

Deciphering the Physiological Role of Methylglyoxal in Rice

Thesis Submitted

to

Jawaharlal Nehru University
New Delhi, India

For the Award of the Degree of

Doctor of Philosophy

by

Brijesh Kumar



Plant Stress Biology Group

International Centre for Genetic Engineering and Biotechnology

New Delhi, India

2021



INTERNATIONAL CENTRE FOR GENETIC ENGINEERING AND BIOTECHNOLOGY

ICGEB, Aruna Asaf Ali Marg,
New Delhi-110067, India
<http://www.icgeb.res.in>

Tel: 91-11-26742358/60/61
91-11-26741007
Fax: 91-11-26742316
icgeb@icgeb.res.in

Certificate

This is to certify that the research work embodied in this thesis entitled **"Deciphering the physiological role of methylglyoxal in rice"** has been carried out by Brijesh Kumar, under my supervision at the Plant Stress Biology Group, International Centre for Genetic Engineering and Biotechnology, New Delhi for the award of the degree of Doctor of Philosophy. The presented research work is original and no part or full of this thesis has been submitted for the award or any other degree or diploma to any other university.

Dr. Sneh Lata Singla-Pareek

Group Leader
Plant Stress Biology Group
ICGEB, New Delhi-110067
India

Dr. Dinakar M. Salunke

Director
ICGEB, New Delhi-110067
India **DR. DINAKAR M. SALUNKE**
Director
International Centre for
Genetic Engineering & Biotechnology
New Delhi

Declaration

I hereby declare that this research work embodied in this thesis titled **"Deciphering the physiological role of methylglyoxal in rice"** has been carried out by me under the supervision of Dr. Sneha Lata Singla-Pareek in Plant Stress Biology Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, India.

Date: 29/12/2021

Place: New Delhi



Brijesh Kumar

Pre-doctoral Fellow
Plant Stress Biology Group
ICGEB
New Delhi-110067
India

Acknowledgements

To begin, I'd want to express my gratitude to Dr. Sneha Lata Singla-Pareek for allowing me to work in her lab for my PhD. She provided me with the flexibility to pursue the questions I was passionate about, and her mentoring style helped me become a far better problem solver than I could have imagined, a talent that will serve me well for the rest of my life. I'd also want to thank Dr. Suresh Nair and Prof. Ashwani Pareek, members of my advisory committee, for reading this thesis and providing me with helpful advice during my time as a PhD student.

I'd want to express my gratitude to all former and present members of the Plant Stress Biology Group, especially Charan ma'am and Brijesh Gupta sir. I'd also want to express my gratitude to the other Plant Group PIs at ICGEB, New Delhi, who volunteered to let me utilise their lab resources. PhD is difficult, but this incredible and supportive community of intelligent individuals made it easy and fun. I am so grateful to have had the opportunity to go through this experience with such wonderful friends.

I am grateful to Dr. Dinakar M. Salunke, Director of the ICGEB, for providing all of the essentials and amenities required for a fruitful working environment. I'd also want to thank the ICGEB administration for aiding me in my academic pursuits in many ways. I appreciate Rajeev Ji, Ramesh Ji, Harivansh Ji, and the rest of the Green House crew for providing the greatest care for my plants and maintaining the Green House operations. Special thanks to Baldev, Shafiq and Ranjit for all the assistance in the lab and fields.

I'd want to convey my thanks to Ashish Sir for his assistance in every manner he could. Manjari ma'am, Rathore, Surabhi and Yajna deserve special gratitude for making the lab enjoyable and memorable. Thank you, Deepti and Subia, for all of the memorable occasions at ICGEB and for always being a rock of support. I am grateful for your finest support, affection, and those joyful days, and I cheer for our relationship.

Shweta has been a wonderful friend and now life partner to me. Without your love and support, I would not have been able to finish my PhD. Thank you also to my parents and sisters for supporting me to follow my aspirations. Thank you, GOD, for providing me with such wonderful friends and family.

Finally, thank you to the Department of Biotechnology for funding my PhD at ICGEB, New Delhi through the DBT-JRF/SRF Fellowship.

List of Abbreviations, Acronyms and Symbols

S. No.	Abbreviations, Acronyms and Symbols	Full Name
1	~	Approximately
2	aa	Amino acid
3	APS	Ammonium persulphate
4	At	Arabidopsis thaliana
5	ATP	Adenosine triphosphate
	Amp	Ampicillin
6	β -ME	Beta-mercaptoethanol
7	bp	Base pair
8	BSA	Bovine Serum Albumin
9	cDNA	Complementary deoxyribonucleotide
10	C-terminal	Carboxy-terminal
11	Da	Dalton
12	$^{\circ}$ C	Degree centigrade
13	DNA	Deoxyribonucleic acid
14	dNTP	Deoxy ribonucleotide triphosphate
15	DAPI	4',6-diamidino-2-phenylindole
16	DEPC	Diethylpyrocarbonate
17	DTT	Dithiothreitol
18	<i>E. coli</i>	<i>Escherichia coli</i>
19	EDTA	Ethylenediaminetetra acetic acid
20	EtBr	Ethidium bromide
21	Fw	Forward primer
22	g	gram
23	GO	Glyoxal
24	GLYI	Glyoxalase I
25	GLYII	Glyoxalase II
26	GLYIII	Glyoxalase III
27	hr(s)	Hour(s)
28	His	Histidine
29	H ₂ O ₂	Hydrogen peroxide
30	pI	Isoelectric point
31	IPTG	Isopropyl β -D-1-thiogalactopyranoside
32	KCl	Potassium chloride
33	Kan	Kanamycin
34	Kb	Kilo base pairs
35	kDa	Kilo Dalton
36	LB	Luria-Bertani broth
37	L	Litre
38	MG	Methylglyoxal
39	μ g	Microgram
40	μ l	Micro-litre
41	μ M	Micro-molar
42	ml	Milliliter
43	mM	Millimolar
44	mg	Miligram
45	min	Minute(s)

46	mRNA	Messenger Ribonucleic acid
47	M	Molar
48	mol	Mole
49	ng	Nanogram
50	NAD ⁺	Nicotinamide adenine dinucleotide
51	NADPH	Nicotinamide adenine dinucleotide phosphate reduced
52	NaCl	Sodium chloride
53	N-terminal	Amino terminal
54	ORF	Open Reading Frame
55	OD	Optical density
56	Os	<i>Oryza sativa</i>
57	OsDJ	<i>Oryza sativa</i> DnaJ
58	GSSG	Oxidized Glutathione
59	%	Percentage
60	PBS	Phosphate buffered saline
61	PG	Phenylgloxal
62	PAGE	Polyacrylamide gel electrophoresis
63	PCR	Polymerase Chain Reaction
64	PDB	Protein Data Bank
65	qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
66	GSH	Reduced Glutathione
67	RWC	Relative Water Content
68	Re	Reverse Primer
69	RT-PCR	Reverse transcriptase Polymerase Chain Reaction
70	rpm	Revolutions per min
71	RNA	Ribonucleic acid
72	RT	Room temperature
73	s	Second
74	SLG	S-D-Lactoylglutathione
75	SDS	Sodium dodecyl sulfate
76	TEMED	N, N, N', N'-Tetramethylethylenediamine
77	<i>Taq</i>	<i>Thermus aquaticus</i>
78	TCA	Trichloroacetic acid
79	TFs	Transcription factors
80	V	Volts
81	v/v	volume/volume
82	WT	Wild type

Table of Contents

Chapter 1: INTRODUCTION AND SCOPE OF THESIS	1-5
Chapter 2: REVIEW OF LITERATURE	6-18
2.1 Stress in plants	
2.2 Methylglyoxal in plants	
2.3 Methylglyoxal formation	
2.3.1 MG production from carbohydrates	
2.3.2 MG production from proteins	
2.3.3 MG production from lipids	
2.4 MG toxicity	
2.5 Glyoxalase pathway in plants	
2.5.1 Glyoxalase I	
2.5.2 Glyoxalase II	
2.5.3 Glyoxalase III	
2.6 Other detoxification enzymes	
2.7 Physiological significance of MG	
2.7.1 Positive role: Role in stress signalling	
2.7.2 Negative role: Formation of AGEs, inhibition of cellular proliferation and cell death	
Chapter 3: MATERIALS AND METHODS	19-40
3.1 Bioinformatics Protocols	
3.1.1 Identification and nomenclature of genes	
3.1.2 Sub-cellular localization prediction	
3.1.3 Chromosomal Location and Duplication	
3.1.4 Gene and protein structure of <i>AKR</i> genes	
3.1.5 Multiple sequence alignment and phylogenetic analysis	

- 3.1.6 Identification of interacting partners of OsAKR
- 3.1.7 Promoter analysis for identification of CREs and CpG islands
- 3.1.8 Interactome of AKR proteins in rice
- 3.1.9 Temporal and spatial expression profiling of *OsAKR* genes

3.2 Wet-lab Protocols

- 3.2.1 Chemical reagents and kits
- 3.2.2 Preparation of competent cells
- 3.2.3 Transformation of plasmid DNA
- 3.2.4 Plasmid DNA Isolation
- 3.2.5 Extraction of DNA from agarose gel
- 3.2.6 Restriction Digestion and Ligation
- 3.2.7 Agarose gel electrophoresis
- 3.2.8 Polymerase Chain Reaction
- 3.2.9 Plant growth conditions, stress treatments and sample harvest
- 3.2.10 RNA Isolation
- 3.2.11 Quantification of isolation RNA
- 3.2.12 cDNA Synthesis
- 3.2.13 Expression analysis by quantitative real-time PCR (qRT-PCR)
- 3.2.14 Estimation of methylglyoxal (MG)
- 3.2.15 Glyoxalase enzyme activity
- 3.2.16 GC-MS analysis
- 3.2.17 Confocal microscopy
- 3.2.18 TUNEL (Terminal deoxynucleotidyl transferase-mediated dUTP Nick-End Labelling) assay
- 3.2.19 Octet Assay (Bio-Layer Interferometry)
- 3.2.20 Yeast-two hybrid assay
- 3.2.21 Agrobacterium-mediated transformation of rice (PB1 variety)

- 3.2.22 Screening of transgenic plants by Tissue-PCR
- 3.2.23 Isolation of native total plant protein for glyoxalase enzyme activity
- 3.2.24 Primer tables

Chapter 4: RESULTS AND DISCUSSION (PART-I)

41-64

4.1 Background

4.2 Results

4.2.1 Comparative analysis of spatiotemporal variations in MG levels in salt-sensitive, IR64 and salt-tolerant Pokkali

4.2.1.1 Profiling of MG across various developmental stages and in different tissues

4.2.1.2 Expression profiling of Glyoxalase genes

4.2.1.3 Tissue-specific and developmental variations in glyoxalase enzyme activity

4.2.2 Genome-wide analysis of AKRs in rice: Identification, phylogeny, architecture and transcriptional regulation

4.2.2.1 *In silico* analysis reveals 28 putative AKR genes in rice

4.2.2.2 Physiochemical Features of AKR proteins and their localization

4.2.2.3 Chromosomal Localization and Genome Duplication

4.2.2.4 Phylogenetic analysis of *OsAKR* family

4.2.2.5 Gene and Protein Structures

4.2.2.6 Analysis of CREs in the *OsAKR* gene promoters

4.2.2.7 Elucidating AKR interactome in rice

4.2.2.8 Temporal, spatial and stress-based expression profiling of *OsAKR* genes

4.2.2.9 Effect of *OsAKR* genes under different stress conditions

4.2.2.10 Comparative response of AKR genes to stress in IR64 and Pokkali

4.2.3 Screening of drought-tolerant rice varieties using MG as a biomarker

4.2.3.1 Screening of gamma-mutant rice plants

4.2.3.2 Evaluation of the stress-tolerance ability of gamma-mutant rice

4.4 Discussion

4.4.1. Spatio-temporal profiling of MG and *Gly* gene expression

4.4.2. Genome-wide analysis of *AKR* family in rice and expression under different stress conditions

4.4.3. MG as a biomarker for stress

4.4 Conclusions

4.4.1 *Gly* genes play an important role in MG detoxification

4.4.2 *AKR* might be an important component of the MG detoxification mechanism in plants

4.4.3 MG as a biomarker for stress tolerance

Chapter 5: RESULTS AND DISCUSSION (PART-II) 65-70

5.1 Background

5.2 Results

5.3 Discussions

5.4 Conclusions

Chapter 6: RESULTS AND DISCUSSION (PART-III) 71-78

6.1 Background

6.2 Results

6.2.1 SEM analysis indicates disruption of cellular organelles at higher MG concentrations

6.2.2 MG causes cell death in a dose-dependent manner

6.2.3 Mitochondrial potential is impaired at higher MG concentrations

6.3 Discussion

6.3.1 Cellular organelles show degradation at higher MG concentrations

6.3.2 MG causes cell death in a dose-dependent manner

6.3.3 Mitochondrial activity is affected at higher MG concentrations

6.4 Conclusions

Chapter 7: RESULTS AND DISCUSSION (PART-IV)

79-98

7.1 Background

7.2 Results

7.2.1 *Gly* genes are differentially regulated as evident from the presence of CRE's responsive to diverse factors

7.2.2 Promoter analysis of *Gly* genes to predict possible MG-response elements (MGREs)

7.2.3 Identification of genes involved in MG signalling using inhibitors

7.2.4 Possible interactors of GLY proteins were identified using the STRINGS database

7.2.5 Identification & validation of interacting proteins using Yeast-2-Hybrid

7.3 Discussions

7.3.1 *OsGly* genes harbour CREs for stress response

7.3.2 ERF proteins may bind to the predicted MGREs and regulate stress gene regulation

7.3.3 MG signalling in rice plants is mediated via Ca²⁺ and MAPK

7.3.4 *In silico* analysis indicates possible GLY interactors

7.4 Conclusions

Chapter 8: SUMMARY AND CONCLUSIONS

99-100

8.1 Summary

8.2 Conclusion and Future Prospectives

Chapter 9: REFERENCES

101-117

Chapter 10: PUBLICATIONS

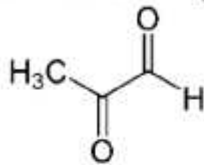
118

Chapter 11: ADDITIONAL FILES

119-130

INTRODUCTION

Methylglyoxal (MG), also known as pyruvaldehyde or 2-oxopropanal, is an organic compound with two carbonyl groups, a ketone and an aldehyde, having the formula $\text{CH}_3\text{C}(\text{O})\text{CHO}$ (Darkin and Dudley, 1913). It was first discovered in 1913 when it was found that it is enzymatically converted to lactic acid. However, its presence in the plant kingdom was reported in the late 20th century.



MG is a common by-product of cellular metabolism. However, its production rate may vary depending on various factors, like physiological conditions (normal or stressed). Major source of its production is the enzymatic conversion (in prokaryotes) (Cooper and Anderson, 1970) as well as non-enzymatic (in eukaryotes) conversion of triose phosphate, dihydroxyacetone phosphate (DHAP) (Philips and Thornalley, 1993). Some amount of MG is also contributed by protein (Degenhardt et al., 1998) and lipid metabolism (Esterbauer et al., 1982). The complete pathway is shown in fig. 1.

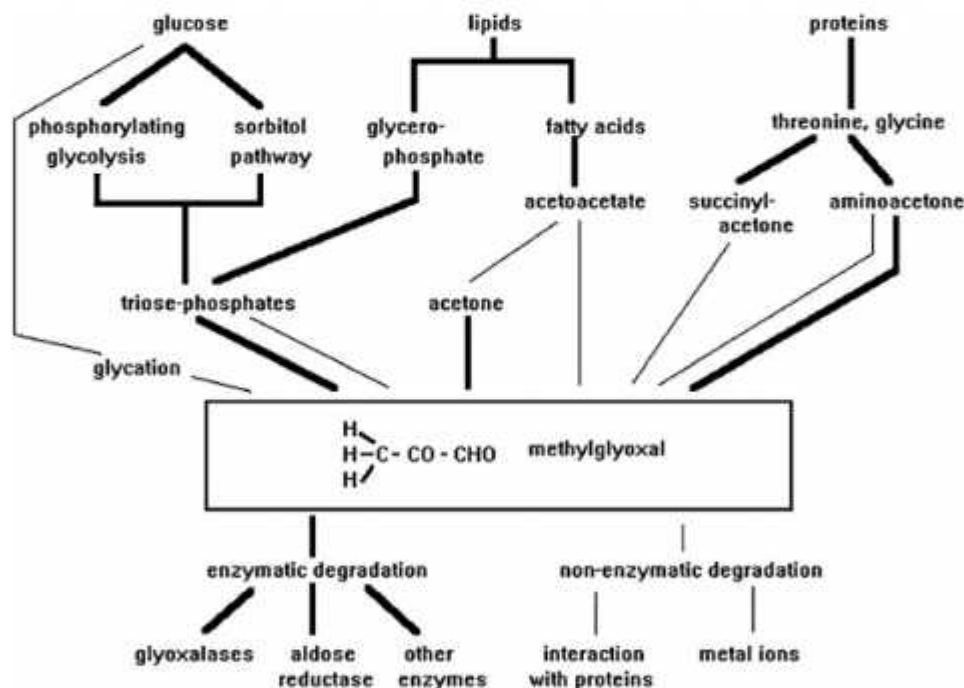


Fig. 1: The pathways of methylglyoxal formation and degradation. Bold lines indicate enzyme-catalyzed pathways (Kalapos, 2008).

MG has been shown to activate the HOG-MAP kinase pathway in yeast cells under osmotic stress (Maeta et al., 2005). It was identified that MG acts as the initiator of signal transduction pathway, and activates Msn2 transcription factor (Maeta et al., 2005). A possible role of MG in signal transduction pathway has been suggested by Kaur et al. (2015), probably by acting as a stress signal molecule (Kaur et al., 2015).

MG is highly reactive in nature due to its two carbonyl groups, keto and aldehyde. It is a potent and highly reactive glycation agent. It induces oxidative stress in cells either directly through increased generation of ROS (reactive oxygen species) or indirectly by modifying amino groups of proteins or DNA molecules forming protein and nucleotide AGEs (Advanced Glycation End Products) (Thornalley et al., 1999). Fig. 2 shows the various ways in which MG is toxic in the biological system.

