work in Al stress tolerance.

Rice is considered an Al tolerant plant compared to other economically important crops such as sorghum (*Sorghum bicolor*), maize (*Zea mays*) and wheat (*Triticum aestivum*), having a wide variation in tolerance levels among different genotypes. These characteristics turn rice into a very suitable crop to study transcriptional patterns in varieties with contrasting responses to Al stress. In addition, it is well known that wild relatives have the potential to introduce desirable traits into crops. From this perspective, *Oryza glumaepatula* which is a wild rice species native to South America and distributed in a wide range of environments, has a high capacity to tolerate different stress conditions (Brondani et al., 2005). Posso et al., (2013) identified *O. glumaepatula* genotypes with contrasting tolerance to Al stress condition, designating OG97 as a tolerant genotype and OG131 as a susceptible one. Despite the potential of wild rice varieties in uncovering novel mechanisms associated with Al tolerance, no transcriptomic studies involving wild rice species have been conducted so far.

In this chapter, transcriptional responses were evaluated for *O. sativa* and *O. glumaepatula* genotypes after a 10-day period of Al exposure, which represents an adaptation phase for tolerant genotypes. The analyses revealed several DEGs indicating a coordinated action of exclusion and detoxification mechanisms in Al responses of *O. sativa* and *O. glumaepatula*. A core set of 28 genes that were consistently differentially expressed across all genotypes and species was identified, suggesting their involvement in Al exposure responses. Additionally, 14 DEGs exclusive to Al-tolerant varieties were identified, which play crucial roles in understanding the adaptive strategies adopted by tolerant genotypes under toxic conditions. Many of these genes are novel and uncharacterized, presenting potential as key regulators directly associated with Al tolerance responses and serving as candidates for the development of improved rice varieties with enhanced agronomic traits.

4.2 Materials and Methods

4.2.1 Plant material and Al treatment

Four rice genotypes were selected for global transcriptomic analysis under Al stress conditions. For *Oryza sativa,* the Al-tolerant genotype Azucena (AZU) and the Al-susceptible genotype BGI9311 (BGI) were selected. For *Oryza glumaepatula,* the Al-tolerant genotype OG97 and the Al-susceptible OG131 were chosen (Supplementary table 4.1). The tolerance levels of *O. glumaepatula* genotypes were previously characterized in the plant physiology laboratory of ICESI University, Cali, Colombia. Seedlings were germinated on agarose medium with a 12-h photoperiod and temperature of 28°C, conditions that were kept for 7 days. After this period, seedlings were grown in a hydroponic medium with a *Kimura B* solution (pH 6) and Arnon micronutrients for 3 weeks. In order to include a plant adaptation period to low pH as a basal condition for all treatments, three days before the Al induction, solutions pH was progressively dropped to 4 for all plants. Subsequently, for the Al treatment, seedlings were subjected to Al stress conditions for 10 days in the *Kimura B* solution supplemented with Al_3Cl-100 uM while controls remained in the original hydroponic culture. Finally, roots were collected from treatment and control conditions and stored at -80°C. Total genomic RNA was extracted using the RNAzol protocol with modifications (Supplementary data1). For the extraction process, roots from five different plants were pooled and considered as a single biological replicate. Genomic RNA quality was assessed on agarose gels and confirmed via Bioanalyzer 2100. Paired-end (2x150pb) RNA sequencing was outsourced to *NovoGene* company using the *Illumina NovaSeq* platform (NovaSeq 6000) (https://www.novogene.com/).

4.2.2 Transcriptomic Analysis

After sequencing raw sequence quality was assessed using *FastQC* (Simon Andrews, 2010). Obtained reads were cleaned removing Illumina adapters via *Trimmomatic V.0.32* (Bolger et al., 2014). Cleaned sequences were aligned to the respective reference genomes with STAR V.2.5 (Dobin et al., 2013). For *O. sativa* the Os-Nipponbare-Reference-IRGSP-1.0 genome from de Rice annotation project database (RAP-DB) (https://rapdb.dna.affrc.go.jp/download/irgsp1.html) was used and for *O. glumaepatula* the reference genome Oryza_glumaepatula_v1.5, downloaded from Ensembl Plants Database was selected (http://plants.ensembl.org/Oryza_glumipatula/Info/Index?db=core). Read count was carried out using the software *featureCounts* v.2.0.3 (Liao et al., 2014) selecting the exon counting option and summarized by gene ID to finally obtain reads number per gene. Differentially expressed genes (DEGs) were identified using the *DESeq2* R package (Love et al., 2014) with default parameters including the Walt test for hypothesis testing and the Benjamini-Hochberg (BH) method for multiple testing correction. To retain only highly up- or down-regulated genes for further analyses, genes were considered differentially expressed if their adjusted p-values were < 0.05 and their absolute Log2 Foldchange (Log2FC) values were ≥ 1.5. All reported statistics in the present study represent the average values calculated from the three biological replicates tested per genotype. Finally, to identify shared DEGs between both rice species, the g:Orth tool from the g:Profiler web service (Raudvere et al., 2019) was used to identify orthologs between *O. glumaepatula* and *O. sativa*.

4.2.3 Functional Analysis

For the functional enrichment analyzes, the g:Profiler web service was used. For each of the analyzed genotypes, up- and down-regulated DEGs were considered independently. Functional enrichments were obtained for biological process, molecular function, and cellular compartment ontologies, and complemented with enriched KEGG metabolic pathways. As a multiple testing correction procedure, the g:SCS (Set Counts and Sizes) parameter was used (Reimand et al., 2007). Only GO terms or KEGG pathways with g:SCS values < 0.05 were considered as enriched features. In addition, the *O. sativa* subsp. *japonica* genome annotation was used as reference for AZU and BGI genotypes and *O. glumipatula* genome for OG97 and OG131 genotypes. The background gene list used for functional analysis was integrated for genes with expression data in at least two samples per genotype (available p-value). Finally, the tool REVIGO (Reduce and Visualize Gene Ontology) (Supek et al., 2011) was used to group the Gene Ontology terms with redundant GO levels. The aim of this tool is to find a single representative GO term for highly similar GO clusters. The default C-factor parameter (0.7) was used, which is considered a conservative value, although it maintains strong statistical support.

4.3 Results

4.3.1 Transcriptome Analysis

To analyze transcriptional responses in contrasting rice genotypes under Al stress conditions, high-throughput RNA-sequencing experiments were carried out. After raw data filtering, the number of total reads per treatment ranged from 20.6 to 22.2 million reads for both species. Similarly, the percentage of uniquely mapped reads to the selected reference genomes ranged from 67 to 83% while the reads assigned to genes varied from 61 to 86% (Supplementary Table 1).

Using the expression data, a PCA analysis was performed for all the samples within each analyzed species. Specific clusters were formed where control replicates grouped together as well as treatment replicates (Figures 4.1A and 4.1B). These results confirmed that sampled replicates behaved as expected, supporting the biological variability during experimentation, and thus, subsequent analyses can be considered reliable.

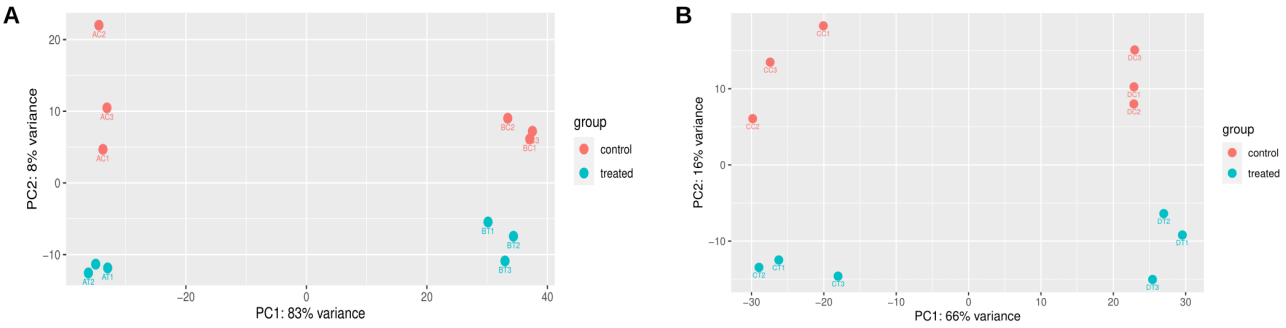


Figure 4.1. PCA analysis for all samples within each analyzed species. For all biological replicates, a principal component analysis (PCA) was performed using the normalized transcripts for each gene. Specific clusters were formed where control (C) replicates grouped together as well as treatment (T) replicates in **A)** *Oryza sativa*, AZU (AC and AT) and BGI (BC and BT) and **B)** *Oryza glumaepatula*, OG97 (CC and CT) and OG131 (DC and DT).

4.3.2 Differentially Expressed Genes (DEGs) Identified for All the Analyzed Genotypes

As a result of the differential expression analysis, Log2FC values were obtained for all the analyzed genotypes. An evaluation was made on the number of genes retained when different Log2FC thresholds (>0.5, 1, 1.5 and 2) were established (Supplementary figure 4.1). The number of DEGs resulting from contrasting the tolerant and susceptible

genotypes decreased when the Log2FC threshold stringency was increased, especially for *O. glumaepatula*. For further analysis, genes with absolute Log2FC values ≥ 1.5 in combination with an adjusted *p*-value < 0.05 were classified as DEGs. These values varied from -7.9 to 8.7 for AZU, from -7.3 to 7.4 for BGI, from -7.4 to 11.2 for OG97, and from -6.4 to 8.3 for OG131 (Figure 4.2A). Tolerant genotypes showed a broader Log2FC range than susceptible ones for both species.

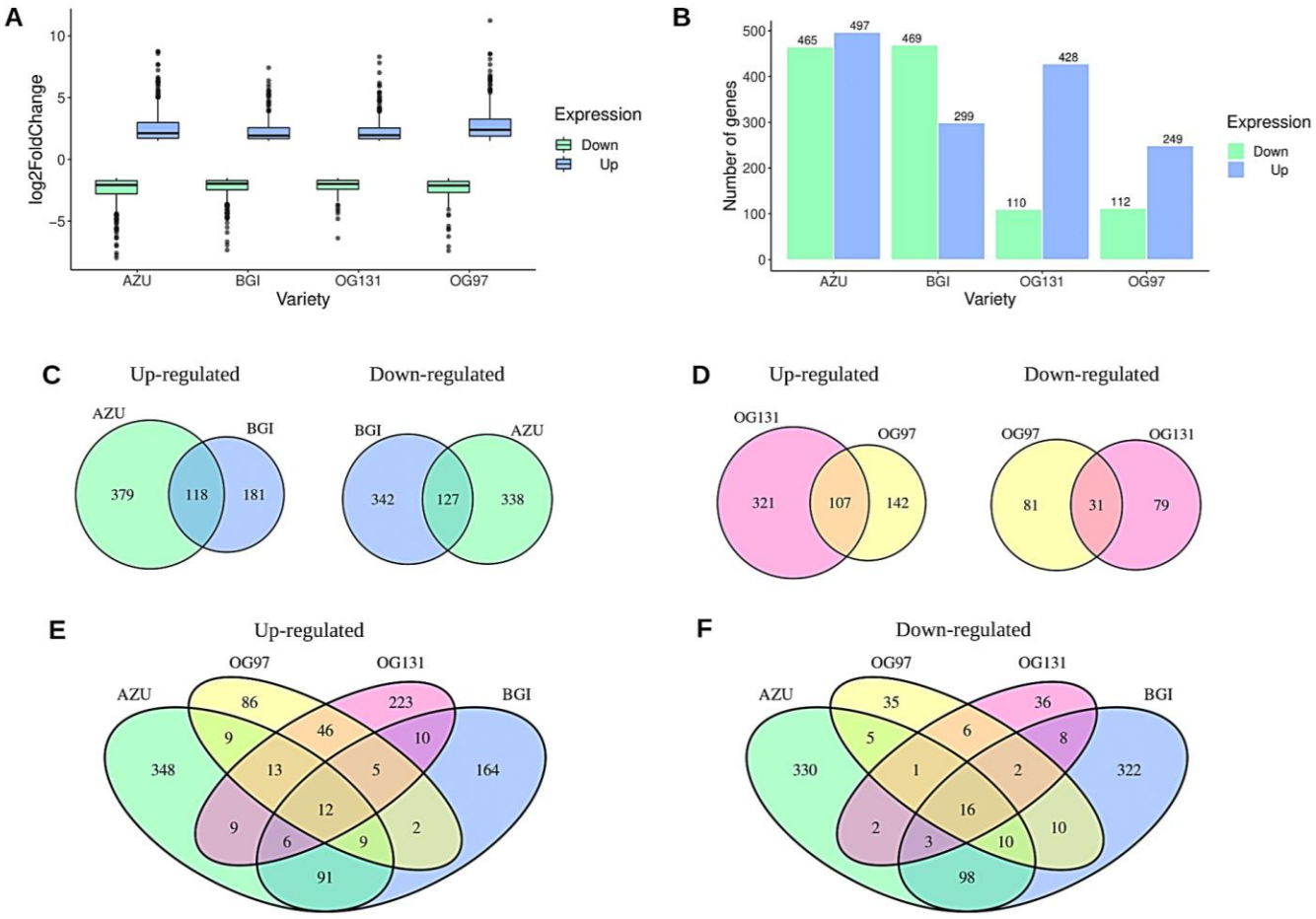


Figure 4.2. Differentially Expressed Genes (DEGs) identified for all the analyzed genotypes. **A)** Fold change patterns of differentially expressed genes (*padj*<0.05) when contrasting control vs Al-treated genotypes in *Oryza sativa* (AZU and BGI) and *Oryza glumaepatula* (OG131 and OG97) species. **B)** Number of DEGs (*padj*<0.05) for the analyzed genotypes. **C)** Venn diagrams for shared up-regulated and down-regulated genes between *O. sativa* genotypes. **D)** Venn diagrams for shared up-regulated and down-regulated genes between *O. glumaepatula* genotypes. **E)** Venn diagrams for shared up-regulated genes in all genotypes and **F)** Venn diagrams for shared down-regulated genes in all genotypes.

For the AZU genotype, a total of 962 DEGs were identified when contrasting control against Al-stress condition (497 up- and 465 down-regulated genes, supplementary table 4.2), 768 DEGs were detected for BGI (299 up- and 469 down-regulated genes, supplementary table 4.3), for the OG97 genotype, 361 DEGs were identified (249 up- and 112 down-regulated genes, supplementary table 4.4), and for the OG131 genotype, 538 DEGs were found (428 up- and 110 down-regulated genes, supplementary table 4.5). For AZU, OG97 and OG131 genotypes, a greater number of up-regulated genes was observed, in contrast to the BGI genotype where an increased number of down-regulated genes was detected (Figure 2B). In the wild-tolerant *Oryza glumaepatula* genotype (OG97), the lowest number of DEGs among all genotypes were identified.

4.3.3 Comparison of DEGs between Cultivated and Wild-rice Genotypes

Once DEGs were retrieved, different pairwise comparisons were established (Figures 4.2C-F). Accordingly, in *O. sativa*, 118 up-regulated genes were shared between AZU and BGI, and 127 down-regulated genes were common. In *O. glumaepatula*, 107 up-regulated genes were shared between OG97 and OG131, while 31 down-regulated genes were found to be in common. To establish proper comparisons between both rice species and based on information retrieved from the Ensembl Plants database, DEGs identifiers from *O. glumaepatula* were translated to their respective orthologs in *O. sativa*. It was possible to detect 74% orthologs between commercial and wild rice. After interspecific comparisons, 12 up- and 16 down-regulated genes shared among all the rice genotypes were identified (Figures 4.2E and 4.2F, Tables 4.1 and 4.2). Similarly, when contrasting only tolerant genotypes (AZU and OG97), 9 up- and 5 down-regulated genes were detected (Figures 4.2E and 4.2F, Table 4.3). Finally, for the 28 shared DEGs among all genotypes, a correlation analysis was performed implementing the Pearson Correlation Coefficient (PCC) calculated using Log2FC values. Similar correlation values were generated for all the genotypes with a slightly higher similarity between tolerant and susceptible genotypes inside species (0.95-PCC) than tolerant genotypes between both species (0.93-PCC) (Supplementary Figure 4.2).

Table 4.1. Up-regulated genes shared between *Oryza sativa* and *Oryza glumaepatula* genotypes

Gene ID	Gene name	Description	Gene Ontology associated terms (BP, CC, or MF)
Os04g0538900	*OsGLYI5*	Glyoxalase/bleomycin resistance protein/dioxygenase domain containing protein	
Os07g0214900	*OsCHS2*	Similar to Chalcone synthase C2 (EC 2.3.1.74) (Naringenin-chalcone synthase C2)	Biosynthetic process, Acyltransferase activity, Transferring groups other than amino-acyl groups
Os09g0435700		Hypothetical conserved gene	Hydrolase activity
Os01g0595600		Alpha/beta hydrolase fold-1 domain containing protein	Hydrolase activity
Os01g0343100		Protein of unknown function DUF594 family protein	Membrane
Os12g0614500		Similar to Lipase family protein	Lipid metabolic process, phospholipase activity
Os12g0614100		Similar to Lipase family protein	Lipid metabolic process, phospholipase activity
Os01g0106400	*OsIRL*	Similar to Isoflavone reductase homolog IRL (EC 1.3.1.-)	Oxidoreductase activity
Os09g0448200	*OsHAK17*	Similar to High-affinity potassium transporter	Potassium ion transport, Potassium ion transmembrane transporter activity, Integral component of membrane
Os05g0463000	*OsRLCK188*	Similar to Receptor protein kinase-like protein	Protein phosphorylation, ATP binding, Plasma membrane
Os12g0108500	*OsFbox636*	Similar to Leucine Rich Repeat family protein, expressed	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process,protein binding, SCF ubiquitin ligase complex
Os01g0919100	*OsFRDL4*	Aluminum-induced citrate transporter, Aluminum tolerance	Transmembrane transport, xenobiotic transmembrane transporter activity, Membrane

Table 4.2. Down-regulated genes shared between *O. sativa* and *O. glumaepatula* genotypes

Gene ID	Gene name	Description	Gene Ontology associated terms (BP, CC, or MF)
Os11g0249000		AP2/ERF family protein, Abiotic stress respons	Defense response, ADP binding
Os01g0609300	*OsPDR9*	F-box domain, cyclin-like domain containing protein	Transmembrane transport, ATP hydrolysis activity, integral component of membrane
Os10g0525500	*OsGSTU21*	Conserved hypothetical protein	Glutathione metabolic process, Glutathione transferase activity, Cytoplasm
Os09g0367700	*OsGSTU5*	Similar to GST6 protein (EC 2.5.1.18)	Glutathione metabolic process, Glutathione transferase activity, Cytoplasm
Os04g0556400		Similar to UDP-glycosyltransferase UGT93B9	Hexosyltransferase activity, Intracellular membrane-bounded organelle
Os07g0638400	*Os1-Cys*	Similar to 1-Cys peroxiredoxin	Hydrogen peroxide catabolic process, Antioxidant activity,nucleus
Os08g0153900		Similar to nodulin-like protein	Integral component of membrane
Os07g0241500	*OsUGT710C2*	Zinc finger, C2H2 domain containing protein	UDP-glycosyltransferase activity, Intracellular membrane-bounded organelle
Os11g0145200	*OsUGT75E1*	UDP-glucuronosyl/UDP-glucosyltransferase family protein	UDP-glycosyltransferase activity, Intracellular membrane-bounded organelle
Os09g0518200	*OsSGT*	UDP-glucuronosyl/UDP-glucosyltransferase family protein	UDP-glycosyltransferase activity, Intracellular membrane-bounded organelle
Os01g0557100		Alpha/beta hydrolase family protein	jasmonic acid metabolic process, methyl indole-3-acetate esterase activity
Os03g0838800		Similar to Legumain	Response to acidic pH, Nucleic acid binding
Os03g0226200	*HB2*	Similar to Polyketide reductase	Oxygen binding
Os02g0227500		Protein kinase, catalytic domain domain containing protein	Protein phosphorylation, ATP binding, Plasma membrane
Os03g0571900	*PEZ1*	Similar to transparent testa 12 protein	Transmembrane transport, Antiporter activity, Integral component of membrane
Os07g0625400	*OSK28*	Conserved hypothetical protein	Ubiquitin-dependent protein catabolic process, Cullin family protein binding, Cytoplasm

4.3.4 Comparative Functional Enrichment

To characterize the functional roles of identified DEGs in rice crops under Al-exposure, a functional enrichment analysis was performed using the *g:Profiler* tool (Supplementary Table 4.6). From this analysis, the unique and shared enriched functional categories were identified for all genotypes studied (Figure 4.3). Functional enrichment profiles were different between tolerant and susceptible genotypes, showing a higher number of unique categories than shared.

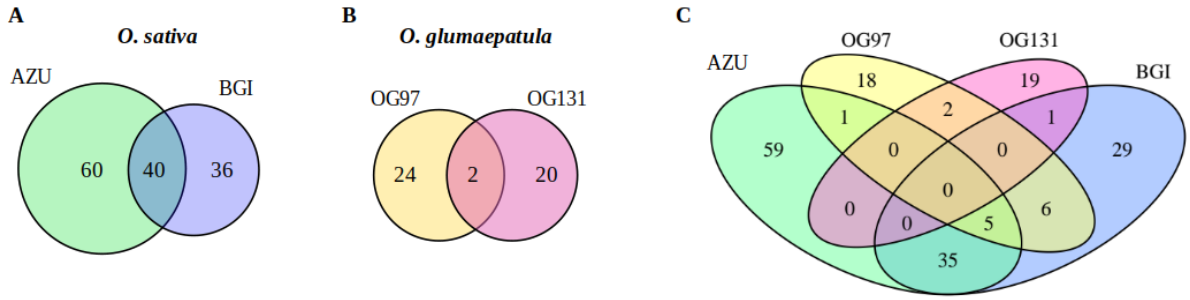


Figure 4.3. Comparative functional counts for analyzed genotypes. To characterize the functional roles of DEGs identified in rice crops under Al-exposure, a functional enrichment analysis using the *g:Profiler* was performed. **A)** Number of enriched categories shared between *Oryza sativa* genotypes (AZU and BGI). **B)** Number of enriched categories shared between *Oryza glumaepatula* genotypes (OG97 and OG131). **C)** Number of enriched categories shared among all analyzed genotypes for both rice species.

Interaction plots were generated using the REVIGO tool based on the biological processes (BP) associated with each genotype. This application removes redundant GO terms enabling the construction of clean and meaningful functional associations. For cultivated rice, the functional analyses revealed 19 BPs exclusively enriched for the AZU genotype, classified into eight functional subclusters, four derived from differential up-regulated genes and four subclusters obtained from differential down-regulated genes (Figures 4.4A-B). Remarkably, for BGI, unique categories derived from up-regulated genes were not identified given that all categories were shared with the AZU genotype. In contrast, 10 unique categories resulting from down-regulated genes were retrieved (Figure 4.4C). After comparing the functional categories shared by *O. sativa* genotypes, seven BPs were retrieved, mainly associated with the photosynthetic process, and associated with differentially down-regulated expressed genes (Figure 4.4D). Categories derived from shared up-regulated genes were not obtained.

In the case of *O. glumaepatula*, it was difficult to derive functional information, most likely due to its genome remaining poorly annotated. Nevertheless, four functional

clusters were identified for the OG97 genotype, three obtained from differential down-regulated genes and one subcluster obtained from differential up-regulated genes (Figures 4E and F). For the OG131 genotype, two exclusively enriched BPs were identified, namely the "transmembrane transport" and the "response to aluminum ion".

As previously mentioned, 12 up- and 16 down-regulated genes shared among all analyzed rice genotypes were identified, forming a cluster of 28 genes which structure a core machinery that is triggered upon Al induction. To address the role of these central genetic components, a functional characterization was performed. The categories from the Gene Ontology (GO) associated with "Catalytic activity", "Primary metabolic processes", "Transmembrane transport", and "Cellular anatomical entity" were derived from differentially up-regulated genes (Table 4.1). In the same way, categories linked with differentially down-regulated genes were associated with "Transmembrane transport", "Metabolic processes", "Catalytic activity", "Response to stimulus", and "Oxygen binding" (Table 4.2). Additionally, for the 14 DEGs shared between both tolerant genotypes but not differentially expressed in the susceptible genotypes, categories were recovered according to their functional characterization that related to "Transmembrane transport", "Metal ion transport", "oxidoreductase activity" and "Phytohormone metabolic processes" (Table 4.3).

4.3.5 Genes Previously Associated with Al Stress in Rice

In order to contrast the results with studies previously conducted on rice crops under Al stress conditions, the DEGs identified in this current study were compared to a list of 288 Al stress responsive genes, generated via massive RNA sequencing technology (Arbelaez et al., 2017; Arenhart et al., 2014; Tyagi et al., 2020; Zhang Zhong, et al., 2019). A match was found with 95 of these previously reported genes: 52 for BGI, 62 for AZU, 23 for OG131, and 17 genes for OG97 (Supplementary table 4.7). After contrasting this gene list with scientific literature, genes with an experimentally proven role in Al tolerance were identified. Seven differentially expressed genes were found in this study with Log2FC values between 0.5 and 1.5, and three genes with Log2FC values > 1.5. Similarly, six common differentially expressed genes in all evaluated rice genotypes had been previously reported in other studies: *Os12g0108500 (OsFbox636), Os01g0106400, Os01g0609300 (OsPDR9), Os12g0614500, Os01g0919100 (OsFRDL4), Os03g0233900 (HB1)* (Table 4.4).

Table 4.3. Functional characterization of DEGs shared between tolerant genotypes

Gene ID	Status	Gene Description	Gene Ontology associated terms (BP, CC, or MF)
Os03g0807900	Down	PAMP (pathogen-associated molecular pattern)-responsive transrepressor, Defense response	Chaperone-mediated protein folding, chloroplast stroma, protein folding chaperone
Os01g0940800	Down	Similar to IN2-1 protein	Carbohydrate metabolic process, anchored component of plasma membrane, hydrolase activity
Os05g0244700	Down	Glycoside hydrolase, family 17 protein	Carboxylic acid metabolic process, cytosol, branched-chain-amino-acid transaminase activity
Os04g0665200	Down	SABATH family protein, IAA homeostasi	Auxin homeostasis, methyltransferase activity
Os04g0660900	Down	TGF-beta receptor, type I/II extracellular region family protein	Transmembrane transport, plasma membrane, transmembrane transporter activity
Os03g0182000	Up	Similar to flavin-dependent monooxygenase 1	NADP binding
Os04g0288100	Up	Similar to Auxin-binding protein ABP20	Extracellular region, nutrient reservoir activity
Os10g0555700	Up	Beta-expansin	Sexual reproduction, extracellular region
Os04g0629200	Up	Cupredoxin domain containing protein	Electron transport chain, anchored component of plasma membrane, electron transfer activity
Os01g0249800	Up	Heavy metal-associated domain, HMA domain containing protein	Metal ion transport, metal ion binding
Os04g0666800	Up	Similar to H1005F08.21 protein	Phenylpropanoid biosynthetic process, apoplast
Os01g0722300	Up	Myb transcription factor domain containing protein	Cell differentiation, nucleus, sequence-specific DNA binding
Os09g0518750	Up	Conserved hypothetical protein	
Os01g0187600	Up	Similar to Cytokinin dehydrogenase 1	Cytokinin metabolic process, Cytokinin dehydrogenase activity

Table 4.4. Genes previously associated with Al-tolerance mechanisms in rice and detected as DEGs in the present study. Values in parentheses represent Log2 FoldChange values for each genotype

Gene name	Function	Tolerance mechanism	Log2FoldChange value per genotype	Reference
NRAT1/ NRAMP4	Transports trivalent Al ion, required for a prior step of final Al detoxification through sequestration of Al into vacuoles	Inclusion	AZU (2.6), BGI (1), OG97 (1.5), OG131 (1.6)	Xia et al. 2010
OsALS1	Responsible for sequestration of Al into the vacuoles	Inclusion	AZU (0.82)	Huang et al. 2011
OsSTAR1	Form a complex that functions as an ABC transporter, which is required for detoxification of Al in rice	Exclusion	AZU (1.72), BGI (1.19), OG97 (1.63), OG131 (2.38)	Huang et al. 2009
OsSTAR2	Form a complex that functions as an ABC transporter, which is required for detoxification of Al in rice	Exclusion	OG131 (1.6)	Huang et al. 2009
OsFRDL4	Al-induced citrate transporter localized at the plasma membrane of rice root cells and is one of the components of high Al tolerance in rice	Exclusion	AZU (4.1), BGI (5.1), OG97 (7.2), OG131 (6.2)	Yokosho et al. 2011
OsEXPA10	Expressed in the root tips are required for the root cell elongation, but the contribution of this gene to high Al tolerance in rice is small.	Exclusion	AZU (0.8), BGI (1.1)	Che et al. 2016
OsFRDL2	OsFRDL2 is involved in the Al-induced secretion of citrate, its contribution to high Al tolerance is relatively small in rice	Exclusion	AZU (1), BGI (0.67)	Yokosho et al. 2016
WRKY22	Promote Al-induced increases in OsFRDL4 expression, thus enhancing Al-induced citrate secretion and Al tolerance in rice	Transcription factor	OG131 (1.5)	Li, 2017
ASR5	Regulates the expression of different genes that collectively contribute to the protection of the cell in response to aluminum stress.	Transcription factor	OG131 (1.9), OG97 (1.18), AZU (1.14), BGI (0.66)	Arenhart et al. 2014

4.4 Discussion

Multiple stressors affect rice production, with Al identified as one of the most significant, resulting in substantial yield losses when combined with acid soils. Among cereals, rice has been reported to be one of the most Al-tolerant crops when compared with maize, sorghum, and wheat (Famoso et al., 2010). To develop a comprehensive description of the molecular mechanism underlying Al-tolerance in rice, the transcriptome of two cultivated and two wild rice genotypes with contrasting responses to Al-exposure were explored. The transcriptional data allowed cluster the samples according to experimental treatments (control and Al stress). Similarly, tolerant, and susceptible genotypes were also grouped together, demonstrating the correspondence of the transcriptional data with genotypic identity for both evaluated species.

For all analyzed genotypes, several genes were differentially expressed under the condition of Al stress. Susceptible genotypes showed a higher number of DEGs compared to the tolerant ones. Additionally, the difference in the number of DEGs between the tolerant and susceptible genotypes is reduced when the stringency level based on the Log2FC value is more restrictive, suggesting that for susceptible genotypes, Al stress generates fluctuations in a high number of genes, but expression changes are mild. These results have been reported in previous studies where an increased Al toxicity is reflected in massive transcriptional responses, particularly in the susceptible genotypes (Maron et al., 2008). Similar observations were also reported for rice under salinity stress (Walia et al., 2005, 2007) and for other crops under abiotic stresses such as cold and salinity (Chopra et al., 2015; Rahman et al., 2014; Shen et al., 2014), where a large number of transcriptional changes were found for susceptible genotypes during abiotic stress adaptation times. Interestingly, in the present research the transcriptional responses were evaluated after 10 days of exposure to Al stress, and thus it can be stated that gene expression changes are mediated by this adaptation phase in the evaluated genotypes. From this perspective, Kawasaki et al., (2001) evaluating rice genotypes with contrasting sensitivity to salt stress, observed that the number of DEGs after 1h of the stress increased for the tolerant genotype, but after 3h, these DEGs were progressively reduced declining over time (7 days). In contrast, in the susceptible genotype, this response was delayed, resulting in many down-regulated genes prior plants death, suggesting that this delay might underlie the ineffective response of susceptible genotypes to the stress condition. The transcriptional changes observed are consistent with this behavior, in which the susceptible genotypes displayed a higher number of deregulated genes in the evaluated adaptive phase, indicating the massive physiological changes that susceptible genotypes are tuning to cope with Al conditions.

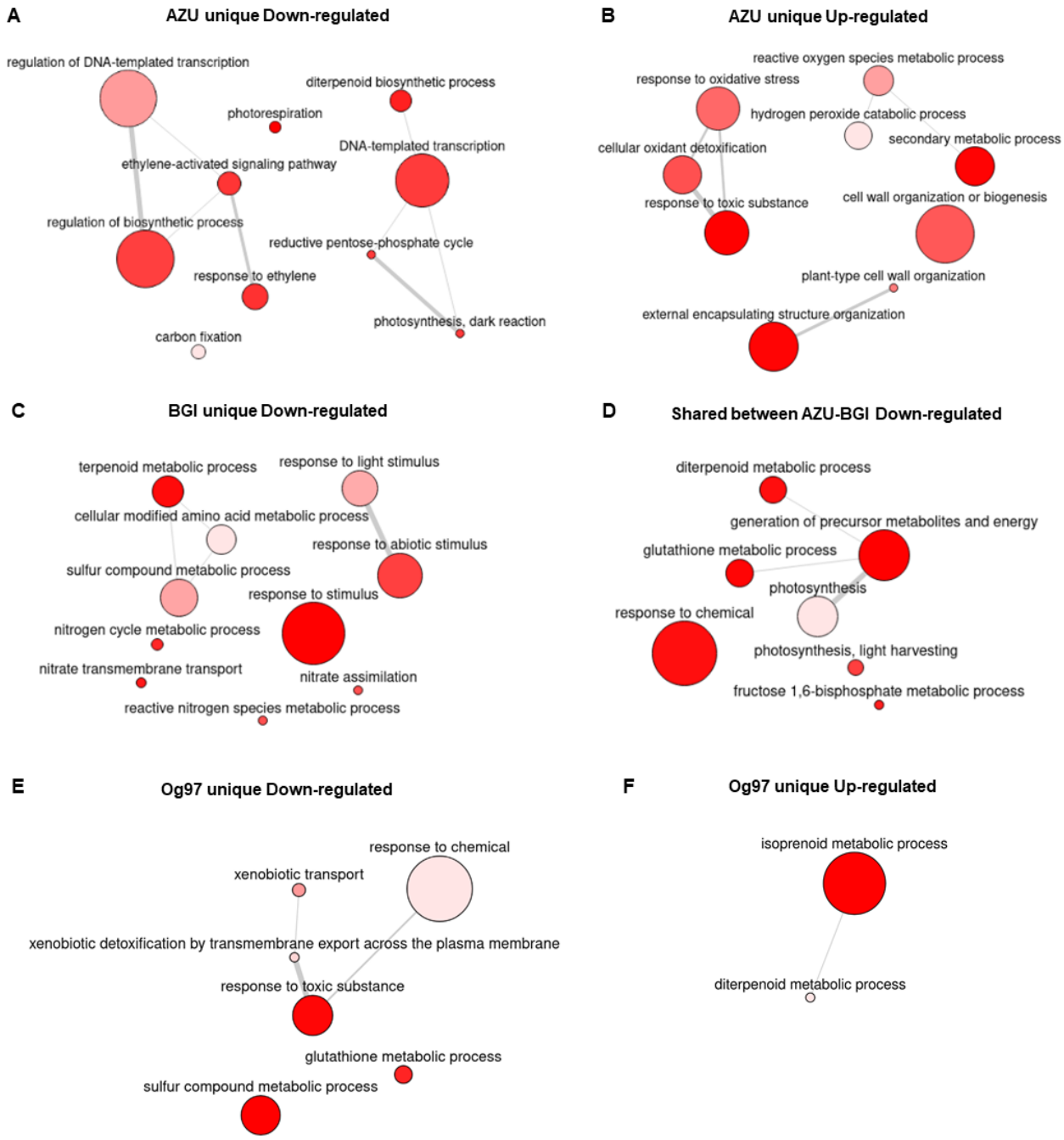


Figure 4.4. Functional characterization for analyzed genotypes. Biological processes (BP) linked to identified DEGs were determined **A.** Exclusively enriched BP for the AZU genotype when down-regulated genes are functionally analyzed; **B.** Exclusively enriched BP for the AZU genotype when up-regulated genes are functionally analyzed; **C.** Exclusively enriched BP for the BGI genotype when down-regulated genes are functionally characterized; **D.** Shared enriched BP for the *Oryza sativa* genotypes; **E.** Exclusively enriched BP for the Og97 genotype when down-regulated genes are functionally characterized, and **F.** Exclusively enriched BP for the Og97 genotype when up-regulated genes are functionally analyzed. Circle sizes correspond to Log-values for the depicted GO Term.

Among the reported DEGs for the evaluated genotypes, several genes previously linked to Al responses, were identified. This is key evidence that reinforces the idea of a common set of genes in rice synergistically acting in response to Al. Interestingly, the wild genotypes showed a smaller fraction of genes shared with previous reports, suggesting a divergence in the genetic machinery that is triggered upon Al stress induction between wild and cultivated rice species. However, it is also possible that during the orthologs search between *O. sativa* and *O. glumaepatula,* information from genes that have not been well characterized in the wild rice had been lost, reducing the potential number of genes from which to infer functional associations.

Al detoxification has been associated with two main processes: exclusion mechanism, where Al ions are blocked before entering the root apex, and detoxification, in which plants tolerate Al accumulation either in the root cell wall or in the symplast, where it is chelated and sequestered. In this sense, these results revealed contrasting patterns of gene expression that encompass both tolerance mechanisms. This evidence reinforces the notion that Al tolerance in rice crops is an additive process involving multiple genes and mechanisms that operate in an orchestrated fashion (Kochian et al., 2015).

Differential expression of *OsSTAR1, OsSTAR2, OsFDRL4, OsFDRL2* and *OsEXPA10* genes involved in the external detoxification mechanism is reported here. Complementarily, *OsSTAR1* and *OsSTAR2* genes are involved in cell wall modification. *OsSTAR1* encodes an ATP-binding domain while *OsSTAR2* encodes a membrane-binding domain of a bacterial-type ABC transporter (Arenhart et al., 2014; Huang et al., 2009). This carrier transports UDP-glucose from the cytoplasm to the cell wall, reducing the binding ability of Al to the cell wall (Huang et al., 2009). In rice, overexpression of *OsSTAR1* and *OsSTAR2* genes has been found to be associated with reduced Al accumulation in the cell wall (Zhu et al., 2018).

The overexpression of the *OsEXPA10* gene in response to Al stress induction has been previously reported in rice (Che et al., 2016). In fact, expression changes of this gene were specifically and rapidly induced by Al exposure. *OsEXPA10* is involved in root cell elongation by mediating cell wall loosening, and thus affects Al binding. Additionally, *OsFRDL4* and *OsFRDL2* are two Al-induced citrate transporters belonging to the multidrug and toxic compound extrusion (MATE) family, located at the plasma membrane and at cytosol vesicles (Jingguang et al., 2020). It is well known that organic acid secretion is a mechanism associated with Al-tolerance, given that organic acids can chelate Al^{3+} ions, forming non-toxic compounds that cannot enter root tip cells. Interestingly, both the expression of *OsFRDL2*, a citrate transporter, and *OsEXPA10* induction are rapidly and specifically induced by Al. In this study they presented low Log2FC values and were only differentially expressed for AZU and BGI genotypes. Nevertheless, previous studies have

also reported the minor effects of these genes on Al tolerance suggesting that they are probably not present in wild genotypes (Che et al., 2016; Yokosho et al., 2011).

NRAT1 and *ALS1* genes are related to the detoxification mechanism. *NRAT1* was differentially expressed for all the genotypes, with higher values in the AZU genotype (Log2FC 2.6) than in the other varieties. In contrast, *ALS1* was differentially expressed only in AZU (Log2FC 0.82). *NRAT1* is a member of the Nramp gene family that transports Al^{3+} ions into the root cytoplasm (Xia et al., 2010). After the introduction of Al^{3+} ions, the transport process is mediated by the ABC transporter Al-sensitive 1 (*OsALS1*), which is a half-size ABC transporter localized at the tonoplast. It is interesting to find a tolerance mechanism involving the introduction of toxic substances into the cell, but it could be considered a strategy to reduce the high levels of Al accumulation in the cell wall and their detrimental effects. Similar to *OsSTAR1* and *OsSTAR2* expression behavior, *NRAT1* was differentially expressed in all genotypes. *ALS1*, however, was only expressed in AZU, a tolerant genotype, suggesting that adapted genotypes are better able to cope with the stress, triggering more specific mechanisms than susceptible genotypes.

The involvement of previously studied Al-responsive genes in the tolerance mechanism linked with Al detoxification has been discussed. Thus, it is proposed that these already known mechanisms coincide with the canonical strategies reported for Al tolerance in cultivated and wild rice genotypes. However, in this study 28 genes were differentially expressed for all genotypes after long-term Al exposure. Most of these genes have not been directly linked to Al-stress response, but some have been previously associated with responses to different stress conditions such as nitrogen starvation (*Os01g0595600*), salinity stress (*Os04g0538900*), drought (*Os09g0448200, Os11g0145200),* sulfur starvation (*Os01g0106400*), anaerobic germination (*Os07g0638400*) and wounding stress (Baldoni et al., 2021; Hsieh et al., 2018; Lunde et al., 2008; Mustafiz et al., 2011; Seto et al., 2011; Yang et al., 2019), indicating that they are indeed stress-associated genes. In addition, this group of 28 genes are associated with functional categories such as “Metabolism”, “Transport activity”, “Glutathione metabolic process”, “Plant hormone signaling”, and “Oxidative stress” that relate to stress conditions in plants. From this perspective, the evidence gathered supports the idea that, regardless of tolerance levels, a core set of genes exist which is acting response to develop long-term tolerance responses to Al stress. Reinforcing this hypothesis, other researchers have also put forward the idea of a central responsive machinery that is activated by different stress conditions in rice (Cohen & Leach, 2019).

Functional characterization revealed the association of the core of genes identified with oxidative stress responses. Such responses are due to an excess of Reactive Oxygen Species (ROS), produced in the context of heavy metal (HM) stress that can disrupt the redox status of cells, resulting in the oxidative stress of exposed cells, and leading to

membrane dismantling, biological macromolecule deterioration, ion leakage, lipid peroxidation and DNA-strand cleavage (Shahid et al., 2014). In addition, the up-regulation of oxidative stress-responsive genes under Al exposure has been shown in several plant species, including tea (Li et al., 2017), maize (Maron et al., 2008) and rice (Tsutsui et al., 2012). Another function related to the characterized central machinery is the Glutathione metabolic process. Notably, when the functional enrichment analysis was performed, this category was enriched for all genotypes, suggesting glutathione metabolism plays an important role in the basal response that is proposed. Glutathione associated processes have been determined as one of the central mechanisms that regulate ROS levels in plants, but glutathione metabolism is also associated with the detoxification of the glycating agent methylglyoxal, a toxic product of glycolysis (Foyer & Noctor, 2011). Similar to ROS, increased methylglyoxal accumulation is a common outcome in plants exposed to environmental constraints such as salinity, drought, cold, and heavy metals (Dorion et al., 2021). Within these functional categories two specific down-regulated genes (*Os10g0525500* and *Os09g0367700*) were found in all the genotypes analyzed. It should be noted that enriched categories are being reported after 10 days of stress exposure and therefore the down-regulation observed is related to a compensation process after several days of stress conditions.

From these findings, it is proposed that there is a common machinery that regulates Al stress responses in rice, and, furthermore, that there is an additional core of specific genes and mechanisms that control Al stress responses in tolerant genotypes. In addition, 14 DEGs shared by tolerant genotypes but not differentially expressed in susceptible ones were reported. Previous work linked three of these genes to stress responses. The most interesting of these is the Os01g0249800 gene that is a cation transporter under cadmium stress conditions (Chen et al., 2023). Subsequently, as a result of the functional characterization of these 14 DEGs, several GO categories were identified such as “Transmembrane transport”, “Metal ion transport” and “phytohormone signaling”, that have been previously reported as specific processes associated with stress signaling in plants (Jingguang et al., 2020). As part of the mechanism that involves phytohormone signaling activation, ethylene regulation has been potentially implicated in plant adaptation or tolerance to toxic metals (Asgher et al., 2014; Khan & Khan, 2014). Recent studies have shown that tolerant genotypes activate genes that encode components of the ethylene signaling pathway (Cao et al., 2014; Fu et al., 2014; Guan et al., 2015). Likewise, Keunen and collaborators described a close relationship between ethylene and glutathione metabolism during HM stress responses (Keunen et al., 2016). This group of alternative core genes associated with specific responses observed only in the tolerant genotypes are suggested as potential key factors involved in the inter-specific tolerance mechanism in rice. The lack of information around these genes could be indicative of novel

tolerance mechanisms that have yet to be studied, especially given the focus on *O. glumaepatula*, a wild species that remains poorly studied.

Concluding Remarks

In this chapter, transcriptional responses of tolerant and susceptible rice genotypes were analyzed after 10 days of Al exposure. Based on the evidence gathered, a model to further understand the distinct mechanisms triggered by Al exposure in rice varieties was proposed (Figure 4.5). The model suggests that upon Al stress induction, susceptible genotypes transcribe a greater number of DEGs, suggesting that the drastic physiological changes observed in susceptible genotypes is associated with a greater transcriptional variation, indicative of they attempt to adapt to cope with Al conditions. Nevertheless, this transcriptional response is not effective and susceptible genotypes exhibit detrimental phenotypes. In contrast, it is suggested that in tolerant genotypes a distinct response occurs after prolonged exposure to Al that is mediated by a smaller set of genes conferring the ability to adapt to the stress condition. This set of genes is directly associated with biological processes such as transmembrane and metal ion transport, plant hormone signaling, and secondary metabolism-associated genes. Some of these molecules have not been previously characterized and could be considered novel regulators directly linked with Al tolerance responses, and promising candidates for the generation of agronomically improved rice varieties.

A group of 28 genes that were differentially expressed in all genotypes were described, putting them forward as core components in the functional response of rice cultivars to Al stress conditions, regardless of their tolerance levels. These genes are involved in glutathione metabolism, plant hormone signaling and oxidative responses, underscoring the relevance of these processes in the generation of specific Al-tolerance mechanisms. In addition, it is worth noting the significant transcriptional induction of *NRAT1* and *FRDL4* genes in all genotypes. These two genes participate in exclusion and detoxification mechanisms respectively, suggesting that both tolerance mechanisms for Al detoxification are active in the analyzed genotypes. This is of particular interest in the case of wild genotypes that have not been previously studied, as possible tolerance mechanisms operating in wild rice to Al stress can now be highlighted.

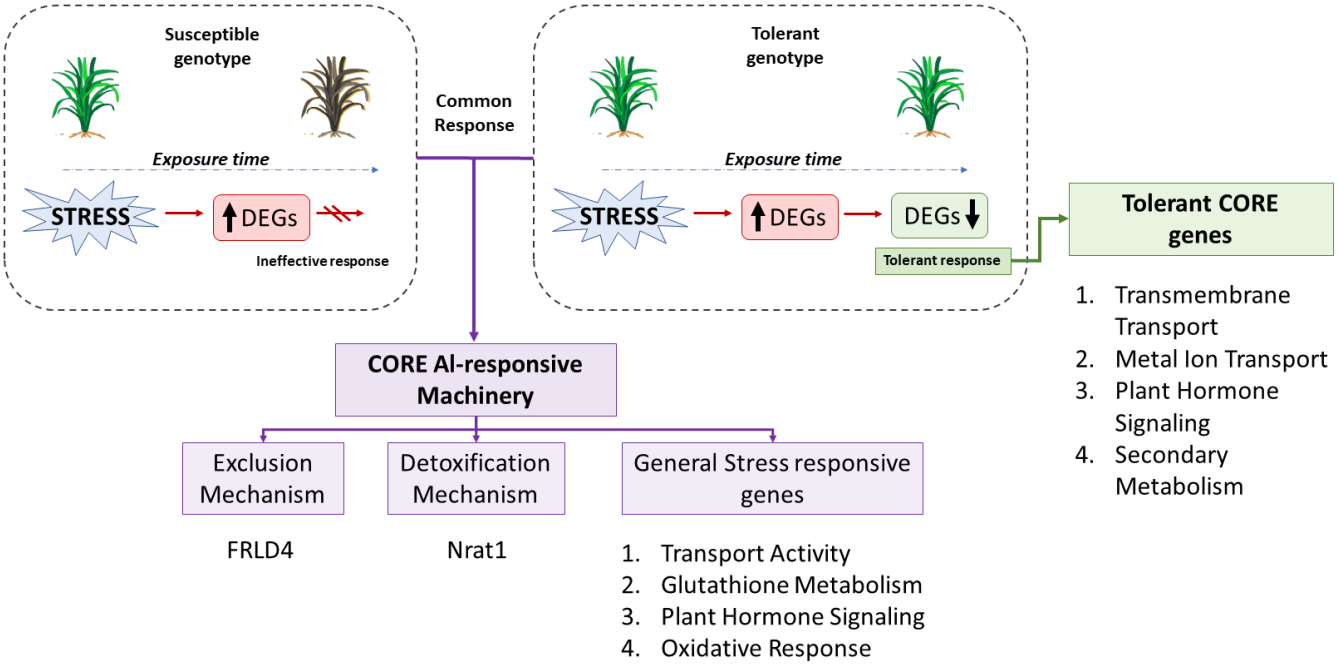


Figure 4.5. Suggested model that explains Al tolerance mechanism exhibited by analyzed genotypes. A proposed model is presented to explain the potential mechanism of Al tolerance exhibited by the analyzed genotypes. Purple rectangles enclosed the core Al-response genes and mechanisms that are present in the analyzed genotypes, regardless of their tolerance levels. This core machinery includes well-known genes, such as *FRDL4* and *NRAT1*, which are involved in the exclusion and detoxification mechanisms, respectively. The green box on the right includes functional categories associated with a set of genes and mechanisms that are differentially expressed only in tolerant genotypes.

Chapter 5

DNA Methylation Changes Associated with Gene Expression in Cultivated (*Oryza sativa*) and Wild Rice (*Oryza glumaepatula*) Genotypes Exposed to Aluminum Stress Conditions

Chapter Summary

DNA methylation is the most studied epigenetic mark, and it has been involved in TEs silencing, chromosomal interactions, trait inheritance and regulation of gene expression. In addition, several studies have considered DNA methylation as a potential mechanism of plant response to stress conditions, showing specificity for different types of stresses. However, so far there are no studies of DNA methylation patterns associated with Al stress conditions in rice, which is one of the most critical abiotic stresses affecting rice crops in acid soils. Thus, in this chapter, DNA methylation variation in response to long-term Al toxicity was evaluated in Al-tolerant and susceptible genotypes of cultivated and wild rice species. In addition, the relationship between DNA methylation patterns and gene expression was evaluated to identify the possible role of DNA methylation in gene regulation under Al stress conditions. According to these results, the overall genome-wide level of variation is similar across the genome and among genotypes, showing a trend of hypomethylation for CG and CHG sequence contexts and hypermethylation for CHH sequence context for stress samples. Similarly, the distribution of differentially methylated regions is similar among genotypes and species, suggesting a common response of rice to Al-stress conditions. Several of these reported DMRs are related to genes, a small portion of which were also differentially expressed, suggesting that some specific genes are associated with epigenetic regulation under Al-stress conditions. To reinforce this hypothesis, a functional analysis of genes differentially methylated and differentially expressed genes (DMG-DEGs) was performed resulting in several enriched categories related to stress response in plants. Finally, potential genes involved in the mechanisms of Al tolerance in rice that are being epigenetically regulated were identified. These DEGs with potential epigenetic regulation represent key genes with a view to their potential use in breeding susceptible genotypes.

Keywords: *Abiotic stress, Aluminum tolerance, Bisulfite sequencing, DNA methylation, Differential expression, Methylome, RNA seq, Transcriptome.*

Connection to Previous Chapter

In chapter 4, transcriptional analysis of *O. sativa* and *O. glumaepatula* under prolonged Al exposure conditions was performed. As a result, several differentially expressed genes (DEGs) were identified for all genotypes studied. However, the question remains open as to the role of DNA methylation in the transcriptional regulation of these genes under Al stress conditions. To answer that question, in this chapter, variation in the methylome of tolerant and susceptible rice genotypes in response to Al stress was evaluated and the results were contrasted with the differential expression results of the previous chapter.

5.1 Introduction

Plants are constantly confronted with challenging environmental conditions and have evolved different physiological and genetic defense strategies to survive. In addition to the elucidation of the genetic basis, signal transduction and regulatory mechanisms underlying stress responses, a growing number of studies have demonstrated the important role of epigenetics in plant responses to abiotic stress. Likewise, it has been evidenced that epigenetic mechanisms play crucial roles in the formation of stress memory, which can be inherited by the offspring of stress-treated plants (Chang et al., 2020). Therefore, elucidating the epigenetic codes of plant stress responses could be of great importance for breeding stress-tolerant crops.

Epigenetic mechanisms, including DNA histone modifications and DNA methylation, are heritable changes that do not affect DNA sequence and define chromatin structure and accessibility. DNA methylation is considered the most studied epigenetic mark, and is considered a relatively stable, heritable, and transgenerational mark. This epigenetic modification has been implicated in TEs silencing, chromosomal interactions, trait inheritance and DNA regulation.

Stress-induced DNA methylation changes have been extensively investigated under various stress conditions in many plants, showing specificity to different types of stresses (Sun et al., 2022). Most abiotic stress-induced DNA methylation modifications are transient and return to initial levels after removal of the stress; however, studies have also shown that some short- or long-term memory effect could be induced in plants (Verhoeven et al., 2010). Thus, DNA methylation has been considered a mechanism by which plants adapt to abiotic stress (Akhter et al., 2021). In rice several studies have reported variations of DNA methylation patterns in response to stress conditions, as well as a possible role of DNA methylation in gene responsive gene regulation. Details of these findings is described in depth in Chapter 2.

In Chapter 3 of this thesis, several evidence on the possible role of pre-established DNA methylation patterns on the Al tolerance levels of rice genotypes grown under control conditions were presented. However, the next step is to evaluate the DNA methylation changes generated in response to the stress condition. So far, there have been some studies evaluating the effect of heavy metals on DNA methylation in plants (Choi & Sano, 2007; Ou et al., 2012), but there are no Al-related epigenetic studies in rice. Thus, in this chapter, whole genome bisulfite sequencing (WGBS) and RNA-sequencing (RNAseq) data was used to identify the patterns of DNA methylation changes under Al exposure and their correlation with gene expression regulation.

5.2 Materials and Methods

5.2.1 Plant Material and Al treatment

The rice genotypes used for DNA methylation experiments were the same analyzed in chapter 3 for transcriptomic analysis (Supplementary table 3.1). For *Oryza sativa,* the Al-tolerant genotype Azucena (AZU) and the Al-susceptible genotype BGI9311 (BGI) were selected. For *Oryza glumaepatula,* the Al-tolerant genotype OG97 and the Al-susceptible OG131 were selected. Plant growth and Al treatment conditions are described in section 3.2.1. Total genomic DNA was extracted from frozen root tissue by CTAB 2X protocol with modifications (Maropola et al., 2015). For the extraction process, roots from five different plants were pooled and considered as a single sample. Genomic DNA quality was evaluated on agarose gels and DNA quantity was measured using a Nanodrop spectrophotometer (Thermo Scientific).

5.2.2 Whole-Genome Bisulfite Sequencing (WGBS), Read Alignment, and Data Imputation

Bisulfite-seq (BS-seq) libraries and paired-end (2x150pb) DNA sequencing were made from genomic DNA isolated from *O. sativa* and *O. glumaepatula* seedlings roots by NovoGene-Korea using an Illumina NovaSeq platform (NovaSeq 6000) (https://www.novogene.com/). Three samples were sequenced per experiment.

The FastQC tool (Simon Andrews, 2010) was used to perform basic statistics on the quality of the raw reads. Then, sequencing adapters and low-quality data (Phred score < 30) were removed by Trimmomatic (V.0.32) (Bolger et al., 2014). The reads were mapped to the Os-Nipponbare-Reference-IRGSP-1.0 (https://rapdb.dna.affrc.go.jp/download/irgsp1.html) genome for *O. sativa* and the Oryza_glumaepatula_v1.5 (http://plants.ensembl.org/Oryza_glumipatula/Info/Index?db=core) genome for *O. glumaepatula* using Bismark (v.16.3) (Krueger & Andrews, 2011) with default parameters. In contrast to Chapter 3, this chapter uses the reference genome available for each species, since comparisons of the same regions of the genome will not be made, but rather a comparison of the final results.

Only the uniquely aligned reads were maintained and all the samples were de-duplicated using the Bismark deduplication module. Finally, cytosine-level methylation calls were obtained for CG, CHG and CHH sites with METHimpute (Taudt et al., 2018), which is a Hidden Markov Model for inferring the methylation status/level of individual cytosines, even in the presence of low sequencing depth and/or missing data. All WGBS data processing was performed using the MethylStar pipeline (Shahryary et al., 2020).

5.2.3 Genome-Wide Methylation Patterns

A comparative methylome analysis was done between Al-stress and control samples for all the rice genotypes using the methylated cytosines (mCs) counting throughout the complete genome. The rice genome was divided into windows of 100 kb and for each window the number of mCs per sequence context (CG, CHG, and CHH) was computed for each sample. In addition, the difference in mCs between the stress and control samples per window was calculated (Average values for the three replicates were used to compare). Then, exponential smoothing with $\alpha = 0.1$ was applied to the calculated difference to remove noise associated with the abrupt change in the count of mCs in adjacent windows.

5.2.4 Identification of Differentially Methylated Regions (DMRs) and Differentially Methylated genes (DMGs)

To identify DMRs, the *jDMR* R package was used (https://github.com/jlab-code/jDMR), a heuristic DMR caller that produces methylation calls for bins based on a Hidden Markov Model (HMM). In this experiment, the genome was divided into windows of 500, 400, and 100pb for the CG, CHG, and CHH sequence context respectively and only bins with at least five cytosines were retained. The methylation status of a given region was classified as either unmethylated or methylated. Pairwise comparisons between the control and treatment groups were performed by generating a DMR matrix containing region calls corresponding to each sample and each bin. Finally, non-polymorphic regions were excluded, and adjacent bins were merged to obtain the list of DMRs.

To characterize the distribution of DMRs along the genome their overlap with transposable elements (TEs), genes, promoters, and intergenic regions was evaluated. An overlapping of 2pb for each functional feature was considered as a positive result. Annotation files for genes and TEs in gff3 format were downloaded from Ensembl Plants (http://plants.ensembl.org/info/data/ftp). TEs annotation was unavailable for *O. glumaepatula.* Promoters were defined arbitrarily as regions 2 kb ahead of the TSS. Intergenic regions were defined as all regions in the genome that were not covered by any of the above annotations. Therefore, genes with DMRs overlapping their body or promoter region were considered Differentially Methylated Genes (DMGs).

5.2.5 Functional Enrichment Analysis

g:Profiler web service was used for functional enrichment analysis of DMGs shared only between tolerant genotypes (AZU and OG97) but not present in susceptible ones. Functional enrichments were obtained for biological processes (BP), molecular function

(MF), cellular compartment (CC), and KEGG metabolic pathways. g:SCS (Set Counts and Sizes) method was used as a multiple testing correction procedure because it takes into consideration the directed acyclic graph structure of the GO terms to obtain adjusted p-values (Reimand et al., 2007). Only GO terms or KEGG pathways with g:SCS values < 0.05 were considered enriched. Finally, the REVIGO (Reduce and Visualize Gene Ontology) tool (Supek et al., 2011) was used for summarizing the Gene Ontology terms by removing redundant GO terms. The default factor C = 0.7 was used, which is considered a conservative value, maintaining strong statistical support.

5.3 Results

5.3.1 Consistent DNA Methylation Changes in Wild and Cultivated Rice Genotypes under Aluminum Exposure

In this chapter, the base resolution methylome of one Al-susceptible and one Al-tolerant genotype for *O. sativa* and *O. glumaepatula* species grown under stress conditions was generated. As a result, a total of 55 million raw reads on average were generated for *O. sativa* and 58 million for *O. glumaepatula.* After the mapping of reads, the genome coverage ranged from 88% to 96% for *O. sativa* and 65% to 69% for *O. glumaepatula* with sequencing depth ranging between 14 and 20X. The bisulfite conversion rate for all the libraries was above 99.5%. The depth and quality of the sequencing were enough to ensure a high-quality genome-wide methylation analysis in all the samples. Regarding the overall methylation results, a higher number of mCs was found in the CG context, followed by CHG and CHH context. Summary statistics for each sample can be found in Supplementary Table 5.1. The overall statistics of *O. sativa* were better than *O. glumaepatula*, which may be due to a better quality of the reference genome.

To analyze the distribution of DNA methylation in CG, CHG, and CHH contexts, a genome-wide profile of mCs counts using 100kb static windows was plotted for stress and control samples of *O. sativa* and *O. glumaepatula*. Figure 1 presents the methylation profiles of chromosome 1 in O. sativa (Figures 5.1A and B) and O. glumaepatula (Figures 5.1C and D). A similar pattern was observed across all chromosomes with consistent behavior observed in CG and CHG contexts, while the opposite trend was observed in the CHH context. The graphs reveal significant differences between control and stress samples, particularly in O. glumaepatula genotypes, suggesting a widespread shift in the rice genotypes' methylome under Al-exposure. However, the results for the CHH context in O. sativa genotypes were inconclusive due to ambiguity in one of the replicates.

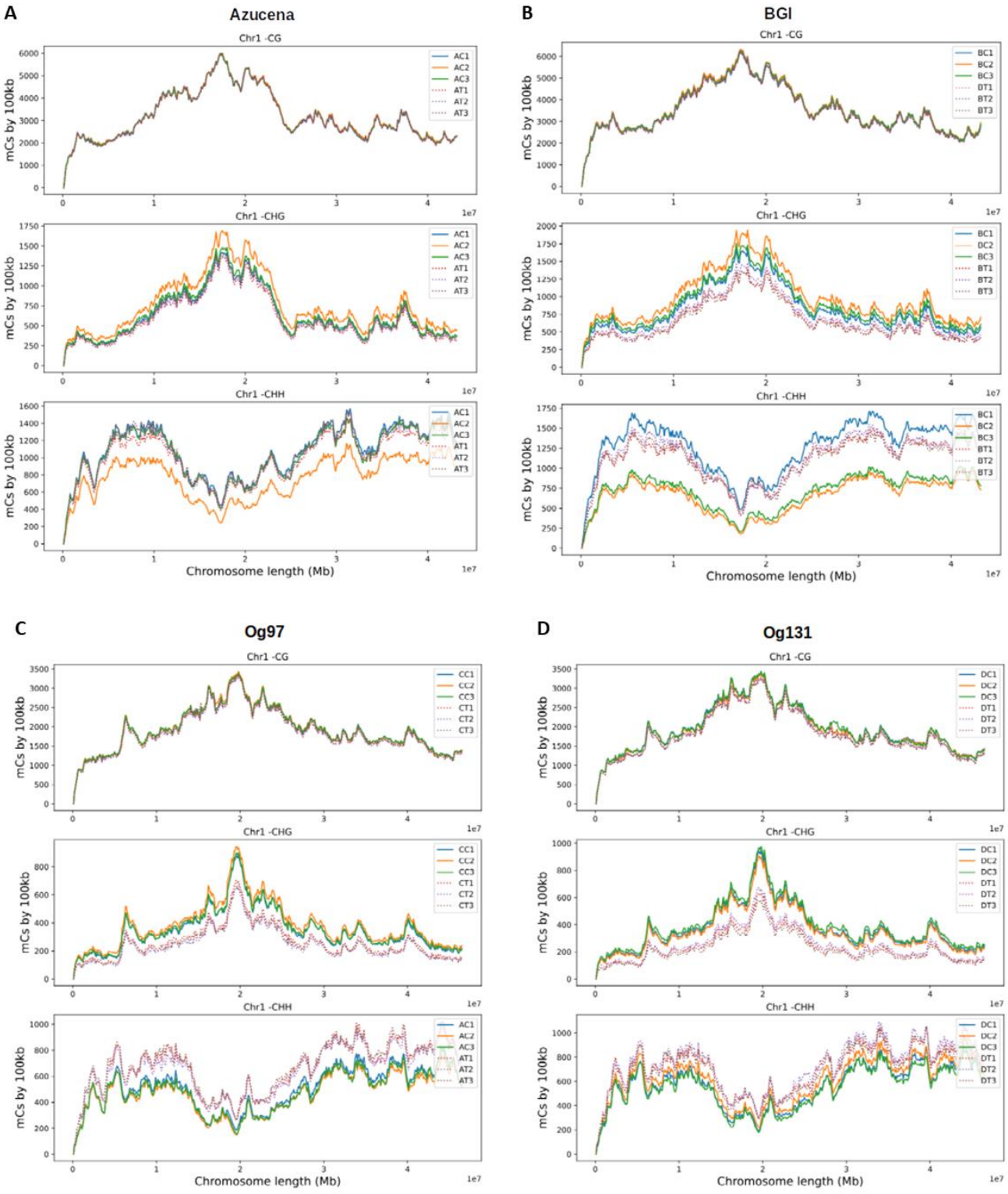


Figure 5.1. Number of methylated cytosines (mCs) along the genome for **A.** AZU, **B.** BGI, **C.** OG97 and **D.** OG131 genotypes. The rice genome was divided into windows of 100 kb and for each one of them, the number of mCs per sequence context (CG, CHG, and CHH) for each sample was computed. The plain lines correspond to the control samples and the dotted lines correspond to the samples under stress conditions.

The difference in the mCs counts per window between control and stress conditions was analyzed. Figure 5.2 shows the methylation profile for chromosome 1 in all analyzed genotypes as an example of the obtained results. A similar pattern was observed across all chromosomes. Interestingly, in both *O. sativa* and *O. glumaepatula* species, the susceptible genotype exhibited the highest values for the CG and CHG contexts and the lowest values for the CHH context. It is possible to say this based in the percentage of windows where the susceptible genotype showed a greater change compared to the tolerant genotype. In *O. sativa,* BGI showed a higher variation in a greater percentage of windows (CG: 56%, CHG: 99%, CHH: 93%) compared to AZU (CG: 44%, CHG: 1%, CHH: 7%). In *O. glumaepatula*, OG131 displayed a higher variation in a greater percentage of windows for CG (80%) and CHG (80%) compared to OG97 (CG:20%, CHG:20%), but the opposite trend was observed in the CHH context (OG131: 35%, OG97: 65%). Notably, CG and CHG context showed mainly positive differences, indicating generalized hypomethylation in the genome. In contrast, only negative differences were observed in the CHH context, indicating generalized CHH hypermethylation in the rice genome under stress exposure. These results suggest uneven effects of Al stress on DNA methylation in the three contexts.

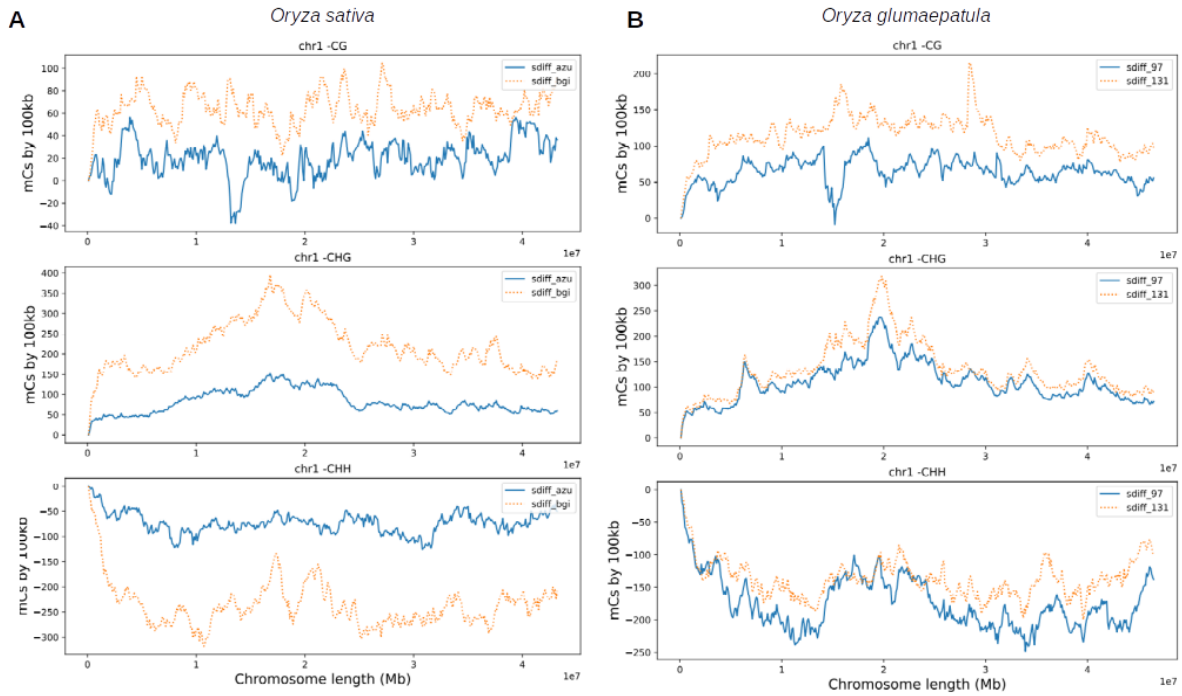


Figure 5.2. Calculated difference in the number of mCs between control and stress conditions for **A.** *Oryza sativa* genotypes, AZU and BGI, and **B.** *Oryza glumaepatula* genotypes, OG97 y OG131. The control sample was used as the reference point. A positive value indicates hypomethylation which represents a decrease in methylation levels compared to the control condition. Conversely, negative values indicate hypermethylation, which represents an increase in methylation levels compared to the control condition, in response to the treatment.

To analyze specific regions in the genome where the significative changes are occurring under Al exposure, DMRs between control and stress conditions were identified for all the rice genotypes. For AZU 1759 CG-DMRs (828 hyper and 931 hypomethylated), 2639 CHG-DMRs (633 hyper and 2006 hypomethylated), and 10980 CHH-DMRs (2328 hyper and 8652 hypomethylated) were found (Figure 3A). For BGI 1852 CG-DMRs (1126 hyper and 726 hypomethylated), 2429 CHG-DMRs (552 hyper and 1877 hypomethylated), and 13588 CHH-DMRs (212 hyper and 11467 hypomethylated) were found (Figure 3A). For OG97 3335 CG-DMRs (907 hyper and 2428 hypomethylated), 5325 CHG-DMRs (471 hyper and 4854 hypomethylated), and 14267 CHH-DMRs (1581 hyper and 12687 hypomethylated) were found (Figure 5.3B). For OG131, 14814 CG-DMRs (992 hyper and 13822 hypomethylated), 9614 CHG-DMRs (705 hyper and 8909 hypomethylated), and 26855 CHH-DMRs (983 hyper and 25872 hypomethylated) were found (Figure 3B). Overall, there are mainly hypo-DMRs for all the genotypes and sequence contexts and the number of DMRs is higher in the susceptible genotypes. The high number of hypomethylated CHH-DMRs is interesting considering that CHH positions along the genome tended to be hypermethylated under Al exposure (Figures 5.1 and 5.2).

Finally, DMR annotation was done to characterize its distribution along the genome. An overlapping of 2bp between the DMR and the functional element (gene, promoter, TEs) was considered a positive result for the DMR annotation. According to the results, there exists a common pattern for DMRs location among susceptible and tolerant genotypes for both rice species (Figure 5.3C and 5.3D), where most of the DMRs overlap with TEs. In the same way, the different locations of hypo and hyper-DMRs showed the same proportions. For *O. glumaepatula* the coordinates of TEs in the genome are not available so only genes, promoters, and "others" as the excluded category were considered. It is remarkable that most of the hyper-DMRs for the CG context in both genotypes were in the genes or gene promoters instead of TEs.

5.3.2 Al-exposure Alters DNA Methylation Patterns of Genes in Rice

Genes with DMRs located within their body or promoter region (2kb upstream) were considered DMR-associated genes (DMGs). A total of 3979 DMGs were identified in AZU, 4440 in BGI, 6175 in OG97, and 13482 in OG131, considering the overlaps with the promoter region. Likewise, 1765 DMGs were identified in AZU, 2049 in BGI, 4275 in OG97, and 8940 in OG131 considering the overlaps with the gene-body region. The distribution of hypo-DMGs and hyper-DMGs per context is depicted in Figures 5.4A and B. Notably, the CHH context exhibited the highest number of hypo-DMGs, and there was a greater number of DMGs for the susceptible genotype in both species.