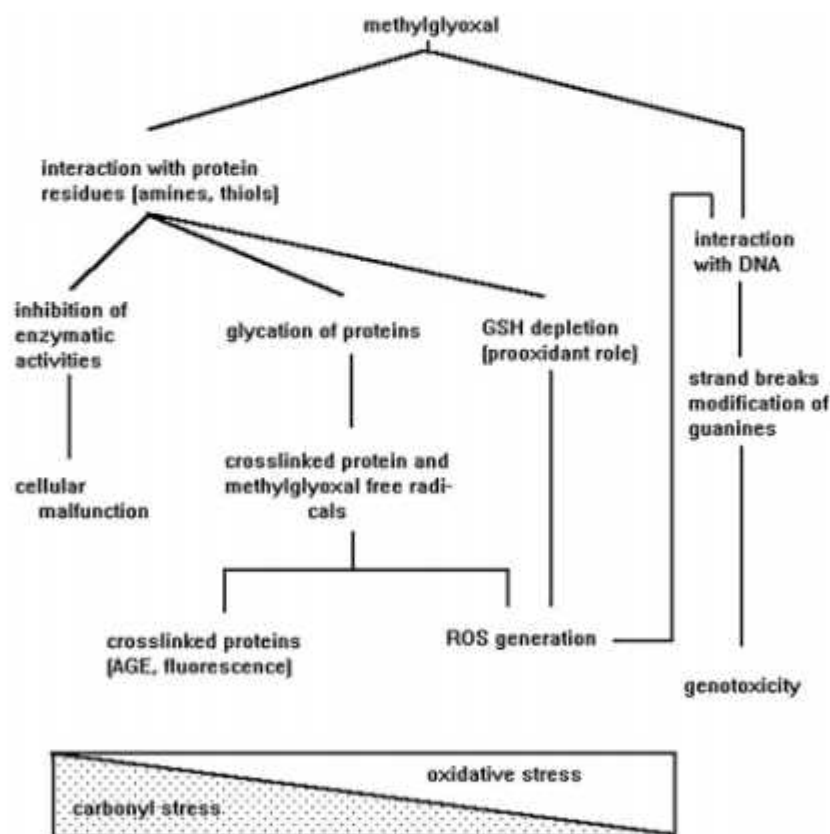


Fig. 2: Mechanism of methylglyoxal-mediated toxicity (Kalapos, 2008).

Several reports have indicated that levels of MG increase under stress conditions, and therefore it is important to understand it at the physiological level to devise strategies to mitigate the adverse effects of MG. All living systems, including both plants and animals, have evolved several detoxification mechanisms to combat the so-called ‘MG stress’ developed under such conditions. The major pathway for detoxification of MG is through a two-step enzyme catalyzed glyoxalase system, comprising GlyI (glyoxalase I, also abbreviated Glo1) and GlyII (glyoxalase II, also abbreviated Glo2) enzymes that uses hemi-thioacetal formed from the spontaneous combination of MG and glutathione (GSH) as a substrate to yield D-lactate, thereby regenerating GSH in the process (Kaur et al., 2014). Other enzymes for MG detoxification include aldo-keto/aldehyde reductases and dehydrogenases.

The glyoxalase pathway in plants operates in both cytosol and mitochondria, with GlyI enzymes being cytosolic since their substrate MG is produced mostly as a by-product of the cytosol-based glycolytic pathway. However, peroxisomal localization of GlyI has also been reported in *Arabidopsis*. On the other hand, GlyII proteins are present in both cytosol and mitochondria. These enzymes are believed to be key players in the plant stress response, and their overexpression confers significant tolerance to multiple stresses such as salinity (Singla-Pareek et al., 2004) and heavy metal stress (Singla-Pareek et al., 2006). In animal systems, MG has been reported to induce apoptosis in sarcoma cells (Ghosh et al., 2011). It kills most of the cell types when accumulated to high levels.

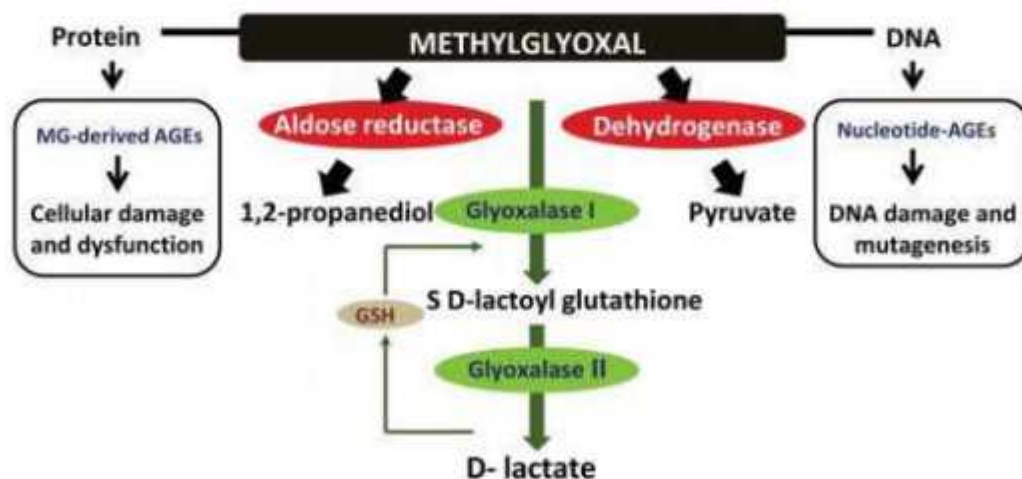


Fig. 3: Methylglyoxal detoxification systems in living organisms (Kaur et al., 2015).

Though MG was discovered almost a century ago, a lot still remains to be discovered regarding its function. It is known that MG affect cell functions and leads to the formation of AGEs, which in turn causes cell death. However, it is still not clear whether the cell death is either due to necrosis or programmed cell death, and what's the exact mechanism of cell death. Reports suggest that MG leads to cell death through ROS-mediated mitochondrial dysfunction and oxidative stress. However, since MG reacts with DNA as well, there is a possibility of cell death due to DNA damage.

MG has been shown to affect cell organelles such as mitochondria. It has been found that MG leads to increased ROS and lactate production, along with decrease in mitochondrial membrane potential (MMP) and intracellular ATP levels (Arriba et al., 2007). Ghosh et al. (2011) have shown MG-induced alteration in the structure of mitochondria and inhibition of mitochondrial complex I (Ghosh et al., 2011). MG has even been reported to induce stress in endoplasmic reticulum, leading to activation of unfolded protein response, which in turn leads to excessive ROS production (Palsy et al., 2014).

A lot of questions still remains unanswered. MG has always been seen as a toxin having a negative role. Is it possible, that MG also has a good side and is beneficial to the plants? My quest during this PhD will be to investigate various aspects of MG, and get a holistic view of effects of MG in the cell, as well as in the whole plant system. For this study, the following objectives are being proposed.

OBJECTIVES

1. Studying variations in MG and glyoxalase levels at different developmental stages in a tissue-specific manner in rice
2. Investigating systemic perturbations in rice caused by exogenous MG treatment
3. Assessing MG-mediated damage on cellular structure and functioning in rice
4. Exploring the role of MG in signaling

WORK PLAN

- 1. MG and glyoxalase profiling in a tissue-specific manner at different developmental stages:** Endogenous MG levels will be measured in rice in different plant tissues at various stages of development through MG derivatization-based assay. Next, transcript levels will be measured to identify specific glyoxalase genes induced at each developmental stage in different tissues through qRT-PCR. Total glyoxalase activity will also be measured, through spectrophotometry-based assays, at each stage in order to investigate the link between MG levels and glyoxalase activity.
- 2. Studying systemic variations in rice due to MG using omics approach:** Changes in metabolite profile of rice will be analyzed at different concentrations of MG. Global changes will be monitored at low MG concentrations (nearing its physiological concentration) in order to investigate changes related to its possible role as a signaling molecule. At higher doses, perturbations related to MG-mediated toxicity will be studied.
- 3. Studying the effect of MG at cellular and molecular level:** Using microscopic techniques, various cell organelles will be observed for any changes in morphology or other characteristics upon exogenous MG treatment. Further, MG-mediated DNA damage will be studied in rice roots using various techniques, such as TUNEL assay.
- 4. Studying the role of MG as a signal molecule:** Upstream sequences of MG-responsive genes, identified through analysis of MG-treated rice transcriptome, will be screened for conserved motifs that can act as possible MG-response elements. Further, the specific signaling pathway through which the MG signal may be transduced will also be investigated. For this, the effect of MG on expression of various signaling pathway genes will be studied using inhibitors of specific signaling pathways.

REVIEW OF LITERATURE

Introduction

Agriculture in the 21st century has become challenging due to many factors, major of which is climate change. Rising global temperatures, coupled with erratic precipitation, has drastically affected the productivity of crop plants. By the end of the 21st century, global temperatures are predicted to increase by 1.5°C, and extreme scenarios may lead to greater than 2°C increase if significant actions are not taken (Intergovernmental Panel on Climate Change, 2013). The situation has been aggravated by increasing population, decreasing agricultural land due to urbanization and scarcity of water. Immediate steps are required to improve the productivity of important crop species such as rice, corn, and wheat, to meet the increasing demands of the world population.

The first green revolution was the combined result of plant breeding techniques and nitrogen fertilizers. However, these techniques have their limitations. Breeding takes a lot of time, whereas fertilizers are costly as well as pose a significant threat to the environment (Kant et al., 2011). Hence, there is a need for biotechnological interventions to fast track the process of making stress-tolerant and high yielding crops. However, to achieve this objective, we must first understand the innate plant mechanisms which have enabled them to survive through the course of evolution. Thereafter, we can tweak the various mechanisms using biotechnological tools to boost crop productivity as well as stress tolerance.

2.1 Stress in plants

Plants are exposed to climatic fluctuations which vary with time and geographical location. Plant growth and productivity are negatively affected if the environmental conditions are beyond the physiologically favourable range. Plant stress can be defined as a state where the plant is growing in non-ideal growth conditions that increase the demands made upon it (Mosa *et al.*, 2017). As a result, there are deficiencies in growth and crop yields, or even permanent damage or death if the stress is beyond the tolerance limit of the plant. Plant stresses are classified into two broad categories, biotic and abiotic.

Biotic stress is caused by living organisms that share their environment with plants, such as pathogens and pests, interacting with them in various ways, including herbivory. Abiotic stress is usually caused by different environmental factors (such as drought, salinity, extremes of temperature and irradiance), pollutants (heavy metals and xenobiotic compounds) and soil

characteristics (extremes of pH) (Rout and Das, 2013). Plants have evolved over the years and have acclimatised to different environmental conditions throughout the world. As a result, different plants have different optimum conditions for growth. Hence, stress is a relative concept. Conditions which may be unfavourable for one may be normal for another.

When faced with a shorter duration of moderate stress, plants usually recover upon restoration of optimum conditions. However, when faced with severe stress, plants suffer from reduced flowering and heavy yield losses, coupled with induction of senescence leading to plant death. Based on tolerance limits, plants are classified either as susceptible or tolerant. Stress-tolerant plants adopt two different strategies for survival under stress conditions, i.e., stress avoidance and stress-acclimation. Under stress-avoidance strategy, plants use novel strategies they have acquired over time to survive under stress conditions. For instance, plants restrict water loss through decreased stomatal conductance and changes in leaf morphology and/or orientation. Further, they also undergo osmotic adjustments or changes in tissue elasticity (Touchette *et al.*, 2009).

The current climatic conditions are such that they favour pathogens which negatively affect plant growth and productivity. The physiological cost of adapting to these changing environmental conditions is huge, and often cuts down the resources available for the production of biomass and seeds, which in turn affects the yield. Further, the combinatorial effect of various stress factors often leads to a trade-off, causing an appropriate response against one stress while enhancing susceptibility to the other (Zhang *et al.*, 2020). The local crop varieties are faced with new climate-driven migration of pests and pathogens. As a consequence, there is a clash between attempts to boost yields and yield-threatening environmental challenges. Hence, to sustain and improve crop yields, it is imperative to decipher the stress response mechanism of plants. Most of the stress tolerance mechanisms have been studied in model crops and there is no clarity regarding their operation in other crop plants. Hence, there is a need for translation of research from model plants to crops, or from one crop to another.

2.2 Methylglyoxal (MG) in plants

Several reactive metabolites are formed in the cellular environment as a result of various metabolic processes. Some are extremely toxic at higher than threshold levels. Stress is known to increase the levels of these reactive metabolites which imposes toxicity in the cellular milieu. Based on the main element in the metabolite, these reactive metabolites are known by different

names, such as reactive oxygen species (ROS), reactive carbonyl species (RCS), reactive nitrogen species (RNS) and reactive sulphur species (RSS). Among RCS, more than twenty have been identified till date in living organisms (Kosmachevskaya *et al.*, 2015) including methylglyoxal (MG), glyoxal (GO), 3- deoxyglucosone (3-DG), phenylglyoxal (PG) and hydroxyl-pyruvaldehyde, which are the major ones (Racker, 1951).

MG has two carbonyl groups, i.e., a ketone and an aldehyde (P J Thornalley, 1996). Also known as pyruvaldehyde or 2-oxopropanal, it was first discovered in 1913, but was reported in plants in the late 20th century. It is a common by-product of cellular metabolism and is ubiquitous throughout the tree of life, from the primitive archaea to the complex plants and animals. It is constantly present under all physiological conditions, normal as well as stressed. Under normal conditions, cells have a basal level of MG, whereas, under stressed conditions, the level of MG is highly elevated. MG production occurs mainly through the glycolytic pathway where enzymatic (only prokaryotes), as well as non-enzymatic reactions, lead to the conversion of triose phosphates to MG (Cooper and Anderson, 1970; Philips and Thornalley, 1993). Protein and lipid metabolism also contribute to MG production to some extent (Degenhardt *et al.*, 1998; Esterbauer *et al.*, 1982).

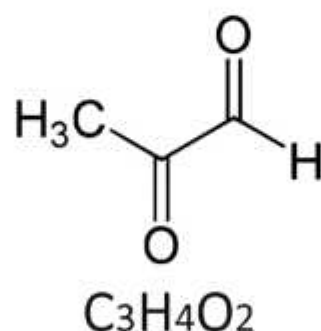


Figure 2.1: Chemical structure of methylglyoxal.

2.3 Methylglyoxal formation

2.3.1 MG production from carbohydrates

The glycolytic pathway is a major source of MG production from carbohydrates, both enzymatically as well non-enzymatically. The non-enzymatic breakdown of triose phosphates of the glycolytic pathway, such as dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP), leads to the formation of MG. Triosephosphate isomerase (TPI; EC 5.3.1.1) which catalyses the interconversion of DHAP and GAP, is known to play an important

role in MG formation. At physiological pH, these triose phosphates have a high tendency to form enediolate phosphate by the loss of α -carbonyl protons (Richard, 1984). This is followed by the non-enzymatic β -elimination of the phosphate group from enediolate phosphate to form MG. Methylglyoxal synthase, present in prokaryotes also catalyses the formation of MG, from DHAP by eliminating an inorganic phosphate in the glycolytic pathway. This glycolytic bypass accounts for 40% of the total glycolytic flux in *Desulfovibrio gigas* (Fareleira *et al.*, 1997). After initial isolation from *E. coli* (Hopper and Cooper, 1972), MG synthase was later isolated from other organisms as well, such as *Saccharomyces cerevisiae*, *Pseudomonas saccharophila*, goat liver and *Clostridium acetobutylicum* (Cooper, 1974; Ray and Ray, 1981; Murata, Fukuda, Watanabe, *et al.*, 1985; Huang *et al.*, 1999). However, MG synthase is yet to be reported from plants. Metabolites, such as 3-phosphoglycerate, phosphoenolpyruvate, inorganic phosphate and pyrophosphate, are known to negatively regulate MG synthase.

2.3.2 MG production from proteins

MG is also formed from protein during its metabolism. Glycine and threonine catabolism yield amino-acetone (Kalapos, 1999) which is then converted by the enzyme semicarbazide-sensitive amine oxidase (SSAO; EC 1.4.3.6) to MG (Lyles and Chalmers, 1992). Further investigations revealed monoamine oxidase (MAO) as the enzyme responsible for MG formation from amino acetone (Urata and Granick, 1963). Studies in *S. cerevisiae* and *Staphylococcus aureus* have confirmed the presence of the MAO enzyme (Murata *et al.*, 1986). MG accumulation was also observed in *S. cerevisiae* when fed with L-threonine as a nitrogen source (Murata *et al.*, 1986). However, this enzyme hasn't been reported in plants yet.

2.3.3 MG production from lipids

Under pathological conditions, ketosis and diabetic ketoacidosis also lead to the generation of MG. In diabetic patients, MG is mainly formed from ketone bodies (Nemet *et al.*, 2006). MG is also formed via lipo-peroxidation (Esterbauer *et al.*, 1982). In the gluconeogenic pathway, the conversion of acetone to MG is catalysed by cytochrome P450 using NADPH (Casazza *et al.*, 1984; Koop and Casazza, 1985). The cytochrome P450 has been identified as both acetone and acetol monooxygenases in liver microsomes (Koop and Casazza, 1985). Acetone is converted by acetone mono-oxygenase to acetol which is subsequently converted to MG by acetol mono-oxygenase (Casazza *et al.*, 1984). MG is also formed through the Maillard reaction where reducing sugars react with amines (Thornalley *et al.*, 1999).