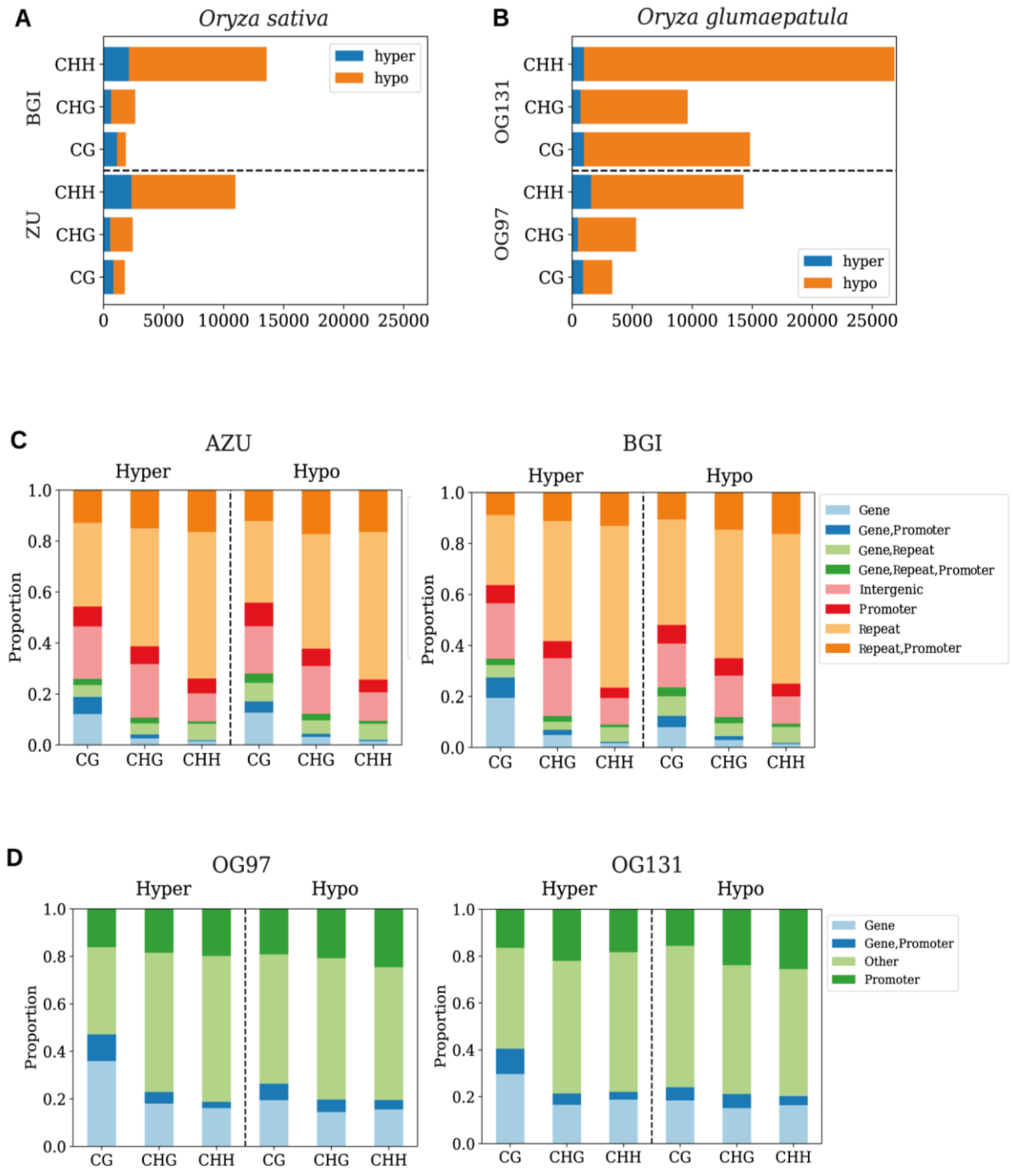


Figure 5.3. DMR statistics for *Oryza* genotypes **A**. Number of hypo and hyper DMRs reported for *Oryza sativa* and **B.** *Oryza glumaepatula*. **C.** DMR annotation for *O. sativa* genotypes, (AZU and BGI) based on functional features including Genes, Promoter (-2Kb), Repeats and Intergenic regions. Some DMRs may overlap with multiple features. **D.** Annotation of DMRs for *O. glumaepatula* genotypes (OG97 and OG131). The functional features considered were Genes, Promoter region (-2Kb) and Other, which includes any other location in the genome.

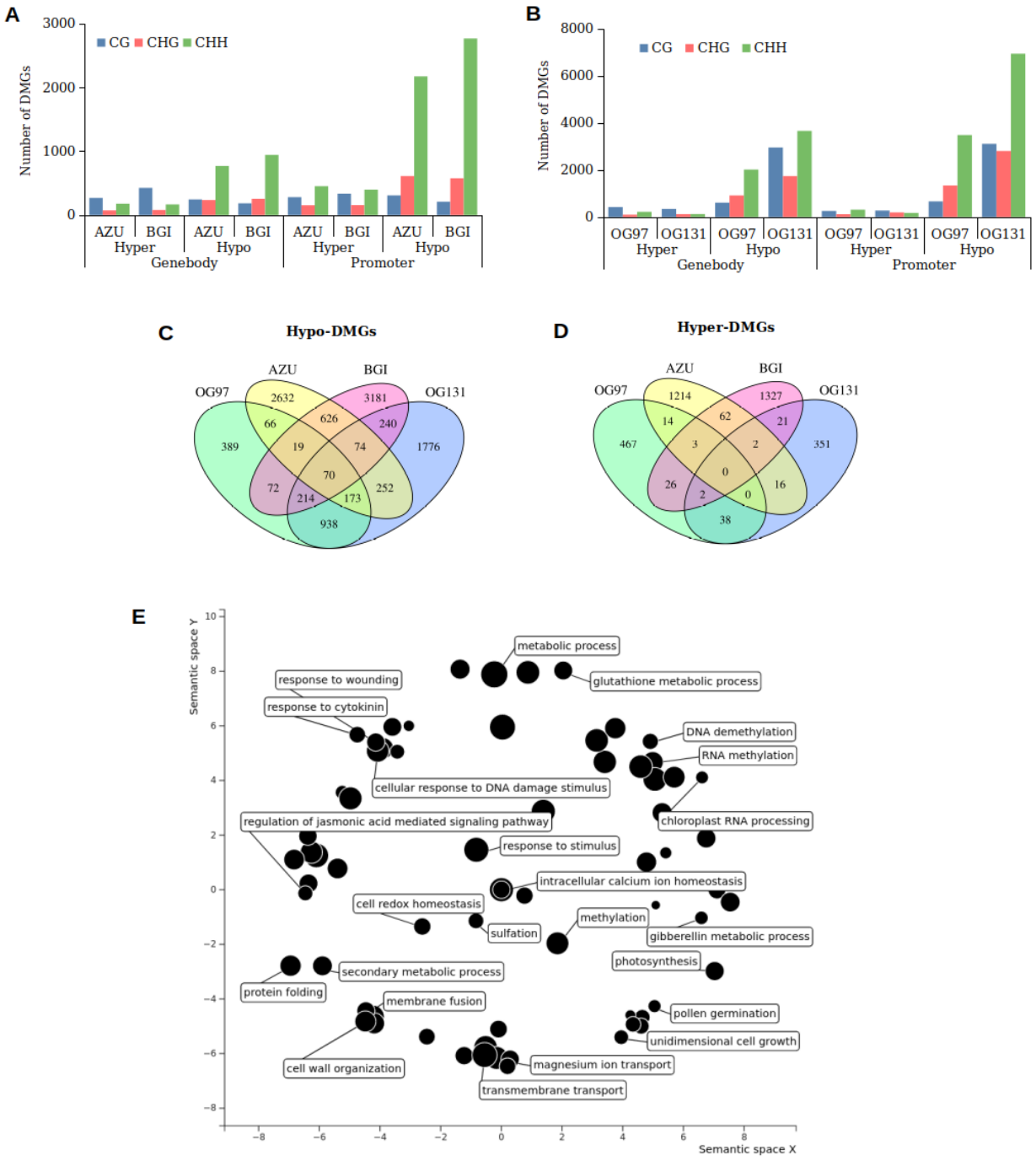


Figure 5.4. Differentially methylated genes (DMGs) for all the analyzed rice genotypes. **A.** Number of DMGs for *Oryza sativa* and **B.** *Oryza glumaepatula*. **C.** Number of shared hyper and **D.** hypo DMGs among rice genotypes**. E.** Relationship among functional categories associated with DMGs unique to tolerant genotypes. The axes in the plot have no intrinsic meaning. Revigo uses Multidimensional Scaling (MDS) to reduce the dimensionality of a matrix of the GO terms pairwise semantic similarities. The guiding principle is that semantically similar GO terms should remain close together in the plot.

Functional enrichment analysis was performed for hyper and hypo-DMGs of all genotypes. No results were obtained for *O. sativa* genotypes. However, several categories were enriched for *O. glumaepatula* genotypes. Eight functional categories were enriched for OG97 (two biological processes (BP), three molecular functions (MF), one cellular component (CC), and two metabolic pathways (KEGG)) of which seven were for hypo-DMGs. On the other hand, 26 functional categories were enriched for OG131 (six BP, 16 MF, two CC, and two KEGG) of which 25 were associated with hypo-DMGs (Supplementary table 5.2). Among the enriched categories it is worth highlighting the categories "lipid metabolic process", "transmembrane transport", "catalytic activity", "oxidoreductase activity" and "Diterpenoid biosynthesis" that were shared between both genotypes.

Finally, 70 DMGs shared among all the genotypes were identified, suggesting a possible role of these genes in the epigenetic regulation of rice response to Al-stress. Likewise, 80 DMGs were identified only in the tolerant varieties turning these DMGs into potential tolerance genes that are being regulated through DNA methylation (Figures 5.4C and D). Functional characterization of these 80 DMGs showed a relationship with several categories such as "Membrane transport", "Cell wall" and "Regulation of plant hormones" for the hyper-DMGs and categories such as "DNA demethylation", "Transmembrane transport", "Response to stimulus", "Cellular response to DNA damage stimulus", "Regulation of plant hormones", among others for hypo-DMGs (Supplementary table 5.3). The relationship of the different ontology terms associated with these genes based on their semantic similarity was evaluated using the Revigo web platform. The guiding principle of the resulting scatterplot is that semantically similar GO terms should remain close together in the plot (Figure 5.4E). It is worth noting the formation of functional clusters of genes that are affected by DNA methylation in both tolerant genotypes.

5.3.3 Correlation Between DNA Methylation and Gene Expression Under Al-Stress Conditions

Although several genes exhibited differential DNA methylation under Al-exposure conditions, it remains unknown whether these changes affect their transcriptional activity. Therefore, a comparison was made between the Differentially expressed genes (DEGs) obtained in chapter 4 to identify genes that showed both differential methylation and expression (DMG-DEGs) under Al conditions. For detailed methods and results regarding the transcriptomic analysis, please refer to Chapter 4.

Subsequently, the comparison between DMGs and DEGs revealed that 13% of the total DEGs were also identified as DMGs, while 3% of the DMGs overlapped with the DEGs (Figure 5.5A and B). These findings suggest that alterations in DNA methylation within the promoter or gene-body region do not necessarily correlate with changes in gene expression. However, it is noteworthy that a subset of genes may be regulated by DNA methylation. Importantly, the number of DMGs does not necessarily correspond to the

number of DMG-DEGs, indicating the existence of specific mechanisms that orchestrate the impact of DNA methylation on gene expression.

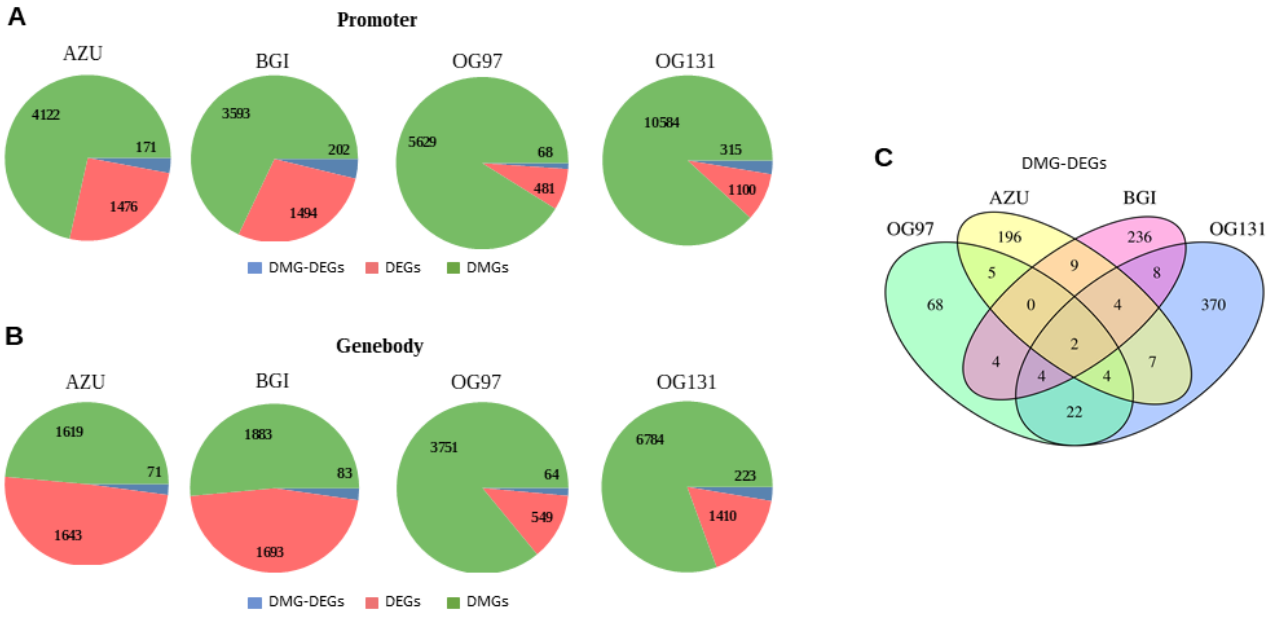


Figure 5.5. Differentially methylated and expressed genes (DMG-DEGs) in *Oryza sativa* and *Oryza glumaepatula* genotypes. DEGs: differentially expressed but not differentially methylated, DMGs: differentially methylated but not differentially expressed and DMG-DEGs: differentially methylated and expressed genes. **A.** Number of DEGs, DMGs and DMG-DEGs differentially methylated in the promoter region. **B.** Number of DEGs, DMGs and DMG-DEGs differentially methylated in the genebody region. **C.** Venn diagram of DMG-DEGs shared among all the genotypes.

To gain more insights into the possible role of DNA methylation in the regulation of gene expression under Al-exposure in rice, DMG-DEGs will be the focus of analysis. As a first approach, whether there is a directional regulation of gene expression it was explored, meaning that hyper-DMRs cause the downregulation of genes or vice versa. Therefore, the Log2FC value of genes was plotted against the DMR meth level difference between control and stress samples for *O. sativa* (Supplementary figure 5.1) and *O. glumaepatula* (Supplementary figure 5.2). As a result, no clear trend was found from the results, but a higher number of up-regulated rather than down-regulated genes associated with DMRs was identified.

In Chapter 4 of this dissertation, a list of genes previously associated with Al-tolerance was presented, which were identified as DEGs in this study (Table 4.4). Then, to gain further insights into the epigenetic regulation of Al stress-responsive genes, the methylation status of these genes across all genotypes was analyzed. Notably, DMRs in the promoter or genebody regions of *Nrat1, OsSTAR1* and *FRDL4* genes were observed, with

distinct sequence contexts (Table 5.1). These findings suggest that DNA methylation might play a role in regulating these genes; however, additional studies are required to substantiate this conclusion.

Table 5.1. Methylation status of genes experimentally linked to aluminum stress response in literature.

Gene	Transcriptional status	Methylation status	DMR location	DMR sequence context	Genotype
Nrat1	Up	Hyper	Promoter	CG	OG131
OsSTAR1	Up	Hypo	Promoter	CG	OG131, AZU
			Genebody	CHH	OG131, AZU
FRDL4	Up	Hypo	Genebody	CG	OG131
			Promoter	CHG	

A comparison was conducted between DMG-DEGs across all the genotypes was done. A total of 939 DMG-DEGs were identified, out of which 69 genes (7%) were found to be shared between at least two genotypes, while 870 genes (93%) are unique to specific genotypes. The next step involved assessing whether there were enriched functional categories for all the identified DMG-DEGs. As a result, eight functional categories were found to be enriched in AZU, with the highest number of genes falling into the “Oxidoreductase activity” category. Similarly, eight categories were enriched in BGI, with the highest number of genes in the “Response to stimulus” category. Additionally, OG97 displayed enrichment in seven categories, with the highest number of genes in the "Oxidoreductase activity" category, while OG131 exhibited enrichment in ten categories, with the highest number of genes in the "Transmembrane transport" category. (Table 5.2).

The tolerant genotypes AZU and OG97 displayed shared enriched functional categories including “Tetrapyrrole binding”, “Oxidoreductase activity” and “Heme binding”. Interestingly, unlike DMGs, DMG-DEGs do show enriched categories for both AZU and BGI genotypes. Notably, the categories "Transmembrane transport" and "Oxidoreductase activity" were enriched for both DMGs and DMG-DEGs in OG97 and OG131 genotypes.

Furthermore, two DMG-DEGs shared among all genotypes were identified (Figure 5.5C), namely *Os01g0343100,* characterized as an integral component of the membrane (CC), and *Os04g0556400,* involved in hexosyltransferase and UDP-glycosyltransferase activity (MF), and characterized as an intracellular membrane-bounded organelle. Additionally, five DMG-DEGs were identified specifically in the tolerant genotypes (Figure 5.5C). These genes are associated with functional categories such as "Ion transport,"

"Glutathione metabolic process," and "Integral component of membrane," among others (Table 5.3). These findings suggest that these genes may play crucial roles in the epigenetic regulation of rice Al-tolerance levels.

5.3.4 Al exposure Alters Expression of Genes Related to DNA Methylation

Among the differentially expressed genes, those associated with DNA methylation changes in rice will be selected for analysis. Several methyltransferases have been identified in this regard. OsMET1-1 and OsMET1-2 belong to the MET family and are essential for maintaining DNA methylation marks. CMT2 and CMT3 encode chromomethylases (CMT) methyltransferases responsible for CHG methylation (Lanciano & Mirouze, 2017). Finally, DRM1, DRM2 and DRM3 are the domains rearranged methyltransferases responsible for DNA CHH-specific methylation (Cao & Jacobsen, 2002). No orthologous genes for these proteins were found in the *O. glumaepatula* genotypes and no differential expression was observed for these DNA methyltransferases genes (Figure 5.6A). The recently discovered gene coding for the demethylase responsible for DNA demethylation, *OsROS1*, was also evaluated (Agius et al., 2006). This gene showed significant up-regulation only in OG131. Finally, the genes encoding RNA-dependent RNA polymerases (RDR1, RDR2, RDR3 and RDR6) were analyzed, which play a role in generating 24 nt small RNAs (sRNAs) (Matzke & Mosher, 2014). Among these, only the gene OsRDR1 exhibited significant up-regulation in BGI (Figure 5.6B)**.** These results indicate that Al-stress modulates the expression of key genes involved in the dynamic regulation of DNA methylation in rice, although to different extents.

Table 5.2. Functional enriched categories for differentially methylated and expressed genes (DMG-DEGs) under Al-exposure.

Genotype	Source	Term name	Number of genes	Other genotypes sharing the category
AZU	BP	Hydrogen peroxide catabolic process	7	
	CC	Extracellular region	19	
	MF	Tetrapyrrole binding	18	OG97
	MF	Oxidoreductase activity	35	OG97, OG131
	MF	Heme binding	16	OG97
	MF	Monooxygenase activity	12	
	KEGG	Phenylpropanoid biosynthesis	9	

	KEGG	Biosynthesis of secondary metabolites	18	BGI
BGI	BP	Cellular modified amino acid metabolic process	8	
	BP	Glutathione metabolic process	6	
	BP	Response to stimulus	54	
	BP	Export across plasma membrane	5	OG97
	MF	Glutathione transferase activity	6	
	MF	Betaine-homocysteine S-methyltransferase activity	2	
	KEGG	Biosynthesis of secondary metabolites	24	
	KEGG	Lysine degradation	3	
OG97	BP	Export across plasma membrane	4	BGI
	BP	Transmembrane transport	15	OG131
	MF	Xenobiotic transmembrane transporter activity	4	
	MF	Heme binding	10	AZU
	MF	Oxidoreductase activity	20	AZU, OG131
	MF	Tetrapyrrole binding	10	AZU
	MF	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	9	
OG131	BP	Secondary metabolic process	14	
	BP	Phenylpropanoid metabolic process	11	
	BP	Secondary metabolite biosynthetic process	10	
	BP	Protein retention in Golgi apparatus	3	
	BP	Transmembrane transport	35	OG97
	CC	Extracellular region	29	
	CC	Apoplast	13	
	MF	ABC-type transporter activity	8	
	MF	Oxidoreductase activity	44	OG97, AZU
	KEGG	Fatty acid degradation	4	

Table 5.3. Functional characterization of DMG-DEGs shared between Al-tolerant genotypes, AZU and OG97, under Al exposure.

Gene stable ID	Gene name	GO domain	GO term name
Os02g0198700	*OsSub12*	BP	Proteolysis
		MF	Serine-type peptidase activity
		MF	Serine-type endopeptidase activity
		MF	Peptidase activity
		MF	Hydrolase activity
Os03g0575200	*OsHAK16*	BP	Ion transport
		BP	Potassium ion transport
		BP	Potassium ion transmembrane transport
		MF	Potassium ion transmembrane transporter activity
		CC	Integral component of membrane
		CC	Membrane
Os10g0525500	*OsGSTU21*	BP	Glutathione metabolic process
		MF	Glutathione transferase activity
		MF	Transferase activity
		CC	Cytoplasm
Os12g0137700		BP	Sulfation
		MF	Sulfotransferase activity
		MF	Transferase activity
		CC	Cytoplasm
Os12g0637100	*OsPAP10c*	MF	Metal ion binding
		MF	Acid phosphatase activity
		MF	Hydrolase activity

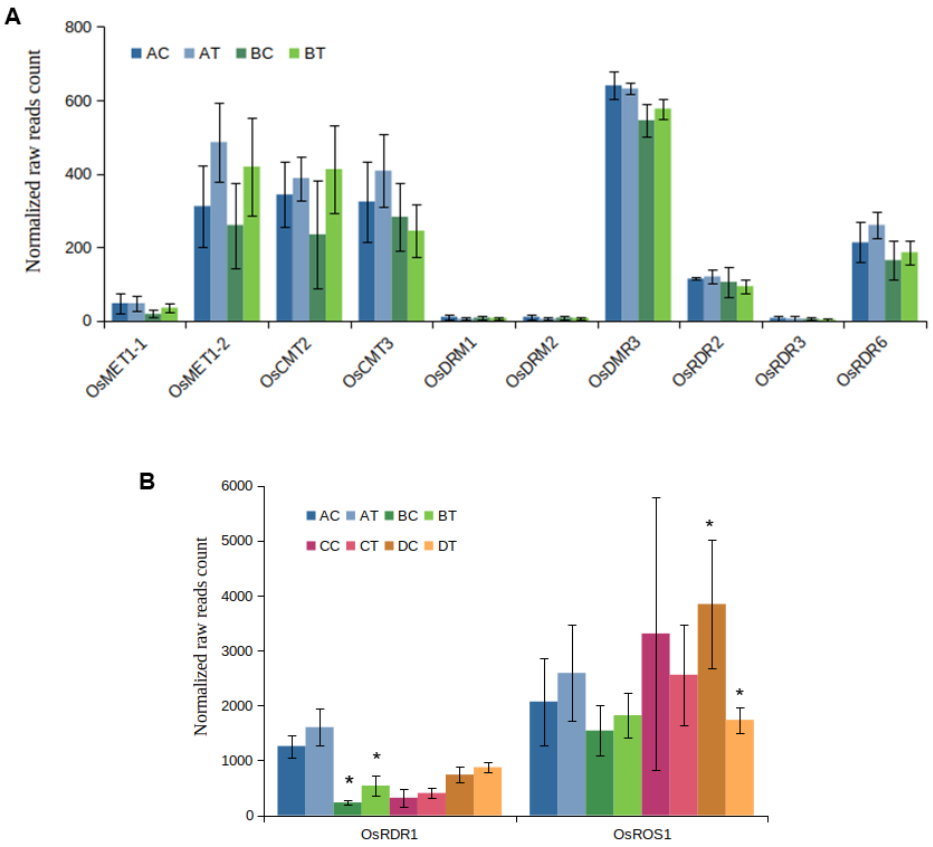


Figure 5.6. Expression values for DNA-methylation and demethylation enzymes. **A.** Normalized reads count values for enzymes identified only in *O. sativa* genotypes and **B.** Normalized reads count values for enzymes genes identified in both *O. sativa* and *O. glumaepatula*. AC: AZU control, AT: AZU treatment, BC: BGI control, BT: BGI treatment, CC: OG97 control, CT: OG97 treatment, DC: OG131 control, DT: OG131 treatment.

5.4 Discussion

As sessile organisms, plants undergo biotic and abiotic stresses and need to evolve diverse physiological mechanisms to cope with them through regulation of gene expression in their genome. Mechanisms such as DNA methylation and demethylation of cytosines are believed to play a key role in this adjustment (Zhang et al., 2018). So far, several studies have reported that genome-wide changes in DNA methylation occur in response to environmental stress (Chang et al., 2020), but the role of these changes in the plant's stress response is still poorly understood. Similarly, there are some previous studies on DNA methylation patterns related to Al-stress exposure as describeb in Chapter 1. However, no studies have been reported in rice plants exposed to Al-stress conditions. Therefore, in this chapter, DNA methylation changes in response to Al-stress conditions were evaluated

in rice cultivated and wild genotypes, as well as its relationship with gene expression using whole genome bisulfite sequencing data.

According to the results, the Al-stress exposure causes genome-wide DNA methylation changes for all the sequence contexts compared with control conditions. It is interesting that variation is homogeneous along the genome and among genotypes. Overall, hypomethylation was observed for CG and CHG context, with bigger changes for CHG. This is in accordance with the idea that plant CHG methylation is more prone to perturbation by environmental stresses than CG methylation (Boyko & Kovalchuk, 2008). In fact, some studies have reported hypomethylation only in the CHG methylation under heavy metal (HM) conditions (Ou et al., 2012). In the same way, various recent studies have shown that HM and other environmental challenges tend to reduce global DNA methylation (Choi & Sano, 2007; Feng et al., 2016; Marconi et al., 2013; Ou et al., 2012; Wang et al., 2011). It is possible that hypomethylation is an effect of the stress condition on DNA e.g. by the production of Reactive Oxygen Species (ROS). These elements are common products of HM stress (Shahid et al., 2014) and possess endonuclease activity and hence can cause DNA damage including double-stranded breaks (DSBs). Jing et al., (2022) demonstrated that ROS play an important role in mediating HM-induced changes in DNA methylation in plants. In fact, other studies have also shown that ROS can result specifically in DNA demethylation (Aina et al., 2004; Bernardo et al., 2017). A possible explanation to this phenomenon is that ROS-induced DNA damage may interfere with the capability of DNA as an acceptor for the methyl groups and result in passive loss of methylation at the affected sites (Jing et al., 2022; Ou et al., 2012).

Genome-wide methylation changes between control and stress conditions, as well as the number hypo-DMR regions reported for both rice species, showed a higher magnitude in the susceptible genotype compared to the tolerant one. This supports the previous hypothesis that tolerant genotypes are more efficient in avoiding genome hypomethylation. The results reported by Aina et al., (2004) also support this idea, as they report DNA hypomethylation in clover (*Trifolium repens* L.) and hemp (*Cannabis sativa* L.) exposed to chromium, nickel, and cadmium, with slightly higher methylation levels observed in the partially tolerant hemp compared to clover.

Importantly, the results highlight the methylation status of plants after 10 days of Al exposure, which is supposed to be an adaptation point for tolerant genotypes. Notably, during the recovery periods, tolerant genotypes exhibited a significantly more efficient reversion of hypomethylation compared to the susceptible genotypes. This finding is in line with the results reported by Marconi et al., (2013), demonstrating that tolerant genotypes display a superior ability to recover from hypomethylation, while the methylation levels in the recovery period remain similar to those of the susceptible variety.

For the CHH context, there was an oppositive trend of global hypermethylation. These opposite DNA methylation patterns among the different sequence contexts were also reported by (Feng et al., 2016), who evaluated the DNA methylation variation in rice plants exposed to Cadmium (Cd). These results suggest that HM stress unevenly affects DNA methylation in the three contexts. As discussed in chapter 1, hypermethylation could act as a defense mechanism to counteract the negative effect of stress conditions. It is possible that CHH methylation is more related to this global protective function than CG and CHG methylation especially because there is a positive correlation between CHH methylation and genes density, as well as higher levels of CHH methylation are present in cytosines closer to genes (Chapter 1). Interestingly, previous studies evaluating DNA methylation under Al-exposure, have reported both hyper or hypomethylation depending on the species or metal doses (Sun et al., 2022). Likewise, multiple studies have shown that patterns in methylation levels vary by HM type, and in some cases conflicting results (Cong et al., 2019; Fan et al., 2020; Jing et al., 2022). This suggests that HM-induced DNA methylation in plants is a relatively complicated process and accordance to the results could also vary according to the methylation sequence context, a factor that could be hidden in several studies that do not use massive sequencing techniques which allows us to identify changes with one nucleotide resolution.

Although DNA methylation changes occur globally along the genome, specific regions that are differentially methylated between control and Al stress conditions were identified. The distribution of DMRs along the genome showed a consistent pattern across genotypes, suggesting a common regulatory mechanism for DNA methylation changes. In particular, a majority of DMRs in CG, CHG and CHH contexts were reported to overlap with TEs including both TEs and gene promoter regions in several cases. It is well known that DNA methylation plays an important role in regulating TEs silencing and thus regulates the genome stability (Law & Jacobsen, 2010; Zhang et al., 2018). However, methylation changes associated with TEs can also impact gene expression, as TEs can influence gene regulation through disruption of cis-regulatory sequences, chromatin alteration, and providing novel regulatory information. (Hirsch & Springer, 2017). Therefore, it was proposed that changes in DNA methylation changes linked to TEs may indirectly affect the expression of Al-responsive genes. For example, Le et al., (2014) reported that DNA demethylases play a role maintaining or positively regulating stress response genes involved in *Fusarium oxysporum* resistance. Interestingly, the downregulated stress-response genes were enriched for short transposable element sequences in their promoters, which showed localized DNA methylation changes in a triple DNA demethylase mutant, RDD (ros1, dml2, dml3), and a general reduction in CHH methylation. These results suggest that RNA-directed DNA methylation (RdDM), responsible for CHH methylation, may participate in DNA demethylase-mediated regulation of stress response genes. Interestingly, according to the results described in this chapter, the *O. sativa*

susceptible genotype (BGI) showed differential expression for the *OsRDR1* gene involved in RdDM pathways. In addition, although no differential expression was shown in the other genotypes, there is also a tendency for up-regulation of the gene that might have decreased over the course of the treatment. Together, these results support the possible role of CHH methylation in the response of plants to stressful conditions.

Furthermore, a higher number of DMGs was found, mostly located in the promoter region rather than the genebody. This result has also been reported previously, where the gene regulatory region has been stated as the major target of methylation/demethylation (Cao & Jacobsen, 2002). Several studies on the HM stress response have reported specific changes in DNA methylation associated with the promoter regions of metal detoxification genes (Feng et al., 2016; Shafiq et al., 2019). Notably, Shafiq et al. (2019) reported hypomethylation of some selected metal detoxification transporters only for the HM (Cd, Zn, Pb) tolerant maize variety. Then, to evaluate the association between methylation changes and specific Al-stress response in this study, a functional enrichment analysis for all the DMGs was performed. As a result, *O. sativa* genotypes did not show any enriched categories, suggesting that DNA methylation changes are occurring for genes involved in several functional networks. However, for *O. glumaepatula* several enriched categories related to stress response were found such as “transmembrane transport”, “oxidoreductase activity” and “ion binding” suggesting a more stress-directed methylation response in the wild-type species than in the cultivated species. To further explore the potential mechanisms of Al tolerance associated with DNA methylation changes the DMGs unique to tolerant genotypes (AZU and OG97) were identified. These tolerant-DMGs were found to be associated with various functional categories related to Al stress, such as “glutathione metabolic process”, “cellular response to DNA damage”, “DNA demethylation”, “response to stimulus”, “transmembrane transport”, “cell wall organization” and “cell redox homeostasis”. These findings further reinforce the earlier proposition that the tolerance mechanism in rice may be associated with DNA demethylation.

Among the identified tolerant-DMGs, the promoter region of the REPRESSOR OF SILENCING 1 (*ROS1*) gene, which encodes a demethylase responsible for DNA demethylation (Agius et al., 2006), showcased hypomethylation in both tolerant genotypes. Pertinently, Lei et al., (2015) reported a unique target sequence for RNA-directed DNA methylation (RdDM) within the *ROS1* promoter region, where DNA methylation plays a critical role in maintaining proper active DNA methylation. They suggested that this specific region acts as a "methyl stat" that senses DNA methylation levels, subsequently regulating DNA methylation through the control of ROS1 expression. Hence, the observed hypomethylation in the *ROS1* promoter region exclusively in the tolerant genotypes, but not in the susceptible ones, provides compelling evidence of a mechanism employed by tolerant genotypes to circumvent hypomethylation under Al-stress conditions. Furthermore, down-regulation of *ROS1* expression was observed solely

in the susceptible genotype OG131 (*O. glumaepatula*), potentially indicating a delayed response of this species to Al-stress conditions. Then, variations in genes encoding epigenetic-modifying proteins could likely account for the distinctive DNA methylation patterns observed in other genes under Al stress.

Profiling both methylome and transcriptome provides a comprehensive analysis of the effect of DNA methylation on gene expression under Al-stress conditions. The findings revealed that DNA methylation variation may be associated with changes in transcript levels in a small group of genes. Interestingly, functional enrichment analysis for this group of DMG-DEGs showed several categories involved in Al stress response, such as "Transmembrane transport" and "Oxidoreductase activity", suggesting that DMG-DEGs predominantly include genes involved in regulating the response to Al-stress. Similar studies conducted on rice cultivars have also reported a limited number of DMG-DEGs in response to stress conditions (Li et al., 2020; Rajkumar et al., 2019). However, they also reported a correlation between DMG-DEGs and stress-related functions. Moreover, it has been reported in several studies that genes involved in heavy metal response could be regulated by DNA methylation changes (Choi & Sano, 2007; Feng et al., 2016). Indeed, Jing et al., (2022) showed that exposure to manganese (Mn) and cadmium (Cd) induces dose-dependent DNA methylation changes at specific sites, suggesting that site-specific methylation patterns, rather than overall methylation levels, play a key role in plant responses to different heavy metals. These findings further support the hypothesis that DNA methylation changes driven by Al stress play a significant role, as several genotype-specific DMG-DEGs were observed that were independently enriched in Al-related functional categories. This suggests that DNA methylation contributes to the plant's response to Al exposure conditions.