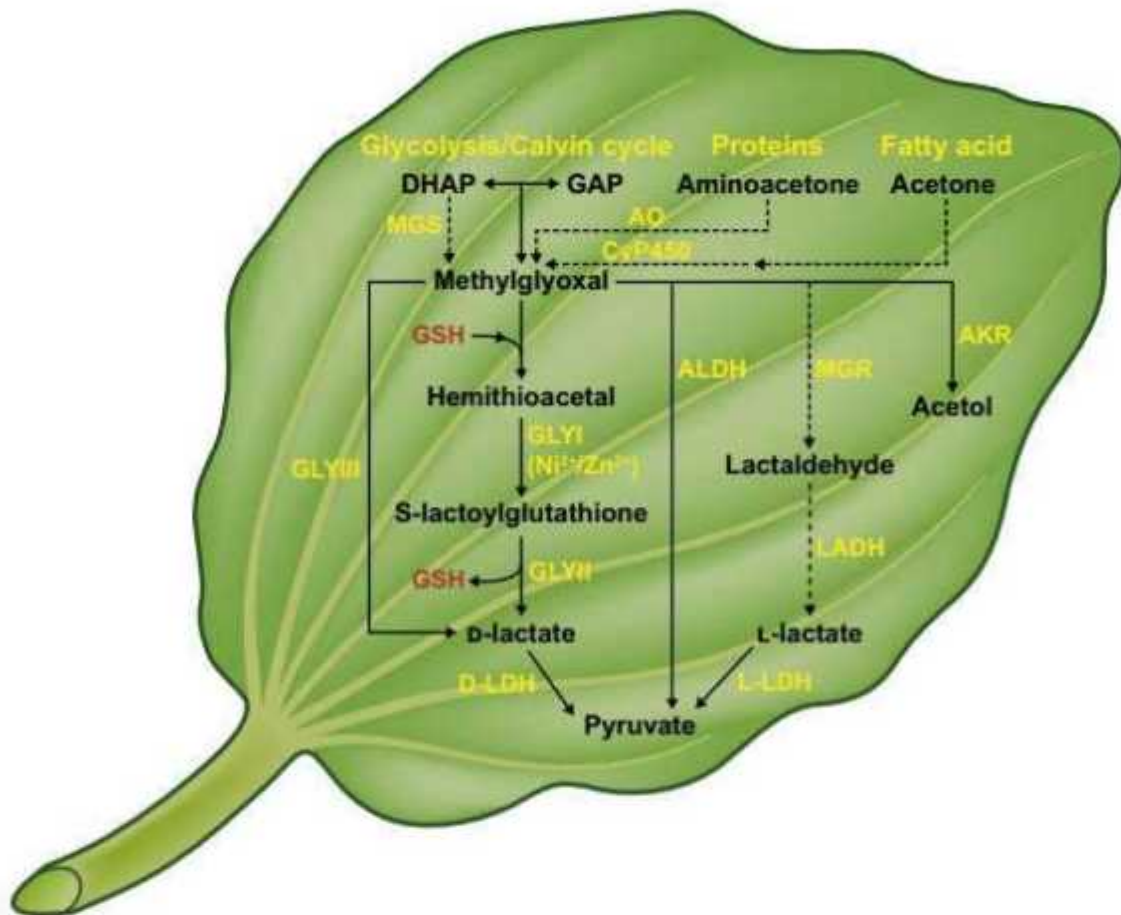


Figure 2.2: Various pathways for MG metabolism in living systems. The major route of MG breakdown in plants is the glyoxalase pathway comprising the enzymes glyoxalase I (GLYI; lactoylglutathione lyase, EC 4.4.1.5) and glyoxalase II (GLYII; hydroxyacylglutathione hydrolase, EC 3.1.2.6). MG spontaneously reacts with glutathione (GSH) to form a hemithioacetal adduct which is metabolized by GLYI to form S-lactoylglutathione (SLG). GLYI is a metalloenzyme requiring divalent metal ions, either Ni^{2+} or Zn^{2+} , for activation. Subsequently, GLYII converts SLG to D-lactate, recycling GSH in the process. The D-lactate so formed is finally converted to pyruvate through D-lactate dehydrogenase (D-LDH), which then enters the Krebs cycle. GLYIII enzymes (EC 4.2.1.130) can directly convert MG to D-lactate in one step without requiring GSH and divalent metal ions and no SLG is formed in the process. Besides GLYIII activity, these enzymes also possess other catalytic functions such as chaperone and protease activities. Aldoketo reductases (AKRs), aldehyde dehydrogenases (ALDHs) and methylglyoxal reductases (MGRs) can also metabolize MG. While AKRs form acetol from MG, ALDHs form pyruvate and MGRs convert MG to lactaldehyde which is subsequently converted to L-lactate via lactaldehyde dehydrogenases (LADH). L-Lactate is finally converted to pyruvate through the enzyme L-lactate dehydrogenase (L-LDH). The pyruvate thus formed can enter the Krebs cycle. Unlike the glyoxalase pathway, these enzymes have broad substrate specificity and can catalyse reduction of various ketones and aldehydes. The pathways shown by dotted lines have not yet been established in plant systems.

2.4 MG toxicity

Glycation is the non-enzymatic post-translational modification of proteins. Dicarboxyls, such as MG, with a free keto and aldehyde group, are potent glycation agents. It is a strong electrophile and tends to form covalent adducts with proteins and nucleic acids (Speer *et al.*, 2003; Li *et al.*, 2008). It rapidly reacts with proteins and sugars to form unstable Schiff base compounds. These convert to more stable Amadori products, which further undergo a series of complex and irreversible reactions to form advanced glycation end products (AGEs) (Thornalley, 1996). MG is responsible for the formation of the major quantitative AGE, hydroimidazolone MG-H1.

2.5 Glyoxalase pathway in plants

The cytotoxic nature of MG necessitates its detoxification to prevent cellular damage. There are many detoxification pathways active within the cell for the efficient removal of MG (Figure 2.2). However, the glyoxalase pathway, comprising of glyoxalase I (GLYI) and glyoxalase II (GLYII) enzymes, is considered to be the major route of MG detoxification in plants. Additionally, glyoxalase III (GLYIII) enzymes also exist in living systems which catalyses the one-step conversion of MG. MG formed enzymatically as well non-enzymatically during the glycolytic pathway is converted by GLYI-II or GLYIII to D-lactate. This D-lactate is further converted to pyruvate by D-lactate dehydrogenase, and thus, acts as a shortcut for pyruvate formation from DHAP, especially during phosphate starvation conditions (Cooper and Anderson, 1970).

2.5.1 Glyoxalase I

The GLYI (Lactoylglutathione lyase, EC 4.4.1.5) enzyme has been extensively characterized in different organisms (Kaur *et al.*, 2017). Hemithioacetal formed from the spontaneous or enzymatic reaction between GSH and MG, is converted to S-lactoylglutathione (SLG) by GLYI (Thornalley, 1990; Kammerscheit *et al.*, 2020). It is the rate-limiting enzyme in the glyoxalase pathway (Ferguson *et al.*, 1998).

GLYI enzymes are metallozyme, requiring divalent ions for optimum activity (Kaur *et al.*, 2017). Generally, GLYI enzymes are homodimers containing a metal-binding centre (Sousa Silva *et al.*, 2013; Cameron *et al.*, 1997) but lately monomeric forms have also been reported, mainly in plants (Deponte *et al.*, 2007; Frickel *et al.*, 2001; Kaur *et al.*, 2017). The metal ion selectivity of GLYI depends on the active site geometry and is usually of two types depending

on the GLYI enzyme size (Himo and Siegbahn, 2001). The shorter GLYI proteins (~150 aa) require $\text{Ni}^{2+}/\text{Co}^{2+}$, whereas the longer GLYI proteins (~180 aa) require Zn^{2+} ions (Sukdeo and Honek, 2007). Despite the difference in size, all GLYI proteins share the conserved $\beta\alpha\beta\beta\beta$ structural motif (Kaur *et al.*, 2017). The metal protein complex forms an octahedral geometry with the help of four amino acids and two water molecules, which is necessary for GLYI enzyme activity (He *et al.*, 2000).

2.5.2 Glyoxalase II

The GLYII (Hydroxyacylglutathione hydrolase, EC 3.1.2.6) enzyme catalyses the conversion of SLG to D-lactate while regenerating GSH. This enzyme of the metallo- β -lactamase superfamily has been well characterized in many organisms, including prokaryotes as well as eukaryotes (Sousa Silva *et al.*, 2013; Kaur *et al.*, 2017). It has a highly conserved THxHxDH motif which forms the binuclear metal centre (Cameron *et al.*, 1997; Silva *et al.*, 2008). Besides being highly specific for the glutathione group of the substrate, GLYII also metabolizes other thioesters (Vander Jagt, 1993). Like GLYI, metal ions are essential for the enzyme activity of GLYII and facilitate substrate binding (Zang *et al.*, 2001; Campos-Bermudez *et al.*, 2010).

2.5.3 Glyoxalase III

The GLYIII (EC 4.2.1.130) enzyme is a novel glyoxalase enzyme that catalyzes the detoxification of MG directly to D-lactate, without requiring GSH (Misra *et al.*, 1995; Lee *et al.*, 2012). It is ubiquitous across all organisms and belongs to the DJ-1/Pfpl/ThiJ superfamily (Ghosh *et al.*, 2016). These enzymes are catalytically less efficient compared to the conventional GLYI & II enzymes (Silva, *et al.*, 2013; Ghosh *et al.*, 2016). The enzymatic reaction is irreversible and the process has been described based on the structure of AtDJ-1D (Kwon *et al.*, 2013). Briefly, the thiol group of a cysteine residue, similar to glutathione in the GLYI/GLYII system, reacts with the aldehyde group of MG to form hemithioacetal intermediate (Enzyme-Cys-MG). Subsequently, glutamic acid near the cysteine residue abstracts a proton from hemithioacetal to form an enediol intermediate, followed by the formation of S-lactoylcysteine due to proton transfer by the glutamic acid residue. This intermediate is hydrolysed to finally release D-lactate. The whole process is a single-step irreversible reaction. The importance of the sulfhydryl group for GLYIII enzymatic activity has been well established. Oxidizing agents, such as superoxide and hydrogen peroxide, can decrease the GLYIII activity (Okado-Matsumoto and Fridovich, 2000). Similarly, mutation

studies in different organisms have also established the importance of the sulfhydryl group (Subedi *et al.*, 2011; Lee *et al.*, 2012; Hasim *et al.*, 2014).

2.6 Other detoxification enzymes

The enzyme MG reductase (MGR) catalyses the conversion of MG to L-lactaldehyde in the presence of NADPH (Murata, Fukuda, SIMOSAKA, *et al.*, 1985; Aguilera and Prieto, 2001). It has been isolated from various organisms, such as goat liver (Ray and Ray, 1984) and *Saccharomyces cerevisiae* (Murata, Fukuda, Watanabe, *et al.*, 1985).

Aldehyde dehydrogenases (ALDHs) are a big superfamily of different types of enzymes that are involved in the detoxification of endogenous as well as exogenous aldehydes. They catalyse the oxidation of aldehydes to carboxylic acids using either NAD⁺ or NADP⁺ as a cofactor. The NADH or NADPH is regenerated during this process and thus act as a major source of reducing equivalents necessary for maintaining cellular homeostasis (Brocker *et al.*, 2012). Under stress conditions, ALDHs act as “aldehyde scavengers” and protect the cell from damage (Bartels and Sunkar, 2005).

Aldo-keto reductases (AKRs) are a family of NADPH-dependent monomeric enzymes, which are oxidoreductases having similar physico-chemical properties. They catalyze the reduction of aldehydes and ketones to their corresponding alcohols. These enzymes have a broad range of substrate specificity, but have preferential substrate specificity. There are three distinct types of enzymes, which fall in the family of AKRs: (a). aldehyde reductases (EC 1.1.1.2), (b). aldose reductases (1.1.1.21), and (c). carbonyl reductase (EC 1.1.1.184). (Wermuth, 1985). In plants, they are involved in diverse metabolic reactions, like detoxification of reactive aldehydes, osmolyte biosynthesis, membrane transport and secondary metabolism (Sengupta *et al.*, 2015).

2.7 Physiological significance of methylglyoxal

The physiological effect of MG is mostly studied at higher levels when it becomes cytotoxic and starts to affect the growth and development of the organism. However, it is now known that MG is maintained at a threshold level even under normal conditions. MG synthase enzyme is known to catalyse MG formation from DHAP in prokaryotes (Ferguson *et al.*, 1998). This indicates that MG may have beneficial roles as well. Hence, the physiological role of MG can be broadly classified under two categories: positive (Figure 2.3) and negative role (Figure 2.4).

2.7.1 Positive role: Role in stress signalling

Plants are sessile in nature and hence, have no option to escape from environmental fluctuations. Hence, plants have evolved various mechanisms to overcome these unfavourable situations. The foremost requirement to mount an appropriate response is the perception of stress. Plants have evolved different sensory mechanisms to perceive stress stimuli and mount an appropriate cellular response. Various sensors located on the cell surface or cytoplasm perceives the stress stimuli and passes them to the nucleus through various signal transduction pathways. Hence, the signalling pathways act as a connecting link between stress perception and downstream cellular response.

The generation of MG via the ubiquitous energy-generating pathway raises the question regarding its role in the cellular system. MG levels in different plants vary from 30 to 75 μM and are known to increase up to six times under different stress conditions (Yadav *et al.*, 2005). However, these are overestimated most probably due to triosephosphate degradation to MG during sample processing (McLellan *et al.*, 1992) or interference of peroxidases (Thornalley and Rabbani, 2014). The actual level of MG in plants has been estimated to be around 1-4 μM (Rabbani and Thornalley, 2014). Importantly, MG levels never fall below a threshold level

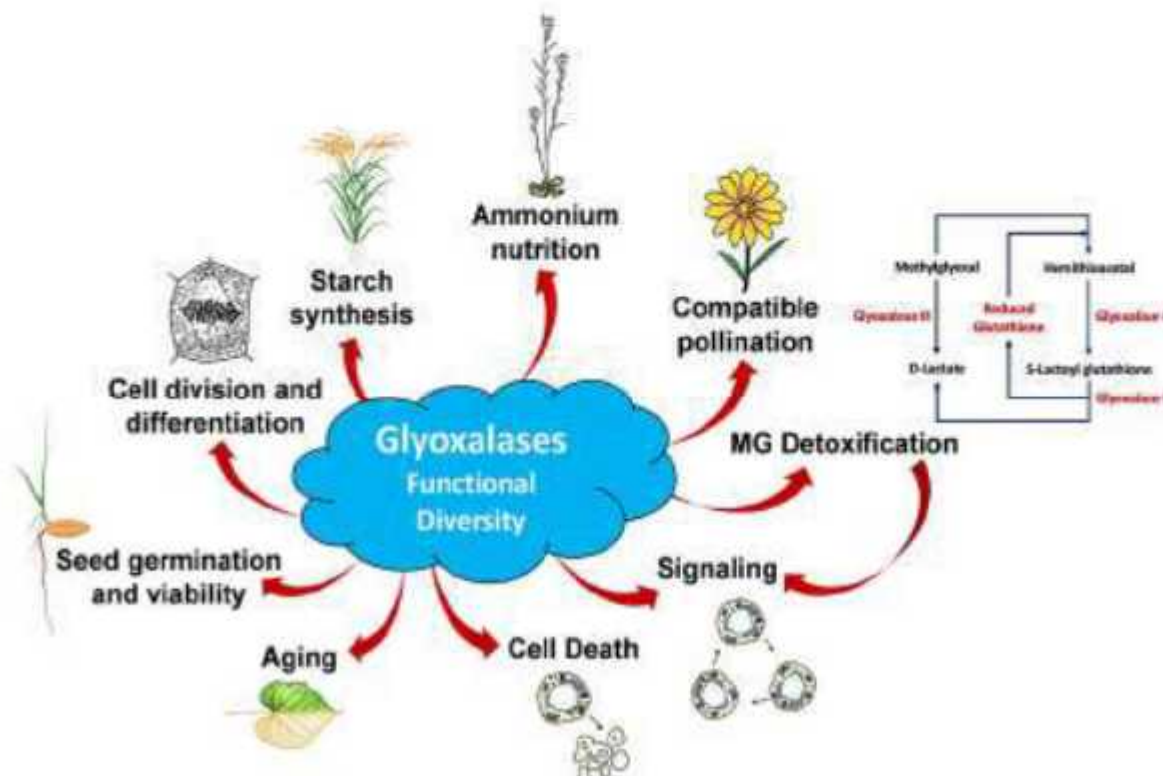


Figure 2.3: Physiological roles of methylglyoxal.

despite the presence of several detoxification mechanisms. These basal MG levels are maintained even in transgenic plants overexpressing glyoxalase genes. This indicates some important role of MG in the physiology of plants. It's quite possible that MG acts as a stress signal and regulates the expression of stress-responsive genes.

The presence of MG synthase enzyme indicates that MG is required in the system and hence, must have a beneficial role (Ferguson *et al.*, 1998). In bacteria, MG production has been proposed as a mechanism to regulate metabolic imbalance by controlling the influx of carbon (Booth *et al.*, 2003). In *E. coli*, MG controls signalling through Ca^{2+} ions. It induces Ca^{2+} transients by opening up Ca^{2+} channels in a dose-dependent manner (Campbell *et al.*, 2007). Similarly, in yeast, MG activates Ca^{2+} uptake and stimulates the calcineurin/Crz1-mediated Ca^{2+} -signalling pathway (Maeta *et al.*, 2005). A possible signalling pathway for the induction of the *GlyI* gene by MG under high osmotic stress has been suggested in *Saccharomyces cerevisiae* (Maeta *et al.*, 2005; Inoue *et al.*, 1998). Osmotic stress induces an increase in MG levels which is scavenged by *GlyI* gene expression via the HOG-MAPK pathway (Inoue *et al.*, 1998). Further, MG acts as a signal initiator and initiates signal transduction during oxidative stress via activation of the high-osmolarity glycerol (HOG)-mitogen-activated protein kinase (MAPK) cascade (Maeta *et al.*, 2005). Among the two osmosensors, Sln1 and Sho1, which are known to function upstream of the HOG-MAPK cascade, the Sln1 branch is responsible for the activation of the MG-induced HOG-MAPK pathway. Further, MG also activates the Msn2 transcription factor. Similarly, in *Saccharomyces pombe*, MG initiates the activation of stress-induced protein kinase. However, the MG signal is directed to the SAPK pathway via Mcs4 by an alternate module (Klahre *et al.*, 1998).

In plants, the exogenous application of MG regulates the expression of many stress-induced genes, including *GlyI* and *GlyII* genes (Majláth *et al.*, 2020). However, the mechanism behind MG-induced gene expression is yet to be understood. Nevertheless, NAC-overexpression has been shown to induce the expression of many genes, including the *GlyI* gene (Fujita *et al.*, 2004). Similarly, exogenous application of MG was also found to increase the expression of NAC transcription factors in rice (Kaur *et al.*, 2015). Hence, the relationship between NAC transcription factors and *Gly* genes needs to be explored further. Similarly, overexpression of many stress-responsive genes also resulted in the expression of the *Gly* gene (Yadav *et al.*, 2008).

In plants, MG levels are regulated by various detoxification mechanisms, and this helps to regulate various physiological processes. For instance, in *Brassica napus*, the energy requirement of the growing pollen tube is met through increased glycolysis which leads to a surge in MG levels in stigma (Sankaranarayanan *et al.*, 2015). This increased MG is regulated via enhanced GLYI activity, which makes it necessary for pollen germination on stigma. Hence, MG indirectly controls pollen germination. Similarly, in rice, GLYI is involved in compound starch granule formation and starch synthesis in the endosperm (You *et al.*, 2019).

2.7.2 Negative role: Formation of AGEs, inhibition of cellular proliferation and cell death

The high reactivity of MG allows modification of macromolecules such as DNA, RNA and proteins to form AGEs (Thornalley, 2008). In animal systems, many diseases, such as diabetes and cancer, has been attributed to post-translational protein modifications caused by high levels of MG (Ramasamy *et al.*, 2006). During protein glycation, usually during “early glycation”, the reversible interaction of reducing sugars (aldoses and ketoses) with amino groups, lead to the formation of aldoamines and ketoamines, respectively. These further form aldimines and ketoimines (Schiff bases), which undergo Amadori and Hens rearrangements, to yield the first relatively stable intermediate products of glycation, termed as “early glycation products” (Bilova *et al.*, 2016). These products, similar to free sugars, undergo autoxidation readily to form highly reactive α -dicarbonyl compounds (GO, MG, etc.). Two major advanced glycation pathways, depending on the structure of the carbohydrate moiety, i.e., free sugar or protein-bound early glycation products, have been classified, namely “oxidative glycosylation” and “glycoxidation”, respectively (Wolff and Dean, 1987; Wells-Knecht *et al.*, 1995).

AGEs (Advanced glycation end-products) are a group of highly heterogenic compounds which vary greatly in their stability (Bilova *et al.*, 2016). Lysine-derived AGEs include N ϵ -(carboxymethyl)lysine (CML) (Ahmed *et al.*, 1986), N ϵ -(carboxyethyl)lysine (CEL) (Ahmed *et al.*, 1997), ϵ -(2-formyl-5-hydroxymethyl-pyrrolyl)-L-norleucine (pyrraline) (Foerster and Henle, 2003) and glyceraldehyde-derived pyridinium compound (GLAP) (Usui *et al.*, 2007). The reaction of arginine with glyoxal formed 1-(4-amino-4-carboxybutyl)2-imino-5-oxoimidazolidine (Glarg) (Schwarzenbolz *et al.*, 1997), which upon hydrolysis at 37°C formed N δ -carboxyethylarginine (CMA) (Glomb and Lang 2001). Arginine reacts with MG to form N δ -(5-hydroxy-4,6-dimethylpyrimidine-2-yl)-L-ornithine (argpyrimidine, Argpyr) (Shipanova *et al.*, 1997) and N δ -(4-carboxy-4,6-dimethyl-5,6-dihydroxy-1,4,5,6-tetrahydropyrimidine-2-yl)-L-ornithine (tetra-hydroargpyrimidine, TH-Argpyr) (Oya *et al.*, 1999). MG was shown to

form MG-derived hydroimidazolones (NG-Hs), with N δ -(5-methyl-4-oxo-5-hydroimidazolinone-2-yl)-L-ornithine (MG-H1) being the major derivative (Henle *et al.*, 1994). Another MG-Hs, i.e., 2-amino-5-(2-amino-4-hydro-4-methyl-5-imidazol-1-yl)pentanoic acid (MG-H3) forms carboxyethyl-L-arginine (CEA) upon hydrolysis (Gruber and Hofmann, 2005).

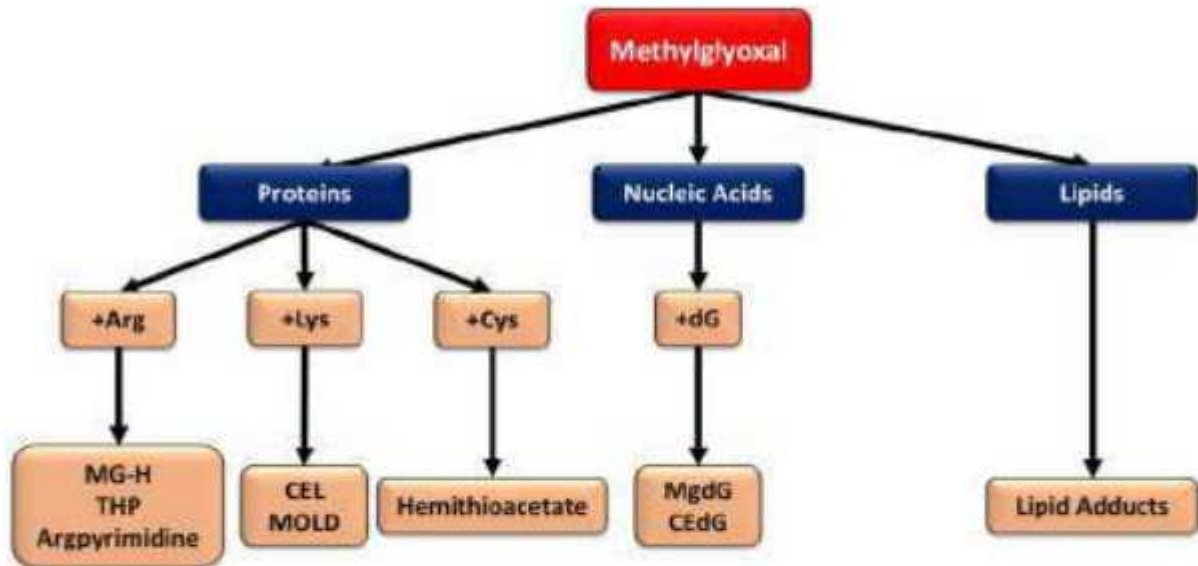


Figure 2.4: Various glycation products formed by methylglyoxal. MG reacts with proteins, nucleic acids and lipids to form toxic adducts.

In humans, AGEs of different types, mostly CML, hydroimidazolones and pentosidine, act as a ligand for pattern recognition receptors known as the receptor for AGEs (RAGEs) and trigger inflammatory diseases, such as atherosclerosis and type 2 diabetes mellitus (Uribarri *et al.*, 2005; Nettleton *et al.*, 2007). The serum and urinary concentration of AGEs were found to be independent of consumption of thermally processed food and vegetarians, especially long-term vegetarians, were found to have higher plasma levels of CML and fluorescent AGEs, despite lower levels of lysine- and arginine- containing proteins in vegetarian diets (Šebeková *et al.*, 2001; Krajčovičová-Kudláčková *et al.*, 2002; De la Maza *et al.*, 2007). This indicates higher glycation of proteins (early and advanced) in plants. Hence, it's obvious for plants to have efficient anti-glycation mechanisms, such as GLYI & II (Singla-Pareek *et al.*, 2008), ribulosamine/erythrosamine-3-kinase (Fortpied *et al.*, 2005), and acylamino acid-releasing enzyme (Yamauchi *et al.*, 2003). Environmental stresses cause an increase in the generation of AGEs in plants, leading to their accumulation in different plant tissues, which changes the plant

proteome and hence, the physiological state of the plant (Bechtold *et al.*, 2009; Bilova *et al.*, 2016). In *Arabidopsis*, MG levels are enhanced during ammonium toxicity due to the formation of MG-derived AGEs (MAGEs) (Borysiuk *et al.*, 2018).

3. MATERIALS AND METHODS

3.1 Bioinformatics Protocols

3.1.1 Identification and nomenclature of genes

For the identification of rice *AKR* genes, a well-characterized *AKR* gene from rice, *OsAKR1* (Os01g0847600) (Turóczy *et al.*, 2011) was used to search for homologous sequences using blastn and blastp in the RGAP database (Kawahara *et al.*, 2013) using default settings. The identified proteins were confirmed for the presence of AKR domain (PF03514.14) using Pfam (El-Gebali *et al.*, 2019) and SMART (Letunic and Bork, 2018; Letunic *et al.*, 2015; Schultz *et al.*, 2000).

3.1.2 Sub-cellular localization prediction

CELLO online server (Yu *et al.*, 2004; Yu *et al.*, 2006) was used to predict the localization of OsAKR proteins in the cell.

3.1.3 Chromosomal Location and Duplication

The location of the various *AKR* genes on the various rice chromosomes was predicted using the online MapGene2Chromosome Tool (Jiangtao *et al.*, 2015).

3.1.4 Gene and protein structure of *AKR* genes

The online server Gene Structure Display Server (GSDS 2.0) (Hu *et al.*, 2015) was used to identify the gene structure (introns and exons) of the *OsAKR* genes.

Domain architecture of the OsAKR proteins was inferred from the Pfam (El-Gebali *et al.*, 2019) and SMART (Letunic and Bork, 2018; Schultz *et al.*, 2000; Letunic *et al.*, 2015) results. Domain positions were recorded from Pfam results. Also, the proteins were analyzed for the presence of conserved motifs using the MEME software (Bailey *et al.*, 2009).

3.1.5 Multiple sequence alignment and phylogenetic analysis

Protein sequences of identified *OsAKR* genes were aligned using ClustalW in MEGA 7.0 (Kumar *et al.*, 2016) and then used for phylogenetic analysis using the neighbour-joining method in MEGA 7.0 (Kumar *et al.*, 2016) with 1000 bootstrap replicates. The tree was further processed in iTol (Letunic and Bork, 2016).

3.1.6 Identification of interacting partners of OsAKR

STRING database (Szklarczyk *et al.*, 2019) was used to identify the protein-protein interaction of OsAKRs with a confidence score of 0.150. Gene co-expression and co-occurrence analysis were also derived from the STRING database.

3.1.7 Promoter analysis for identification of CREs and CpG islands

1kb sequence upstream of the transcription start site was extracted from the RGAP database (Kawahara *et al.*, 2013) and used as a promoter sequence. This was further used in the PlantCARE database (Magali *et al.*, 2002) to identify the cis-regulatory elements (CREs) present. The same promoters were also analysed for the presence of CpG islands using the CpG island search database (<http://dbcat.cgm.ntu.edu.tw/>). Minimum 200 bp length was used to search for CpG islands with GC content set at 50 %, and observed/expected CpG ratio as 0.60.

3.1.8 Interactome of AKR proteins in rice

The interactome of the AKR proteins was investigated using the STRINGS database (STRING: functional protein association networks (string-db.org)). The AKR proteins were fed into the database and *Oryza sativa* was selected as the species. All interaction sources, i.e., text mining, experiments, databases, co-expression, neighbourhood, gene fusion and co-occurrence, were selected with medium confidence (0.4).

3.1.9 Temporal and spatial expression profiling of OsAKR genes

For microarray-based expression profiling, publicly available expression data for the Rice (51 K Affymetrix gene chips, respectively) were used. For studying expression patterns at various developmental stages and in different plant tissues, the normalized and curated log₂-transformed signal intensity values on the respective arrays were retrieved using the Genevestigator database tool (Grennan, 2006; Hruz *et al.*, 2008) (<https://www.genevestigator.com/gv/plant.jsp>) using default parameters as done previously. We could find specific microarray probe sets for *OsAKR14*. For analysing expression levels under different biotic and abiotic stresses, the relative signal ratio values for each of the genes were retrieved using Genevestigator and the log₂ transformed fold change values were calculated. Heat maps were generated with Multi Experiment Viewer software (Saeed *et al.*, 2006) (<http://www.tm4.org/mev.html>) with average linkage hierarchical clustering either using Euclidean or Pearson correlation as the distance metric, as the case may be.

3.2 Wet-lab Protocols

3.2.1 Chemical reagents and kits

Table 3.1. Different chemicals, kits and reagents used in the study.

S. No.	Materials	Source
Plant material		
1	<i>Oryza sativa</i> L. sp <i>indica</i> cv PB1	JNU, New Delhi
2	Pokkali rice	JNU, New Delhi
3	IR64 Gamma-mutant rice lines	JNU, New Delhi
Microbial materials		
4	<i>Escherichia coli</i> (Top10)	Invitrogen, USA
5	<i>Agrobacterium tumefaciens</i> (LBA4404)	ICGEB, New Delhi
Vectors		
6	pET28a	Addgene, USA
7	pCAMBIA1391Z	CAMBIA, Australia
8	Y-2-H Cloning vectors	ICGEB, New Delhi
Reagents and Kits		
10	<i>Taq</i> DNA Polymerase	HIMEDIA, India
11	Phusion High-Fidelity DNA Polymerase	Thermo Fisher Scientific, USA
12	GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific, USA
13	QIAquick Gel Extraction Kit	QIAGEN, Germany
14	Phusion Site-Directed Mutagenesis Kit	Thermo Fisher Scientific, USA
15	1kb DNA ladder	Thermo Fisher Scientific, USA
16	Prestained and Unstained Protein Marker	Genetix, India
18	TRIzol	Thermo Fisher Scientific, USA
19	Bradford Reagents	Amresco, USA
20	RevertAid H Minus First Strand cDNA Synthesis Kit	Thermo Fisher Scientific, USA

21	SYBR® Green Real-Time PCR Master Mixes	Applied Biosystems, USA
22	REExtract-N-Amp™ Tissue PCR Kit	Sigma-Aldrich, USA
23	Restriction enzymes	Thermo Fisher Scientific, USA; NEB, UK
24	T4 DNA ligase	NEB, UK
25	Molecular Biology Chemicals	Sigma-Aldrich, USA; AMRESCO, USA; Amersham, UK; Merck, Germany; HIMEDIA, India

3.2.2 Preparation of competent cells

The chemically competent *E. coli* (Top10/BL21) cells were prepared by chilled CaCl₂ treatment method with some modifications as previously described by Sambrook and Russell (2001). In brief, the bacterial cells were streaked from the glycerol stock on the LB agar plate without any selection. The plate was incubated at 37°C for 12 hrs and a single colony was inoculated in 5 mL LB broth (no selection) as a primary culture, and allowed to grow at 37°C for overnight with shaking at 180 rpm. The next day, secondary culture (LB broth, 50 mL) was inoculated with 1% of primary culture without any selection. It was allowed to grow at 37°C and 180 rpm till OD_{600 nm} reached 0.4-0.6. The culture was chilled on ice for 30 min and transferred to an Oakridge tube. The culture was then centrifuged at 3000 rpm for 5 min at 4°C. The supernatant was discarded and the pellet was gently dissolved in ice-cold sterile 20 mL of 100 mM CaCl₂ and centrifuged at 3000 rpm for 5 min at 4°C. The pellet was again gently dissolved in ice-cold sterile 20 mL of 100 mM CaCl₂ and the suspension was incubated on ice for 2 hrs. After 2 hrs the suspension was centrifuged at 3000 rpm for 5 min at 4°C and the supernatant was decanted. The loose pellet was re-suspended in 1.5 mL of ice-cold 100 mM CaCl₂ by gently swirling the tube and was incubated on ice for 12 hrs. The next day, an equal volume of 50% glycerol was added and mixed gently. The competent cells were made into aliquots of 100 µL in pre-chilled micro-centrifuge tubes and snap-frozen with liquid nitrogen. The frozen cells were stored at -80°C freezer till further use.

3.2.3 Transformation of plasmid DNA

For each transformation, one vial of competent cells was removed from the -80°C freezer and kept on ice for 10 min to thaw completely. 10 μL (100 ng) of ligated DNA/plasmid or 1 μL plasmid (100 ng) was added (avoiding up and down mixing with pipette tip) to the cells and allowed to incubate on ice for 15 min. Then, the cells were subjected to heat shock by placing the culture tubes in a water bath for 90 sec at 42°C . Subsequently, the tubes were kept on ice for 10 min and added 800 μL of LB broth. The cells were placed in a shaking incubator at 180 rpm for 1 hr at 37°C . The tubes were then centrifuged at 3000 rpm for 5 min at room temperature. The supernatant was discarded leaving 100 μL of LB broth medium in the tube in which the transformed cells were resuspended and plated on LB agar plates supplemented with appropriate antibiotics using a sterile spreader and incubated overnight at 37°C . The appearing transformed colonies on the plates were screened by colony PCR with the specific primer sets.

3.2.4 Plasmid DNA Isolation

Isolation of plasmid DNA from bacterial cells was carried out by using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). All the steps were carried out at room temperature unless mentioned. The bacterial cells were streaked from the glycerol stock on LB agar plate supplemented with appropriate selection antibiotics. The plate was kept at 37°C for 12 hrs and a separate single bacterial colony inoculated on 10 mL LB broth medium supplemented with the appropriate antibiotics and allowed to grow at 37°C for 12 hrs at 180 rpm. The bacterial cells were pelleted down by centrifugation at 5000 rpm for 5 min. The supernatant was removed completely and the pellet was re-suspended completely in 250 μL of re-suspension solution. After mixing properly, the cell suspension was transferred to a micro-centrifuge tube. 250 μL of lysis solution was added and gently mixed by inverting the tube 4-6 times until the solution appeared slightly clear. After that, 350 μL of neutralization solution was added, mixed thoroughly by inverting the tube 4-6 times and incubated in ice for 5 min. The bacterial lysate was centrifuged at 12000 rpm for 5 min and the supernatant was transferred into the supplied GeneJET spin column and centrifuged at 12000 rpm for 1 min. The flow-through was discarded and returned the column into the same collection tube. The GeneJET spin column was washed twice with the addition of wash buffer (500 μL) and centrifugation at 12000 rpm for 1 min. Additional centrifugation with an empty column was given to remove the residual wash solution. The GeneJET spin column was transferred into a fresh micro-centrifuge tube. After that, 50 μL of elution buffer or nuclease-free water was added to the centre of the GeneJET

spin column membrane and incubated for 2 min. For elution of plasmid, the GeneJET spin column was centrifuged at 12000 rpm for 2 min at room temperature. The isolated plasmid DNA was quantified by measuring its absorbance at 260 nm and stored in a -20°C freezer for further use. DNA sequencing was carried from a DNA sequencing facility at Macrogen Inc., South Korea.

3.2.5 Extraction of DNA from agarose gel

Gel elution and purification of PCR amplified DNA fragments or plasmids were performed by using QIAquick Gel Extraction Kit (QIAGEN, Germany) according to the manufacturer's protocol. Briefly, DNA fragment of interest was excised from the agarose gel with a clean, sharp scalpel and kept into a fresh micro-centrifuge tube. The excised gel slice within the micro-centrifuge tube was weighed and three volumes of QG buffer (3 volumes Buffer QG to 1 volume gel) was added and incubated at 50°C for 10 min (until the gel slice gets completely dissolved). After that, equal gel volume isopropanol was added to the sample and mixed properly. The dissolved gel with buffer was passed through the QIAquick column and centrifuged at 13000 rpm for 1 min at room temperature. Flow-through was discarded and the column was washed with 500 µL of PE buffer twice by centrifuging at 13000 rpm for 1 min at room temperature. Another centrifugation with an empty column was done to remove the residual wash solution. Then added 50 µL of EB buffer (10 mM Tris-HCl, pH 8.5) or nuclease-free water was added to the centre of the QIAquick membrane and incubated for 2 min at room temperature. The column was centrifuged at 13000 rpm for 1 min at room temperature and collected the DNA in a new Eppendorf tube. The purified DNA was quantified by measuring absorbance at 260 nm and stored at -20°C for further use.

3.2.6 Restriction Digestion and Ligation

Restriction digestion of the target DNA (insert and plasmid) was performed for the cloning with various restriction enzymes. The reaction mixture contained 2 µg of plasmid DNA or eluted DNA insert, 1 µL of restriction enzymes (10 U/µL, Fast digest, Thermo Fisher Scientific, USA) and 1X reaction buffer with the final volume of 20 µL, and incubated at 37°C for 30 min in a water bath. For single restriction site cloning, the vector was further treated with alkaline phosphatase (1 U/µL, Fast AP, Thermo Scientific) and incubated at 37°C for 10 min followed by inactivation at 75°C for 5 min. The complete digestion of the target DNA was checked on 1% agarose gel containing ethidium bromide (10 µg/mL) and visualized on a UV transilluminator. The digested DNAs (plasmid and insert) were purified from gel using

QIAquick Gel Extraction Kit (QIAGEN, Germany). For ligation, 1:3 vector/insert ratio was prepared in 1 μ L of T4 DNA ligase (400 U/ μ L, NEB, England), 1X reaction buffer in 20 μ L reaction volume and incubated at 16°C for 12-15 hrs.