To investigate DNA methylation as a key factor in Al tolerant genotypes of rice, DMG-DEGs unique in both tolerant genotypes was selected. It is interesting to note that all the identified DMG-DEGs were related to stress response. For instance, the gene *OsHAK16* encodes a potassium transporter. Intracellular potassium homeostasis undergoes significant alterations during abiotic and biotic stresses, which is crucial for optimal plant metabolic functions. Membrane transporters play a vital role in controlling potassium uptake, efflux, and intracellular relocation (Shabala & Pottosin, 2014). Similarly, the gene *OsPAP10c* encodes an acid phosphatase enzyme that hydrolyzes soil and plant phosphoesters and anhydrides, to release inorganic phosphate for plant acquisition. Acid phosphatases are an important group of enzymes that hydrolyze soil and plant phosphoesters and anhydrides thus releasing inorganic phosphate for plant uptake. Phosphorus is an essential nutrient element that can alleviate exposure to potentially toxic levels of Al, not only by direct immobilization and detoxification of Al through its complexation with phosphorus in the rhizosphere and root tissues but also by stimulating the exudation of Al-chelating organic acids. Interestingly, it has been reported that P-

efficient genotypes exhibit higher Al tolerance compared to P-inefficient genotypes (Wang et al., 2023). Although there are no specific reports on the involvement of *OsHAK16* and *OsPAP10c* in Al tolerance in rice, it is hypothesized that the up-regulation of these DNA methylation-regulated genes enhances tolerance levels in rice plants. Further investigation of these tolerant DMG-DEGs, which show potential epigenetic regulation, is worthwhile in order to explore their potential utility in improving susceptible genotypes.

Concluding Remarks

In conclusion, this study revealed similar genome-wide variation in DNA methylation across rice genotypes of both *O. sativa* and *O. glumaepatula,* with a consistent trend of hypomethylation in the CG and CHG sequence context, and hypermethylation in the CHH sequence context. Hypomethylation appeared to be a common response to Al stress, potentially influenced by reactive oxygen species (ROS) interactions with DNA. Conversely, hypermethylation in the CHH context may serve a protective role in preventing DNA damage. Importantly, tolerant genotypes exhibited lower levels of hypomethylation, suggesting that this could be a contributing strategy to their higher Al tolerance. Furthermore, the distribution of DMRs between the stress and control condition also showed a consistent pattern across all the genotypes, with a higher number of DMRs associated with TEs. These results together suggest the existence of a shared genome-wide response in the rice methylome under Al exposure. However, specific DNA methylation responses also exist. Supporting this notion, a group of genes that were both differentially methylated and expressed (DMG-DEGs) were identified, predominantly associated with stress-responsive functions. This suggests the presence of epigenetic regulation of these genes in response to Al exposure. Moreover, a set of DMG-DEGs unique to Al-tolerant genotypes were discovered, which are functionally related to previously known stress pathways such as phosphatase activity and potassium transporters. These genes have received limited attention in previous studies, suggesting their potential involvement in epigenetically regulated high levels of Al tolerance. This study represents the first exploration of epigenetic regulation under Al stress conditions in rice, providing fundamental knowledge towards understanding the diverse regulatory factors influencing Al tolerance. Further research is warranted to investigate the precise mechanisms underlying the epigenetic control of Al tolerance and the potential application of these findings in improving crop resilience to Al stress.

Chapter 6

Methylation in the CHH Context Allows to Predict Recombination in Rice: Another Function of DNA Methylation in Rice

This chapter was previously published as:

Peñuela, M., **Gallo-Franco, J.J.,** Finke, J., Rocha, C., Gkanogiannis, A., Ghneim-Herrera, T., Lorieux, M. (2022) Methylation in the CHH Context Allows to Predict Recombination in Rice. Int. J. Mol. Sci. 2022, 23, 12505.

Chapter Summary

DNA methylation is the most studied epigenetic trait. It is considered a key factor in regulating plant development and physiology and has been associated with the regulation of several genomic features, including transposon silencing, regulation of gene expression, and recombination rates. Nonetheless, understanding the relation between DNA methylation and recombination rates remains a challenge. This work explores the association between recombination rates and DNA methylation for two commercial rice varieties. The results show negative correlations between recombination rates and methylated cytosine counts for all contexts tested at the same time, and for CG and CHG contexts independently. In contrast, a positive correlation between recombination rates and methylated cytosine count is reported in CHH contexts. Similar behavior is observed when considering only methylated cytosines within genes, transposons, and retrotransposons. Moreover, it is shown that the centromere region strongly affects the relationship between recombination rates and methylation. Finally, machine learning regression models are applied to predict recombination using the count of methylated cytosines in the CHH context as the entrance feature. These findings shed light on the understanding of the recombination landscape of rice and represent a reference framework for future studies in rice breeding, genetics, and epigenetics.

Keywords: *Bisulfite sequencing, DNA methylation, Machine learning, Modeling, Prediction, Recombination*

Connection to Previous Chapters

Up to chapter 5 of this dissertation, the possible role of DNA methylation in the stress response of tolerant and susceptible rice genotypes was discussed. However, it is crucial to understand all the factors influencing such methylation patterns and their indirect or direct impacts on plant breeding strategies. One such factor is the relationship between DNA methylation and recombination rates, which plays a pivotal role in generating tolerant varieties. Therefore, as the final chapter of this thesis, the analysis of DNA methylation was extended to explore its relationship with recombination prediction. The bisulfite sequencing data obtained in Chapter 3 served as the basis for this analysis. By incorporating DNA methylation as an input trait for recombination prediction, this chapter provides valuable insights into the broader implications of epigenetics in plant breeding. It offers a different perspective and sheds light on the intriguing contrast observed in the methylation patterns of the CHH sequence context compared to the CG and CHG contexts discussed in previous chapters. The findings presented in this final chapter contribute to a deeper understanding of the intricate relationship between DNA methylation, recombination, and the generation of genetically diverse and stress-tolerant rice varieties.

6.1 Introduction

Meiotic recombination is recognized as a key process in genetics. During this process, maternally and paternally inherited homologous chromosomes exchange information by gene conversion or crossing over to create novel allelic combinations. Recombination is widely recognized for its role in promoting diversity to respond to continually shifting environments, in addition to preventing the build-up of genetic load by decoupling linked deleterious and beneficial variants (Rodgers-Melnick et al., 2015). However, meiotic recombination between homologous chromosomes is restricted by the number and location of crossover sites per chromosome. The crossover distribution and frequency along the genome are uneven, especially in plants (Lambing et al., 2017). Sites with high recombination rates have been linked to subtelomeric regions that are generally hypomethylated and have high gene and DNA transposon frequencies. In contrast, recombination is suppressed in the centromeric region, which is characterized by high frequencies of long terminal repeat retroelements and few genes (Henderson, 2012).

The role of chromatin structure and DNA methylation in determining recombination rates has been previously reported. For example, high levels of histone H3 acetylation in Arabidopsis mutants were associated with changes in the crossover frequencies (Perrella et al., 2010). Likewise, studies using met1 and ddm1 mutants, which are globally hypomethylated, showed regional remodeling of crossover frequencies with increased recombination in chromosome arms and decreased recombination in the pericentromeric region (Melamed-Bessudo & Levy, 2012; Mirouze et al., 2012). However, understanding how the DNA methylation patterns affect the recombination rates remains an open challenge.

Identifying factors influencing the meiotic recombination rates is important for breeders interested in transferring genes from one variety to another through crosses. Thus, developing new allelic combinations that allow breeders to meet the needs present in agricultural systems. Recently, several studies have addressed this issue and have developed different types of strategies to discover where crossovers occur most frequently and try to predict them. For example, Liu et al. (2016) developed a predictor of recombination hot/cold spots in yeast using a machine learning approach combined with principal component analysis. Demirci et al. (2018) explored DNA sequence and shape features to train machine learning models for predicting crossover occurrence in Arabidopsis (*Arabidopsis Thaliana*), maize (*Zea mays*), tomato (*Solanum lycopersicum*), and rice (*Oryza sativa*). Moreover, Adrion et al. (2020) used recurrent neural networks, a deep learning method for estimating genome-wide recombination in a natural population of African *Drosophila melanogaster*.

In recent years, rice has been a model monocotyledonous plant for several research approaches. However, few studies have analyzed methylation patterns in relation to recombination rates in rice. For instance, Habu et al. (Habu et al., 2015) developed an experiment crossing methylated and unmethylated rice varieties and concluded that the position and frequency of meiotic recombination in rice centromeric heterochromatin are regulated by the epigenetic state of the chromatin. Likewise, Choi & Purugganan (2018) explore how transposable elements interact with host plant epigenetics. They suggest that high levels of methylation at these elements have a role in suppressing deleterious ectopic recombination. Nevertheless, none of these studies have explored in detail how methylation contexts are related with recombination rates.

In this chapter, the relationship between chromosomal recombination rates and DNA methylation is explored by using *Oryza sativa* as a model. The objectives of this study are as follows: (i) To estimate the correlation between recombination and DNA methylation across all sequence contexts, (ii) to describe the effect of methylation within genes, transposons, and retrotransposons with respect to recombination, and (iii) to implement a machine learning model to predict recombination based on methylation data. The results of this study provide evidence that recombination can be characterized by methylation patterns specifically in the CHH context, regardless of their location within or inside genes, transposons, and retrotransposons. Based on these results, it is proposed the utilization of machine learning models to predict chromosomal recombination rates in rice cultivars by leveraging CHH methylation data.

6.2 Materials and Methods

6.2.1 Recombination Rates

The recombination rates values used for comparisons in this chapter were obtained from Peñuela et al. 2023. The methods used are described briefly below. The recombination rates were assessed in a population of 212 F11 recombinant inbred lines (RILs) derived from a cross between the rice varieties IR64 (indica group) and Azucena (tropical japonica group). These RILs were generated through a single seed descent method. To obtain genotypic information, shallow Illumina sequencing with a coverage of approximately 2× was performed on the population, followed by an imputation process using NOISYmputer. To calculate local recombination rates, sliding windows of 100 kb were employed, and the rates were expressed in centimorgans per base pair (cM/bp). This analysis was conducted using MapDisto software, allowing for the estimation of recombination rates at a fine-scale resolution.

6.2.2 Plant Material and Growth Conditions for Methylation Experiment

Seeds of rice varieties IR64 and Azucena were germinated and grown in a growth chamber at 30 °C and 12:12 dark/light conditions for 10 days. Seedlings were transferred to a hydroponic medium with a Kimura B solution (pH 7) and Arnon micronutrients. Roots from three weeks-old seedlings were collected and stored at −80 °C. Total genomic DNA was extracted from frozen root tissue by CTAB 2X protocol with modifications (Maropola et al., 2015). Genomic DNA quality was evaluated on agarose gels, and DNA quantity was measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

6.2.3 Whole-Genome Bisulfite Sequencing and Data Analysis

Bisulfite-seq (BS-seq) libraries were made from genomic DNA isolated from IR64 and Azucena seedling roots. DNA from three independent seedlings for each genotype was pooled as one sample and sequenced. Bisulfite conversion of DNA, library construction, and sequencing were performed by CD Genomics (CD Genomics Inc., Shirley, New York, NY, USA). Raw data are available in the GenBank repository for IR64 (Accession number: SRR20325840) and Azucena (Accession number: SRR20325842). Basic statistics on the quality of the raw reads was done with the FastQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 5 September 2021)). Sequencing adapters and low-quality data of the sequencing data were removed by Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic (accessed 21 Novem-ber 2021)). Cleaned data were aligned to the reference genomes reported in the GenBank repository for IR64 (Accession number: RWKJ00000000) and Azucena (Accession number: PKQC000000000) using Bismark v.0.16.3 (Krueger & Andrews, 2011) with default parameters. Only uniquely aligned reads were maintained. Methylation calling data obtained from Bismark were used for further analysis.

6.2.4 Comparison between Recombination Rates and Methylation Patterns

To compare the methylation patterns with the local recombination rates, the genomes were divided into 100 kb windows. Within each window, the number of cytosines exhibiting a methylation level greater than 75% was calculated separately for the CG, CHG, and CHH sequence contexts. To mitigate the influence of noise arising from abrupt changes in the count of methylated cytosines between adjacent windows, exponential smoothing with a smoothing parameter (α) of 0.1 was applied to both the recombination and methylation data. This smoothing process helped to achieve a more consistent

representation of the underlying trends. Subsequently, a Pearson correlation analysis per chromosome was developed to evaluate the linear relationships between the recombination rates and the methylation patterns of both varieties.

6.2.5 Functional Evaluation

Pearson correlation analyses were conducted to examine the relationship between the number of genes, transposons, and retrotransposons and the recombination landscape of the chromosome. This analysis aimed to investigate whether these genomic elements were associated with variations in recombination rates. Subsequently, the start and end coordinates of these functional elements were utilized to extract the count of methylated cytosines within each element. New correlation analyses were then performed to explore the trends between methylated cytosines in each sequence context (CG, CHG, and CHH) within these functional elements and their relationship with recombination. A differentiation between the centromere and non-centromere regions was also included.

6.2.6 Machine Learning Modeling

To assess the usefulness of methylation in predicting chromosome recombination, different machine learning approaches were explored. The total counts of methylated cytosines in windows of 100 kb belonging to the CG, CHG, and CHH contexts for each genotype were evaluated as features for machine learning modeling using the Shapley package (https://shap.readthedocs.io/en/latest/index.html; accessed on 2 February 2022). Subsequently, the performance of different machine learning models was evaluated using the LazyPredict package (https://pypi.org/project/lazypredict/; accessed on 2 February 2022). Exponential smoothing with $\alpha = 0.1$ was applied to the data input before training the model and another one to the model output with $\alpha = 0.3$. The coefficient of determination R2 and the root of the mean square error RMSE were used to evaluate the performance of the models. MSE was used for predictions. Pearson correlation analyses were also performed to discover general linear trends between the predictions and the experimental data. The resulting best model was fitted, and the information from the twelve chromosomes of one variety was used as a training dataset to predict the recombination rates in each of the twelve chromosomes of the other variety. All these analyses and the previous ones were run in Python.

6.3 Results and Discussion

In this study, the correlation between recombination rates and the methylated cytosine counts for all chromosomes in two rice cultivars is evaluated (Figures 6.1 and 6.2). The

correlation values are, on average, −0.44 ± 0.17 for all chromosomes of both varieties, with higher values in the centromere region. Similar results in rice were previously described by Yan et al. (2010), revealing that DNA methylation patterns in the centromere are shaped by the DNA sequence and the centromeric domains. Habu et al. (2015) described how artificial chromatin modification can vary the frequency of meiotic recombination. Overall, high levels of methylation in heterochromatin regions near the centromeres have been reported as a common pattern, where meiotic recombination is repressed. In the same way, recombination-free regions around centromeres are likely to be important for normal centromere function during meiosis (Habu et al., 2015; Yan et al., 2005).

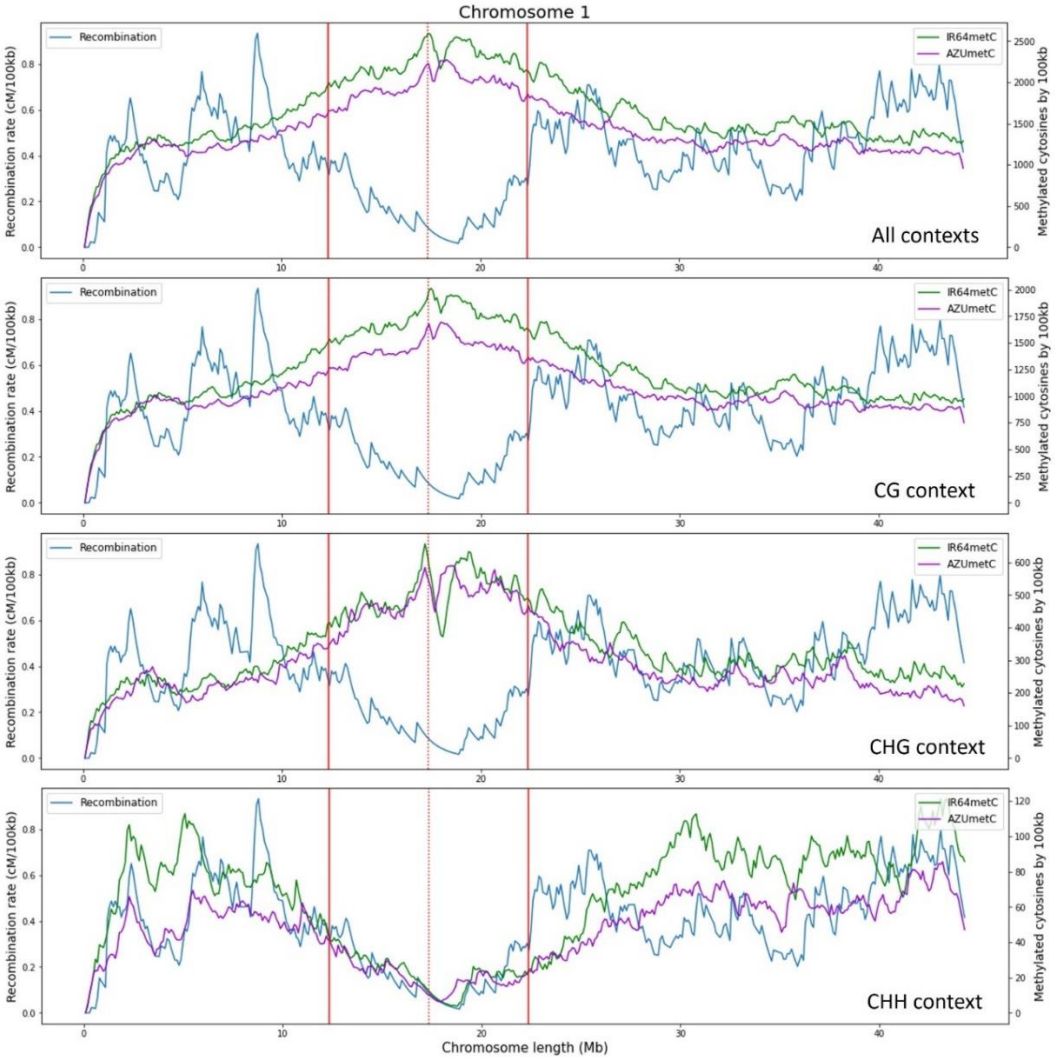


Figure 6.1. Recombination and methylated cytosines through Chromosome 1 for the rice varieties IR64 and Azucena. The centromere is represented by a red dotted line and the influence of the centromere region by solid red lines.

	All Contexts		CG Context		CHG Context		CHH Context	
Chromosome	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena
1	-0.17	-0.41	-0.43	-0.44	-0.45	-0.54	0.62	0.66
2	-0.34	-0.50	-0.42	-0.41	-0.52	-0.61	0.68	0.77
3	-0.54	-0.57	-0.56	-0.50	-0.66	-0.64	0.67	0.72
4	-0.55	-0.63	-0.55	-0.61	-0.68	-0.72	0.65	0.81
5	-0.50	-0.57	-0.59	-0.53	-0.70	-0.71	0.82	0.84
6	-0.35	-0.47	-0.44	-0.42	-0.55	-0.59	0.78	0.86
7	-0.47	-0.35	-0.46	-0.23	-0.53	-0.48	0.40	0.71
8	-0.42	-0.64	-0.62	-0.60	-0.67	-0.75	0.90	0.89
9	0.06	-0.11	-0.13	-0.16	-0.25	-0.34	0.65	0.78
10	-0.42	-0.55	-0.47	-0.45	-0.58	-0.65	0.62	0.71
11	-0.59	-0.64	-0.55	-0.54	-0.66	-0.69	0.73	0.82
12	-0.45	-0.60	-0.53	-0.44	-0.59	-0.67	0.81	0.87

Figure 6.2. Correlations between recombination rates and the count of methylated cytosines for rice varieties IR64 and Azucena. Blue and red colors correspond to positive and negative correlations, respectively. The higher the correlation value, the higher the color intensity.

By evaluating the CG and CHG methylation contexts independently, a decrease in recombination rates with increasing methylated cytosines is reported. On the contrary, methylated cytosines in the CHH context increase with recombination rates showing a positive correlation (Figure 6.3). The opposite relationship between the methylation contexts of CG and CHH has been reported in rice by Li et al. (2012), who identified the tendency to-wards hypermethylation in CG context, but hypomethylation in CHH.

The positive correlation between methylated cytosine count and recombination rates observed in the context of CHH is not clear when all methylation contexts are assessed together because the total number of methylated cytosines in the CG and CHG contexts was higher. This trend is observed for both varieties, IR64 and Azucena, where the methylation data and the alignment process have been obtained independently. The positive relationship between the CHH methylated cytosine count and recombination rates has been reported by Rodgers-Melnick et al. (2015), who include the CHH methylation as a feature of a linear model to predict recombination in maize. It is unclear what role methylated cytosines play in the CHH context with respect to recombination.

Variability in DNA methylation can be heritable or reversible, and this can allow for phenotypic variation and rapid response to environmental changes. Even the degree of intraspecies epigenomic diversity can be correlated with climate and geographic origin (Lanciano & Mirouze, 2017). It has been reported that CHH methylation could be related to fruit size in apples (Daccord et al., 2017) and silencing transposons in sugar beets (Zakrzewski et al., 2017). A potential role in *A. thaliana* seed dormancy, with increases in

CHH methylation in seeds during seed development and a decrease during germination, has also been reported in (Zhang et al., 2018). These observations suggest the multiple roles that CHH methylation can play in plant genomes. Recently, Wang et al. (Wang et al., 2022) reported that CHH methylation levels are higher in rice reproductive organs, such as panicles and pistils, than in seedlings, suggesting a positive feedback loop between DNA methylation and RNA-directed DNA methylation activity involved in sexual reproduction.

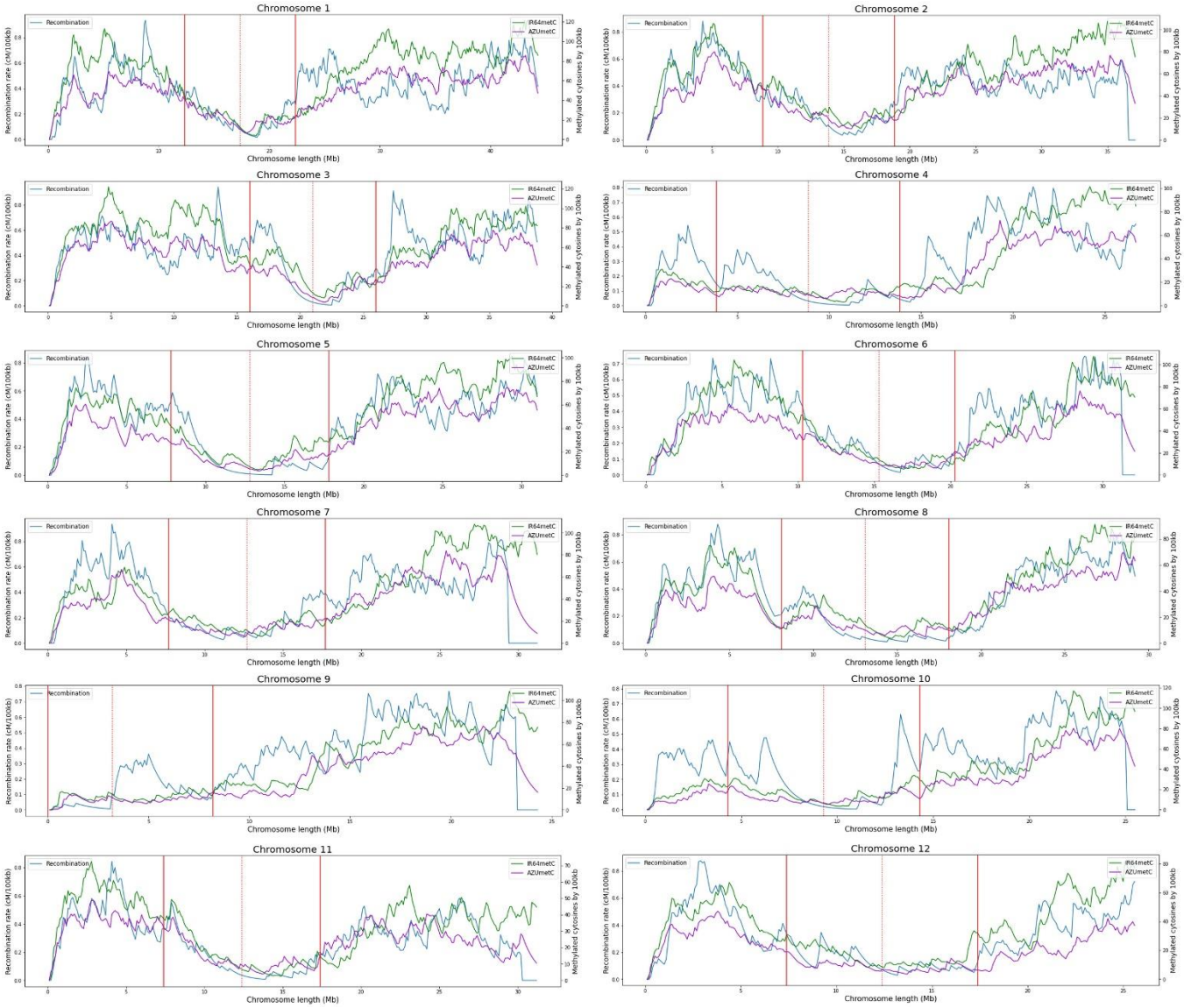


Figure 6.3. Distribution of methylated cytosines in CHH context in the twelve rice chromosomes for the IR64 and Azucena varieties, in comparison with the chromosomal recombination between these two varieties. The centromere is represented by a red dotted line and the influence of the centromere region in recombination by solid red lines.

The functional analysis performed with annotation data of genes, transposons, and retrotransposons for each variety, shows that the increment in the number of genes per window is correlated with recombination rates in the chromosomes of both varieties (Figures 6.4 and 6.5). This positive trend has been previously evidenced in Drosophila, *A. thaliana*, yeast, finches, monkeyflowers, and dogs, with recombination hotspots typically located near the promoter regions of genes (Kent et al., 2017) and observed in the euchromatic regions of maize (Anderson et al., 2006). In contrast, a negative correlation between the number of transposons and retrotransposons has been found with respect to recombination rates across all chromosomes for both rice varieties. This can be explained by the abundance of such elements near the centromere where recombination rates are low. Similar results have been found by Tian et al. (2009), who suggested that the rice genome is organized along recombinational gradients due to the negative correlation of recombination with transposable elements and positive one with gene densities.

Recombination tends to occur within and near genes and away from transposable elements. This may reflect the passive effects of recombination initiating in open chromatin (Kent et al., 2017). Recent analyses of the localization of recombination at the fine scale tend to show negative correlations with local densities of repetitive elements. Actually, strong recombination suppression and a large accumulation of transposable elements are usual in peri-centromeric regions (Kent et al., 2017). For rice, this pattern is shared between japonica and indica groups (Tian et al., 2009). There remains uncertainty about the directionality of cause and effect, the extent to which the correlation is driven by associations of both recombination and transposable elements with other factors, or why patterns differ among species and types of repetitive elements (Kent et al., 2017).

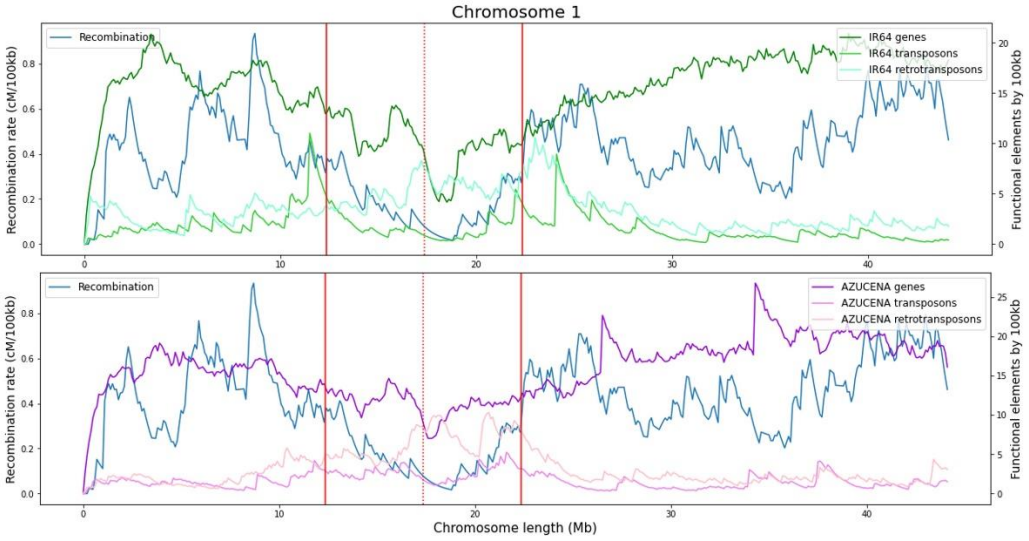


Figure 6.4. Genes, transposons, and retrotransposons compared to cross over recombination through chromosome 1 for rice varieties IR64 and Azucena. The centromere is represented by a red dotted line and the influence of the centromere region in recombination by solid red lines.

	Complete chromosome						Chromosome arms						Centromere region					
	Genes		Transposons		Retrotransposons		Genes		Transposons		Retrotransposons		Genes		Transposons		Retrotransposons	
Chromosome	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena
1	0.61	0.49	0.12	-0.34	-0.24	-0.52	0.25	0.13	0.06	-0.10	0.18	0.00	0.63	0.52	0.78	0.45	-0.44	-0.33
2	0.73	0.70	0.13	0.05	-0.39	-0.65	0.47	0.54	0.29	0.27	0.02	0.04	0.81	0.52	0.29	0.49	-0.70	-0.64
3	0.66	0.69	0.14	0.24	-0.57	-0.63	0.41	0.39	0.22	0.16	-0.04	0.07	0.72	0.76	0.03	0.42	-0.46	-0.88
4	0.55	0.63	0.10	-0.32	-0.60	-0.73	0.25	0.49	-0.07	-0.47	-0.27	-0.51	0.84	0.54	0.46	0.08	-0.71	-0.42
5	0.73	0.79	-0.29	-0.43	-0.75	-0.79	0.48	0.55	-0.13	0.06	-0.24	-0.33	0.60	0.70	0.28	-0.13	-0.65	-0.71
6	0.79	0.77	0.24	-0.26	-0.47	-0.62	0.71	0.69	0.05	-0.16	0.01	-0.12	0.75	0.71	0.36	-0.49	-0.52	-0.75
7	0.74	0.80	-0.09	-0.49	-0.68	-0.74	0.62	0.67	0.09	-0.22	-0.03	-0.19	-0.16	0.33	-0.35	0.11	-0.66	-0.74
8	0.84	0.84	-0.03	-0.05	-0.46	-0.61	0.83	0.75	-0.14	0.10	-0.26	-0.32	0.44	0.54	0.41	-0.07	-0.05	-0.46
9	0.79	0.80	-0.18	-0.27	-0.37	-0.73	0.65	0.67	-0.50	-0.36	-0.65	-0.78	0.12	0.22	-0.39	-0.09	0.73	0.51
10	0.74	0.64	-0.16	0.00	-0.65	-0.57	0.63	0.67	-0.19	-0.19	-0.01	-0.28	0.75	0.26	0.32	0.04	-0.76	-0.45
11	0.76	0.76	-0.10	-0.09	-0.67	-0.67	0.60	0.68	-0.30	0.03	-0.24	-0.04	0.80	0.86	0.06	0.36	-0.81	-0.80
12	0.88	0.80	-0.24	0.27	-0.43	-0.44	0.82	0.71	-0.38	0.19	0.12	-0.15	0.72	0.85	0.12	0.69	-0.56	-0.10

Figure 6.5. Correlations between recombination rates and the number of genes, transposons, and retrotransposons rates for rice varieties IR64 and Azucena. Blue and red colors correspond to positive and negative correlations, respectively. The higher the correlation value, the higher the color intensity.

The count of methylated cytosines was assessed within genes, transposons, and retrotransposons and compared to recombination rates. The analysis showed that methylated cytosine count in genes, transposons, and retrotransposons is negatively correlated with recombination rates when evaluated for all contexts together. This indicates that methylation inside these entities is higher when recombination is lower. The same negative trend is observed when methylated cytosines are analyzed in CG and CHG contexts. Methylation events in transposons and retrotransposons are associated with the prevention of their expression and movement in chromosomes, which can be damageable to the organism and even deleterious (Ahmed et al., 2011; Kent et al., 2017). It should be noted that these methylation events can also affect surrounding genomic regions (Ahmed et al., 2011), potentially influencing the methylation status of nearby genes. In genes, methylation usually occurs at the promoters or within the body of the transcribed gene, inhibiting their expression (Zhang et al., 2018). However, the methylated cytosines in the CHH context are also positively correlated with the recombination rates. This is a consequence of low CHH methylation near the centromere region and greater presence in the chromosome arms. Gallo-Franco et al. (Gallo-Franco et al., 2020) reported high CHH methylation levels of transposable elements close to genes in rice, which supports the conclusion of Martin et al. (Martin et al., 2021) for grass species that long genes and genes close to transposable elements tend to have CHH islands more frequently. It could be hypothesized that the presence of these CHH islands is promoting the positive correlation between methylation and recombination in gene-rich regions.

Chromosomal regions close to the centromere have a high incidence on DNA methylation. When only the chromosome arms are evaluated, correlation trends change,

from being high negative to being negative, for all contexts evaluated together and for the CG and CHG contexts evaluated independently (Figure 6.6). For CHH methylation, the markedly positive correlation also decreases but is still positive. In the context centromere regions are evaluated, negative correlations are evidenced in all contexts when they are evaluated together and for CG and CHG contexts independently. These results are in agreement with the reported importance of DNA methylation for plant chromosomal interactions in peri-centromeric regions (Kawashima & Berger, 2014). They also agree with the results obtained by Habu et al. (2015), who indicate that the position and frequency of meiotic recombination in the centromeric heterochromatin of rice are regulated by the epigenetic state of the chromatin. With respect to methylation in CHH contexts, the correlation of the centromere region is positive but weaker than that of the whole chromosome (Figure 6.6).

	All Contexts						CG Context					
	Complete Chromosome		Chromosome arms		Centromere region		Complete Chromosome		Chromosome arms		Centromere region	
Chromosome	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena
1	-0.17	-0.41	0.26	0.25	-0.22	-0.68	-0.43	-0.44	0.21	0.23	-0.78	-0.83
2	-0.34	-0.50	0.22	0.21	-0.86	-0.86	-0.42	-0.41	0.22	0.33	-0.84	-0.89
3	-0.54	-0.57	0.25	0.22	-0.77	-0.80	-0.56	-0.50	0.14	0.22	-0.85	-0.82
4	-0.55	-0.63	-0.23	-0.22	-0.71	-0.59	-0.55	-0.61	-0.25	-0.26	-0.19	-0.61
5	-0.50	-0.57	0.09	0.07	-0.78	-0.69	-0.59	-0.53	0.06	0.10	-0.67	-0.32
6	-0.35	-0.47	0.22	0.16	-0.79	-0.69	-0.44	-0.42	0.12	0.23	-0.83	-0.66
7	-0.47	-0.35	0.39	0.46	-0.22	-0.36	-0.46	-0.23	0.08	0.49	-0.37	-0.11
8	-0.42	-0.64	-0.25	-0.31	-0.59	-0.69	-0.62	-0.60	-0.31	-0.31	-0.70	-0.45
9	0.06	-0.11	0.01	0.04	0.62	0.67	-0.13	-0.16	-0.32	-0.13	0.67	0.63
10	-0.42	-0.55	0.15	0.15	-0.74	-0.82	-0.47	-0.45	0.03	0.05	-0.60	-0.60
11	-0.59	-0.64	-0.24	-0.24	-0.78	-0.87	-0.55	-0.54	-0.16	-0.15	-0.84	-0.91
12	-0.45	-0.60	-0.25	-0.28	-0.38	-0.69	-0.53	-0.44	-0.09	-0.12	-0.64	-0.78
	CHG Context						CHH Context					
	Complete Chromosome		Chromosome arms		Centromere region		Complete Chromosome		Chromosome arms		Centromere region	
Chromosome	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena	IR64	Azucena
1	-0.45	-0.54	0.17	0.08	-0.43	-0.74	0.62	0.66	0.25	0.35	0.77	0.72
2	-0.52	-0.61	0.10	0.02	-0.88	-0.87	0.68	0.77	0.36	0.55	0.80	0.74
3	-0.66	-0.64	0.00	0.04	-0.82	-0.84	0.67	0.72	0.38	0.43	0.79	0.88
4	-0.68	-0.72	-0.31	-0.45	-0.73	-0.61	0.65	0.81	0.48	0.74	0.64	0.63
5	-0.70	-0.71	-0.07	-0.13	-0.80	-0.75	0.82	0.84	0.59	0.63	0.68	0.80
6	-0.55	-0.59	-0.05	-0.06	-0.83	-0.75	0.78	0.86	0.60	0.73	0.86	0.88
7	-0.53	-0.48	-0.01	0.27	-0.28	-0.42	0.40	0.71	-0.03	0.52	0.60	0.62
8	-0.67	-0.75	-0.31	-0.48	-0.67	-0.75	0.90	0.89	0.85	0.81	0.69	0.82
9	-0.25	-0.34	-0.23	-0.19	0.63	0.68	0.65	0.78	0.41	0.66	-0.21	-0.10
10	-0.58	-0.65	0.00	-0.18	-0.76	-0.83	0.62	0.71	0.46	0.65	0.73	0.71
11	-0.66	-0.69	-0.28	-0.34	-0.84	-0.89	0.73	0.82	0.49	0.67	0.93	0.90
12	-0.59	-0.67	-0.15	-0.40	-0.49	-0.73	0.81	0.87	0.68	0.80	0.73	0.83

Figure 6.6. Correlations between recombination rates and the count of methylated cytosines for complete chromosomes, chromosome arms, and centromere region of rice varieties IR64 and Azucena. Blue and red colors correspond to positive and negative correlations, respectively. The higher the correlation value, the higher the color intensity.

The contributions of methylation in CG, CHG, and CHH contexts to predict recombination as features of machine learning models are assessed using the Shapley package. The results show a great contribution of CHH for the prediction of recombination and a low contribution of CG and CHG for both varieties (Supplementary figure 6.1). This agrees with the fact that the CHH context has the highest correlation values with respect to chromosome recombination rates, while the CG and CHG contexts have lower correlations. The Shap summary plot also shows the same trend, evidencing the strongest effect on recombination when the CHH values are higher.

Subsequently, the methylated cytosine count in the CHH context is used as a unique feature to evaluate regression algorithms of machine learning, because the performance of the model decreases when the other features are considered. The evaluation is carried out independently for each variety using the Lazy Predict package. The results show that the Extra Trees algorithm performed the best prediction (R2 = 0.57, RMSE = 0.01 for IR64; R2 = 0.69, RMSE = 0.01 for Azucena). Thus, this algorithm is used to develop training and subsequent predictions.

Predictions on Azucena's chromosomes, by training the Extra Trees algorithm with information from IR64, give an R2 of 0.32 ± 0.13 and an MSE of 0.02 ± 0.00, on average. Meanwhile, predictions on IR64's chromosomes by training the Extra Trees algorithm with information from Azucena give an R2 of 0.21 ± 0.21 and an MSE of 0.03 ± 0.00, on average. In both cases, the average correlation values between predictions and recombination rates are 0.67 ± 0.06 for Azucena and 0.65 ± 0.07 for IR64, evidencing a positive trend (Table 6.1, Figure 6.7).

Several studies have focused on predicting recombination using machine learning. For example, Liu et al. (2016) combined support vector machines with consensus feature dinucleotide-based autocross covariance to predict the recombination of hot/cold spots in yeast. Demirci et al. (2018) used features, such as gene annotation, propeller, and helical twist, AT/TA dinucleotides, and CA dinucleotides to train machine learning models for predicting crossover occurrences in Arabidopsis, maize, rice, and tomato. More recently, Adrion et al. (2020) proposed an approach to predict the recombination landscape in African populations of *Drosophila melanogaster* using deep learning with recurrent neural networks. For all cases, the results have been satisfactory according to the specific objective of each study, which demonstrates the power of machine learning approaches to predict complex traits such as chromosomal recombination.

The Extra Trees regression model makes it possible to predict chromosomal recombination using a single feature: The CHH methylated cytosine count. It is possible due to the high correlation between this feature and the recombination rates, which behaved similarly in all chromosomes. The model was trained on a dataset of one variety and was tested on the other, performing two independent tests and finding that results were consistent (Figure 6.8). This opens a door for future studies. The evidence suggests that these models can be used to predict chromosomal recombination rates in any variety of

Oryza sativa rice. This is because the two varieties used in this study, IR64 and Azucena, are highly distant genetically, belonging to the indica and japonica groups, respectively.

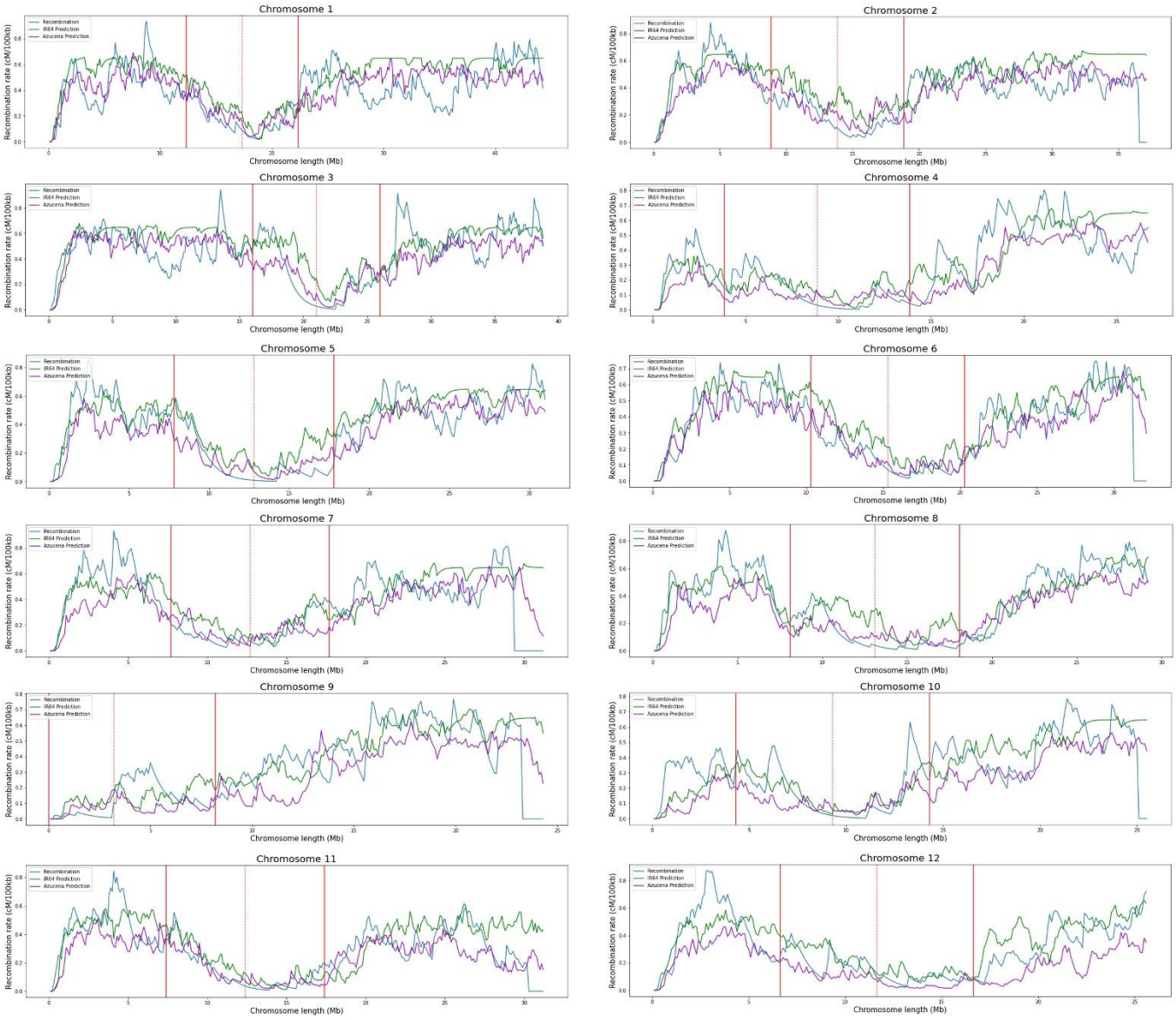


Figure 6.7. Recombination predictions between IR64 and Azucena varieties by the Extra Trees machine learning model using the count of methylated cytosines in the CHH context as a feature. Predictions on the IR64 manifold are made using Azucena methylation as the training dataset, and predictions on the Azucena manifold are made using IR64 methylation as the training dataset. The centromere is represented by a red dotted line and the influence of the centromere region in recombination by solid red lines.

Table 6.1. Performance of chromosome recombination rates predictions of IR64 and Azucena rice varieties using the Extra Trees model trained with CHH methylation data.

Chromosome	IR64			Azucena		
	R^2	Correlation	MSE	R^2	Correlation	MSE
1	0.00	0.63	0.03	0.44	0.67	0.02
2	0.04	0.66	0.03	0.53	0.73	0.01
3	0.37	0.70	0.02	0.49	0.72	0.02
4	0.44	0.72	0.02	0.60	0.81	0.01
5	0.59	0.81	0.02	0.67	0.84	0.01
6	0.44	0.78	0.02	0.68	0.82	0.01
7	0.16	0.53	0.04	0.50	0.73	0.02
8	0.71	0.85	0.01	0.67	0.88	0.02
9	0.32	0.65	0.03	0.50	0.75	0.02
10	0.41	0.70	0.02	0.28	0.69	0.03
11	0.30	0.70	0.02	0.52	0.77	0.01
12	0.54	0.77	0.01	0.35	0.85	0.02

Concluding Remarks

This chapter reveals that methylated cytosines in the CHH context positively correlate with recombination rates in the twelve rice chromosomes of the IR64 and Azucena varieties. Conversely, a negative correlation is observed for CG and CHG contexts, as well as for all three methylation contexts combined. In addition, functional analysis showed that genes were positively correlated with recombination rates, unlike transposons and retrotransposons, which showed a negative correlation. The correlation between methylation and recombination suggests the same trends for the entire genome with respect to only methylation in genes, transposons, and retrotransposons. The influence of the centromere on methylation patterns and its correlation with recombination rates was evident, supporting the hypothesis that the position and frequency of meiotic recombination in rice centromeric heterochromatin are regulated by the epigenetic state of the chromatin. Finally, a machine learning model was proposed and trained using the CHH methylated cytosine count to predict recombination rates, which obtained consistent results in two independent data sets. This suggests that the extraction of methylation data and the use of machine learning models in future studies is a promising path to focus on predicting recombination rates using the count of CHH-methylated cytosines in rice as a feature.

Chapter 7

Conclusion and Future Work

This chapter presents the conclusion of this doctoral dissertation, followed by a discussion on research directions opened by it.

7.1 Conclusion

According to the in-depth literature review, several epigenetic mechanisms, especially DNA methylation, have been implicated in the response of plants to HM stress. However, existing evidence points to a complex influence of DNA methylation on the response to Al-induced stress in a species-dependent manner, as well as depending on the type of HM, the dose or intensity of the stress condition, and the time of exposure. Thus, because of this literature review, two main epigenetic strategies underlying the HM stress response were proposed: (i) DNA methylation as a mechanism to protect plants from possible DNA damage caused by metal ions through random DNA methylation along the genome, and (ii) the epigenetic variation used for the regulation of transposon and stress-responsive genes.

Based on the information available at the time of writing this dissertation, there is no report in the literature on the role of DNA methylation in the response to Al stress in rice crops. Therefore, this study is considered a pioneer in the subject. Thus, the first step to understand the possible role of DNA methylation in the response to Al stress in rice crops was to compare the methylome of tolerant and susceptible varieties in a cultivated rice species and a wild one under control conditions. As a conclusion of this first approach, several species-specific genomic regions were reported evidencing the evolutionary history they have undergone independently. However, regions in the genome characteristic of the tolerant varieties were also report despite being different species, which have been related to the response to Al stress, making them regions or genes of

tolerance to Al toxicity potentially regulated by DNA methylation. It should be noted that these are patterns already established in tolerant genotypes under control conditions, which could influence the ability of tolerant genotypes to respond rapidly to stress conditions.

As a next step in seeking to understand the epigenetic patterns of Al tolerance, the pre-established marks in the different genotypes were no longer evaluated, but rather the changes in DNA methylation that occurred in direct response to the stress condition. As a conclusion of this analysis, it is reported that the patterns of DNA methylation changes along the genome are similar for all genotypes in response to the stress condition, with a tendency to hypomethylation in the CG and CHG contexts and hypermethylation in the CHH context. Thus, it is proposed that there is a generalized effect of Al stress on rice plants methylome. From these results, the existence of a hypomethylation phenomenon as an indirect result of stress damage and hypermethylation possibly associated with genome protection is proposed. These conclusions are reinforced by the fact that tolerant genotypes showed less variation than susceptible genotypes.

Furthermore, methylome variation associated with specific regions of the genome was also found. To assess the effect of these variations on gene regulation, both differentially methylated and differentially expressed genes were identified in response to stress. As a conclusion of these analyses, a specific group of genes were reported which show changes in their methylation profile and that are also differentially expressed in response to Al stress. Most of these genes are functionally related to plant tolerance or detoxification mechanisms under HM toxicity, making them potential epigenetically regulated stress tolerance genes. In the first place, it is proposed that there is a general response at the functional level, since most of these genes show a general response in both tolerant and susceptible varieties. However, it was also reported a group of differentially methylated genes expressed only in tolerant varieties, which have not been previously studied or implicated in Al tolerance mechanisms, probably involved in novel tolerance mechanisms, especially considering that a wild variety is being studied.

As a general conclusion of this doctoral thesis, it is proposed that DNA methylation is a mechanism potentially involved in the response to Al stress in rice cultivars showing a global response that is more efficient in tolerant varieties, but also a specific response possibly associated with gene regulation mechanisms of response to Al toxicity. Although this response also occurs in a generalized manner in susceptible and tolerant genotypes, there is a unique regulation of tolerant varieties that lays the foundation for future breeding strategies. Overall, the results of this dissertation represent an important advance in the state of the art, as well as in the biological underpinnings of the role of methylation in the HM stress response in rice cultivars.

7.2 Future Work

This doctoral thesis represents the first known study on the role of epigenetics in the response of plants to Al stress. The findings presented in this dissertation establish a fundamental basis for understanding Al tolerance, while also raising compelling questions that pave the way for future research and exploration. Naturally, all the knowledge gained in basic science will need to be extrapolated to the agricultural field, which entails an entirely new branch of study.

However, several perspectives of great interest for the continuity of this work are as follows:

- Expand this study on a larger scale by introducing additional rice varieties with different ranges of tolerance to Al stress.
- Continue the characterization of the transcriptional profiles for all the genes identified in this study, elucidating their potential involvement in the response to Al stress and providing insights into the genetic basis of Al tolerance.
- It is necessary to conduct follow-up studies to assess the stability of the reported epigenetic changes across generations. Epigenetic marks have the potential to become heritable, affecting genotypes and influencing phenotypic responses. Evaluating the transgenerational stability of these epigenetic changes will provide insights into their long-term implications.
- Numerous additional experiments could be done to further characterize the whole landscape of DNA methylation effects in Al response in plants including the evaluation of different durations of Al exposure, examination of plant responses to stress and subsequent recovery, among others.
- As a long-term perspective with a more practical application, conduct genetic editing on genes with potential responses to Al stress to confirm their active role in tolerance, for example, through gene knockout experiments in different varieties. And targeting epigenetic marks to evaluate their direct effect on the expression of Al-responsive genes.

In summary, future research should focus on expanding the scope of varieties studied, characterizing new genes, investigating the direct effects of epigenetic marks, and assessing the stability of epigenetic changes across generations. These efforts will contribute to a deeper understanding of the complex interplay between genetics and epigenetics in plant responses to Al stress and pave the way for improved breeding strategies and crop improvement.

References

Adrion, J. R., Galloway, J. G., & Kern, A. D. (2020). Predicting the Landscape of Recombination Using Deep Learning. *Molecular Biology and Evolution*, *37*(6), 1790-1808. https://doi.org/10.1093/molbev/msaa038

Agius, F., Kapoor, A., & Zhu, J.-K. (2006). Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proceedings of the National Academy of Sciences*, *103*(31), 11796-11801. https://doi.org/10.1073/pnas.0603563103

Agnieszka, N. (2018). The influence of Al3+ on DNA methylation and sequence changes in the triticale (× Triticosecale Wittmack) genome. *Journal of Applied Genetics*, *59*(4), 405-417. https://doi.org/10.1007/s13353-018-0459-0

Ahmed, I., Sarazin, A., Bowler, C., Colot, V., & Quesneville, H. (2011). Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. *Nucleic Acids Research*, *39*(16), 6919-6931. https://doi.org/10.1093/nar/gkr324

Aina, R., Sgorbati, S., Santagostino, A., Labra, M., Ghiani, A., & Citterio, S. (2004). Specific hypomethylation of DNA is induced by heavy metals in white clover and industrial hemp. *Physiologia Plantarum*, *121*(3), 472-480. https://doi.org/10.1111/j.1399-3054.2004.00343.x

Akhter, Z., Bi, Z., Ali, K., Sun, C., Fiaz, S., Haider, F. U., & Bai, J. (2021). In Response to Abiotic Stress, DNA Methylation Confers EpiGenetic Changes in Plants. *Plants*, *10*(6), 1096. https://doi.org/10.3390/plants10061096

Anderson, L. K., Lai, A., Stack, S. M., Rizzon, C., & Gaut, B. S. (2006). Uneven distribution of expressed sequence tag loci on maize pachytene chromosomes. *Genome Research*, *16*(1), 115-122. https://doi.org/10.1101/gr.4249906

Arbelaez, J. D., Maron, L. G., Jobe, T. O., Piñeros, M. A., Famoso, A. N., Rebelo, A. R., Singh, N., Ma, Q., Fei, Z., Kochian, L. V., & McCouch, S. R. (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*(4), e00014. https://doi.org/10.1002/pld3.14

Arenhart, R. A., Bai, Y., Valter de Oliveira, L. F., Bucker Neto, L., Schunemann, M., Maraschin, F. dos S., Mariath, J., Silverio, A., Sachetto-Martins, G., Margis, R., Wang, Z.-Y., & Margis-Pinheiro, M. (2014). New Insights into Aluminum Tolerance in Rice: The ASR5 Protein Binds the STAR1 Promoter and Other Aluminum-Responsive Genes. *Molecular Plant*, *7*(4), 709-721. https://doi.org/10.1093/mp/sst160

Arenhart, R. A., Schunemann, M., Bucker Neto, L., Margis, R., Wang, Z.-Y., & Margis-Pinheiro, M. (2016). Rice *ASR1* and *ASR5* are complementary transcription factors regulating aluminium responsive genes: *ASR5* and *ASR1* regulate genes in response to Al. *Plant, Cell & Environment*, *39*(3), 645-651. https://doi.org/10.1111/pce.12655

Arif, N., Yadav, V., Singh, S., Singh, S., Ahmad, P., Mishra, R. K., Sharma, S., Tripathi, D. K., Dubey, N. K., & Chauhan, D. K. (2016). Influence of High and Low Levels of Plant-Beneficial Heavy Metal Ions on Plant Growth and Development. *Frontiers in Environmental Science*, *4*. https://doi.org/10.3389/fenvs.2016.00069

Asgher, Mohd., Khan, N. A., Khan, M. I. R., Fatma, M., & Masood, A. (2014). Ethylene production is associated with alleviation of cadmium-induced oxidative stress by sulfur in mustard types differing in ethylene sensitivity. *Ecotoxicology and Environmental Safety*, *106*, 54-61. https://doi.org/10.1016/j.ecoenv.2014.04.017

Baldoni, E., Frugis, G., Martinelli, F., Benny, J., Paffetti, D., & Buti, M. (2021). A Comparative Transcriptomic Meta-Analysis Revealed Conserved Key Genes and Regulatory Networks Involved in Drought Tolerance in Cereal Crops. *International Journal of Molecular Sciences*, *22*(23), 13062. https://doi.org/10.3390/ijms222313062

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., & Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature*, *480*(7376), 245-249. https://doi.org/10.1038/nature10555

Bednarek, P. T., Orłowska, R., & Niedziela, A. (2017). A relative quantitative Methylation-Sensitive Amplified Polymorphism (MSAP) method for the analysis of abiotic stress. *BMC Plant Biology*, *17*(1), 79. https://doi.org/10.1186/s12870-017-1028-0

Bender, J. (1998). Cytosine methylation of repeated sequences in eukaryotes: The role of DNA pairing. *Trends in Biochemical Sciences*, *23*(7), 252-256. https://doi.org/10.1016/S0968-0004(98)01225-0

Bender, J. (2004). DNA METHYLATION AND EPIGENETICS. *Annual Review of Plant Biology*, *55*(1), 41-68. https://doi.org/10.1146/annurev.arplant.55.031903.141641

Bernardo, S., Dinis, L.-T., Luzio, A., Pinto, G., Meijón, M., Valledor, L., Conde, A., Gerós, H., Correia, C. M., & Moutinho-Pereira, J. (2017). Kaolin particle film application lowers oxidative damage and DNA methylation on grapevine (Vitis vinifera L.). *Environmental and Experimental Botany*, *139*, 39-47. https://doi.org/10.1016/j.envexpbot.2017.04.002

Bhatia, H., Khemka, N., Jain, M., & Garg, R. (2018). Genome-wide bisulphite-sequencing reveals organ-specific methylation patterns in chickpea. *Scientific Reports*, *8*(1), 9704. https://doi.org/10.1038/s41598-018-27979-w

Bojórquez-Quintal, E., Escalante-Magaña, C., Echevarría-Machado, I., & Martínez-Estévez, M. (2017). Aluminum, a Friend or Foe of Higher Plants in Acid Soils. *Frontiers in Plant Science*, *8*, 1767. https://doi.org/10.3389/fpls.2017.01767

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Boyko, A., Blevins, T., Yao, Y., Golubov, A., Bilichak, A., Ilnytskyy, Y., Hollander, J., Meins, F., & Kovalchuk, I. (2010). Transgenerational Adaptation of Arabidopsis to Stress Requires DNA Methylation and the Function of Dicer-Like Proteins. *PLoS ONE*, *5*(3), e9514. https://doi.org/10.1371/journal.pone.0009514

Boyko, A., & Kovalchuk, I. (2008). Epigenetic control of plant stress response. *Environmental and Molecular Mutagenesis*, *49*(1), 61-72. https://doi.org/10.1002/em.20347

Bräutigam, K., & Cronk, Q. (2018). DNA Methylation and the Evolution of Developmental Complexity in Plants. *Frontiers in Plant Science*, *9*, 1447. https://doi.org/10.3389/fpls.2018.01447

Brondani, R. P. V., Zucchi, M. I., Brondani, C., Rangel, P. H. N., Oliveira Borba, T. C. D., Rangel, P. N., Magalhães, M. R., & Vencovsky, R. (2005). Genetic Structure of Wild Rice Oryza

Glumaepatula Populations in Three Brazilian Biomes Using Microsatellite Markers. *Genetica*, *125*(2-3), 115-123. https://doi.org/10.1007/s10709-005-4916-4

Cao, F., Chen, F., Sun, H., Zhang, G., Chen, Z.-H., & Wu, F. (2014). Genome-wide transcriptome and functional analysis of two contrasting genotypes reveals key genes for cadmium tolerance in barley. *BMC Genomics*, *15*(1), 611. https://doi.org/10.1186/1471-2164-15-611

Cao, X., & Jacobsen, S. E. (2002). Role of the Arabidopsis DRM Methyltransferases in De Novo DNA Methylation and Gene Silencing. *Current Biology*, *12*(13), 1138-1144. https://doi.org/10.1016/S0960-9822(02)00925-9

Cao, Y., Lou, Y., Han, Y., Shi, J., Wang, Y., Wang, W., & Ming, F. (2011). Al toxicity leads to enhanced cell division and changed photosynthesis in Oryza rufipogon L. *Molecular Biology Reports*, *38*(8), 4839-4846. https://doi.org/10.1007/s11033-010-0618-9

Chang, Y., Zhu, C., Jiang, J., Zhang, H., Zhu, J., & Duan, C. (2020). Epigenetic regulation in plant abiotic stress responses. *Journal of Integrative Plant Biology*, *62*(5), 563-580. https://doi.org/10.1111/jipb.12901

Che, J., Tsutsui, T., Yokosho, K., Yamaji, N., & Ma, J. F. (2018). Functional characterization of an aluminum (Al)-inducible transcription factor, ART2, revealed a different pathway for Al tolerance in rice. *New Phytologist*, *220*(1), 209-218. https://doi.org/10.1111/nph.15252

Che, J., Yamaji, N., Shen, R. F., & Ma, J. F. (2016). An Al-inducible expansin gene, OsEXPA10 is involved in root cell elongation of rice. *The Plant Journal*, *88*(1), 132-142. https://doi.org/10.1111/tpj.13237

Chen, J., Cao, H., Chen, D., Kuang, L., & Wu, D. (2023). Transcriptome-wide analysis of m6A methylation reveals genetic responses to cadmium stress at germination stage in rice. *Environmental and Experimental Botany*, *205*, 105130. https://doi.org/10.1016/j.envexpbot.2022.105130

Chen, Q., Li, W., Tan, L., & Tian, F. (2021). Harnessing Knowledge from Maize and Rice Domestication for New Crop Breeding. *Molecular Plant*, *14*(1), 9-26. https://doi.org/10.1016/j.molp.2020.12.006

Chen, Z. C., Yamaji, N., Motoyama, R., Nagamura, Y., & Ma, J. F. (2012). Up-Regulation of a Magnesium Transporter Gene *OsMGT1* Is Required for Conferring Aluminum Tolerance in Rice. *Plant Physiology*, *159*(4), 1624-1633. https://doi.org/10.1104/pp.112.199778

Cheng, T.-F., Choudhuri, S., & Muldoon-Jacobs, K. (2012). Epigenetic targets of some toxicologically relevant metals: A review of the literature: Epigenetic effects of metals. *Journal of Applied Toxicology*, *32*(9), 643-653. https://doi.org/10.1002/jat.2717

Cheng, Z., Buell, C. R., Wing, R. A., Gu, M., & Jiang, J. (2001). Toward a Cytological Characterization of the Rice Genome. *Genome Research*, *11*(12), 2133-2141. https://doi.org/10.1101/gr.194601

Choi, C.-S., & Sano, H. (2007). Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. *Molecular Genetics and Genomics*, *277*(5), 589-600. https://doi.org/10.1007/s00438-007-0209-1

Choi, J. Y., & Purugganan, M. D. (2018). Evolutionary Epigenomics of Retrotransposon-Mediated Methylation Spreading in Rice. *Molecular Biology and Evolution*, *35*(2), 365-382. https://doi.org/10.1093/molbev/msx284

Chopra, R., Burow, G., Hayes, C., Emendack, Y., Xin, Z., & Burke, J. (2015). Transcriptome profiling and validation of gene based single nucleotide polymorphisms (SNPs) in sorghum

genotypes with contrasting responses to cold stress. *BMC Genomics*, *16*(1), 1040. https://doi.org/10.1186/s12864-015-2268-8

Cochrane, T.T. (1979). An ongoing appraisal of the savanna ecosystems of Tropical America for beef cattle production. En *Seminar on Pasture Production in Acid Soils of the Tropics (1978, Cali, Colombia). Pasture production in acid soils of the tropics: Proceedings.: Vol. Beef Program* (pp. 1-12).

Cohen, S. P., & Leach, J. E. (2019). Abiotic and biotic stresses induce a core transcriptome response in rice. *Scientific Reports*, *9*(1), 6273. https://doi.org/10.1038/s41598-019-42731-8

Cong, W., Miao, Y., Xu, L., Zhang, Y., Yuan, C., Wang, J., Zhuang, T., Lin, X., Jiang, L., Wang, N., Ma, J., Sanguinet, K. A., Liu, B., Rustgi, S., & Ou, X. (2019). Transgenerational memory of gene expression changes induced by heavy metal stress in rice (Oryza sativa L.). *BMC Plant Biology*, *19*(1), 282. https://doi.org/10.1186/s12870-019-1887-7

Cortijo, S., Wardenaar, R., Colomé-Tatché, M., Gilly, A., Etcheverry, M., Labadie, K., Caillieux, E., Hospital, F., Aury, J.-M., Wincker, P., Roudier, F., Jansen, R. C., Colot, V., & Johannes, F. (2014). Mapping the Epigenetic Basis of Complex Traits. *Science*, *343*(6175), 1145-1148. https://doi.org/10.1126/science.1248127

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E. A., Gouzy, J., Rees, D. J. G., Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., … Bucher, E. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*, *49*(7), 1099-1106. https://doi.org/10.1038/ng.3886

Delhaize, E., Ma, J. F., & Ryan, P. R. (2012). Transcriptional regulation of aluminium tolerance genes. *Trends in Plant Science*, *17*(6), 341-348. https://doi.org/10.1016/j.tplants.2012.02.008

Demirci, S., Peters, S. A., de Ridder, D., & van Dijk, A. D. J. (2018). DNA sequence and shape are predictive for meiotic crossovers throughout the plant kingdom. *The Plant Journal*, *95*(4), 686-699. https://doi.org/10.1111/tpj.13979

do Amaral, M. N., Arge, L. W. P., Benitez, L. C., Danielowski, R., Silveira, S. F. da S., Farias, D. da R., de Oliveira, A. C., da Maia, L. C., & Braga, E. J. B. (2016). Comparative transcriptomics of rice plants under cold, iron, and salt stresses. *Functional & Integrative Genomics*, *16*(5), 567-579. https://doi.org/10.1007/s10142-016-0507-y

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. https://doi.org/10.1093/bioinformatics/bts635

Dorion, S., Ouellet, J. C., & Rivoal, J. (2021). Glutathione Metabolism in Plants under Stress: Beyond Reactive Oxygen Species Detoxification. *Metabolites*, *11*(9), 641. https://doi.org/10.3390/metabo11090641

Dubin, M. J., Mittelsten Scheid, O., & Becker, C. (2018). Transposons: A blessing curse. *Current Opinion in Plant Biology*, *42*, 23-29. https://doi.org/10.1016/j.pbi.2018.01.003

Ezaki, B., Higashi, A., Nanba, N., & Nishiuchi, T. (2016). An S-adenosyl Methionine Synthetase (SAMS) Gene from Andropogon virginicus L. Confers Aluminum Stress Tolerance and Facilitates Epigenetic Gene Regulation in Arabidopsis thaliana. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01627

Fahad, S., Adnan, M., Noor, M., Arif, M., Alam, M., Khan, I. A., Ullah, H., Wahid, F., Mian, I. A., Jamal, Y., Basir, A., Hassan, S., Saud, S., Amanullah, Riaz, M., Wu, C., Khan, M. A., & Wang, D. (2019).

Major Constraints for Global Rice Production. En *Advances in Rice Research for Abiotic Stress Tolerance* (pp. 1-22). Elsevier. https://doi.org/10.1016/B978-0-12-814332-2.00001-0

Famoso, A. N., Clark, R. T., Shaff, J. E., Craft, E., McCouch, S. R., & 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 Physiology*, *153*(4), 1678-1691. https://doi.org/10.1104/pp.110.156794

Famoso, A. N., Zhao, K., Clark, R. T., Tung, C.-W., Wright, M. H., Bustamante, C., Kochian, L. V., & McCouch, S. R. (2011). Genetic Architecture of Aluminum Tolerance in Rice (Oryza sativa) Determined through Genome-Wide Association Analysis and QTL Mapping. *PLoS Genetics*, *7*(8), e1002221. https://doi.org/10.1371/journal.pgen.1002221

Fan, S. K., Ye, J. Y., Zhang, L. L., Chen, H. S., Zhang, H. H., Zhu, Y. X., Liu, X. X., & Jin, C. W. (2020). Inhibition of DNA demethylation enhances plant tolerance to cadmium toxicity by improving iron nutrition. *Plant, Cell & Environment*, *43*(1), 275-291. https://doi.org/10.1111/pce.13670

Feng, S. J., Liu, X. S., Tao, H., Tan, S. K., Chu, S. S., Oono, Y., Zhang, X. D., Chen, J., & Yang, Z. M. (2016). Variation of DNA methylation patterns associated with gene expression in rice (*Oryza sativa*) exposed to cadmium: DNA methylation in Cd-exposed rice. *Plant, Cell & Environment*, *39*(12), 2629-2649. https://doi.org/10.1111/pce.12793

Feng, S., Jacobsen, S. E., & Reik, W. (2010). Epigenetic Reprogramming in Plant and Animal Development. *Science*, *330*(6004), 622-627. https://doi.org/10.1126/science.1190614

Ferreira, L. J., Donoghue, M. T. A., Barros, P., Saibo, N. J., Santos, A. P., & Oliveira, M. M. (2019). Uncovering Differentially Methylated Regions (DMRs) in a Salt-Tolerant Rice Variety under Stress: One Step towards New Regulatory Regions for Enhanced Salt Tolerance. *Epigenomes*, *3*(1), 4. https://doi.org/10.3390/epigenomes3010004

Finnegan, E. (2001). Is plant gene expression regulated globally? *Trends in Genetics*, *17*(7), 361-365. https://doi.org/10.1016/S0168-9525(01)02319-8

Foyer, C. H., & Noctor, G. (2011). Ascorbate and Glutathione: The Heart of the Redox Hub. *Plant Physiology*, *155*(1), 2-18. https://doi.org/10.1104/pp.110.167569

Fryzova, R., Pohanka, M., Martinkova, P., Cihlarova, H., Brtnicky, M., Hladky, J., & Kynicky, J. (2017). Oxidative Stress and Heavy Metals in Plants. En P. de Voogt (Ed.), *Reviews of Environmental Contamination and Toxicology Volume 245* (Vol. 245, pp. 129-156). Springer International Publishing. https://doi.org/10.1007/398_2017_7

Fu, S.-F., Chen, P.-Y., Nguyen, Q. T. T., Huang, L.-Y., Zeng, G.-R., Huang, T.-L., Lin, C.-Y., & Huang, H.-J. (2014). Transcriptome profiling of genes and pathways associated with arsenic toxicity and tolerance in Arabidopsis. *BMC Plant Biology*, *14*(1), 94. https://doi.org/10.1186/1471-2229-14-94

Fuchs, E. J., Meneses Martínez, A., Calvo, A., Muñoz, M., & Arrieta-Espinoza, G. (2016). Genetic diversity in *Oryza glumaepatula* wild rice populations in Costa Rica and possible gene flow from *O. sativa*. *PeerJ*, *4*, e1875. https://doi.org/10.7717/peerj.1875

Galindo-González, L., Sarmiento, F., & Quimbaya, M. (2018). Shaping Plant Adaptability, Genome Structure and Gene Expression through Transposable Element Epigenetic Control: Focus on Methylation. *Agronomy*, *8*(9), 180. https://doi.org/10.3390/agronomy8090180

Gallo-Franco, J. J., Sosa, C. C., Ghneim-Herrera, T., & Quimbaya, M. (2020). Epigenetic Control of Plant Response to Heavy Metal Stress: A New View on Aluminum Tolerance. *Frontiers in Plant Science*, *11*, 602625. https://doi.org/10.3389/fpls.2020.602625

Garg, R., Narayana Chevala, V., Shankar, R., & Jain, M. (2015). Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response. *Scientific Reports*, *5*(1), 14922. https://doi.org/10.1038/srep14922

Gent, J. I., Ellis, N. A., Guo, L., Harkess, A. E., Yao, Y., Zhang, X., & Dawe, R. K. (2013). CHH islands: De novo DNA methylation in near-gene chromatin regulation in maize. *Genome Research*, *23*(4), 628-637. https://doi.org/10.1101/gr.146985.112

Greco, M., Chiappetta, A., Bruno, L., & Bitonti, M. B. (2012). In Posidonia oceanica cadmium induces changes in DNA methylation and chromatin patterning. *Journal of Experimental Botany*, *63*(2), 695-709. https://doi.org/10.1093/jxb/err313

Guan, C., Ji, J., Wu, D., Li, X., Jin, C., Guan, W., & Wang, G. (2015). The glutathione synthesis may be regulated by cadmium-induced endogenous ethylene in Lycium chinense, and overexpression of an ethylene responsive transcription factor gene enhances tolerance to cadmium stress in tobacco. *Molecular Breeding*, *35*(5), 123. https://doi.org/10.1007/s11032-015-0313-6