3.2.7 Agarose gel electrophoresis

Nucleic acids (genomic, plasmid or PCR products) were mixed with 6X DNA loading dye (Thermo Fisher Scientific) and were loaded on 0.8-1.2% agarose gel containing 10 μ g/mL EtBr along with 1kb DNA ladder (Thermo Fisher Scientific). The gel electrophoresis was carried out in 1X TAE (40 mM Tris base, 20 mM acetic acid, 1 mM EDTA and pH 8.0) at 5 V/cm using electrophoresis units (Bio-Rad, USA). The resolved DNA was then visualized using gel documentation systems (Alpha imager, USA).

3.2.8 Polymerase Chain Reaction

cDNA/plasmid PCR

The PCR amplification reaction was carried out in a PCR tube containing 100 ng of template DNA, 100 ng of each primer (forward and reverse), 200 μ M dNTPs, 1.5 mM MgCl₂, 1X reaction buffer and 1 U Taq/Phusion DNA polymerase in 25 μ L reaction volume in a thermal cycler.

Table 3.2. The programming of the thermal cycler for the PCR amplification.

Step	Taq polymerase		Phusion polymerase	
	Temperature (°C)	Time	Temperature (°C)	Time
Initial denaturation	95	5 min	98	30 sec
25-35 cycles				
Denaturation	95	30 sec	98	10 sec
Annealing	Vary	30 sec	vary	10 sec
Extension	72	1 min per kb	72	20 sec per kb
Final Extension	72	7 min	72	10 min
Hold	4-16	-	4-6	-

Colony PCR

Colony PCR was carried out to quickly screen for transformed bacterial colonies containing desired recombinant plasmids. For this, a sterile toothpick was used to pick up individual colonies and dipped into each Eppendorf tube containing 20 μ L of sterile water. The same toothpick was streaked onto another LB agar media plate supplemented with appropriate antibiotic(s) and incubated at 37°C for overnight to obtain a master plate. The tubes containing colonies were lysed by heating at 98°C for 10 min and pelleted down the residual colonies and cell debris by a quick short spin. The supernatant (10 μ L) was taken from this mixture and used as the template for PCR amplification in the presence of components and set programmes as described above. The PCR amplicon was observed by agarose gel electrophoresis.

3.2.9 Plant growth conditions, stress treatments and sample harvest

Rice seeds were surface sterilized using 0.25% Bavistin for 20 min, followed by thorough washing with autoclaved distilled water. The seeds were allowed to germinate in germination paper rolls soaked in Yoshida medium (Yoshida et al., 1976) under controlled conditions in a growth chamber (SANYO company) at 28 \pm 2°C temperature with photoperiodic cycle set for 16 hrs light and 8 hrs dark. After ten days, the seedlings were transferred to vermiculite media in small pots for hardening in greenhouses. Thereafter, the plants were transferred to soil pots and allowed to mature. Samples were harvested at appropriate time points. For experiments at the seedling stage, ten days old seedlings were shifted to Yoshida medium and different stresses were imparted for different time durations. Plant tissues were excised, weighed, labelled and wrapped in aluminium foil, frozen in liquid nitrogen, and stored at -80°C for further use.

Table 3.3. Modified Yoshida medium composition used in the study.

S. No.	Reagents	g/L
Macronutrients		
1	NH ₄ NO ₃	91.4
2	K ₂ SO ₄	71.4
3	KH ₂ PO ₄	23.1
4	K ₂ HPO ₄	4.3
5	CaCl ₂ , 6H ₂ O	175.0
6	MgSO ₄ , 7H ₂ O	324.0
Micronutrients		

7	MnCl ₂ , 4H ₂ O	1.5
8	(NH ₄) ₆ Mo ₇ O ₂₄ , 4H ₂ O	0.074
9	H ₃ BO ₃	0.93
10	ZnSO ₄ , 7H ₂ O	0.035
11	CuSO ₄ , 5H ₂ O	0.03
12	FeNa-EDTA	10.5
13	FeSO ₄ (Freshly prepared)	2.5

Dissolved each reagent separately using a magnetic stirrer and 1.25 mL of each stock solution was taken for 1000 mL Yoshida media. The pH was adjusted to 4.5 with diluted HNO₃ before the addition of the micronutrients.

3.2.10 RNA Isolation

Total rice RNA was extracted using TRIzol reagent according to the manufacture's protocol (TRIzol, Thermo Fisher Scientific). DEPC was used at all the necessary steps. 100 mg of leaf tissue was ground to a fine powder in liquid nitrogen using mortar and pestle and added 1 mL of TRIzol reagent. The homogenate was mixed properly and allowed to thaw. The extracts were collected into a 2 mL microcentrifuge tube and 1 mL chloroform (200 µL) was added. The samples were mixed vigorously by hand for 15 sec and kept for an additional 10 min at room temperature. Samples were centrifuged at 12000 X g for 15 min at 4°C. The upper clear aqueous phase was shifted into a fresh micro-centrifuge tube avoiding any contamination from the interphase or the organic layer. Isopropanol (0.6 volume) was added to the aqueous phase and incubated for 10 min at room temperature to precipitate RNA. After that, the samples were centrifuged at 12000 X g for 10 min at 4°C. The RNA pellet was washed with 75% ethanol (prepared in DEPC water) and was centrifuged at 7500 X g for 5 min at 4°C. After 5-10 min of air drying, pellets were re-suspended in 30-40 µL of RNase free water and kept at -80°C for further use.

3.2.11 Quantification of isolated RNA

The yield and quality of RNA were monitored by both spectrophotometric and electrophoresis methods. The concentration of RNA was estimated using the formula ($OD_{260\text{ nm}} = 1$, corresponds to 40µg/mL of RNA). RNA quality was considered to be pure if $A_{260\text{ nm}}/A_{280\text{ nm}} = 2.0$. The integrity of isolated RNA was checked by running on denaturing formaldehyde

(1.85%) agarose gel (1.2 %) in 1X MOPS buffer. Before loading, 500 ng of RNA samples were mixed with 2X RNA loading dye (95% formamide, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF, 0.025% ethidium bromide and 0.5 mM EDTA) and heated at 70°C for 10 min. The samples were chilled on ice and spun down to load on a gel. The gel was run at 5V/cm for 45 min and was visualized using a gel documentation system (Alpha Imager, USA). Intact total RNA run on a denaturing gel appeared as two sharp bands corresponding to 28S and 18S rRNA. The 28S rRNA band was nearly twice intense as the 18S rRNA band which was a good indication for the intact RNA.

3.2.12 cDNA Synthesis

First-strand cDNA was synthesized from total rice RNA using RevertAid™ RNase H minus cDNA synthesis kit (Thermo Fisher Scientific, USA). In brief, a 2 µg RNA template was first treated with 1 U of DNaseI (RNase free) at 37°C for 30 min. DNaseI was inactivated by adding 1 µL of 50 mM EDTA to the mixture and incubated at 65°C for 10 min. RNA was then annealed with 0.5 µg oligo (dT)₁₈ primer (1 µL) at 65°C for 10 min followed by snap cooling. The reaction mixtures were then mixed in a reaction buffer with 1 µL of reverse transcriptase (200 U/µL) and 2 µL dNTP mix (10 mM) along with 1 µL of Ribolock (20 U/µL) to a final volume of 20 µL. The mixtures were incubated at 42°C for 60 min. The reaction was stopped by heating at 70°C for 5 min. The prepared cDNA was stored at -80°C for further use.

3.2.13 Expression analysis by quantitative real-time PCR (qRT-PCR)

Expression analysis was performed as described previously (Tripathi et al. 2015). Primers for qRT-PCR analysis of various genes were designed from the 3' UTR region of the gene using the Primer3plus tool with product length between 99 to 177 bp. The qRT-PCR mixture contained 100 ng of cDNA, 10 µL of 2X SYBR Green PCR Master Mix (Applied Biosystems, USA) and 100 nM of each primer in a final volume of 20 µL. The plate was properly sealed with an optical cohesive cover and then centrifuged briefly to bring down the contents and eliminate any bubbles. qRT-PCR was carried out by using 7500™ Real-Time PCR System and software (Applied Biosystems, USA). The reaction conditions were 95°C (10 min), and 40 cycles of 15 s at 95°C and 1 min at 60°C. The specificity of the amplification was ensured by dissociation curve analysis and agarose gel electrophoresis. The relative expression of each of the genes was calculated using the comparative CT value method (Schmittgenand and Livak, 2008), using eEF as the reference gene (Jain et al. 2006). Three biological and three technical

replicates were analysed for each sample. The mean fold change was calculated and plotted along with corresponding standard deviation values.

3.2.14 Estimation of methylglyoxal

Spectrophotometric assay of MG was done using the protocol of Thornalley and Rabbani, (2014). Fresh tissue (300 mg) was weighed and crushed in mortar-pestle using liquid nitrogen. 2.5 mL of 0.5 M perchloric acid (PCA) was added and mixed well. The mixture was transferred in a micro-centrifuge tube and incubated for 15 min on ice. The extract was centrifuged at 11,000g for 10 min at 4°C. The supernatant was transferred to a fresh tube. Charcoal (10 mg/mL) was added to decolourize the supernatant if coloured. The supernatant was mixed well with charcoal and kept at room temperature for 15 min. The mixture was centrifuged at 11,000g for 10 min. The clear supernatant was then transferred to a fresh tube, and an additional spin was given to remove the residual charcoal from the solution. The solution was neutralized (pH 7.0) using 1 M Na_2HPO_4 which should be added gradually (initially 20 μL , followed by 2 μL at a time). The bubbles of CO_2 gas were allowed to come out. The extract was kept at room temperature for 15 min and centrifuged at 11,000g for 10 min. The supernatant was used for MG estimation.

1 mL of the reaction mixture was prepared by mixing the neutralized supernatant (640 μL) with 1,2-diaminobenzene (250 μL), 5M PCA (100 μL) and 100mM NaN_3 (10 μL). The absorbance was taken initially at 0th hour at 336 nm and set as 'blank'. The same reaction mixture was incubated for 3 hrs at RT, and absorbance was again recorded at 336 nm. A standard curve (Figure 3.1) of different concentrations of MG (10, 25, 50, 100 μM) was made using pure MG.

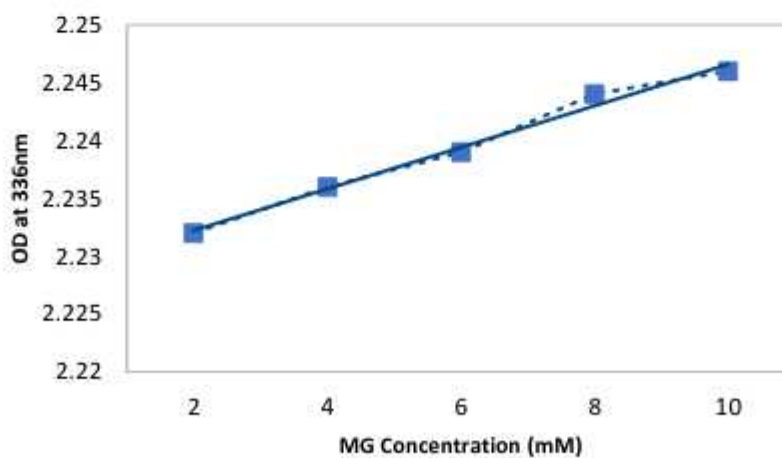


Figure 3.1: Standard curve for MG estimation using a spectrophotometric assay.

3.2.15 Glyoxalase enzyme activity

Glyoxalase I assay

Glyoxalase I assay was done according to Ramaswamy et al. (1983). The purified recombinant protein or total plant extract was used for the determination of the specific enzyme activity in $\mu\text{mol}/\text{min}/\text{mg}$ of protein. The reaction mixture containing 0.1 M sodium phosphate buffer (pH 7.2), 3.5 mM MG and 1.7 mM GSH was prepared and kept in dark for at least 7 min for hemithioacetal formation. 50 μg of protein was added to the reaction mixture up to a final volume of 1 mL. A spectrophotometric time scan was performed for 2 min at 240 nm wavelength. Absorbance was recorded at 10 s intervals and a linear increase in absorbance was observed with time. The molar absorption coefficient of S-D-lactoylglycyl-L-glutathione (SLG) at 240 nm is $3370 \text{ M}^{-1}\text{cm}^{-1}$ and was used to calculate the conversion of hemithioacetal into SLG.

Glyoxalase II assay

GlyII activity was determined according to Marasinghe et al. (2005). A decrease in the absorbance of SLG at 240 nm was monitored to determine the specific enzyme activity in $\mu\text{mol}/\text{min}/\text{mg}$ of protein. 500 ng of purified recombinant protein or 100 μg of total plant protein was added to the reaction mixture containing 10 mM MOPS and 300 μM SLG up to a final volume of 1 mL. A spectrophotometric time scan was performed at 240 nm wavelength for 2 min to determine the change in absorbance. The molar coextinction coefficient of SLG at 240 nm is $3.37 \text{ mM}^{-1}\text{cm}^{-1}$ and was used to calculate substrate depletion.

Glyoxalase III assay

Glyoxalase III activity was determined according to Kwon et al. (2013). A decrease in the absorbance of MG at 540 nm was monitored to determine the specific enzyme activity in $\mu\text{mol}/\text{min}/\text{mg}$ of protein. Protein was added to the reaction mixture containing 10 mM MG, and 20 mM sodium phosphate buffer (pH 6.8), followed by incubation at 37 °C for 30 min. 30 μL of this solution was mixed with 90 μL of a solution containing 0.1% 2,4-dinitrophenylhydrazine, 2 M HCl and 210 μL distilled water. The reaction was stopped by the addition of 420 μL of 10% NaOH. The remaining MG was determined by measuring the absorbance at 540 nm using a spectrophotometer.

3.2.16 GC-MS analysis

The metabolites from shoots and roots of 5-day old rice seedlings, grown on different concentrations of MG, were extracted as described by Soda *et al.*, (2018). The samples were

derivatized by adding 100 μ L N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) and analyzed using GCMS-QP-2010 PLUS (Shimadzu). Metabolites were identified by comparing their retention time and mass spectra with the entries of the NIST14 (<https://www.sisweb.com/manuals/nist.htm>) and Wiley 08 mass spectral libraries (<https://www.sisweb.com/software/wiley-registry.htm>). The data were normalized using ribitol as the internal standard (Soda *et al.*, 2018). Peak identification and spectral deconvolution were done using the GC-MS solution software provided by the manufacturer. Pair-wise comparisons of the metabolite profiles were obtained by comparing the mean standardized peak areas for three replicates.

3.2.17 Confocal microscopy

The tissue samples were cut into small pieces and washed with water into an Eppendorf tube. The recommended concentration of the fluorescent dye solution was taken into an Eppendorf tube and the tissue sample was dipped into it for 5 min. The tissue was then washed thrice with clean water and mounted on a clean glass slide using Prolong Gold Antifade reagent with DAPI (Life Technologies, USA). The slides were visualized using Nikon A1R confocal microscope (Nikon, Japan) using 10X and 20X objectives.

3.2.18 TUNEL (Terminal deoxynucleotidyl transferase-mediated dUTP Nick-End Labelling) assay

Roots of plants were fixed overnight at 4°C with 4% (v/v) paraformaldehyde prepared in 1x PBS, pH 7.4). The next day, the fixed roots were washed twice with 100% ethanol at room temperature and incubated in 70% ethanol for 24 hours. These samples were then incubated in 1x PBS for 1 hour and the solution was changed thrice (after every 20 minutes). Following this, the root samples were incubated in 100 mM sodium citrate buffer (pH 6.0) for 15 minutes followed by an application of microwave (350 W) for 1 minute. The samples were cooled rapidly by adding distilled water. Thereafter, the samples were incubated in permeabilization solution (0.1% Triton X-100 in 100 mM sodium citrate buffer, pH 6.0) at 37°C for 30 minutes. Proteins were digested using proteinase K (20 μ g/mL final concentration, in 10 mM Tris-Cl, pH 7.5 in a total volume of 100 μ L) and incubation at 37°C for 30 minutes. Then the root samples were washed with 1x PBS at room temperature.

For *in situ* TUNEL assay, root tips were cut approximately 1 cm length from the root tip and the TUNEL reactions were performed in microcentrifuge tubes (2 mL) using the DeadEnd™

Fluorometric TUNEL System (Promega, USA) following the manufacturer' instructions. The reaction mixture consisted of 100 μ L equilibration buffer, 10 μ L of fluorescein-12-dUTP - containing nucleotide mix and 1 μ L of recombinant TdT (Terminal deoxynucleotidyl transferase) (Promega, USA). The reaction was carried out at 37°C for 1-hour following which it was stopped by adding 1 mL of 2x Saline-sodium citrate buffer (2x SSC; 300 mM NaCl in 30 mM sodium citrate, pH 7.0). The root tips were stained with 1 μ g/mL of propidium iodide (final concentration prepared in 1x PBS) followed by washing thrice with 1X PBS. The root tips were then mounted onto slides with Prolong Gold Antifade reagent with DAPI (Life Technologies, USA) and the slides were visualized under a confocal microscope (Nikon AIR, Nikon, Japan) using a 20x objective.

3.2.19 Octet Assay (Bio-Layer Interferometry)

Protein-protein interaction was determined using Octet (ForteBio), which is based on the concept of Bio-Layer Interferometry (BLI). It was used to pull out the interacting partners of the target proteins from the crude plant protein extract. Since the target protein had poly-His tag, the Anti-His antibody coated biosensor was used for the experiments. The biosensor was hydrated in a 96-well plate for 10 min in 200 μ L buffer used for protein purification (150 mM Tris-HCl, pH 7.2). The first step was the loading step where the target protein was coated onto the biosensor. The biosensor was dipped in 200 μ L of the target His-tagged protein for 10 min, followed by washing in the Tris-HCl buffer for 5 min. Next the target protein coated biosensor was dipped in the crude protein extract for 10 min, follows by washing in the Tris-HCl buffer to remove unwanted proteins. The bound proteins were eluted in 200 μ L elution buffer (150mM Tris-HCl, pH 7.2). The steps of dipping in crude protein extract to elution was repeated multiple times to concentrate the eluted proteins in the elution buffer. The eluted protein was run on 1.5 % SDS-PAGE gel to confirm the eluted proteins, which was further gel eluted and given for mass spectrometry for identification.

3.2.20 Yeast-two hybrid assay

Yeast-two hybrid assay to determine the interaction of OsDJ-1C & D with various proteins

For yeast-two hybrid assay to check one-to-one interaction, the coding sequence for OsDJ-1C & D were cloned in pGBKT7 vector such that the DNA-binding domain of the GAL4 transcription factor is fused with OsDJ-1C & D (BD). Also, all the four proteins were cloned

in Y2H one-to-one interaction vector pGADT7 (AD). Freshly prepared yeast AH109 competent cells resuspended in PEG/LiAc solution were then co-transformed with the AD-BD pairs and grown on a two-dropout SD media (-Leu-Trp). Colonies were replica-plated onto SD/-Trp/-Leu/-His three-dropout (3DO) selective medium for the selection of interactors. Finally, the interaction was checked by replica plating on a four-dropout medium (SD/-Trp/-Leu/-His/-Ade), containing the chromogenic substrate, for the reporter gene (MEL1 coding for α -galactosidase), X- α -Gal.

3.2.21 Agrobacterium-mediated transformation of rice (PB1 variety)

Rice var. PB-1 was used for the generation of transgenic plants. The tissue culture method as detailed by Sahoo et al (2011) was followed. The temperature in the tissue culture room was maintained at 28°C with 14 hr light and 10 hr dark cycle. Seeds were surface sterilized and inoculated on callus induction media (MCI). The composition of the MCI media is as follows: MS salts containing all vitamins and supplemented with 30 g/L maltose, 0.3 g/L casein hydrolysate, 0.6 g/L L-Proline, 3 mg/L 2, 4-dichlorophenoxyacetic acid (2,4-D), 0.25 mg/L 6-benzylaminopurine (BAP), pH adjusted to 5.8 before autoclaving and gelled with 3.0 g/L phytigel. The inoculated seeds were incubated at 27±1°C in dark. After fourteen days of incubation in dark, callus was selected based on its embryogenic origin. These selected embryogenic calli were divided into approximately three equal parts and sub-cultured again onto a fresh MCI and were further incubated in dark (27±1°C) for 4 days before transformation with *Agrobacterium tumefaciens*.

Preparation of *Agrobacterium tumefaciens* competent cells and transformation

The *Agrobacterium tumefaciens* (LBA4404) cells were streaked from the glycerol stock on an LB agar plate containing rifampicin (10 mg/L) and streptomycin (25 mg/L). The plate was incubated at 28°C for 48 hrs in dark. Then, a single colony was picked and inoculated into 5 mL of YEM broth medium as a primary culture containing the same selection markers. The culture was incubated at 28°C for 48 hrs at 200 rpm in dark. After that, 0.5 mL of primary culture was inoculated into 50 mL of YEM broth medium to initiate secondary culture containing the same concentration of rifampicin and streptomycin and incubated at 28°C and 200 rpm till OD_{600 nm} reached 0.5-0.6. The secondary culture was chilled on ice for 10-20 min and transferred to Oakridge tubes. The tubes were centrifuged at 3000 rpm for 10 min at 4°C. The supernatant was removed and the pellet was dissolved in 10 mL of ice-cold 100 mM CaCl₂.

The cells were incubated on ice for 2 hrs and then were harvested by centrifuging at 3000 rpm for 5 min at 4°C. The pellet was re-suspended in 1 mL of ice-cold 10 mM CaCl₂ by gently swirling the tube. An equal volume of 50% sterile glycerol was added to the suspension and mixed gently. The competent cells were aliquoted into 100 µL in sterile pre-chilled micro-centrifuge tubes and snap-frozen with liquid nitrogen. The frozen cells were stored at -80°C freezer until use.

For transformation, the *Agrobacterium* competent cells otherwise stored at -80°C were first allowed to thaw on ice for 10-15 min. 2 µg of plasmid DNA was added followed by incubating on ice for 30 min. After that, the cells were snap-frozen in liquid nitrogen for 1 min and subsequently, heat-shocked at 37°C for 5 min. The cells were incubated on ice to recover for 10 min. Then 800 µL of YEM broth medium was added and the culture was incubated at 28°C for 3 hrs with shaking at 200 rpm, in dark. Following incubation, the culture was centrifuged at 3000 rpm for 5 min at 4°C and the supernatant was discarded leaving 100 µL in the tube. The bacterial pellet was re-suspended and plated on a YEM plate containing rifampicin (10 mg/L), streptomycin (25 mg/L) and kanamycin (50 mg/L) antibiotics for plasmid selection. The plates were incubated at 28°C for 48 hrs at dark to allow bacterial colonies to form. The colonies were screened by colony PCR with various combinations of primers.

Preparation of *Agrobacterium* cells for the transformation of rice calli

Transformed *Agrobacterium* cells were inoculated in 5 mL of autoclaved YEM broth (HIMEDIA, M1824) supplemented with rifampicin (10 mg/L), streptomycin (25 mg/L) and kanamycin (50 mg/L) using a single colony from a freshly streaked plate. The culture was allowed to incubate at 28°C for 16-20 hrs on a rotatory incubator shaker (Kuhner, Switzerland) at 200 rpm in dark. The secondary culture was inoculated by adding primary inoculum at 1% concentration and grown under similar conditions in a 500 mL flask containing the same antibiotics as used for primary culture. Bacterial cells were allowed to grow till the OD_{600nm} reached ~1.0. Then, *Agrobacterium* cells were pelleted down by centrifugation at 8000 x g for 15 min at 4°C. The pellet was re-suspended in MS re-suspension medium containing MS salts, 68 g/L sucrose, 36 g/L glucose, 3 g/L KCl and 4 g/L MgCl₂, pH 5.2 and 150 µM Acetosyringone. The final cell density of the *Agrobacterium* suspension was adjusted to OD_{600 nm} ~0.3.

Co-cultivation and selection of transformed calli

Sub-cultured embryogenic calli were used for Agro-infection by submerging them in an *Agrobacterium* cell culture (LBA4404) for 20-25 min with intermittent gentle shaking at 50-60 rpm. The Agro-infected calli were dried on autoclaved Whatman No.3 filter paper for 5-10 min and placed on the co-cultivation medium (MCCM) containing MCI, 10 g/L glucose, pH 5.2, 150 μ M Acetosyringone, and kept at $27\pm 1^\circ\text{C}$ in dark for around 48 hrs. Once *Agrobacterium* growth appeared around the calli, they were rinsed with sterile water containing 250 mg/L cefotaxime 8-10 times, followed by drying on sterile Whatman No.3 filter paper and shifted onto the selection medium (MSM; MCI containing 250 mg/L cefotaxime and 1.75 mM/L methylglyoxal) and thus kept for twelve days at $27\pm 1^\circ\text{C}$ in dark. Brown or black calli were discarded after the first selection and only creamish healthy calli were placed onto a fresh MSM media for the second selection and incubated at $27\pm 1^\circ\text{C}$ in dark for another ten days. After the second selection, micro-calli so obtained were finally shifted onto fresh MSM media for the third selection and allowed to proliferate at $27\pm 1^\circ\text{C}$ in dark for five days.

Regeneration of transformed calli

After three rounds of selection, black or brown micro-calli were removed and only creamish coloured granular were placed onto the regeneration media comprising of MS salts, 30 g/L maltose, 2 mg/L kinetin, 0.2 mg/L NAA, pH 5.8. 250 mg/L cefotaxime and 1.75 mM/L methylglyoxal were added after autoclaving. Regeneration media was supplemented with 10 g/L and 8 g/L agarose during the first and second phase of regeneration, respectively. The micro-calli were incubated for seven days at $27 \pm 1^\circ\text{C}$ in dark during the first regeneration. Moreover, micro-calli were shifted onto a fresh second regeneration medium and incubated for four days in light at $27 \pm 1^\circ\text{C}$.

Development of roots

The regenerated shoots were transferred to jam bottles containing rooting medium (1/2 strength MS salt, 30 g/L sucrose, 3 g/L phytagel, pH 5.8; 250 mg/L cefotaxime and 0.75 mM/L methylglyoxal were added after autoclaving) and incubated for root development at $27\pm 1^\circ\text{C}$ for a week. The rooted plantlets were transferred to vermiculite pots for hardening and covered with polythene bags to maintain the moisture for seven days in the greenhouse. After seven days, polythene bags were removed and the plants were transferred to earthen pots. Putative transgenic plants were screened by PCR for the integration of transgene along with the wild type.

3.2.22 Screening of transgenic plants by Tissue-PCR

Putative transgenic rice plants were screened by using the RedExtract-N-Amp Tissue PCR kit (Sigma-Aldrich, USA). In brief, 0.5 to 0.7 cm of leaf disks were cut into small pieces by a sharp scalpel and transferred into a 2 mL micro-centrifuge tube. Extraction solution (50 μ L) was added to the samples and mixed thoroughly by vortexing. After that, the samples were spun down and boiled at 95°C for 10 min. 50 μ L of dilution solution was added to the samples after cooling on ice. The mixtures were vortexed thoroughly to mix them properly. The leaf extract supernatant was used directly as a template for PCR reaction. The PCR reaction was prepared by 10 μ L of RedExtract-N-Amp PCR ready mix, 2 μ L of leaf extract and 400 nM of each primer in a 20 μ L of reaction volume.

3.2.23 Isolation of native total plant protein for glyoxalase enzyme activity

For enzyme activity, total plant protein was extracted under native conditions with special precautions. All the steps were carried out at 4°C. 200 mg of fresh leaf tissue was powdered in liquid nitrogen by using mortar and pestle. The fine powder was homogenized with ice-cold extraction buffer containing 100 mM potassium phosphate buffer- pH 7.0, 50% glycerol, 16 mM MgSO₄ and 0.5 mM PMSF. The homogenate samples were centrifuged at 12000g for 30 min at 4°C. The supernatant was used as a crude extract (enzyme homogenate) for the assay of various enzyme activities after quantification.

3.2.24 Primer tables

All the primers were designed using primer3plus software (<http://primer3plus.com/cgi-bin/dev/primer3plus.cgi>) with default parameters. The uniqueness of each primer in the rice genome was rechecked by primer blast at NCBI database (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). All designed oligonucleotide primers were synthesized from Sigma-Aldrich.

Table 3.4. List of primer pairs used in the various PCR studies.

Gene name	Primer name	Primer sequence
<i>OsGLYI1</i>	OsGLYI1_F	GAGACTTGTGGCATTGCTGA
	OsGLYI1_R	CACCTTAATCATCATCACGAC
<i>OsGLYI2</i>	OsGLYI2_F	TAACAGAGCATGAGGTTTCG

	OsGLY12_R	CAAAGTTGGAACAGATCAGG
<i>OsGLY13</i>	OsGLY13_F	GGATTTTAGCCGTGTATGTTC
	OsGLY13_R	TGGACTTTGGAGTACAGACC
<i>OsGLY14</i>	OsGLY14_F	TTTGCTTGCTGTAACAGTCG
	OsGLY14_R	TATTCATGGGGATATCATGG
<i>OsGLY15</i>	OsGLY15_F	ACGTTCTTATGTCTCTCCTTTG
	OsGLY15_R	CAACTAACGTGTATGCATGG
<i>OsGLY16.1/6.2</i>	OsGLY16.1_F	TAGCAAGGATGGAGTACGTG
	OsGLY16.1_R	TGATCTAGCAAACCACAACC
<i>OsGLY16.3/6.4/6.5</i>	OsGLY16.3_F	ATTTGCATGTCCTACCTCAC
	OsGLY16.3_R	CTTAGTGGAAGGCAATCTG
<i>OsGLY17.1</i>	OsGLY17.1_F	GAAAGTTAGCATTCTGCTG
	OsGLY17.1_R	TCGCTGATTCCATACTTACC
<i>OsGLY17.2</i>	OsGLY17.2_F	CATTCAAAGAGCTTCCATC
	OsGLY17.2_R	GCAAGGATCAGATTAGAAAGC
<i>OsGLY18</i>	OsGLY18_F	CATTTCTTGTTGACAGGTG
	OsGLY18_R	GCAAACAGGGTATCACAATC
<i>OsGLY19.1</i>	OsGLY19.1_F	TGAAGTTGTATGTTGGTATTGG
	OsGLY19.1_R	AGTGGGAGCTTACAGTATTCG
<i>OsGLY19.2</i>	OsGLY19.2_F	TGCTTGTTGTCTTCATAGCC
	OsGLY19.2_R	AGGCTTTCAATTTACCACATC
<i>OsGLY110</i>	OsGLY110_F	CAAGCTTGAGTTTGCATG
	OsGLY110_R	GGAGAACGATCGTAGTAGCC
<i>OsGLY111</i>	OsGLY111_F	GCTAGTACTGTCTGGGTTTGG
	OsGLY111_R	GAAACTCAGCTCAACTGCAC
<i>OsGLY111.2</i>	OsGLY111.2_F	GTGGTGGTTCTTTTCTGATG
	OsGLY111.2_R	GAGAACTGATATCGTGCAATG
<i>OsGLY111.3</i>	OsGLY111.3_F	TTATTATGCCTGAAAGTTCTGAG
	OsGLY111.3_R	GTGCATGCCAAAAGTTATTC
<i>OsGLY111</i>	OsGLY111_F	GGAAACATTCAAGAAAATTATGGAC

	OsGLYIII_F	TGGTGGAGGATCCTGAATAC
<i>OsGLYII2</i>	OsGLYII2_F	CTGTAGGTGGGGCTTACTGC
	OsGLYII2_F	TGCACATCTGATGATGTCTCC
<i>OsGLYII3</i>	OsGLYII3_F	TTACGGAGTAGGCAGTTGGTG
	OsGLYII3_F	TGCAATGAAGTTGATCGAGTG
<i>OsGLYIII1</i>	OsDJ-1A_F	ATGCAGATGTGTGATCTGTGA
	OsDJ-1A_R	TCCATCAATCAATCCATCAGCTA
<i>OsGLYIII2</i>	OsDJ-1B_F	GCCTGTACCTGTGAGTGAT
	OsDJ-1B_R	AGACAGCGTAAACTCACCGA
<i>OsGLYIII3</i>	OsDJ-1C_F	CACAGCTTTGGGCTTTCGTC
	OsDJ-1C_R	CAGCTCCTTCTGCTCGTACA
<i>OsGLYIII4</i>	OsDJ-1D_F	GCGGTTGAGAGACAGTGTGA
	OsDJ-1D_R	CACGGCATGAGAACAACAGC
<i>OsGLYIII5</i>	OsDJ-1E_F	TGAGGCATCCTTTTGCTCGT
	OsDJ-1E_R	ACCCAAAGGTGCAATATCCATC
<i>OsGLYIII6</i>	OsDJ-1F_F	CTCGCCGAGTTCATCGCTC
	OsDJ-1F_R	TGACAGAGAGAACAGAGAAGCA
<i>eEF</i>	eEF_F	TTTCACTCTTGGTGTGAAGCAGAT
	eEF_R	GACTTCCTTCACGATTTTCATCGTAA
<i>Actin</i>	Actin-F	CGGTGTGATGGTTGGTATGG
	Actin-R	GCCTCAGTCAGCAACACAGG
<i>OsMPK3</i>	OsMPK3-F	GCTCCAACCAAGAAGTCTGTC
	OsMPK3-R	AGTCGCAGATCTTGAGG
<i>OsMPK4</i>	OsMPK4-F	CGAGGTCTCCTCCAAGTACG
	OsMPK4-R	GCGAAGCAGCTTGATTTCTC
<i>OsMPK6</i>	OsMPK6-F	AGGTCACCGCCAAGTACAAG
	OsMPK6-R	AGCAGCTTGATCTCCCTGAG
<i>OsMPK14</i>	OsMPK14-F	TCCTGAGTTGCTCCTTTGCT
	OsMPK14-R	CGAGCTTTTGGGTTGTCAAT
<i>OsMPK16-1</i>	OsMPK16-1-F	CTTTGAGCATGTGTCCGATG
	OsMPK16-1-R	TTGGTGCAAATCAGACTCCA
<i>OsMPK16-2</i>	OsMPK16-2-F	TGGAGTCTGATTTGCACCAA

	OsMPK16-2-R	TGCAAGGCCAAAATCACATA
<i>OsMPK17-1</i>	OsMPK17-1-F	TCATCAGAGCGAACGATGAC
	OsMPK17-1-R	TGCAACATAATCCGTCCAAA
<i>OsMPK17-2</i>	OsMPK17-2-F	GCCTGCTAGAGCGTTTAC
	OsMPK17-2-R	CCTTTGTCAGTTTCCTTCG
<i>OsMPK20-1</i>	OsMPK20-1-F	TCGAGGAGGGATTTCAAAGA
	OsMPK20-1-R	CTTCAGGTCCCGGTGATAAA
<i>OsMPK20-2</i>	OsMPK20-2-F	GCTATGGGGTTGTGTGCTCT
	OsMPK20-2-R	ACAATGTCCGGATGCCTTAG
<i>OsMPK20-3</i>	OsMPK20-3-F	5' GATGCTCCGTGCTTTGAAAT
	OsMPK20-3-R	TGCAACATAATCGGTCCAGA
<i>OsMPK20-4</i>	OsMPK20-4-F	TTTCGCTCCCAAGGACT
	OsMPK20-4-R	CCCATTGCACTCAAATTCTC
<i>OsMPK20-5</i>	OsMPK20-5-F	AGAGAACCATCATGCCAACC
	OsMPK20-5-R	GTGTTTTTCGGAGCCATTCAT
<i>OsMPK21-1</i>	OsMPK21-1-F	CGAGTGGCGATCAAGAAGAT
	OsMPK21-1-R	ATGTCCCTGAATTCCCTCCT
<i>OsMPK21-2</i>	OsMPK21-2-F	CCGGAGTTCCTCAGCGAGTA
	OsMPK21-2-R	AGGAGGCGGAGTAGCTTGAT
<i>OsMKK1</i>	OsMKK1-F	ACCATCGGCAAATTCCTGAC
	OsMKK1-R	GAACCAACTGCACGATTCCA
<i>OsMKK3</i>	OsMKK3-F	GTTGAATTCCAGGGTGCATT
	OsMKK3-R	TTCATGCAAGTAGCGCAAAC
<i>OsMKK4</i>	OsMKK4-F	GGACCATCGCCTACATGAGC
	OsMKK4-R	GCGAGTCGGAGTAGCAAAT
<i>OsMKK6</i>	OsMKK6-F	TCCGAGGAAACTGCAGATGA
	OsMKK6-R	TTTGCGAACTGCCTCTTGAA
<i>OsMKK10-1</i>	OsMKK10-1-F	GCCCCTGAGGAGAGAGAGAG
	OsMKK10-1-R	CCGTTGGACTTGCATCATT
<i>OsMKK10-2</i>	OsMKK10-2-F	ACCTCAAGCCGTCGAACCT
	OsMKK10-2-R	CTACCCCGAGGCTCCACAC
<i>OsMKK10-3</i>	OsMKK10-3-F	TGGAAACGAAAAGAAATGGAA