Gullì, M., Marchi, L., Fragni, R., Buschini, A., & Visioli, G. (2018). Epigenetic modifications preserve the hyperaccumulator *Noccaea caerulescens* from Ni geno-toxicity: Hyperaccumulator *Noccaea caerulescens* and Ni Geno-Toxicity. *Environmental and Molecular Mutagenesis*, *59*(6), 464-475. https://doi.org/10.1002/em.22191

Habu, Y., Ando, T., Ito, S., Nagaki, K., Kishimoto, N., Taguchi-Shiobara, F., Numa, H., Yamaguchi, K., Shigenobu, S., Murata, M., Meshi, T., & Yano, M. (2015). Epigenomic modification in rice controls meiotic recombination and segregation distortion. *Molecular Breeding*, *35*(4), 103. https://doi.org/10.1007/s11032-015-0299-0

He, G., Zhu, X., Elling, A. A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., Qi, Y., Chen, R., & Deng, X.-W. (2010). Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. *The Plant Cell*, *22*(1), 17-33. https://doi.org/10.1105/tpc.109.072041

Henderson, I. R. (2012). Control of meiotic recombination frequency in plant genomes. *Current Opinion in Plant Biology*, *15*(5), 556-561. https://doi.org/10.1016/j.pbi.2012.09.002

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Herman, J. J., & Sultan, S. E. (2016). DNA methylation mediates genetic variation for adaptive transgenerational plasticity. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1838), 20160988. https://doi.org/10.1098/rspb.2016.0988

Hirsch, C. D., & Springer, N. M. (2017). Transposable element influences on gene expression in plants. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1860*(1), 157-165. https://doi.org/10.1016/j.bbagrm.2016.05.010

Hollister, J. D., & Gaut, B. S. (2009). Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, *19*(8), 1419-1428. https://doi.org/10.1101/gr.091678.109

Horst, W. J., Wang, Y., & Eticha, D. (2010). The role of the root apoplast in aluminium-induced inhibition of root elongation and in aluminium resistance of plants: A review. *Annals of Botany*, *106*(1), 185-197. https://doi.org/10.1093/aob/mcq053

Hossein Pour, A., Özkan, G., Balpinar Nalci, Ö., & HaliLoğlu, K. (2019). Estimation of genomic instability and DNA methylation due to aluminum (Al) stress in wheat (Triticum aestivum L.) using iPBS and CRED-iPBS analyses. *TURKISH JOURNAL OF BOTANY*, *43*(1), 27-37. https://doi.org/10.3906/bot-1804-23

Hsieh, P.-H., Kan, C.-C., Wu, H.-Y., Yang, H.-C., & Hsieh, M.-H. (2018). Early molecular events associated with nitrogen deficiency in rice seedling roots. *Scientific Reports*, *8*(1), 12207. https://doi.org/10.1038/s41598-018-30632-1

Hu, J., Chen, X., Zhang, H., & Ding, Y. (2015). Genome-wide analysis of DNA methylation in photoperiod- and thermo-sensitive male sterile rice Peiai 64S. *BMC Genomics*, *16*(1), 102. https://doi.org/10.1186/s12864-015-1317-7

Huang, C. F., Yamaji, N., Mitani, N., Yano, M., Nagamura, Y., & Ma, J. F. (2009). A Bacterial-Type ABC Transporter Is Involved in Aluminum Tolerance in Rice. *The Plant Cell*, *21*(2), 655-667. https://doi.org/10.1105/tpc.108.064543

Jackson, N. D., Konar, M., Debaere, P., & Estes, L. (2019). Probabilistic global maps of crop-specific areas from 1961 to 2014. *Environmental Research Letters*, *14*(9), 094023. https://doi.org/10.1088/1748-9326/ab3b93

Jackson, S. A. (2016). Rice: The First Crop Genome. *Rice*, *9*(1), 14. https://doi.org/10.1186/s12284-016-0087-4

Jing, M., Zhang, H., Wei, M., Tang, Y., Xia, Y., Chen, Y., Shen, Z., & Chen, C. (2022). Reactive Oxygen Species Partly Mediate DNA Methylation in Responses to Different Heavy Metals in Pokeweed. *Frontiers in Plant Science*, *13*, 845108. https://doi.org/10.3389/fpls.2022.845108

Jingguang, C., Qi, L., Baiquan, Z., Longbiao, G., & Guoyou, Y. (2020). Progress on Molecular Mechanism of Aluminum Resistance in Rice. *Rice Science*, *27*(6), 454-467. https://doi.org/10.1016/j.rsci.2020.09.003

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*(7), 484-492. https://doi.org/10.1038/nrg3230

Kalisz, S., & Purugganan, M. D. (2004). Epialleles via DNA methylation: Consequences for plant evolution. *Trends in Ecology & Evolution*, *19*(6), 309-314. https://doi.org/10.1016/j.tree.2004.03.034

Kantar, M. B., Sosa, C. C., Khoury, C. K., Castañeda-Álvarez, N. P., Achicanoy, H. A., Bernau, V., Kane, N. C., Marek, L., Seiler, G., & Rieseberg, L. H. (2015). Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (Helianthus annuus L.). *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00841

Kawakatsu, T., Huang, S. C., Jupe, F., Sasaki, E., Schmitz, R. J., Urich, M. A., Castanon, R., Nery, J. R., Barragan, C., He, Y., Chen, H., Dubin, M., Lee, C.-R., Wang, C., Bemm, F., Becker, C., O'Neil, R., O'Malley, R. C., Quarless, D. X., … Zhou, X. (2016). Epigenomic Diversity in a Global

Collection of Arabidopsis thaliana Accessions. *Cell*, *166*(2), 492-505. https://doi.org/10.1016/j.cell.2016.06.044

Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D., & Bohnert, H. J. (2001). *Gene Expression Profiles during the Initial Phase of Salt Stress in Rice*. *13*, 889-905. https://doi.org/10.1105/tpc.13.4.889

Kawashima, T., & Berger, F. (2014). Epigenetic reprogramming in plant sexual reproduction. *Nature Reviews Genetics*, *15*(9), 613-624. https://doi.org/10.1038/nrg3685

Kent, T. V., Uzunović, J., & Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160458. https://doi.org/10.1098/rstb.2016.0458

Keunen, E., Schellingen, K., Vangronsveld, J., & Cuypers, A. (2016). Ethylene and Metal Stress: Small Molecule, Big Impact. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00023

Khan, M. I. R., & Khan, N. A. (2014). Ethylene reverses photosynthetic inhibition by nickel and zinc in mustard through changes in PS II activity, photosynthetic nitrogen use efficiency, and antioxidant metabolism. *Protoplasma*, *251*(5), 1007-1019. https://doi.org/10.1007/s00709-014-0610-7

Kim, J.-M., To, T. K., Ishida, J., Matsui, A., Kimura, H., & Seki, M. (2012). Transition of Chromatin Status During the Process of Recovery from Drought Stress in Arabidopsis thaliana. *Plant and Cell Physiology*, *53*(5), 847-856. https://doi.org/10.1093/pcp/pcs053

Kim, J.-M., To, T. K., Ishida, J., Morosawa, T., Kawashima, M., Matsui, A., Toyoda, T., Kimura, H., Shinozaki, K., & Seki, M. (2008). Alterations of Lysine Modifications on the Histone H3 N-Tail under Drought Stress Conditions in Arabidopsis thaliana. *Plant and Cell Physiology*, *49*(10), 1580-1588. https://doi.org/10.1093/pcp/pcn133

Kimatu, J. N., Diarso, M., Song, C., Agboola, R. S., Pang, J., & Qi, X. (s. f.). *DNA cytosine methylation alterations associated with aluminium toxicity and low pH in Sorghum bicolor*.

Kimatu, J. N., Jiang, L., Ngezahayo, F., Songdi, C., Quan-yuan, Y., Pang, J., & Liu, B. (2013). *ALTERATION IN CYTOSINE DNA METHYLATION PATTERNS AND LEVELS INDUCED BY ALUMINIUM TOXICITY STRESS IN MAIZE VARIETIES*. *2*.

Kochian, L. V., Hoekenga, O. A., & Piñeros, M. A. (2004). HOW DO CROP PLANTS TOLERATE ACID SOILS? MECHANISMS OF ALUMINUM TOLERANCE AND PHOSPHOROUS EFFICIENCY. *Annual Review of Plant Biology*, *55*(1), 459-493. https://doi.org/10.1146/annurev.arplant.55.031903.141655

Kochian, L. V., Piñeros, M. A., Liu, J., & Magalhaes, J. V. (2015). Plant Adaptation to Acid Soils: The Molecular Basis for Crop Aluminum Resistance. *Annual Review of Plant Biology*, *66*(1), 571-598. https://doi.org/10.1146/annurev-arplant-043014-114822

Kou, H. P., Li, Y., Song, X. X., Ou, X. F., Xing, S. C., Ma, J., Von Wettstein, D., & Liu, B. (2011). Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (Oryza sativa L.). *Journal of Plant Physiology*, *168*(14), 1685-1693. https://doi.org/10.1016/j.jplph.2011.03.017

Kovach, M. J., Sweeney, M. T., & McCouch, S. R. (2007). New insights into the history of rice domestication. *Trends in Genetics*, *23*(11), 578-587. https://doi.org/10.1016/j.tig.2007.08.012

Kovalchuk, O., Burke, P., Arkhipov, A., Kuchma, N., James, S. J., Kovalchuk, I., & Pogribny, I. (2003). Genome hypermethylation in Pinus silvestris of Chernobyl—A mechanism for radiation adaptation? *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *529*(1-2), 13-20. https://doi.org/10.1016/S0027-5107(03)00103-9

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571-1572. https://doi.org/10.1093/bioinformatics/btr167

Kumar, S. (2018). Epigenomics of Plant Responses to Environmental Stress. *Epigenomes*, *2*(1), 6. https://doi.org/10.3390/epigenomes2010006

Lambing, C., Franklin, F. C. H., & Wang, C.-J. R. (2017). Understanding and Manipulating Meiotic Recombination in Plants. *Plant Physiology*, *173*(3), 1530-1542. https://doi.org/10.1104/pp.16.01530

Lanciano, S., & Mirouze, M. (2017). DNA Methylation in Rice and Relevance for Breeding. *Epigenomes*, *1*(2), 10. https://doi.org/10.3390/epigenomes1020010

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357-359. https://doi.org/10.1038/nmeth.1923

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*(3), 204-220. https://doi.org/10.1038/nrg2719

Le, T.-N., Schumann, U., Smith, N. A., Tiwari, S., Au, P. C. K., Zhu, Q.-H., Taylor, J. M., Kazan, K., Llewellyn, D. J., Zhang, R., Dennis, E. S., & Wang, M.-B. (2014). *DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis*. 18.

Lee, H.-R., Zhang, W., Langdon, T., Jin, W., Yan, H., Cheng, Z., & Jiang, J. (2005). From The Cover: Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. *Proceedings of the National Academy of Sciences*, *102*(33), 11793-11798. https://doi.org/10.1073/pnas.0503863102

Lei, M., Zhang, H., Julian, R., Tang, K., Xie, S., & Zhu, J.-K. (2015). Regulatory link between DNA methylation and active demethylation in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, *112*(11), 3553-3557. https://doi.org/10.1073/pnas.1502279112

Li, L., He, Y., Zhang, X., Zhang, H., Sun, Z., Li, J., & Hong, G. (2020). Alterations of Rice (Oryza sativa L.) DNA Methylation Patterns Associated with Gene Expression in Response to Rice Black Streaked Dwarf Virus. *International Journal of Molecular Sciences*, *21*(16), 5753. https://doi.org/10.3390/ijms21165753

Li, X., Zhu, J., Hu, F., Ge, S., Ye, M., Xiang, H., Zhang, G., Zheng, X., Zhang, H., Zhang, S., Li, Q., Luo, R., Yu, C., Yu, J., Sun, J., Zou, X., Cao, X., Xie, X., Wang, J., & Wang, W. (2012). Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics*, *13*(1), 300. https://doi.org/10.1186/1471-2164-13-300

Li, Y., Huang, J., Song, X., Zhang, Z., Jiang, Y., Zhu, Y., Zhao, H., & Ni, D. (2017). An RNA-Seq transcriptome analysis revealing novel insights into aluminum tolerance and accumulation in tea plant. *Planta*, *246*(1), 91-103. https://doi.org/10.1007/s00425-017-2688-6

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923-930. https://doi.org/10.1093/bioinformatics/btt656

Lin, H., Ouyang, S., Egan, A., Nobuta, K., Haas, B. J., Zhu, W., Gu, X., Silva, J. C., Meyers, B. C., & Buell, C. R. (2008). Characterization of paralogous protein families in rice. *BMC Plant Biology*, *8*(1), 18. https://doi.org/10.1186/1471-2229-8-18

Lin, H.C., Coe, R.A, Quick, W.P., & A.B. (2019). Sustainable Solutions for Food Security. In A. Sarkar, S. R. Sensarma, & G. W. VanLoon (Eds.). En *Sustainable Solutions for Food Security.* Springer International Publishing.

Liu, B., Liu, Y., Jin, X., Wang, X., & Liu, B. (2016). iRSpot-DACC: A computational predictor for recombination hot/cold spots identification based on dinucleotide-based auto-cross covariance. *Scientific Reports*, *6*(1), 33483. https://doi.org/10.1038/srep33483

Liu, H., Wu, Y., Cao, A., Mao, B., Zhao, B., & Wang, J. (2017). Genome-Wide Analysis of DNA Methylation During Ovule Development of Female-Sterile Rice *fsv1*. *G3 Genes|Genomes|Genetics*, *7*(11), 3621-3635. https://doi.org/10.1534/g3.117.300243

Liu, J., Piñeros, M. A., & Kochian, L. V. (2014). The role of aluminum sensing and signaling in plant aluminum resistance: Al signaling activates plant resistance responses. *Journal of Integrative Plant Biology*, *56*(3), 221-230. https://doi.org/10.1111/jipb.12162

Liu, J., Xu, M., Estavillo, G. M., Delhaize, E., White, R. G., Zhou, M., & Ryan, P. R. (2018). Altered Expression of the Malate-Permeable Anion Channel OsALMT4 Reduces the Growth of Rice Under Low Radiance. *Frontiers in Plant Science*, *9*, 542. https://doi.org/10.3389/fpls.2018.00542

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lunde, C., Zygadlo, A., Simonsen, H. T., Nielsen, P. L., Blennow, A., & Haldrup, A. (2008). Sulfur starvation in rice: The effect on photosynthesis, carbohydrate metabolism, and oxidative stress protective pathways. *Physiologia Plantarum*, *134*(3), 508-521. https://doi.org/10.1111/j.1399-3054.2008.01159.x

Ma, J. F., Shen, R., Zhao, Z., Wissuwa, M., Takeuchi, Y., Ebitani, T., & Yano, M. (2002). Response of Rice to Al Stress and Identification of Quantitative Trait Loci for Al Tolerance. *Plant and Cell Physiology*, *43*(6), 652-659. https://doi.org/10.1093/pcp/pcf081

Magalhaes, J. V., Liu, J., Guimarães, C. T., Lana, U. G. P., Alves, V. M. C., Wang, Y.-H., Schaffert, R. E., Hoekenga, O. A., Piñeros, M. A., Shaff, J. E., Klein, P. E., Carneiro, N. P., Coelho, C. M., Trick, H. N., & Kochian, L. V. (2007). A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nature Genetics*, *39*(9), 1156-1161. https://doi.org/10.1038/ng2074

Mansueto, L., Fuentes, R. R., Borja, F. N., Detras, J., Abriol-Santos, J. M., Chebotarov, D., Sanciangco, M., Palis, K., Copetti, D., Poliakov, A., Dubchak, I., Solovyev, V., Wing, R. A., Hamilton, R. S., Mauleon, R., McNally, K. L., & Alexandrov, N. (2017). Rice SNP-seek database update: New SNPs, indels, and queries. *Nucleic Acids Research*, *45*(D1), D1075-D1081. https://doi.org/10.1093/nar/gkw1135

Mao, C. (2003). Identification of aluminium-regulated genes by cDNA-AFLP in rice (Oryza sativa L.): Aluminium-regulated genes for the metabolism of cell wall components. *Journal of Experimental Botany*, *55*(394), 137-143. https://doi.org/10.1093/jxb/erh030

Marconi, G., Pace, R., Traini, A., Raggi, L., Lutts, S., Chiusano, M., Guiducci, M., Falcinelli, M., Benincasa, P., & Albertini, E. (2013). Use of MSAP Markers to Analyse the Effects of Salt Stress on DNA Methylation in Rapeseed (Brassica napus var. Oleifera). *PLoS ONE*, *8*(9), e75597. https://doi.org/10.1371/journal.pone.0075597

Maron, L. G., Kirst, M., Mao, C., Milner, M. J., Menossi, M., & Kochian, L. V. (2008). Transcriptional profiling of aluminum toxicity and tolerance responses in maize roots. *New Phytologist*, *179*(1), 116-128. https://doi.org/10.1111/j.1469-8137.2008.02440.x

Maropola, M. K. A., Ramond, J.-B., & Trindade, M. (2015). Impact of metagenomic DNA extraction procedures on the identifiable endophytic bacterial diversity in Sorghum bicolor (L. Moench). *Journal of Microbiological Methods*, *112*, 104-117. https://doi.org/10.1016/j.mimet.2015.03.012

Martin, G. T., Seymour, D. K., & Gaut, B. S. (2021). CHH Methylation Islands: A Nonconserved Feature of Grass Genomes That Is Positively Associated with Transposable Elements but Negatively Associated with Gene-Body Methylation. *Genome Biology and Evolution*, *13*(8), evab144. https://doi.org/10.1093/gbe/evab144

Matzke, M. A., & Mosher, R. A. (2014). RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nature Reviews Genetics*, *15*(6), 394-408. https://doi.org/10.1038/nrg3683

Melamed-Bessudo, C., & Levy, A. A. (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proceedings of the National Academy of Sciences*, *109*(16), E981-E988. https://doi.org/10.1073/pnas.1120742109

Mirouze, M., Lieberman-Lazarovich, M., Aversano, R., Bucher, E., Nicolet, J., Reinders, J., & Paszkowski, J. (2012). Loss of DNA methylation affects the recombination landscape in Arabidopsis. *Proceedings of the National Academy of Sciences*, *109*(15), 5880-5885. https://doi.org/10.1073/pnas.1120841109

Mirouze, M., & Paszkowski, J. (2011). Epigenetic contribution to stress adaptation in plants. *Current Opinion in Plant Biology*, *14*(3), 267-274. https://doi.org/10.1016/j.pbi.2011.03.004

Morkunas, I., Woźniak, A., Mai, V., Rucińska-Sobkowiak, R., & Jeandet, P. (2018). The Role of Heavy Metals in Plant Response to Biotic Stress. *Molecules*, *23*(9), 2320. https://doi.org/10.3390/molecules23092320

Murali Achary, V. M., & Panda, B. B. (2010). Aluminium-induced DNA damage and adaptive response to genotoxic stress in plant cells are mediated through reactive oxygen intermediates. *Mutagenesis*, *25*(2), 201-209. https://doi.org/10.1093/mutage/gep063

Mustafa, G., & Komatsu, S. (2016). Toxicity of heavy metals and metal-containing nanoparticles on plants. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1864*(8), 932-944. https://doi.org/10.1016/j.bbapap.2016.02.020

Mustafiz, A., Singh, A. K., Pareek, A., Sopory, S. K., & Singla-Pareek, S. L. (2011). Genome-wide analysis of rice and Arabidopsis identifies two glyoxalase genes that are highly expressed

in abiotic stresses. *Functional & Integrative Genomics*, *11*(2), 293-305. https://doi.org/10.1007/s10142-010-0203-2

Negishi, T., Oshima, K., Hattori, M., Kanai, M., Mano, S., Nishimura, M., & Yoshida, K. (2012). Tonoplast- and Plasma Membrane-Localized Aquaporin-Family Transporters in Blue Hydrangea Sepals of Aluminum Hyperaccumulating Plant. *PLoS ONE*, *7*(8), e43189. https://doi.org/10.1371/journal.pone.0043189

Nguyen, B. D., Brar, D. S., Bui, B. C., Nguyen, T. V., Pham, L. N., & Nguyen, H. T. (2003). Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, ORYZA RUFIPOGON Griff., into indica rice (Oryza sativa L.). *Theoretical and Applied Genetics*, *106*(4), 583-593. https://doi.org/10.1007/s00122-002-1072-4

Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. D., Li, Q., Rohr, N. A., Rambani, A., Burke, J. M., Udall, J. A., Egesi, C., Schmutz, J., Grimwood, J., Jackson, S. A., Springer, N. M., & Schmitz, R. J. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biology*, *17*(1), 194. https://doi.org/10.1186/s13059-016-1059-0

Oono, Y., Yazawa, T., Kanamori, H., Sasaki, H., Mori, S., Handa, H., & Matsumoto, T. (2016). Genome-Wide Transcriptome Analysis of Cadmium Stress in Rice. *BioMed Research International*, *2016*, 1-9. https://doi.org/10.1155/2016/9739505

Ou, X., Zhang, Y., Xu, C., Lin, X., Zang, Q., Zhuang, T., Jiang, L., von Wettstein, D., & Liu, B. (2012). Transgenerational Inheritance of Modified DNA Methylation Patterns and Enhanced Tolerance Induced by Heavy Metal Stress in Rice (Oryza sativa L.). *PLoS ONE*, *7*(9), e41143. https://doi.org/10.1371/journal.pone.0041143

Pecinka, A., & Mittelsten Scheid, O. (2012). Stress-Induced Chromatin Changes: A Critical View on Their Heritability. *Plant and Cell Physiology*, *53*(5), 801-808. https://doi.org/10.1093/pcp/pcs044

Pegoraro, C., da Rosa Farias, D., & de Oliveira, A. C. (2018). Oryza glumaepatula Steud. En T. K. Mondal & R. J. Henry (Eds.), *The Wild Oryza Genomes* (pp. 127-135). Springer International Publishing. https://doi.org/10.1007/978-3-319-71997-9_10

Peñuela M, Riccio-Rengifo C, Finke J, Rocha C, Gkanogiannis A, Wing RA, et al. (2023). Prediction of crossover recombination using parental genomes. *PLoS ONE*. 18(2), e0281804. https://doi.org/10.1371/journal.pone.0281804

Perrella, G., Consiglio, M. F., Aiese-Cigliano, R., Cremona, G., Sanchez-Moran, E., Barra, L., Errico, A., Bressan, R. A., Franklin, F. C. H., & Conicella, C. (2010). Histone hyperacetylation affects meiotic recombination and chromosome segregation in Arabidopsis: Histone acetylation in At meiosis. *The Plant Journal*, *62*(5), 796-806. https://doi.org/10.1111/j.1365-313X.2010.04191.x

Posso, D., Llano, J., Londoño, A., Lentini, Z., & Ghneim-Herrera, T. (2013). *Caracterización de la tolerancia al aluminio en genotipos de la especie silvestre de arroz Oryza glumaepatula Steud.* (pp. 68-75).

Rabinowicz, P. D. (2003). Genes and Transposons Are Differentially Methylated in Plants, but Not in Mammals. *Genome Research*, *13*(12), 2658-2664. https://doi.org/10.1101/gr.1784803

Rahavi, M. (2011). Transgenerational adaptation to heavy metal salts in Arabidopsis. *Frontiers in Plant Science*, *2*. https://doi.org/10.3389/fpls.2011.00091

Rahman, H., Jagadeeshselvam, N., Valarmathi, R., Sachin, B., Sasikala, R., Senthil, N., Sudhakar, D., Robin, S., & Muthurajan, R. (2014). Transcriptome analysis of salinity responsiveness in

contrasting genotypes of finger millet (Eleusine coracana L.) through RNA-sequencing. *Plant Molecular Biology*, *85*(4-5), 485-503. https://doi.org/10.1007/s11103-014-0199-4

Rahman, Md., Lee, S.-H., Ji, H., Kabir, A., Jones, C., & Lee, K.-W. (2018). Importance of Mineral Nutrition for Mitigating Aluminum Toxicity in Plants on Acidic Soils: Current Status and Opportunities. *International Journal of Molecular Sciences*, *19*(10), 3073. https://doi.org/10.3390/ijms19103073

Rajkumar, M. S., Shankar, R., Garg, R., & Jain, M. (2019). *Role of DNA methylation dynamics in desiccation and salinity stress responses in rice cultivars* [Preprint]. Genomics. https://doi.org/10.1101/558064

Rascio, N., & Navari-Izzo, F. (2011). Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Science*, *180*(2), 169-181. https://doi.org/10.1016/j.plantsci.2010.08.016

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191-W198. https://doi.org/10.1093/nar/gkz369

Reimand, J., Kull, M., Peterson, H., Hansen, J., & Vilo, J. (2007). g:Profiler—A web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research*, *35*(suppl_2), W193-W200. https://doi.org/10.1093/nar/gkm226

Richards, E. J. (2006). Inherited epigenetic variation—Revisiting soft inheritance. *Nature Reviews Genetics*, *7*(5), 395-401. https://doi.org/10.1038/nrg1834

Rodgers-Melnick, E., Bradbury, P. J., Elshire, R. J., Glaubitz, J. C., Acharya, C. B., Mitchell, S. E., Li, C., Li, Y., & Buckler, E. S. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences*, *112*(12), 3823-3828. https://doi.org/10.1073/pnas.1413864112

Samac, D. A., & Tesfaye, M. (2003). Plant improvement for tolerance to aluminum in acid soils – a review. *Review of Plant Biotechnology and Applied Genetics*, *75*, 189-207. https://doi.org/10.1023/A:1025843829545

Saze, H., & Kakutani, T. (2007). Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. *The EMBO Journal*, *26*(15), 3641-3652. https://doi.org/10.1038/sj.emboj.7601788

Schmidt, M., Byzova, M., Martens, C., Peeters, M., Raj, Y., Shukla, S., Verwulgen, T., DeBlock, M., & Van Lijsebettens, M. (2018). Methylome and Epialleles in Rice Epilines Selected for Energy Use Efficiency. *Agronomy*, *8*(9), 163. https://doi.org/10.3390/agronomy8090163

Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger, O., Schork, N. J., & Ecker, J. R. (2011). Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. *Science*, *334*(6054), 369-373. https://doi.org/10.1126/science.1212959

Schmohl, N., Pilling, J., Fisahn, J., & Horst, W. J. (2000). Pectin methylesterase modulates aluminium sensitivity in *Zea mays* and *Solanum tuberosum*. *Physiologia Plantarum*, *109*(4), 419-427. https://doi.org/10.1034/j.1399-3054.2000.100408.x

Seto, Y., Hamada, S., Ito, H., Masuta, C., Matsui, H., Nabeta, K., & Matsuura, H. (2011). Tobacco Salicylic Acid Glucosyltransferase Is Active toward Tuberonic Acid (12-Hydroxyjasmonic Acid) and Is Induced by Mechanical Wounding Stress. *Bioscience, Biotechnology, and Biochemistry*, *75*(12), 2316-2320. https://doi.org/10.1271/bbb.110454

Shabala, S., & Pottosin, I. (2014). Regulation of potassium transport in plants under hostile conditions: Implications for abiotic and biotic stress tolerance. *Physiologia Plantarum*, *151*(3), 257-279. https://doi.org/10.1111/ppl.12165

Shafiq, S., Zeb, Q., Ali, A., Sajjad, Y., Nazir, R., Widemann, E., & Liu, L. (2019). Lead, Cadmium and Zinc Phytotoxicity Alter DNA Methylation Levels to Confer Heavy Metal Tolerance in Wheat. *International Journal of Molecular Sciences*, *20*(19), 4676. https://doi.org/10.3390/ijms20194676

Shahid, M., Pourrut, B., Dumat, C., Nadeem, M., Aslam, M., & Pinelli, E. (2014). Heavy-Metal-Induced Reactive Oxygen Species: Phytotoxicity and Physicochemical Changes in Plants. En D. M. Whitacre (Ed.), *Reviews of Environmental Contamination and Toxicology Volume 232* (Vol. 232, pp. 1-44). Springer International Publishing. https://doi.org/10.1007/978-3-319-06746-9_1

Shahryary, Y., Hazarika, R. R., & Johannes, F. (2020). MethylStar: A fast and robust pre-processing pipeline for bulk or single-cell whole-genome bisulfite sequencing data. *BMC Genomics*, *21*(1), 479. https://doi.org/10.1186/s12864-020-06886-3

Shen, C., Li, D., He, R., Fang, Z., Xia, Y., Gao, J., Shen, H., & Cao, M. (2014). Comparative transcriptome analysis of RNA-seq data for cold-tolerant and cold-sensitive rice genotypes under cold stress. *Journal of Plant Biology*, *57*(6), 337-348. https://doi.org/10.1007/s12374-014-0183-1

Simon Andrews. (2010). *FastQC: A quality control tool for high throughput sequence data* (0.11.9) [Java]. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Sokol, A., Kwiatkowska, A., Jerzmanowski, A., & Prymakowska-Bosak, M. (2007). Up-regulation of stress-inducible genes in tobacco and Arabidopsis cells in response to abiotic stresses and ABA treatment correlates with dynamic changes in histone H3 and H4 modifications. *Planta*, *227*(1), 245-254. https://doi.org/10.1007/s00425-007-0612-1

Song, X., & Cao, X. (2017). Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Current Opinion in Plant Biology*, *36*, 111-118. https://doi.org/10.1016/j.pbi.2017.02.004

Stein, J. C., Yu, Y., Copetti, D., Zwickl, D. J., Zhang, L., Zhang, C., Chougule, K., Gao, D., Iwata, A., Goicoechea, J. L., Wei, S., Wang, J., Liao, Y., Wang, M., Jacquemin, J., Becker, C., Kudrna, D., Zhang, J., Londono, C. E. M., … Wing, R. A. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. *Nature Genetics*, *50*(2), 285-296. https://doi.org/10.1038/s41588-018-0040-0

Stroud, H., Ding, B., Simon, S. A., Feng, S., Bellizzi, M., Pellegrini, M., Wang, G.-L., Meyers, B. C., & Jacobsen, S. E. (2013). Plants regenerated from tissue culture contain stable epigenome changes in rice. *ELife*, *2*, e00354. https://doi.org/10.7554/eLife.00354

Sudan, J., Raina, M., & Singh, R. (2018). Plant epigenetic mechanisms: Role in abiotic stress and their generational heritability. *3 Biotech*, *8*(3), 172. https://doi.org/10.1007/s13205-018-1202-6

Sun, M., Yang, Z., Liu, L., & Duan, L. (2022). DNA Methylation in Plant Responses and Adaption to Abiotic Stresses. *International Journal of Molecular Sciences*, *23*(13), 6910. https://doi.org/10.3390/ijms23136910

Sun, Y., Fan, M., & He, Y. (2019). DNA Methylation Analysis of the Citrullus lanatus Response to Cucumber Green Mottle Mosaic Virus Infection by Whole-Genome Bisulfite Sequencing. *Genes*, *10*(5), 344. https://doi.org/10.3390/genes10050344

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). *REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms*.