	OsMKK10-3-R	AGGCTAATCTCAGTGCAAGG
<i>OsAKR3</i>	AKR3_F	GCTGTTCCCTCCCAATGTA
	AKR3_R	CATATCTCTGCTGCCTCTTGC
<i>OsAKR5</i>	AKR5_F	ACTGTCCCTGGGAGCTTCTT
	AKR5_R	CAGAAGCGCAAATACACGAA
<i>OsAKR11</i>	AKR11_F	TGCCATTGCTGCAAATAAAG
	AKR11_R	TTTGCCGGTAAAGATTCTGG
<i>OsAKR16</i>	AKR16_F	AAAATGGAAGAGCGTTGCAC
	AKR16_R	GCGCAAATTAAACATCACGA
<i>OsAKR18</i>	AKR18_F	TTCAGAACAGAACACAGTAGTCAGA
	AKR18_R	TGGCATTGGCAACAACAATA
<i>OsAKR22</i>	AKR22_F	TGGCCCTATGCACCAATAAT
	AKR22_R	CGAATAGAAGAAAGATAACAATTGAGC
<i>Histone</i>	Histone_P_Hind III-F	CCCAAGCTTCAAGGCACTCTCTATGTCAA
	Histone_P_EcoR I-R	CCGGAATTCTCAGATTTAATTTAAACTACATA ATATGCTTG
<i>GATA</i>	GATA_P_HindI II-F	CCCAAGCTTACATACAAATATTTATGGGAGGG AGTT
	GATA_P_EcoR I-R	CCGGAATTCACAAACAATAATTACTCCTTACC C
<i>MYB</i>	MYB_P_HindIII I-F	CCCAAGCTTCCAAGCCAGCCACTGC
	MYB_P_EcoRI- R	CCGGAATTCACTACACTTTGCATGCTCCAG
<i>B3</i>	B3_P_HindIII-F	CCCAAGCTTTCTTTGTACGTCTCTTAAGAATTT ATGG
	B3_P_EcoRI-R	CCGGAATTCTAAACGCACACACATGCTCTC
<i>Hairpin</i>	Hairpin_P_Hind III-F	CCCAAGCTTTCTCAGAACTCATCTCAGGGTT
	Hairpin_P_Eco RI-R	CCGGAATTCGAAGGAACGCACGACAGCTA

Exploring the mechanisms of MG accumulation and detoxification: Tissue and developmental profiling in IR64 and Pokkali

4.1 Background

MG formation and detoxification have been studied extensively in different organisms, including microbes and animals (Rhee *et al.*, 1987; Kalapos and Biology, 1999). However, there are many aspects of MG homeostasis in plants that need investigation. For instance, there aren't yet any reports unravelling catalytic mechanisms for MG formation in plants as against prokaryotes, which possess MG synthase. Functional studies have established the role of MG-scavenging glyoxalase proteins under stress conditions as well as in developmental processes (Deswal *et al.*, 1993; Veena *et al.*, 1999; Yadav *et al.*, 2005; Lin *et al.*, 2010; Singla-Pareek *et al.*, 2003; Singla-Pareek *et al.*, 2006; Singla-Pareek *et al.*, 2008; Ghosh *et al.*, 2016; Gupta *et al.*, 2018; Sankaranarayanan *et al.*, 2015). However, detailed spatiotemporal profiling of MG can give important clues regarding the physiological processes involved in MG homeostasis. In addition, the presence of multiple GLY genes, i.e., 11 GLYI, 3 GLYII (Mustafiz *et al.*, 2011) and 6 GLYIII (Ghosh *et al.*, 2016) in plants, also warrant an in-depth expression analysis along with the determination of their enzymatic activities. Further, equally important is to determine the role of other detoxification enzymes, like aldo-keto reductases (AKRs). Since AKRs are not extensively studied in plants, we do not know how many copies of these genes exist in rice. A genome-wide study is needed to elucidate such information and to further unravel their role and regulation in plants. A better understanding of the MG detoxification process in totality will be beneficial for designing crops with efficient MG detoxification capabilities. Also, the agronomical application of MG as a biomarker can help in the primary screening of stress-tolerant plants from large populations rapidly and cost-effectively.

In the present study, we have carried out the spatiotemporal profiling of MG in two different rice varieties, i.e., salt-sensitive IR64 and salt-tolerant Pokkali. In parallel, we have also analysed the expression of different GLY genes, coupled with glyoxalase enzyme activity in the two rice varieties

4.2 Results

4.2.1 Comparative analysis of spatiotemporal variations in MG levels in salt-sensitive, IR64 and salt-tolerant Pokkali

IR64 and Pokkali rice varieties were grown in pots under similar greenhouse conditions. Samples were harvested from various tissues at different developmental stages (Figure 4.1) and stored at -80°C till further processing. These samples were then used for MG estimation, RNA extraction, and GLY enzyme assays, using the protocols described in the Materials and Methods section (Chapter 3). Three

leaf samples, i.e., flag, second and third, were taken, while the stem samples were divided into top, middle and bottom.

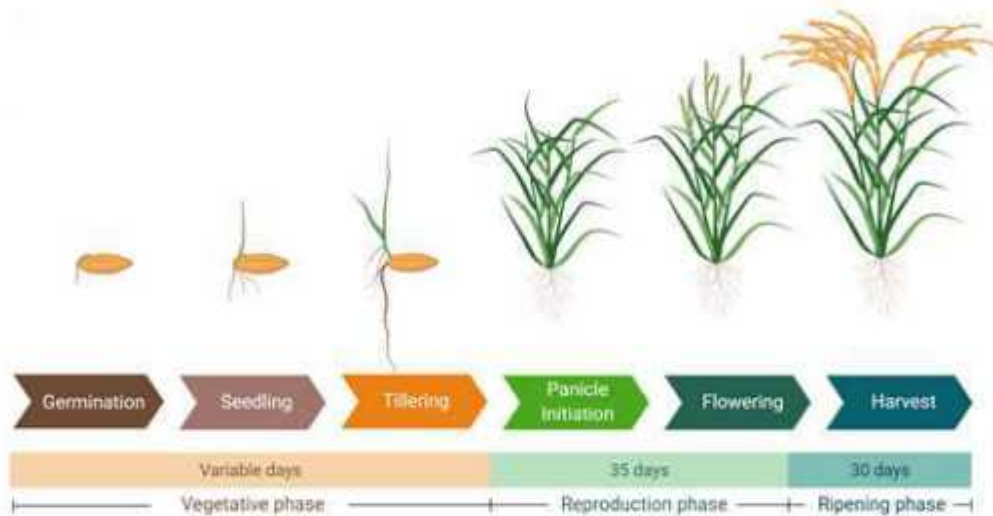


Figure 4.1: Stages in growth and development of rice. In our study, we used tissues from the seedling, tillering, flowering and mature stages of rice for the determination of MG levels and expression profiling of glyoxalases. Among the tissues, we used flag leaf, 2nd leaf, 3rd leaf, top stem, middle stem, bottom stem, root, panicle and seed for profiling.

4.2.1.1 Profiling of MG across various developmental stages and in different tissues

MG levels were estimated in several tissues harvested during different developmental stages and variations were analysed between IR64 and Pokkali (Figure 4.2). At the germination stage, IR64 shoots had around 3-times higher MG levels than found in Pokkali shoots. However, in roots, the MG levels were almost equal in both the genotypes. Notably, roots of both, IR64 and Pokkali, have extremely low MG levels as compared to shoots. Further, at tillering stage, IR64 had higher MG levels in all tissues analysed compared to that in Pokkali. The MG levels in leaves were marginally higher than in stems in both the genotypes, though the difference was more pronounced in IR64. However, during flowering, Pokkali had greater MG levels than IR64 in different tissues (except root and panicle) in contrast to that seen during tillering stage. In the mature stage, the MG levels are similar in all tissues, except flag leaf, root and seed. In mature leaf and seed, IR64 had much higher MG levels, whereas, in the root, Pokkali had higher MG.

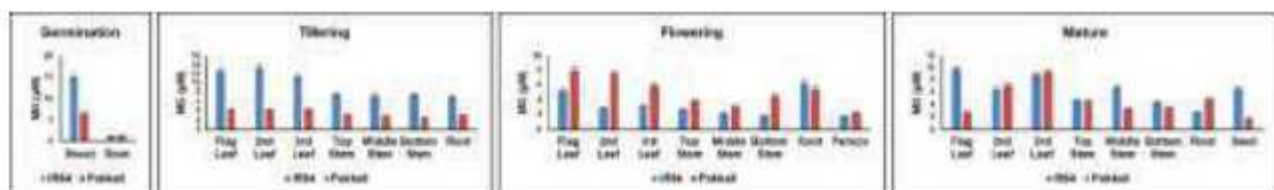


Figure 4.2: Comparison of the spatiotemporal profile of MG in two contrasting rice genotypes, salt-sensitive IR64 and salt-tolerant, Pokkali. Absolute levels of MG were measured in triplicates, as described in Materials and Methods (Section 3.2.13).

4.2.1.2 Expression profiling of Glyoxalase genes

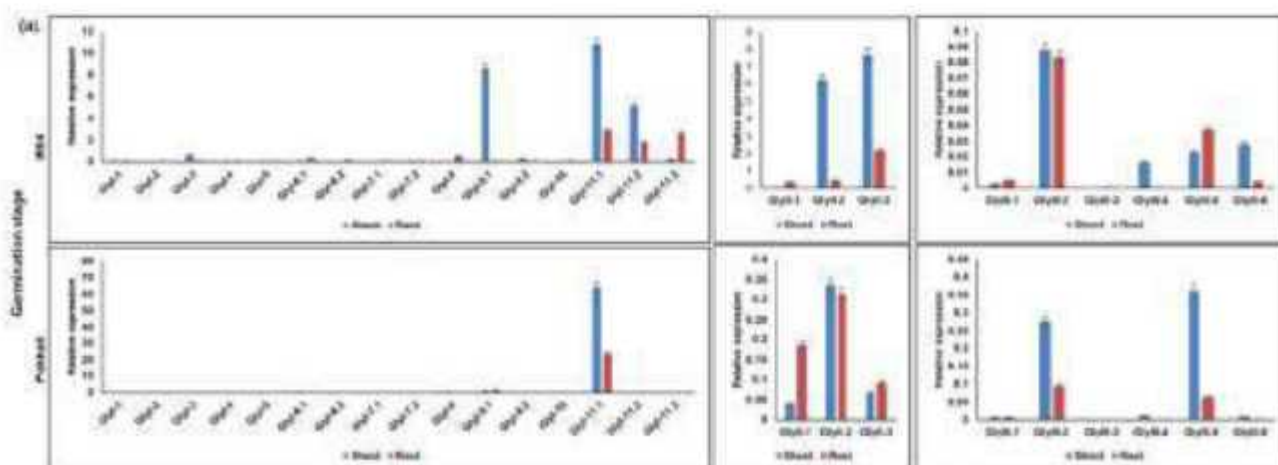
The RNA isolated from various tissues of IR64 and Pokkali was used to make cDNA which was then used to quantify the expression levels of different *Gly* genes (Figure 4.3).

At the germination stage (Figure 4.3a), expression of only a few *GlyI* genes could be detected in both genotypes, with Pokkali exhibiting relatively higher expression of these genes than IR64. Similar was the case for *GlyIII3* genes. In IR64, *GlyI-9.1*, *GlyI-11.1*, *GlyI-11.2*, *GlyII-2*, *GlyII-3* and *GlyIII-2* had higher expression in shoots than in roots. Similarly, Pokkali also had a high expression of *GlyI-11.1*, *GlyII-2*, *GlyIII-2* and *GlyIII-5*. Of which *GlyI-11.1* was expressed the most.

During tillering stage (Figure 4.3b), IR64 showed higher expression of *Gly* genes compared to Pokkali. *GlyI-7.2* and *GlyI-9.1*, along with *GlyIII-2* were expressed at higher levels in IR64 leaves, whereas the stem had a higher expression of *GlyII-2* genes. On the other hand, Pokkali showed higher expression of *GlyI-11.1* in almost all the tissues. While IR64 exhibited higher expression of *GlyII* and *GlyIII* genes.

During the flowering stage (Figure 4.3c), IR64 tissues generally had a higher expression of *Gly* genes compared to Pokkali. Except for *GlyI-6.1* and *GlyI-11.1* which had higher expression in IR64 leaves. Similarly, IR64 leaves also had higher expression of *GlyII-2*, *GlyII-3*, *GlyIII-1* and *GlyIII-2* genes. All the tissues in Pokkali had a higher expression of *GlyI-11.1* and *GlyII-2* genes. Further, Pokkali leaves had a relatively higher abundance of even *GlyIII-2* and *GlyIII-4* genes.

During the maturation stage (Figure 4.3d), *GlyI-11.1* is highly expressed in both IR64 and Pokkali compared to other genes, with its levels being higher in all tissues of Pokkali. Similarly, *GlyII-2* and *GlyII-3* were expressed in almost all the tissues, but higher in IR64 compared to Pokkali. IR64 leaves had a higher expression of *GlyIII-2* and *GlyIII-5*, whereas all the tissues in Pokkali had an abundance of *GlyIII-4*.



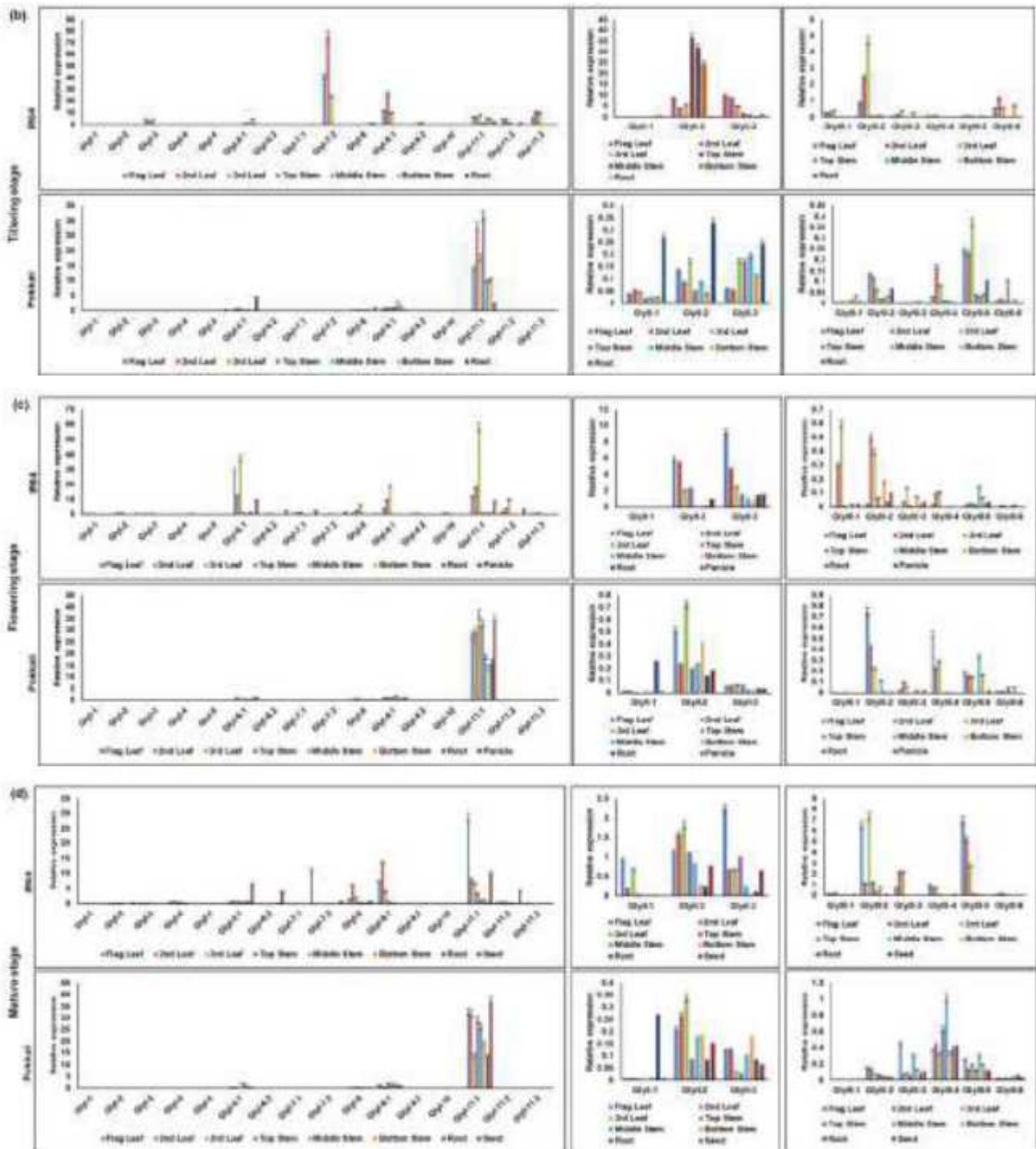


Figure 4.3: Comparative tissue-specific expression profile of *Gly* genes in the two contrasting rice genotypes, i.e., salt-sensitive IR64 and salt-tolerant, Pokkali, at (a). Germination, (b). Tillering, (c). Flowering, and (d). Maturation stage. The relative expression levels of genes are presented using fold-change values ($2^{-\Delta\Delta Ct}$), normalized to the *efl1a* gene.

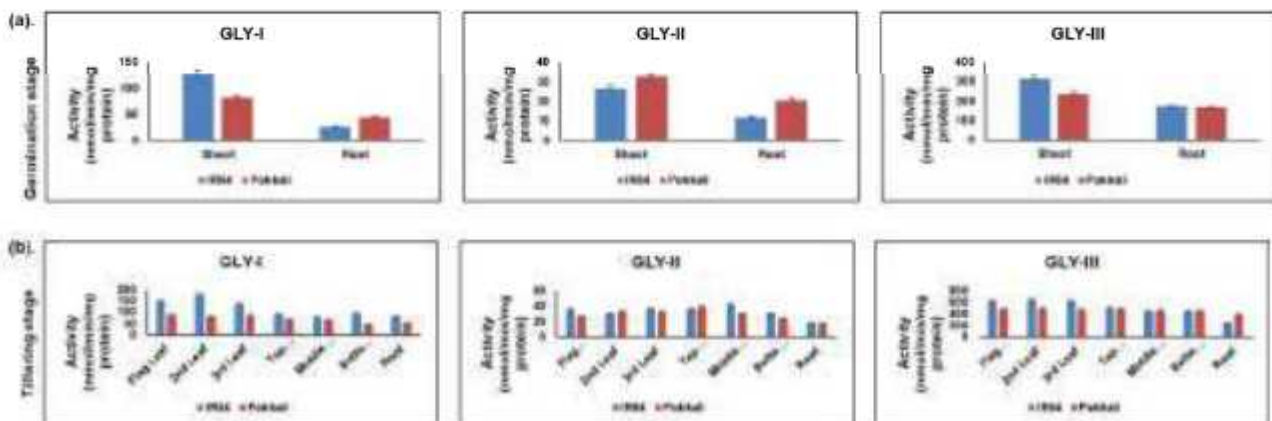
4.2.1.3 Tissue-specific and developmental variations in glyoxalase enzyme activity

The tissue harvested from different developmental stages of IR64 and Pokkali was used to extract crude protein which was then used to measure different types of GLY activities, viz. GLYI, GLYII and GLYIII (Figure 4.4). At the germination stage (Figure 4.4a), GLYI activity was higher in IR64 shoots compared to Pokkali shoots. On the contrary, in the roots of germinating seedlings, IR64 had lower GLYI activity compared to Pokkali. Further, GLYII activity in shoots was not significantly different between the two genotypes, but varied in roots, with Pokkali having slightly higher activity. However, GLYIII activity was higher in IR64 shoots, with no difference in roots, between the genotypes.