Taspinar, M. S., Aydin, M., Sigmaz, B., Yagci, S., Arslan, E., & Agar, G. (2018). Aluminum-Induced Changes on DNA Damage, DNA Methylation and LTR Retrotransposon Polymorphism in Maize. *Arabian Journal for Science and Engineering*, *43*(1), 123-131. https://doi.org/10.1007/s13369-017-2697-6

Taudt, A., Roquis, D., Vidalis, A., Wardenaar, R., Johannes, F., & Colomé-Tatché, M. (2018). METHimpute: Imputation-guided construction of complete methylomes from WGBS data. *BMC Genomics*, *19*(1), 444. https://doi.org/10.1186/s12864-018-4641-x

Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J. L., Jackson, S. A., Gaut, B. S., & Ma, J. (2009). Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Research*, *19*(12), 2221-2230. https://doi.org/10.1101/gr.083899.108

Tsutsui, T., Yamaji, N., Huang, C. F., Motoyama, R., Nagamura, Y., & Ma, J. F. (2012). Comparative Genome-Wide Transcriptional Analysis of Al-Responsive Genes Reveals Novel Al Tolerance Mechanisms in Rice. *PLoS ONE*, *7*(10), e48197. https://doi.org/10.1371/journal.pone.0048197

Turner, B. M. (2009). Epigenetic responses to environmental change and their evolutionary implications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1534), 3403-3418. https://doi.org/10.1098/rstb.2009.0125

Tyagi, W., Yumnam, J. S., Sen, D., & Rai, M. (2020). Root transcriptome reveals efficient cell signaling and energy conservation key to aluminum toxicity tolerance in acidic soil adapted rice genotype. *Scientific Reports*, *10*(1), 4580. https://doi.org/10.1038/s41598-020-61305-7

Ueda, M., & Seki, M. (2020). Histone Modifications Form Epigenetic Regulatory Networks to Regulate Abiotic Stress Response. *Plant Physiology*, *182*(1), 15-26. https://doi.org/10.1104/pp.19.00988

van der Graaf, A., Wardenaar, R., Neumann, D. A., Taudt, A., Shaw, R. G., Jansen, R. C., Schmitz, R. J., Colomé-Tatché, M., & Johannes, F. (2015). Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences*, *112*(21), 6676-6681. https://doi.org/10.1073/pnas.1424254112

Verhoeven, K. J. F., Jansen, J. J., van Dijk, P. J., & Biere, A. (2010). Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytologist*, *185*(4), 1108-1118. https://doi.org/10.1111/j.1469-8137.2009.03121.x

Volkova, P. Yu., Geras'kin, S. A., Horemans, N., Makarenko, E. S., Saenen, E., Duarte, G. T., Nauts, R., Bondarenko, V. S., Jacobs, G., Voorspoels, S., & Kudin, M. (2018). Chronic radiation exposure as an ecological factor: Hypermethylation and genetic differentiation in irradiated Scots pine populations. *Environmental Pollution*, *232*, 105-112. https://doi.org/10.1016/j.envpol.2017.08.123

Wagatsuma, T., Maejima, E., Watanabe, T., Toyomasu, T., Kuroda, M., Muranaka, T., Ohyama, K., Ishikawa, A., Usui, M., Hossain Khan, S., Maruyama, H., Tawaraya, K., Kobayashi, Y., & Koyama, H. (2018). Dark conditions enhance aluminum tolerance in several rice cultivars

via multiple modulations of membrane sterols. *Journal of Experimental Botany*, *69*(3), 567-577. https://doi.org/10.1093/jxb/erx414

Walia, H., Wilson, C., Condamine, P., Liu, X., Ismail, A. M., Zeng, L., Wanamaker, S. I., Mandal, J., Xu, J., Cui, X., & Close, T. J. (2005). Comparative Transcriptional Profiling of Two Contrasting Rice Genotypes under Salinity Stress during the Vegetative Growth Stage. *Plant Physiology*, *139*(2), 822-835. https://doi.org/10.1104/pp.105.065961

Walia, H., Wilson, C., Zeng, L., Ismail, A. M., Condamine, P., & Close, T. J. (2007). Genome-wide transcriptional analysis of salinity stressed japonica and indica rice genotypes during panicle initiation stage. *Plant Molecular Biology*, *63*(5), 609-623. https://doi.org/10.1007/s11103-006-9112-0

Wang, C., Wang, C., Xu, W., Zou, J., Qiu, Y., Kong, J., Yang, Y., Zhang, B., & Zhu, S. (2018). Epigenetic Changes in the Regulation of Nicotiana tabacum Response to Cucumber Mosaic Virus Infection and Symptom Recovery through Single-Base Resolution Methylomes. *Viruses*, *10*(8), 402. https://doi.org/10.3390/v10080402

Wang, L., Zheng, K., Zeng, L., Xu, D., Zhu, T., Yin, Y., Zhan, H., Wu, Y., & Yang, D.-L. (2022). Reinforcement of CHH methylation through RNA-directed DNA methylation ensures sexual reproduction in rice. *Plant Physiology*, *188*(2), 1189-1209. https://doi.org/10.1093/plphys/kiab531

Wang, M., Yu, Y., Haberer, G., Marri, P. R., Fan, C., Goicoechea, J. L., Zuccolo, A., Song, X., Kudrna, D., Ammiraju, J. S. S., Cossu, R. M., Maldonado, C., Chen, J., Lee, S., Sisneros, N., de Baynast, K., Golser, W., Wissotski, M., Kim, W., … Wing, R. A. (2014). The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. *Nature Genetics*, *46*(9), 982-988. https://doi.org/10.1038/ng.3044

Wang, W.-S., Pan, Y.-J., Zhao, X.-Q., Dwivedi, D., Zhu, L.-H., Ali, J., Fu, B.-Y., & Li, Z.-K. (2011). Drought-induced site-specific DNA methylation and its association with drought tolerance in rice (Oryza sativa L.). *Journal of Experimental Botany*, *62*(6), 1951-1960. https://doi.org/10.1093/jxb/erq391

Wang, X., Ai, S., & Liao, H. (2023). Deciphering Interactions between Phosphorus Status and Toxic Metal Exposure in Plants and Rhizospheres to Improve Crops Reared on Acid Soil. *Cells*, *12*(3), 441. https://doi.org/10.3390/cells12030441

Whitten, M. G., Wong, M. T. F., & Rate, A. W. (2000). Amelioration of subsurface acidity in the south-west of Western Australia: Downward movement and mass balance of surface-incorporated lime after 2-15 years. *Soil Research*, *38*(3), 711. https://doi.org/10.1071/SR99054

Wu, P., Liao, C. Y., Hu, B., Yi, K. K., Jin, W. Z., Ni, J. J., & He, C. (2000). QTLs and epistasis for aluminum tolerance in rice (Oryza sativa L.) at different seedling stages: *Theoretical and Applied Genetics*, *100*(8), 1295-1303. https://doi.org/10.1007/s001220051438

Xia, J., Yamaji, N., Kasai, T., & Ma, J. F. (2010). Plasma membrane-localized transporter for aluminum in rice. *Proceedings of the National Academy of Sciences*, *107*(43), 18381-18385. https://doi.org/10.1073/pnas.1004949107

Xing, M.-Q., Zhang, Y.-J., Zhou, S.-R., Hu, W.-Y., Wu, X.-T., Ye, Y.-J., Wu, X.-X., Xiao, Y.-P., Li, X., & Xue, H.-W. (2015). Global Analysis Reveals the Crucial Roles of DNA Methylation during Rice Seed Development. *Plant Physiology*, *168*(4), 1417-1432. https://doi.org/10.1104/pp.15.00414

Xue, Y., Jiang, L., Su, N., Wang, J. K., Deng, P., Ma, J. F., Zhai, H. Q., & Wan, J. M. (2007). The genetic basic and fine-mapping of a stable quantitative-trait loci for aluminium tolerance in rice. *Planta*, *227*(1), 255-262. https://doi.org/10.1007/s00425-007-0613-0

Xue, Y., Wan, J., Jiang, L., Wang, C., Liu, L., Zhang, Y., & Zhai, H. (2006). Identification of Quantitative Trait Loci Associated with Aluminum Tolerance in Rice (Oryza Sativa L.). *Euphytica*, *150*(1-2), 37-45. https://doi.org/10.1007/s10681-006-9089-4

Yamaji, N., Huang, C. F., Nagao, S., Yano, M., Sato, Y., Nagamura, Y., & Ma, J. F. (2009). A Zinc Finger Transcription Factor ART1 Regulates Multiple Genes Implicated in Aluminum Tolerance in Rice. *The Plant Cell*, *21*(10), 3339-3349. https://doi.org/10.1105/tpc.109.070771

Yan, H., Jin, W., Nagaki, K., Tian, S., Ouyang, S., Buell, C. R., Talbert, P. B., Henikoff, S., & Jiang, J. (2005). Transcription and Histone Modifications in the Recombination-Free Region Spanning a Rice Centromere[W]. *The Plant Cell*, *17*(12), 3227-3238. https://doi.org/10.1105/tpc.105.037945

Yan, H., Kikuchi, S., Neumann, P., Zhang, W., Wu, Y., Chen, F., & Jiang, J. (2010). Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice: Genome-wide methylation of rice. *The Plant Journal*, *63*(3), 353-365. https://doi.org/10.1111/j.1365-313X.2010.04246.x

Yang, J., Sun, K., Li, D., Luo, L., Liu, Y., Huang, M., Yang, G., Liu, H., Wang, H., Chen, Z., & Guo, T. (2019). Identification of stable QTLs and candidate genes involved in anaerobic germination tolerance in rice via high-density genetic mapping and RNA-Seq. *BMC Genomics*, *20*(1), 355. https://doi.org/10.1186/s12864-019-5741-y

Yokosho, K., Yamaji, N., & Ma, J. F. (2011). An Al-inducible MATE gene is involved in external detoxification of Al in rice: An Al-inducible MATE gene. *The Plant Journal*, *68*(6), 1061-1069. https://doi.org/10.1111/j.1365-313X.2011.04757.x

Yokosho, K., Yamaji, N., & Ma, J. F. (2014). Global Transcriptome Analysis of Al-Induced Genes in an Al-Accumulating Species, Common Buckwheat (Fagopyrum esculentum Moench). *Plant and Cell Physiology*, *55*(12), 2077-2091. https://doi.org/10.1093/pcp/pcu135

Zakrzewski, F., Schmidt, M., Van Lijsebettens, M., & Schmidt, T. (2017). DNA methylation of retrotransposons, DNA transposons and genes in sugar beet (*Beta vulgaris* L.). *The Plant Journal*, *90*(6), 1156-1175. https://doi.org/10.1111/tpj.13526

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science*, *328*(5980), 916-919. https://doi.org/10.1126/science.1186366

Zhang, H., Lang, Z., & Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, *19*(8), 489-506. https://doi.org/10.1038/s41580-018-0016-z

Zhang, J., Liu, S., Zhang, L., Nian, H., & Chen, L. (2016). Effect of aluminum stress on the expression of calmodulin and the role of calmodulin in aluminum tolerance. *Journal of Bioscience and Bioengineering*, *122*(5), 558-562. https://doi.org/10.1016/j.jbiosc.2016.04.001

Zhang, J., Liu, Y., Xia, E.-H., Yao, Q.-Y., Liu, X.-D., & Gao, L.-Z. (2015). Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proceedings of the National Academy of Sciences*, *112*(50), E7022-E7029. https://doi.org/10.1073/pnas.1515170112

Zhang, P., Ding, Z., Zhong, Z., & Tong, H. (2019). Transcriptomic Analysis for Indica and Japonica Rice Varieties under Aluminum Toxicity. *International Journal of Molecular Sciences*, *20*(4), 997. https://doi.org/10.3390/ijms20040997

Zhang, P., Zhong, K., Zhong, Z., & Tong, H. (2019). Mining candidate gene for rice aluminum tolerance through genome wide association study and transcriptomic analysis. *BMC Plant Biology*, *19*(1), 490. https://doi.org/10.1186/s12870-019-2036-z

Zhang, Q.-J., Zhu, T., Xia, E.-H., Shi, C., Liu, Y.-L., Zhang, Y., Liu, Y., Jiang, W.-K., Zhao, Y.-J., Mao, S.-Y., Zhang, L.-P., Huang, H., Jiao, J.-Y., Xu, P.-Z., Yao, Q.-Y., Zeng, F.-C., Yang, L.-L., Gao, J., Tao, D.-Y., … Gao, L.-Z. (2014). Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proceedings of the National Academy of Sciences*, *111*(46), E4954-E4962. https://doi.org/10.1073/pnas.1418307111

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W.-L., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E., & Ecker, J. R. (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell*, *126*(6), 1189-1201. https://doi.org/10.1016/j.cell.2006.08.003

Zheng, S. J. (2010). Crop production on acidic soils: Overcoming aluminium toxicity and phosphorus deficiency. *Annals of Botany*, *106*(1), 183-184. https://doi.org/10.1093/aob/mcq134

Zheng, X., Chen, L., Li, M., Lou, Q., Xia, H., Wang, P., Li, T., Liu, H., & Luo, L. (2013). Transgenerational Variations in DNA Methylation Induced by Drought Stress in Two Rice Varieties with Distinguished Difference to Drought Resistance. *PLoS ONE*, *8*(11), e80253. https://doi.org/10.1371/journal.pone.0080253

Zheng, X., Chen, L., Xia, H., Wei, H., Lou, Q., Li, M., Li, T., & Luo, L. (2017). Transgenerational epimutations induced by multi-generation drought imposition mediate rice plant's adaptation to drought condition. *Scientific Reports*, *7*(1), 39843. https://doi.org/10.1038/srep39843

Zhu, C. Q., Zhang, J. H., Sun, L. M., Zhu, L. F., Abliz, B., Hu, W. J., Zhong, C., Bai, Z. G., Sajid, H., Cao, X. C., & Jin, Q. Y. (2018). Hydrogen Sulfide Alleviates Aluminum Toxicity via Decreasing Apoplast and Symplast Al Contents in Rice. *Frontiers in Plant Science*, *9*, 294. https://doi.org/10.3389/fpls.2018.00294

Zhu, J.-K. (2016). Abiotic Stress Signaling and Responses in Plants. *Cell*, *167*(2), 313-324. https://doi.org/10.1016/j.cell.2016.08.029

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

Research articles

Gallo-Franco, J.J., Ghneim-Herrera, T., Tobar-Tosse, F., Romero, M., Chaura, J. & Quimbaya, M.A. (2022) Whole-genome DNA methylation patterns of *Oryza sativa* (L.) and *Oryza glumaepatula* (Steud) genotypes associated with aluminum response. Plant Direct: e430.

Peñuela, M., **Gallo-Franco, J.J.**, Finke, J., Rocha, C., Gkanogiannis, A., Ghneim-Herrera, T., Lorieux, M. (2022) Methylation in the CHH Context Allows to Predict Recombination in Rice. International Journal of Molecular Science. 23: 12505.

Gallo-Franco, J.J., Zuluaga-Yusti, I., Restrepo-García, A.M., Zapata-Balanta, S., Gutiérrez, J.P., Sosa, C.C., Ghneim-Herrera, T., and Quimbaya, M. Transcriptional analysis in wild and cultivated rice genotypes identifies core genes and mechanisms associated with aluminum tolerance. Plant Science. Submitted.

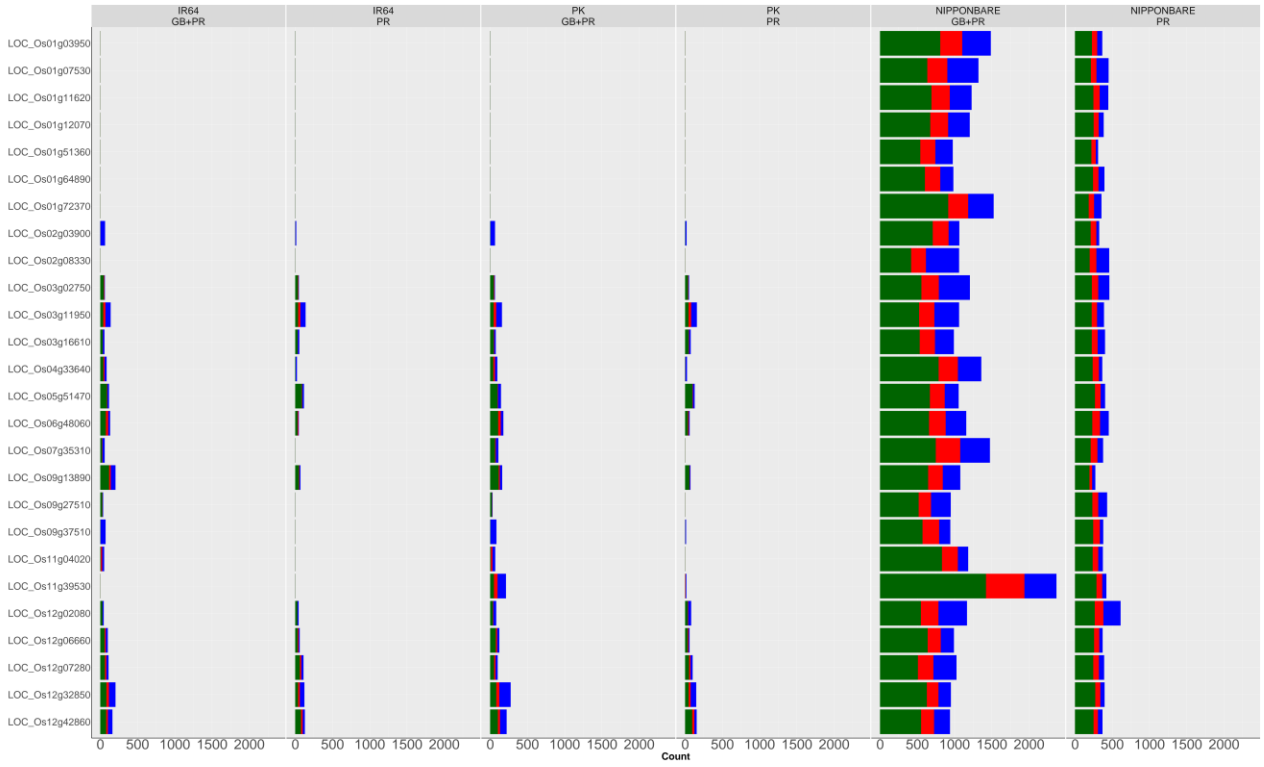
Abstracts, Presented at Scientific Conferences

Gallo-Franco, J.J., Tobar, F., Ghneim-Herrera, T., Quimbaya, M. Whole genome methylation patterns of *Oryza sativa* L. and *Oryza glumaepatula* Steud. genotypes with contrast response to aluminum stress. Virtual Symposium in Plant Omics Sciences (2020)

Gallo-Franco, J.J., Ghneim-Herrera, T., Tobar-Tosse, F., Romero, M., Chaura, J. & Quimbaya, M. Whole-genome DNA methylation patterns of *Oryza sativa* (L.) and *Oryza glumaepatula* (Steud) genotypes associated with aluminum response. EpiPlant, Banyuls–Francia. (2022)

Gallo-Franco, J.J., Tobar, F., Ghneim-Herrera, T., Quimbaya, M. Genome-wide bisulfite sequencing analysis of cultivated and wild rice species reveals epigenome variation in response to aluminum stress. ISCB-LA SOIBIO BioNetMX 2022 (2022) - Award for the best abstract talk

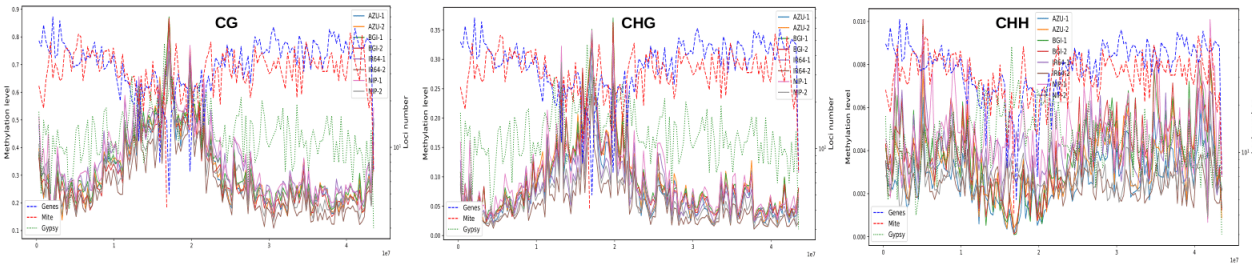
Supplementary Figures



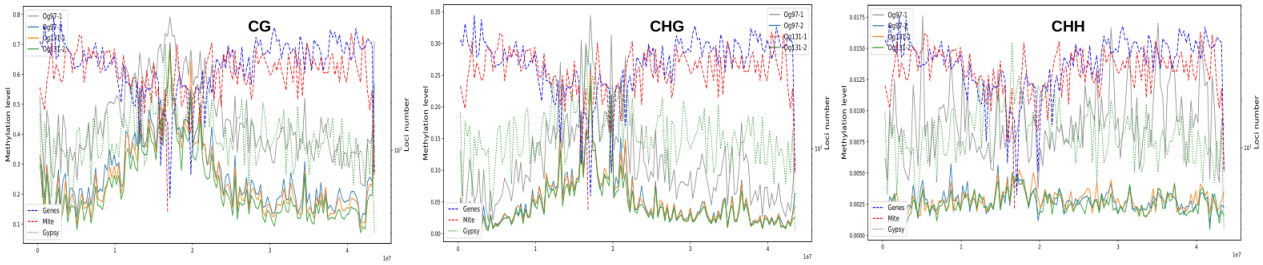
Supplementary figure 2.1. Stacked bar plot showing differences among three different cytosine methylated contexts: CG (blue), CHG (red), and CHH (green) per gene in three different rice varieties. The results are discriminated considering the location of the epigenetic mark, either in the promoter region (PR) or inside the coding region of analyzed genes.

Chromosome 1

Oryza sativa

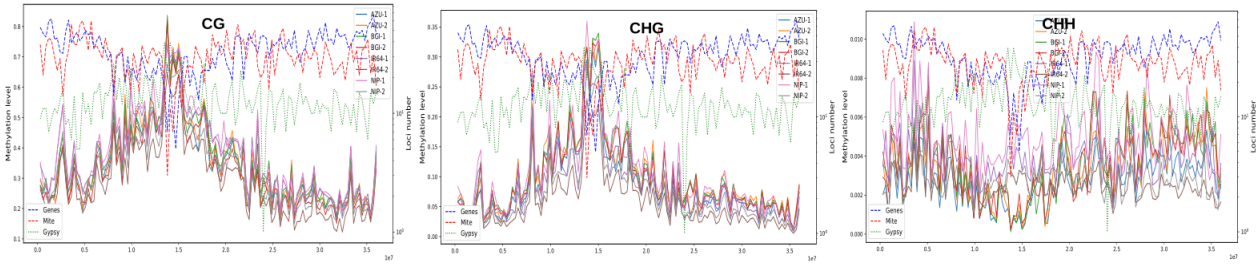


Oryza glumaepatula

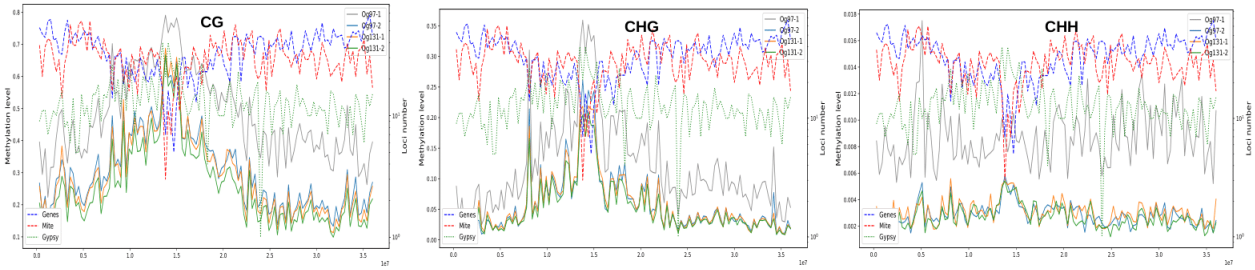


Chromosome 2

O. sativa



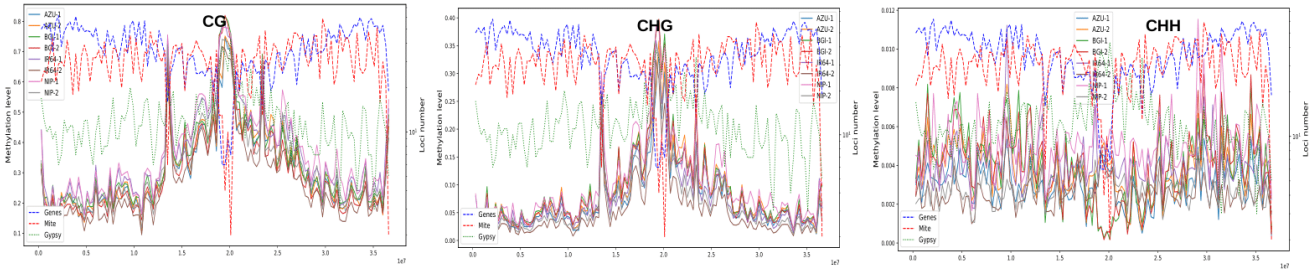
O. glumaepatula



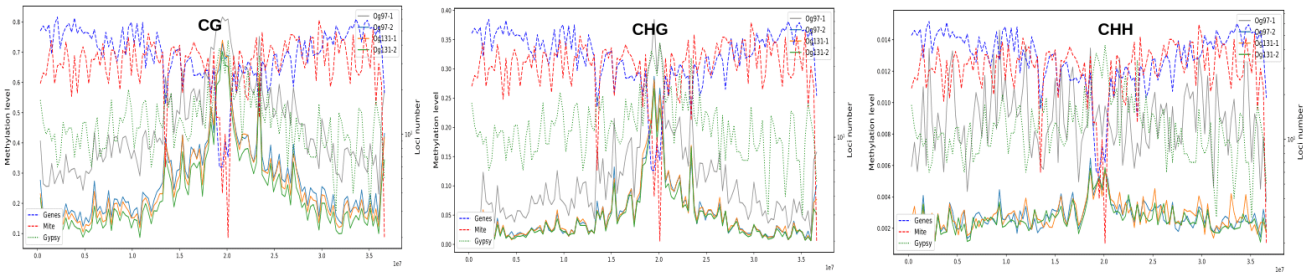
Supplementary figure 3.1. *Cont*

Chromosome 3

O. sativa

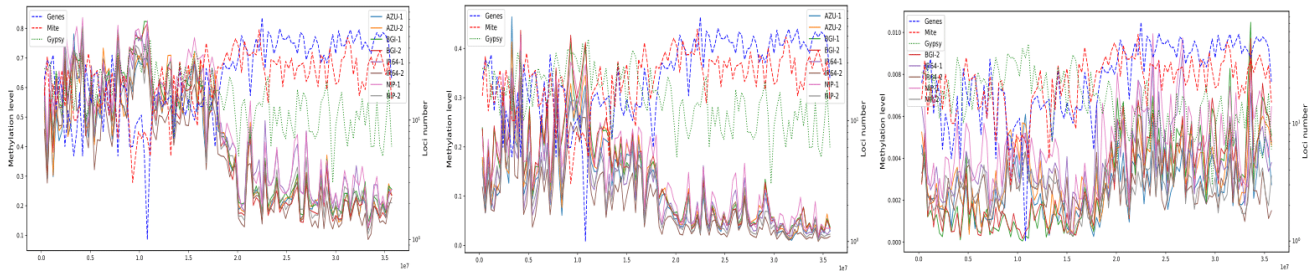


O. glumaepatula

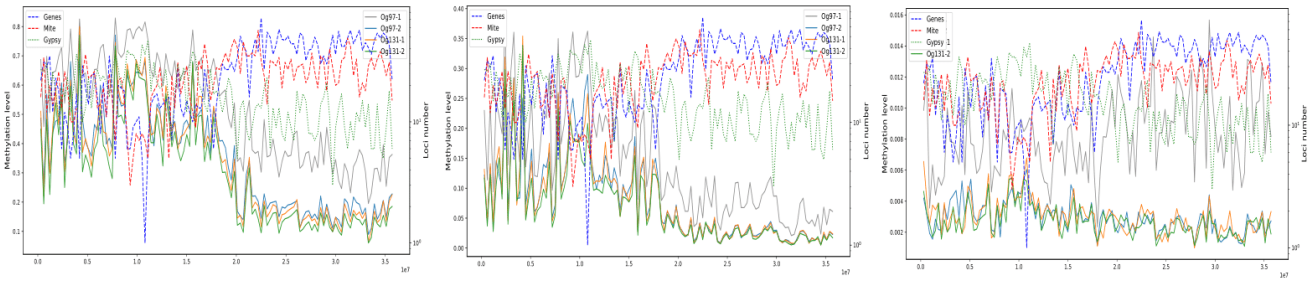


Chromosome 4

O. sativa



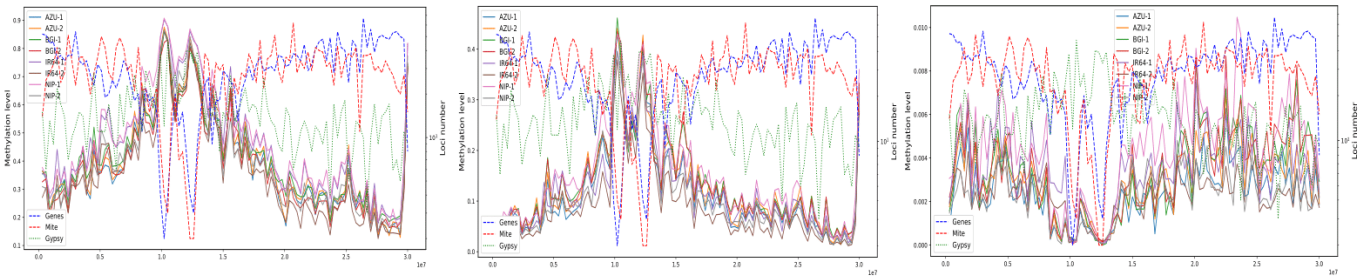
O. glumaepatula



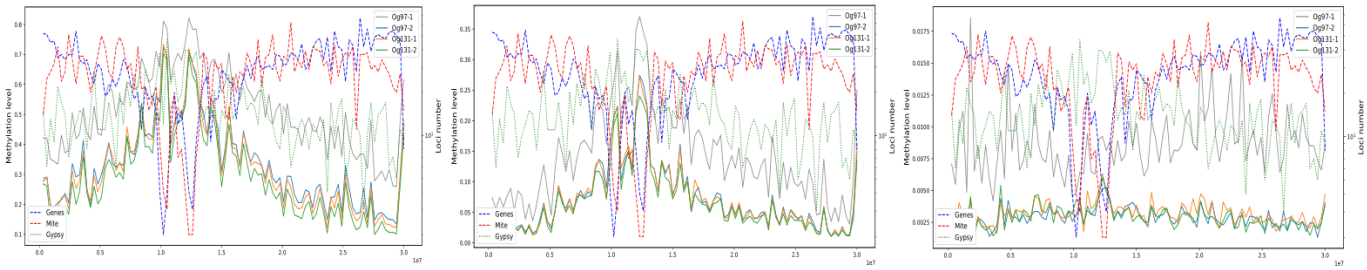
Supplementary figure 3.1. *Cont*

Chromosome 5

O. sativa

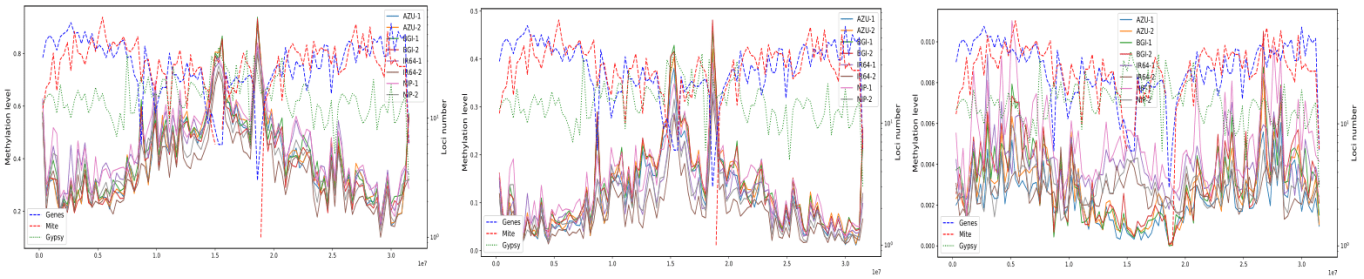


O. glumaepatula

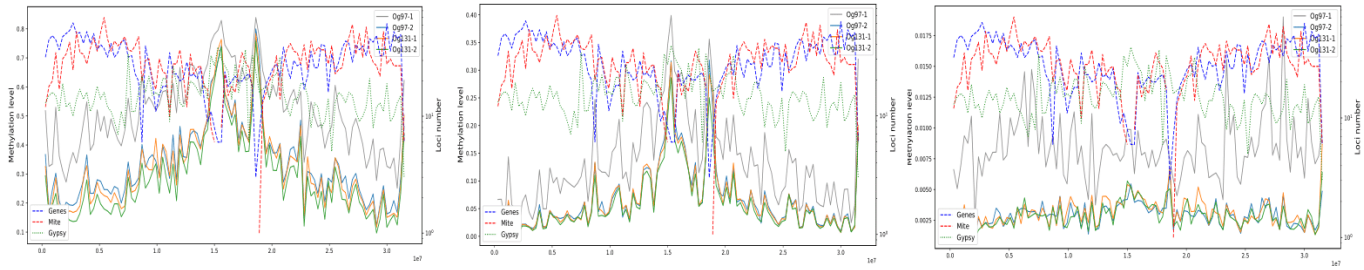


Chromosome 6

O. sativa



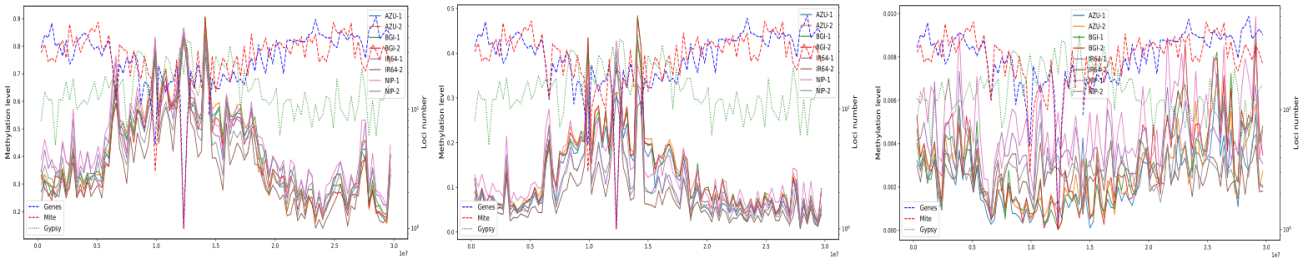
O. glumaepatula



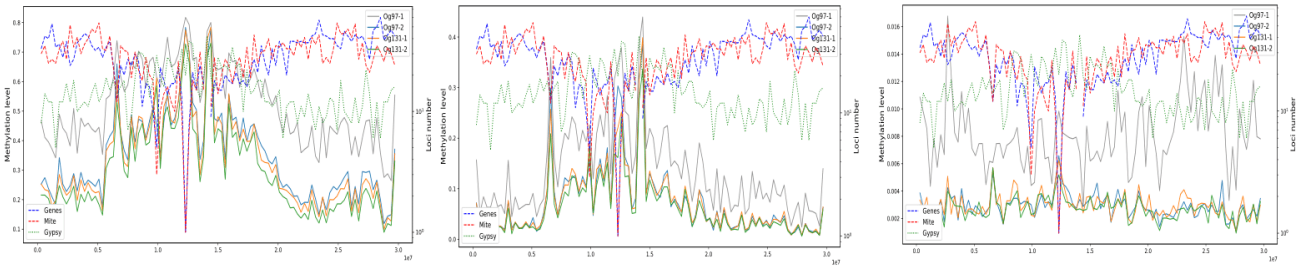
Supplementary figure 3.1. *Cont*

Chromosome 7

O. sativa

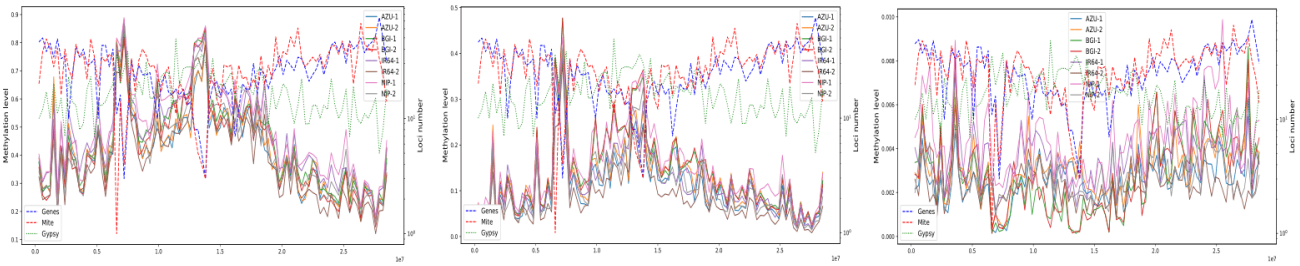


O. glumaepatula

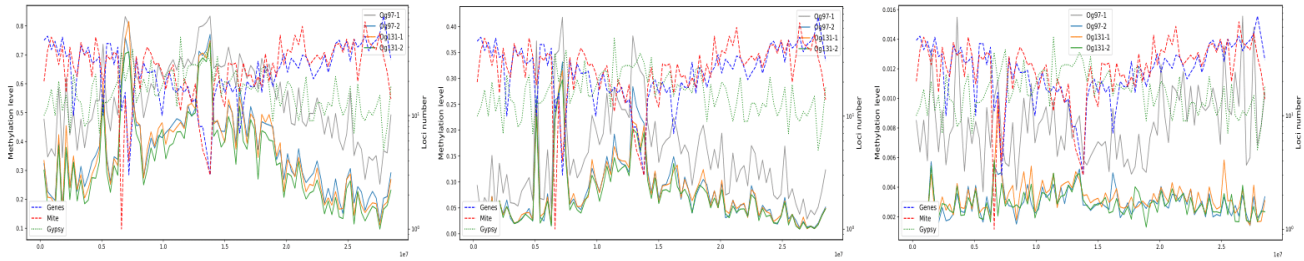


Chromosome 8

O. sativa



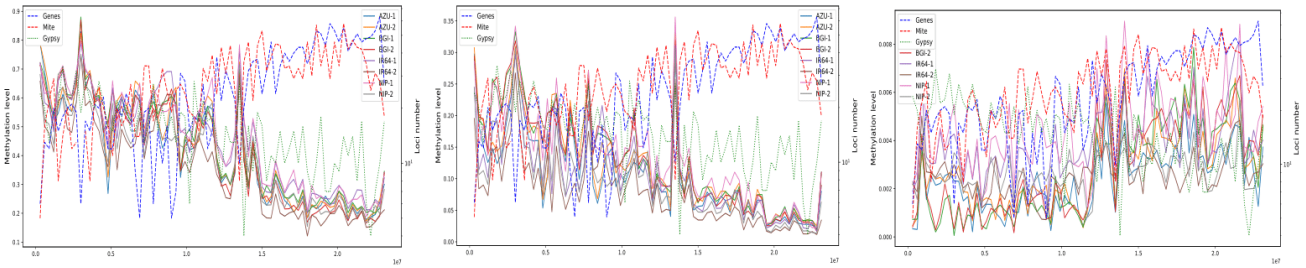
O. glumaepatula



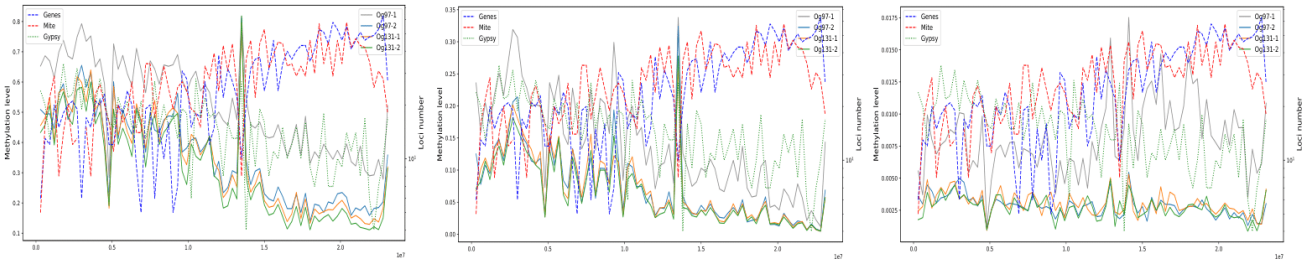
Supplementary figure 3.1. *Cont*

Chromosome 9

O. sativa

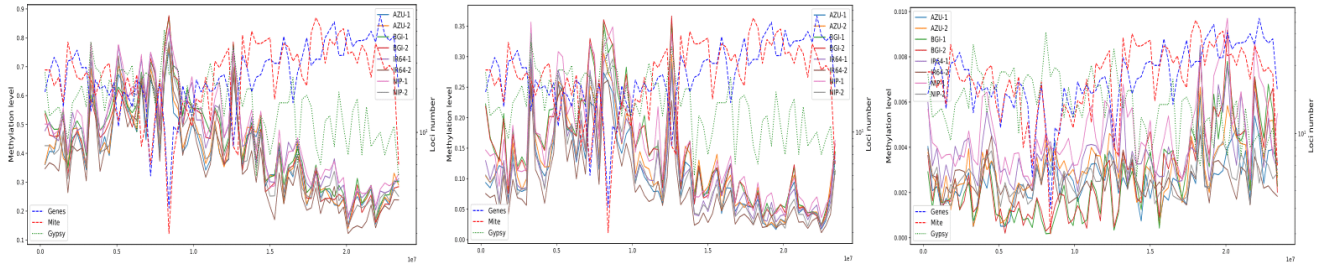


O. glumaepatula

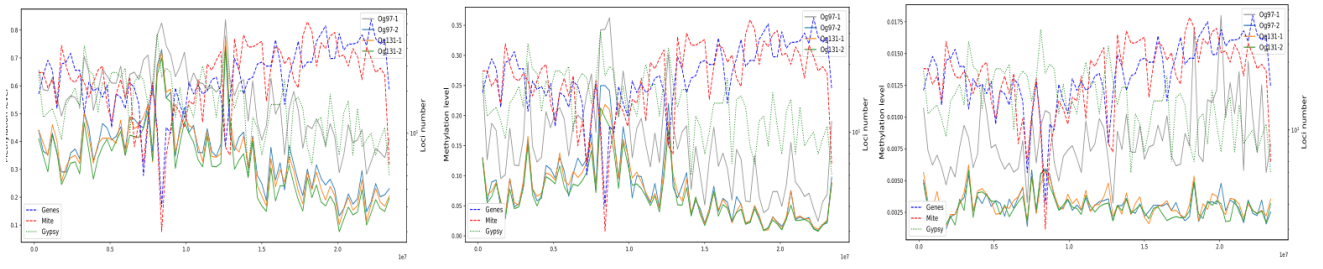


Chromosome 10

O. sativa



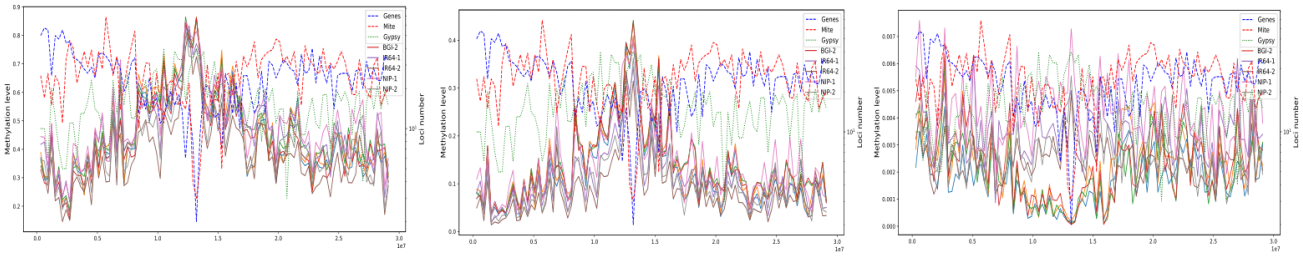
O. glumaepatula



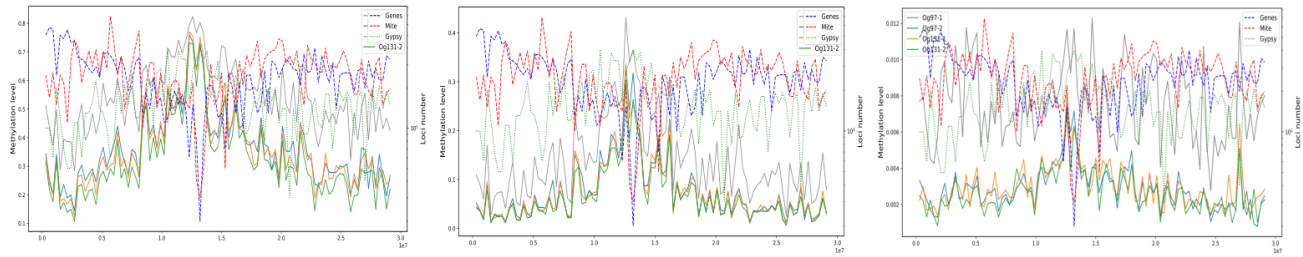
Supplementary figure 3.1 *Cont*

Chromosome 11

O. sativa

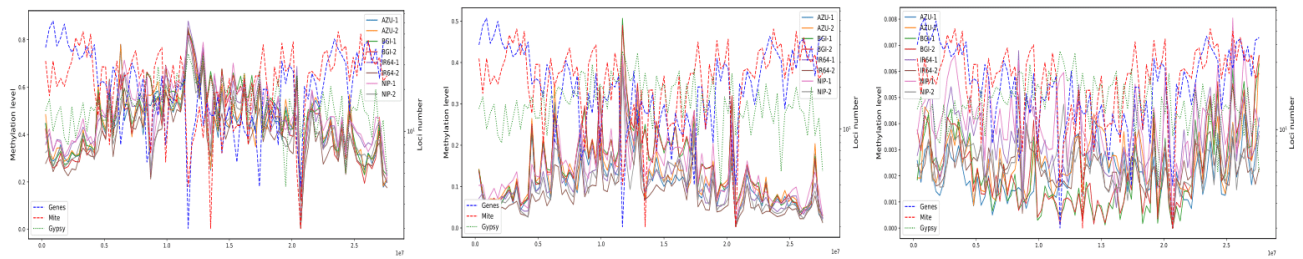


O. glumaepatula

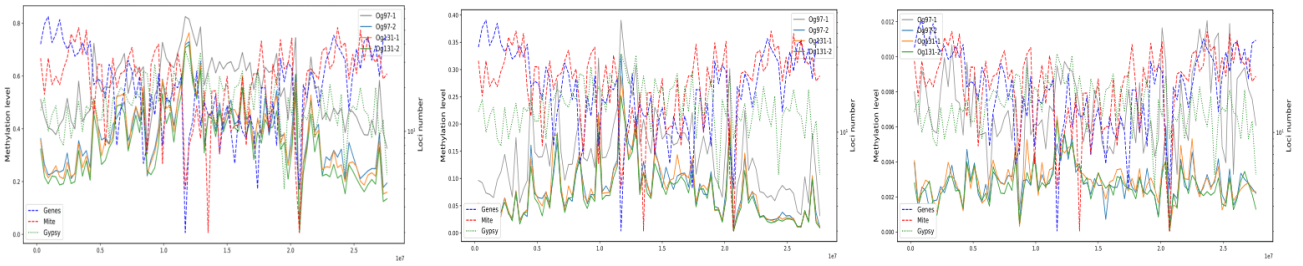


Chromosome 12

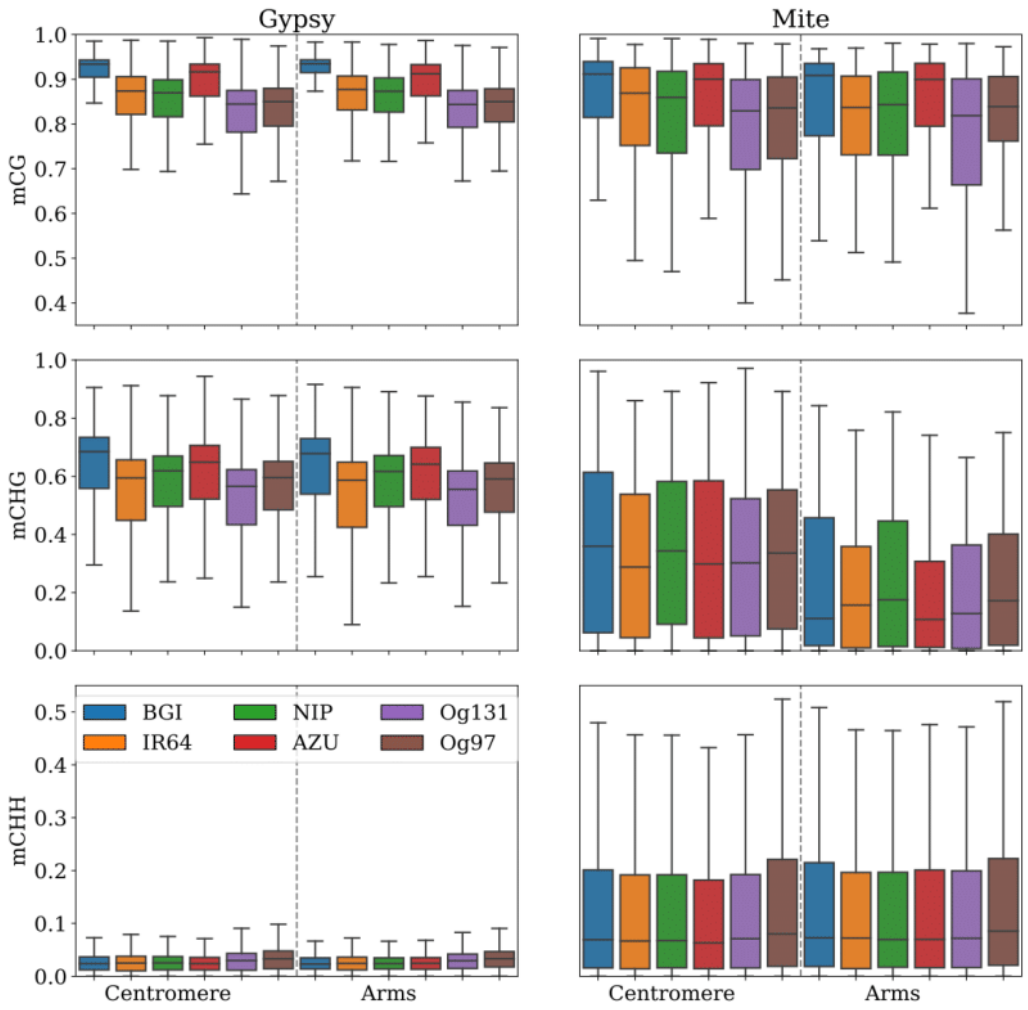
O. sativa



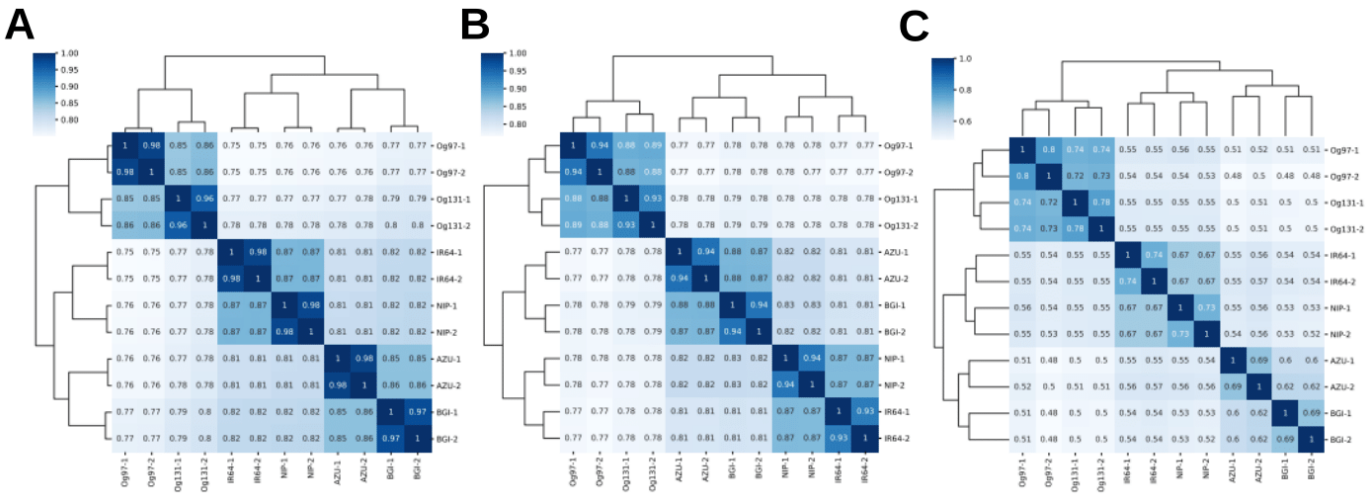
O. glumaepatula



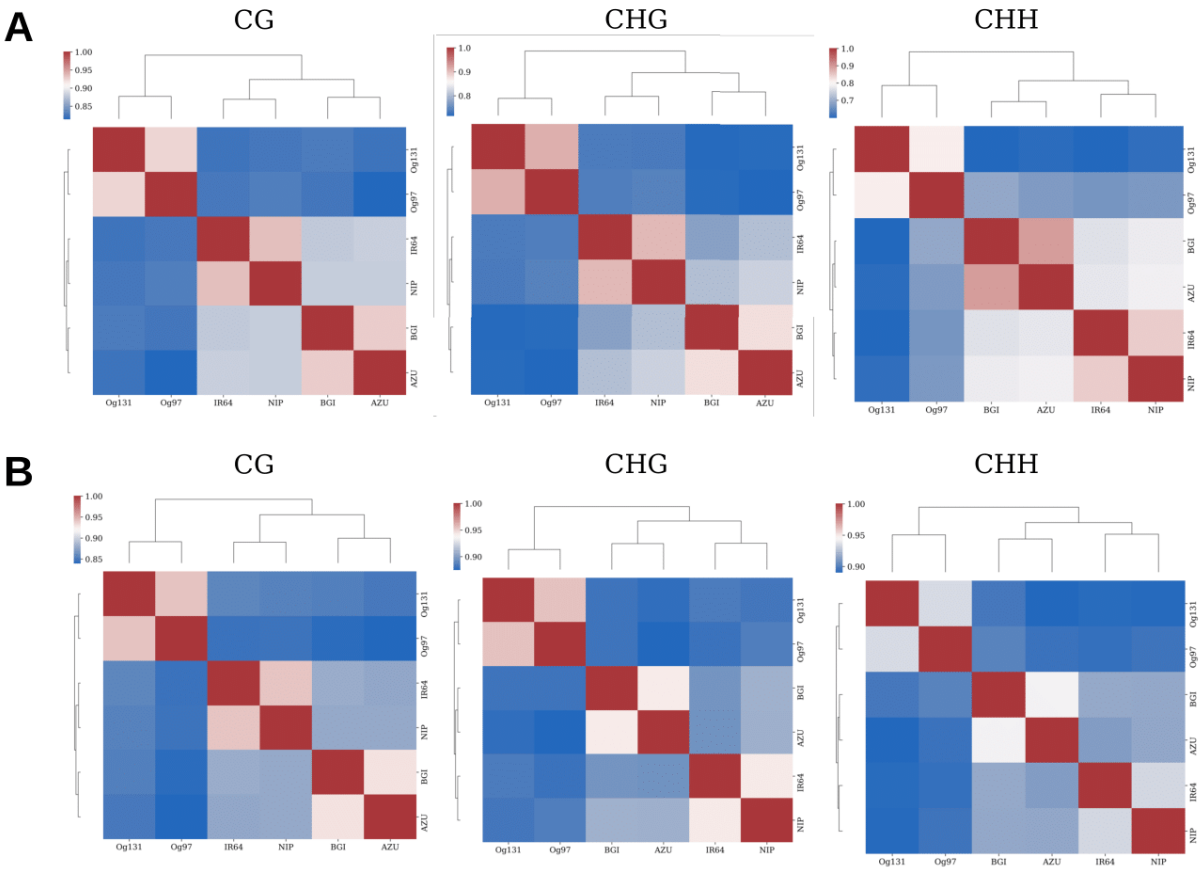
Supplementary figure 3.1. Methylated cytosines (mCs) per window (300Kpbs windows) throughout the genome with respect to the number of genes, Mite TEs and Gypsy TEs for *O. sativa* and *O. glumaepatula.*



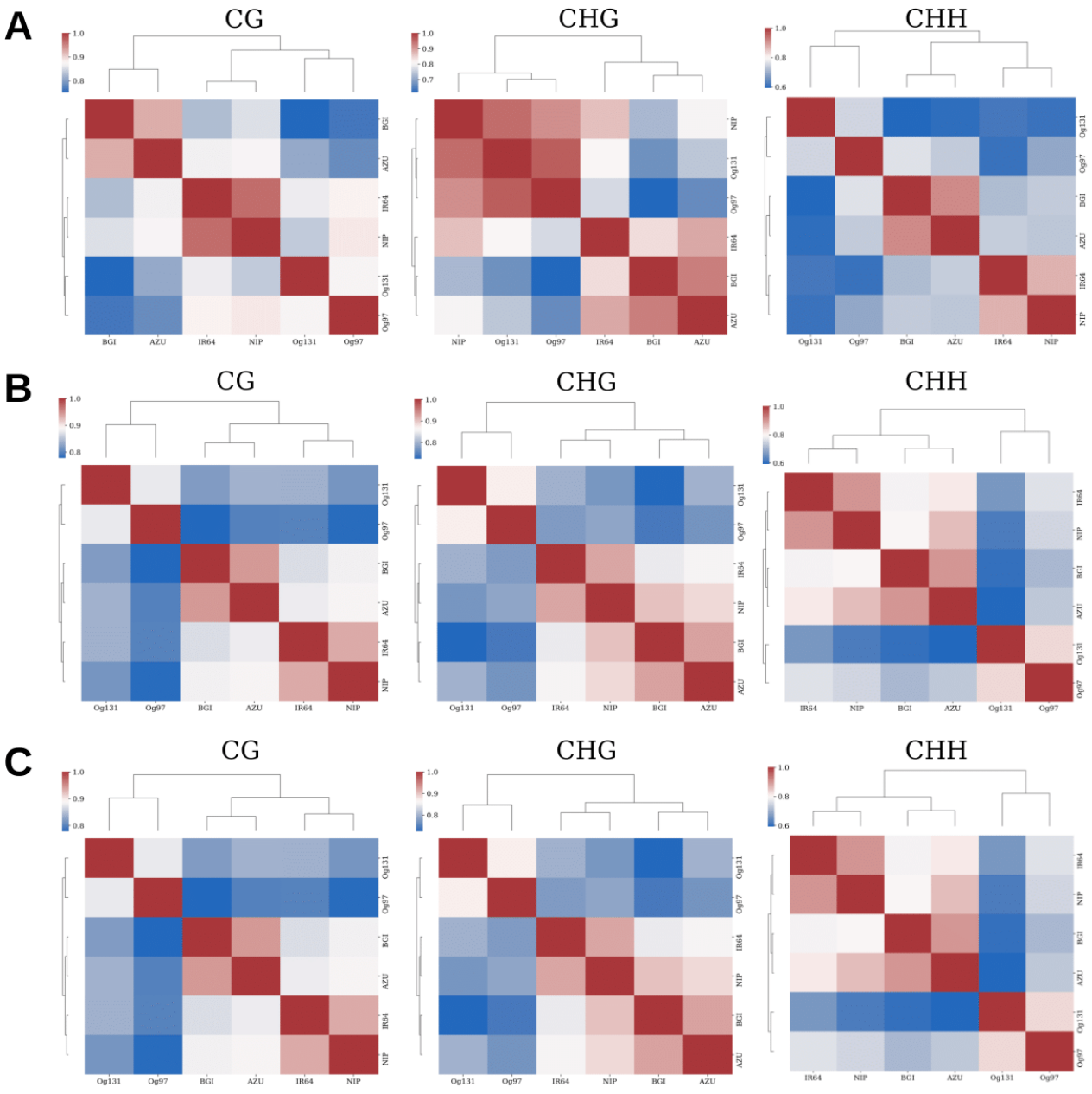
Supplementary figure 3.2. Boxplot showing methylation level variation for Gypsy and Mite TEs in the chromosome arms and the centromeric regions.



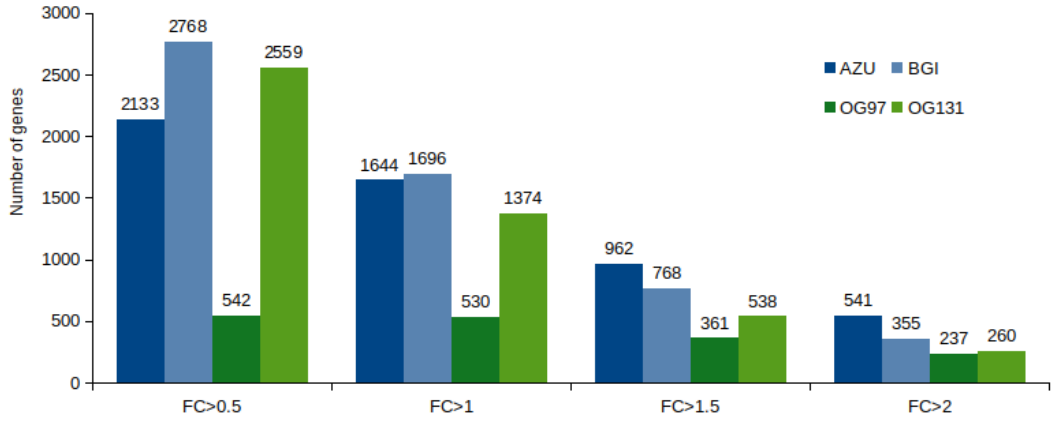
Supplementary figure 3.3. Pearson Correlation Coefficients (PCC) heatmap calculated using the methylation level for each cytosine position along the genome among all Oryza sativa and O. glumaepatula genotypes **A.** Heatmap for the CG context, **B**. for the CHG context and **C.** for the CHH context.



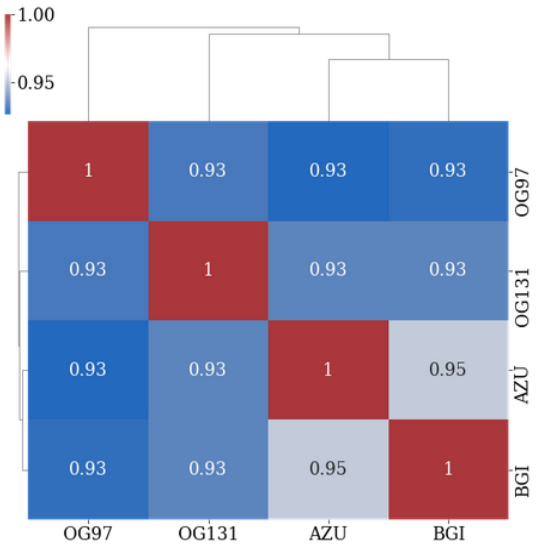
Supplementary figure 3.4. Clustering analysis of rice genotypes according to the Pearson Correlations Coefficients (PCC) among rice samples using methylation levels average for all the sequence contexts of **A.** genes and **B.** TEs.



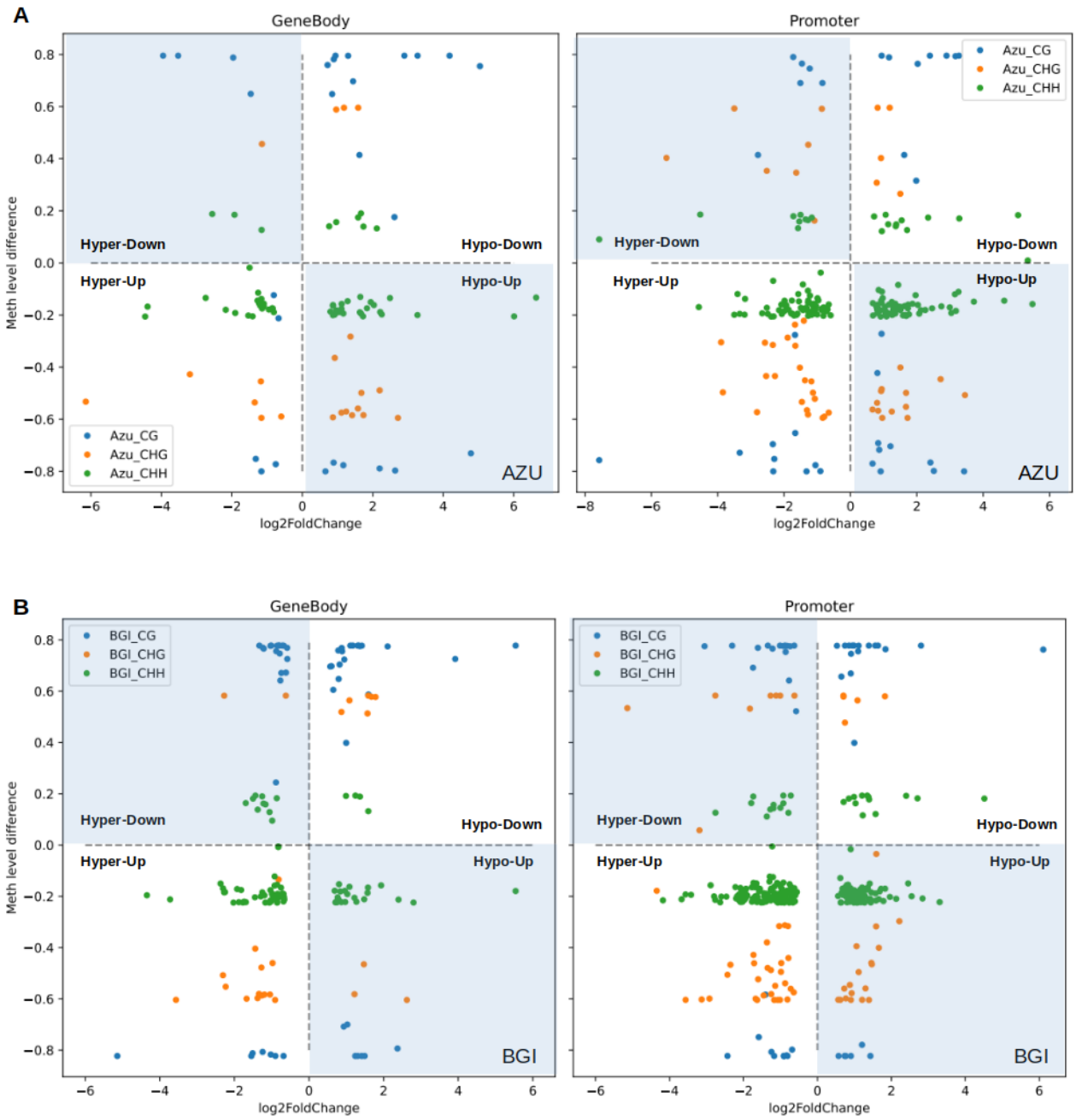
Supplementary figure 3.5. Clustering analysis of rice genotypes according to the Pearson Correlations Coefficients (PCC) among rice samples using the average methylation levels for all the sequence contexts. A) genebody, B) upstream region (-2Kb) and C) downstream region (+2Kb). The analysis was performed using the 250 genes reported as differentially expressed under aluminum exposure in rice (Arbelaez et al., 2017 and Arenhart et al., 2014).



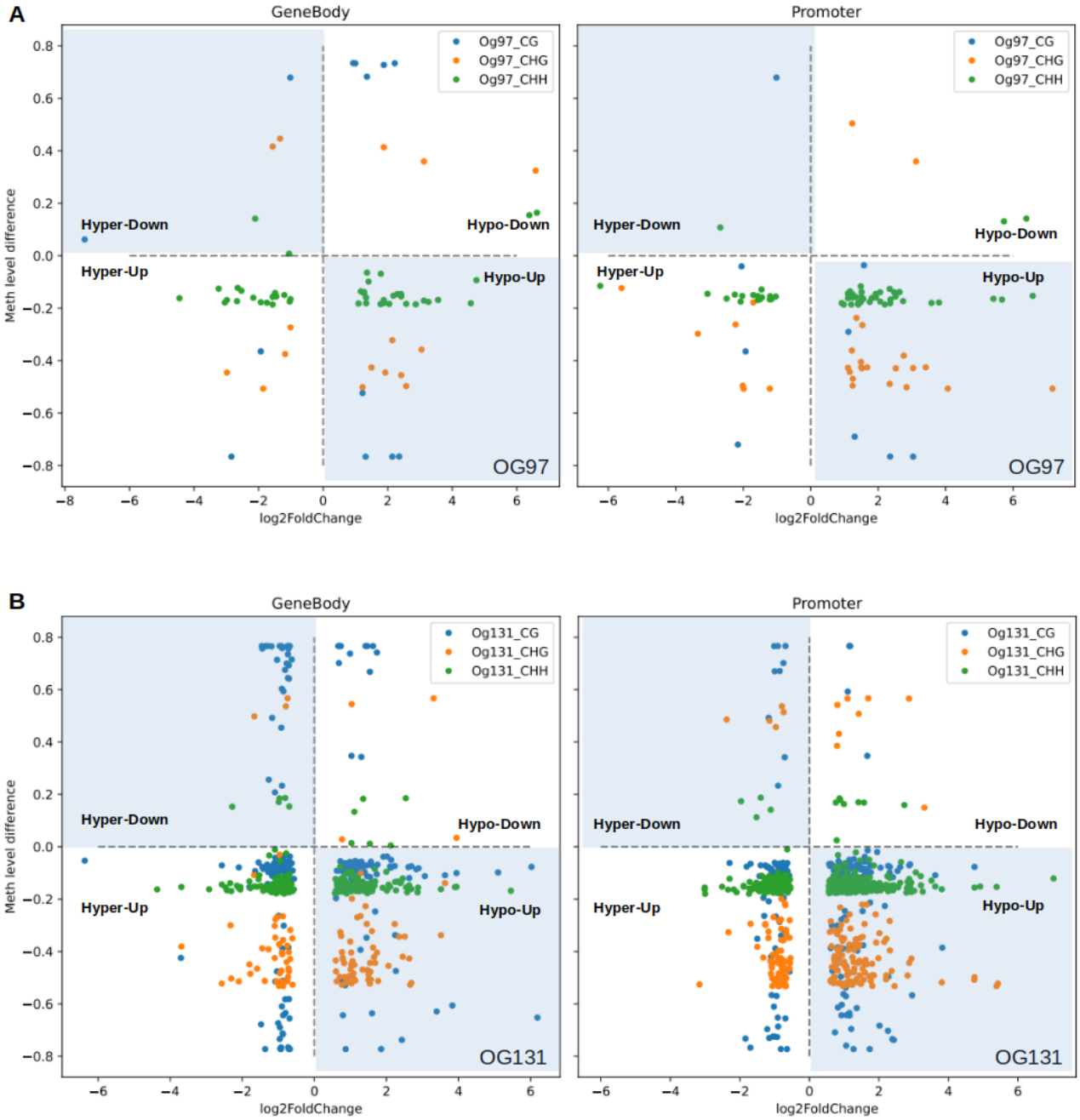
Supplementary figure 4.1. Number of differentially expressed genes with Log2FC values higher than 0.5, 1, 1.5 and 2



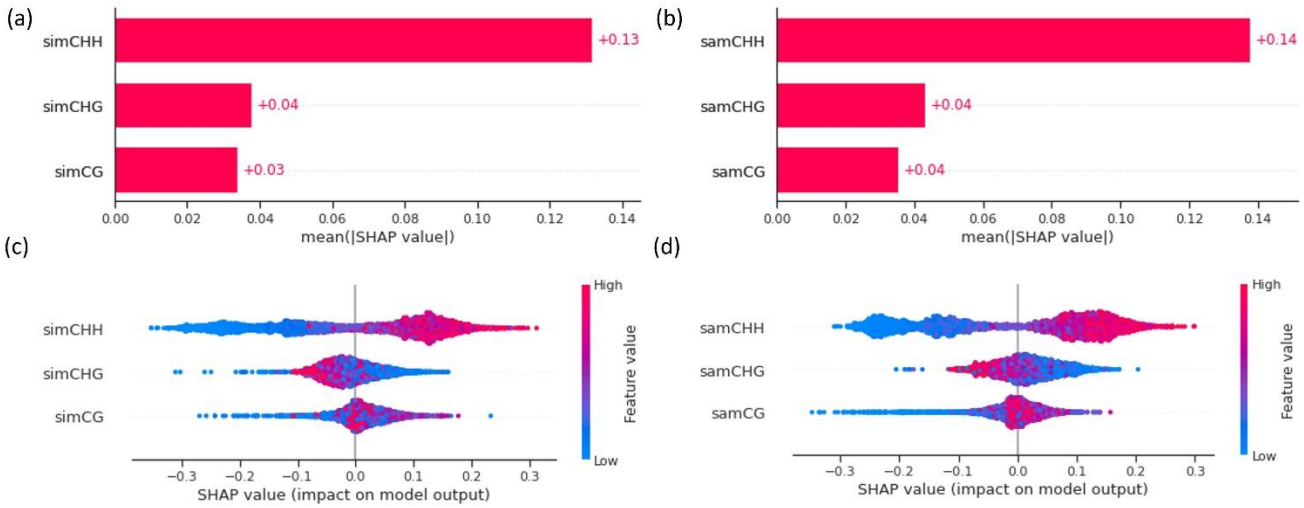
Supplementary figure 4.2. Pearson correlation values among rice genotypes based on the Log2FC values of 28 genes differentially expressed in all the genotypes.



Supplementary figure 5.1. Relationship between changes in methylation levels and changes in gene expression (Log2FC) for differentially methylated and expressed genes (DMG-DEGs) in two genotypes, **A.** Azucena and **B.** BGI, under aluminum exposure. Methylation level change was quantified by comparing control and stress samples for regions identified as DMRs that are associated with the DMG-DEGs.



Supplementary figure 5.2. Relationship between changes in methylation levels and changes in gene expression (Log2FC) for differentially methylated and expressed genes (DMG-DEGs) in two genotypes, **A.** OG131 and **B.** OG97, under aluminum exposure. Methylation level change was quantified by comparing control and stress samples for regions identified as DMRs that are associated with the DMG-DEGs.



Supplementary figure 6.1. Shap values and contributions of features CG, CHG, and CHH to the prediction of re-combination rates, using IR64 and Azucena data: (i) Shap summary plot of features for the IR64 variety; (ii) Shap summary plot of features for the Azucena variety; (iii) Shap values for the IR64 variety; and (iv) Shap values for the Azucena variety.

Supplementary Data 1.

RNA extraction protocol from roots and leaves for complex samples.

1. Add 1mL of Trizol to 200mg of tissue macerated in liquid nitrogen. 2.
2. Vortex for 5 min.
3. Incubate for 2 hours at 4°C and occasionally shake manually.
4. Centrifuge at 14000 rpm for 10min at 4°C.
5. Transfer 900uL of the supernatant to a new tube.
6. Add 500uL of PCI and vortex for 2min.
7. Incubate at 4°C for 10min with manual shaking every 2.5min.
8. Centrifuge at 14000rpm for 10 min at 4°C.
9. Transfer 400uL of the supernatant to a new tube.
10. Add 400uL of chloroform and vortex 30s.
11. Incubate for 3min at 4°C.
12. Centrifuge at 14000rpm for 15 min at 4°C.
13. Transfer the supernatant (approx. 300uL) to a new tube.
14. Add 300uL of chloroform and mix in vortex for 30s.
15. Incubate for 3min at 4°C.
16. Centrifuge at 14000rpm for 10 min at 4°C.
17. Transfer the supernatant (approx. 250uL) to a new tube.
18. Add 500uL of isopropanol and 80uL of 10M LiCl (should be at a final concentration of approximately 1M).
19. Incubate over night at -20°C.
20. Centrifuge at 14000rpm for 15min at 4°C.
21. Discard supernatant.
22. Add 1mL of 75% ethanol and vortex very briefly.
23. Centrifuge at 14000rpm for 10min at 4°C.
24. Discard the supernatant.
25. Add 500uL of 75% ethanol. 26.
26. Centrifuge at 14000rpm for 10min at 4°C.
27. Discard the supernatant
28. Allow the pellet to dry
29. Resuspend the pellet in 50uL of inactivated DEPC water.