At tillering stage (Figure 4.4b), IR64 tissues exhibited greater or similar activities of GLY enzymes compared to Pokkali. The leaves, however, showed greater variations between GLYI & GLYIII activity of IR64 and Pokkali. On the other hand, GLYII activity was similar between the two genotypes. In the case of roots, Pokkali had higher GLYIII activity compared to IR64 whereas GLYI & GLYIII activity was higher in leaves in both IR64 and Pokkali.

During the flowering stage (Figure 4.4c), Pokkali exhibited higher GLYI activity in all tissues, except panicle as compared to IR64. However, GLYII activities were not much difference between the two genotypes, except in the second and third leaves. As against GLYI, GLYIII activity was higher in IR64 plants. The enzyme activity showed greater differences in leaves, followed by stem and root. In panicle, GLYI and GLYII activities were similar between IR64 and Pokkali but GLYIII activity was higher in IR64 panicles.

During the maturation stage (Figure 4.4d), GLYI activity in tissues of IR64 was either slightly higher or equal to Pokkali. While GLYII & GLYIII activities were almost similar in all tissues in both genotypes.



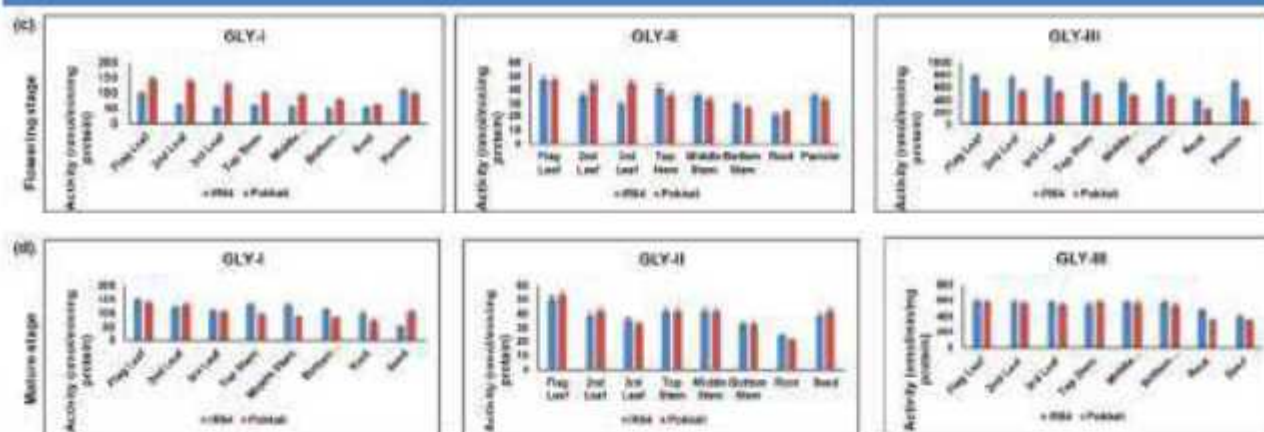


Figure 4.4: Comparative study of the enzymatic activity of GLY proteins in the two contrasting rice varieties, i.e., salt-sensitive IR64 and salt-tolerant, Pokkali, at (a). Germination, (b). Tillering, (c). Flowering, and (d). Maturation stage.

4.2.2 Genome-wide analysis of AKRs in rice: Identification, phylogeny, architecture and transcriptional regulation

AKRs provide an alternate route of MG detoxification. However, they haven't been studied extensively in plants yet. There are a few reports which concentrate only on selected AKRs. For instance, *AKR1* overexpression in rice and tobacco was shown to improve seed longevity during storage by detoxification of reactive carbonyl compounds (Narayana *et al.*, 2017). Another study established the role of *AKR1* in the detoxification of glyphosate (Vemanna *et al.*, 2017). Yet another study, established broad substrate specificity of *AKRs* and indicated their protective role against small and medium-chain aldehydes in rice (Auiyawong *et al.*, 2017). In this section, we have tried to identify all the *AKRs* present in rice and determine the most functionally active ones to be used either alone or in combination with other genes, to improve the stress tolerance of rice. Further, we have tried to analyse the regulation of these *AKR* genes using promoter and interactome studies. The protocols followed for this analysis are given in the Materials and Methods section 3.1 (Chapter 3).

4.2.2.1 *In silico* analysis reveals 28 putative AKR genes in rice

A total of 28 *AKR* genes (Table 4.1) were identified in rice coding for 46 *AKR* proteins (Additional file 1: Table S1) through BLAST search using characterized *AKRs* as query in the RGAP database (Kawahara *et al.*, 2013). These were further confirmed through domain analysis in Pfam database as described in methods section. Nomenclature of the genes was done based on ascending order of their chromosomal location (Table 4.1). The isoforms of a few genes mentioned in the RGAP database were the same with no difference, and hence are errors of that have been mentioned in Table 4.1. All the *AKR* proteins had only the *AKR* domain (PF03514.14) (Additional file 1: Table S2). The *AKR* domains varied in size from 122aa in *AKR12.1* to 336aa in *AKR27.1* protein. However, most of the

proteins, i.e. 50 percent, had domain size between 250 to 300aa. Apart from AKR12.1, only AKR25.4 had an extremely small domain of 126aa. Only 9 proteins had a domain size greater than 300aa.

Table 4.1: List of putative *OsAKR* genes in rice. The accessions highlighted have same coding and protein sequences but were mentioned as different isoforms in the RGAP database.

Name	Gene Locus ID	Protein	Length	
			CDS (bp)	Protein (aa)
OsAKR1	LOC_Os01g43090	LOC_Os01g43090.1	1035	344
OsAKR2 (AKR2)	LOC_Os01g62860	LOC_Os01g62860.1	936	311
OsAKR3 (AKR1)	LOC_Os01g62870	LOC_Os01g62870.1	936	311
		LOC_Os01g62870.2	714	237
OsAKR4 (AKR3)	LOC_Os01g62880	LOC_Os01g62880.1	915	334
		LOC_Os01g62880.2	639	212
OsAKR5	LOC_Os02g03100	LOC_Os02g03100.1	1149	388
		LOC_Os02g03100.2	1128	375
OsAKR6	LOC_Os02g57240	LOC_Os02g57240.1	987	328
OsAKR7	LOC_Os03g13390	LOC_Os03g13390.2	957	318
OsAKR8	LOC_Os03g41510	LOC_Os03g41510.1/2	1218	405
		LOC_Os03g41510.3	1203	400
		LOC_Os03g41510.4	882	293
		LOC_Os03g41510.5/6/7	858	285
		LOC_Os03g41510.8	804	267
OsAKR9	LOC_Os04g08550	LOC_Os04g08550.1	1014	337
OsAKR10	LOC_Os04g26870	LOC_Os04g26870.1	1053	350
OsAKR11	LOC_Os04g26910	LOC_Os04g26910.1	1056	351
OsAKR12	LOC_Os04g26920	LOC_Os04g26920.1	477	158
		LOC_Os04g26920.2	570	189
		LOC_Os04g26920.3	606	201
OsAKR13	LOC_Os04g27060	LOC_Os04g27060.1	1068	355
OsAKR14	LOC_Os04g37470	LOC_Os04g37470.1	1002	333
OsAKR15	LOC_Os04g37480	LOC_Os04g37480.1	972	323
		LOC_Os04g37480.2	855	284
OsAKR16	LOC_Os04g37490	LOC_Os04g37490.1	1074	357
OsAKR17	LOC_Os05g38230	LOC_Os05g38230.1	1005	334
		LOC_Os05g38230.2	981	326
		LOC_Os05g38230.3	594	197
OsAKR18	LOC_Os05g39690	LOC_Os05g39690.1	957	318
OsAKR19	LOC_Os07g04990	LOC_Os07g04990.1	1131	376
		LOC_Os07g04990.2	921	306
OsAKR20	LOC_Os07g05000	LOC_Os07g05000.1	1134	377
OsAKR21	LOC_Os09g39390	LOC_Os09g39390.1	1158	385
OsAKR22	LOC_Os10g02380	LOC_Os10g02380.1	963	320
OsAKR23	LOC_Os10g02480	LOC_Os10g02480.1	993	330
		LOC_Os10g02480.2	822	273

RESULTS AND DISCUSSION (PART-I)				
OsAKR24	LOC_Os10g02490	LOC_Os10g02490.1/2	1029	342
OsAKR25	LOC_Os10g28320	LOC_Os10g28320.1	1032	343
		LOC_Os10g28320.2/3	750	249
		LOC_Os10g28320.4	462	153
OsAKR26	LOC_Os10g37330	LOC_Os10g37330.1	1128	375
OsAKR27	LOC_Os11g42540	LOC_Os11g42540.1	1077	358
		LOC_Os11g42540.2	1020	339
OsAKR28	LOC_Os12g29760	LOC_Os12g29760.1	951	316
		LOC_Os12g29760.2	711	236

4.2.2.2 Physiochemical Features of AKR proteins and their localization

The AKR proteins were analysed using the ProtParam program (<http://web.expasy.org/protparam/>) to determine various properties of proteins such as isoelectric point (pI), and molecular weight (MW) (Additional file 1: Table S3). Almost 70 percent of the proteins were between 300 to 400 aa in length. The largest protein, AKR8.1/2 is 405 aa in length, with a molecular weight of 45.66 kDa, whereas the smallest protein, AKR25.4 is just 153 aa with a molecular weight of 16.77 kDa. The number of acidic AKR proteins (27) is slightly higher than alkaline ones (19), with AKR26.1 having a pI of 9.42 and AKR22.1 and AKR28.1 having a pI of 5.33. Based on the instability index, the ratio of stable (22) to unstable (24) AKR proteins is almost equal to 1. Almost all the AKR proteins are rich in aliphatic amino acids, with their aliphatic index ranging from 78.46 to 99.92. Except for four amino acids, i.e., AKR9.1, 12.2, 28.1 and 28.2, all the other AKR proteins have low hydropathicity scores (GRAVY < 0), indicating their hydrophilic nature.

The localization of AKR proteins was predicted using different *in silico* tools, pSORT, CELLO, TargetP, ChloroP and BUSCA (Additional file 1: Table S4). The maximum number of AKR proteins were found to be localized in the chloroplast (16) and cytoplasm (14). Ten AKR proteins were also predicted to be localized in the nucleus. Only four AKR proteins were predicted to be mitochondrial. Interestingly, four AKR proteins were predicted to be localized in the outer membrane of the chloroplast, while two were predicted to be extracellular. On careful observation, it was found that different splice forms of the same gene had different localizations. For instance, OsAKR3.1, 4.1 and 28.1 were cytoplasmic whereas their corresponding splice forms AKR3.2, 4.2 and 28.2 were nucleus localized. Similarly, AKR8.1/2/3/4/8 were chloroplastic whereas AKR8.5/6/7 were localized in the nucleus.

4.2.2.3 Chromosomal Localization and Genome Duplication

The *AKR* genes are distributed on all chromosomes except 6 and 8 (Figure 4.5). The 4th chromosome has the maximum number of *AKR* genes (8), followed by 5 genes on the 10th and 4 genes on the 1st chromosome. Further, two *AKR*s are present on chromosomes 2, 3, 5, and 7 each, whereas 1 *AKR* gene is present on each of the chromosomes 9, 11 and 12.

Analysis of the various *AKR* genes using the Plant Duplicate Gene Database revealed the presence of a large number of duplications which was further classified in five categories based on their origin (Additional file 1: Table S6). Only two instances of whole-genome duplication (WGD) were found between *AKR2* & 7 and *AKR7* & 22. On the contrary, 26 instances of dispersed duplication were found. *AKR12* and 13 were found to be proximal duplications. 7 gene pairs were found to be a result of transposed duplications. *AKR2* to 4 and *AKR14* to 16 were tandem duplicates. Similarly, gene pairs *AKR11-2*, 19-20 and 23-24 also originated due to tandem duplications.

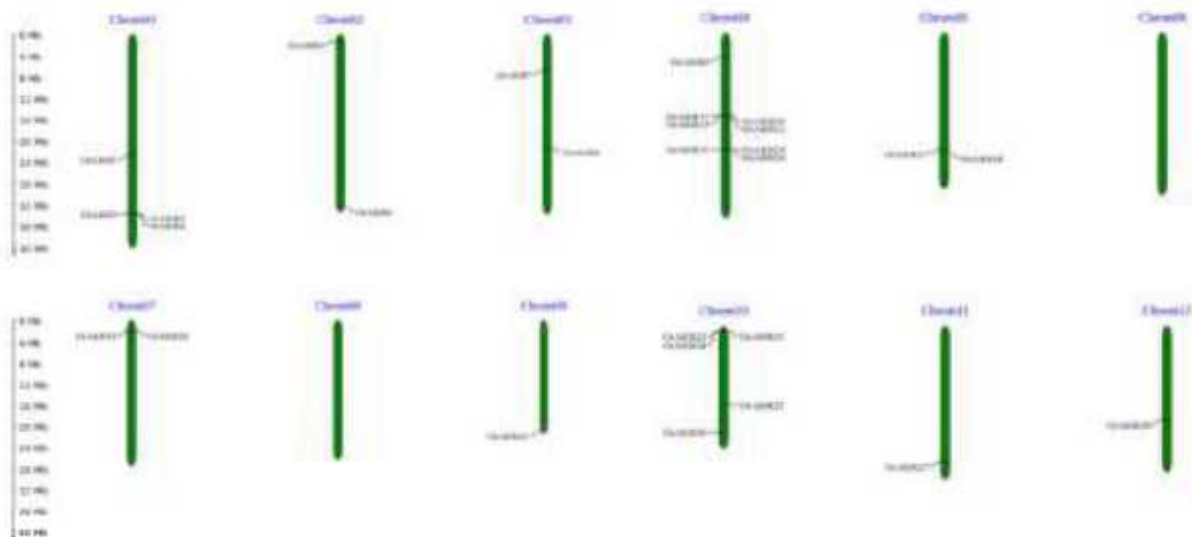


Figure 4.5: Distribution of *OsAKR* gene family members on rice chromosomes.

4.2.2.4 Phylogenetic analysis of *OsAKR* family

To uncover the phylogenetic relationship of the rice *AKR* gene family and enable their classification, a total of 46 *AKR* proteins (Additional file: Table S3) were used to construct a rooted maximum likelihood (ML) tree using MEGA-X (Figure 4.6). Based on phylogenetic analysis and the established nomenclature of the *AKR* superfamily (<https://hosting.med.upenn.edu/akr/nomenclature/>), the *OsAKR* proteins have been classified into nine families (1 - 9) which are further divided into various subfamilies (Fig. 2). Proteins between different families have less than 40% similarity, whereas proteins within the subfamily have more than 60% similarity. Family 1 and 2 have two sub-families,

while 3 and 4 have three subfamilies each. The proteins of families 5 to 9 have similarities of less than 40% between them, and hence, are evolutionary more diverge.

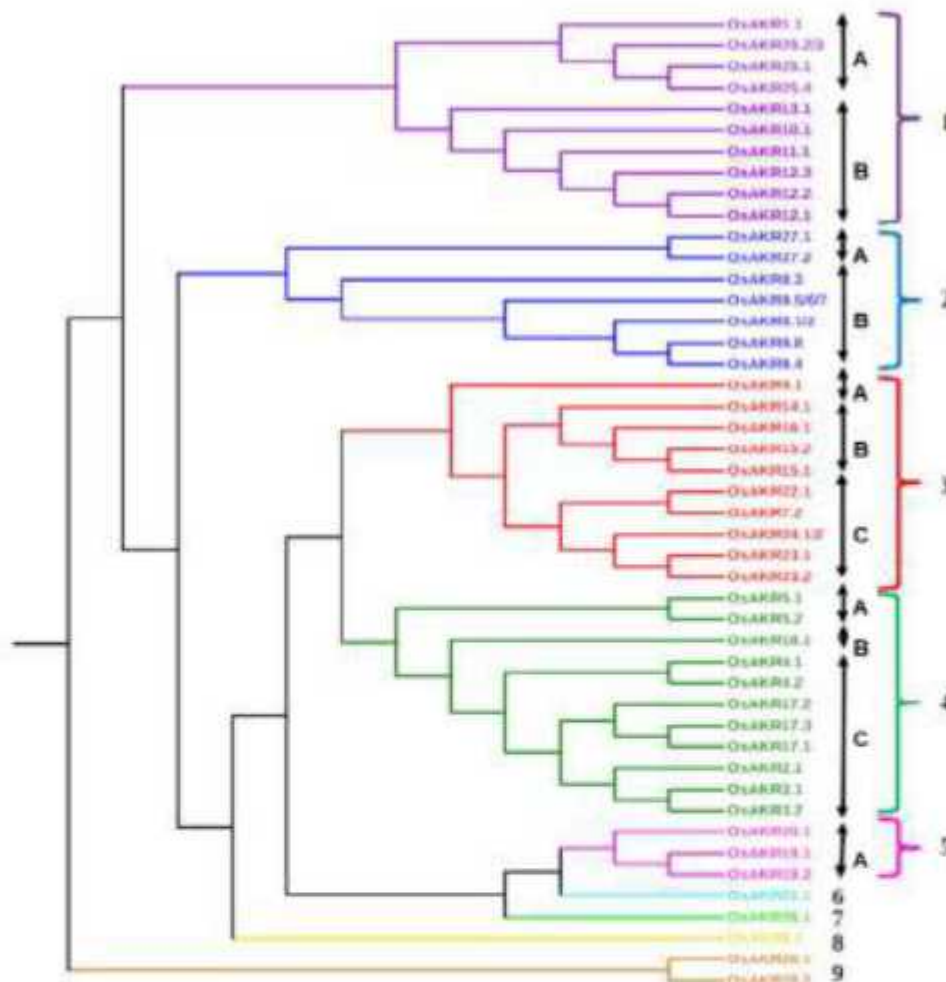


Figure 4.6: Phylogenetic analysis of OsAKR proteins. The maximum likelihood (ML) tree was constructed using MEGA-X with 1000 bootstrap replicates. The OsAKR proteins are divided into nine families (1-9) as per the established system of classification and indicated by different colours.

4.2.2.5 Gene and Protein Structures

AKR gene structure was analysed using the online GSDS server (Hu *et al.*, 2015). The CDS of *OsAKR* genes usually have a large number of exons and introns. Genes with a lesser number of exons have large introns, while genes with more number of exons have smaller introns. A minimum of 2 exons is present in *OsAKR9.1*, while a maximum of 12 exons is present in *OsAKR27.1*. The untranslated regions (UTRs) are usually small in all the genes, except a few where either the 5'- or 3'-UTR is large. *OsAKR14.1* is predicted to lack both 5' and 3' UTRs while *OsAKR5.1* and *5.2* have no 5'-UT region. Further, *OsAKR3.2* has an unusually large 5'-UTR, *OsAKR25.4* has a large 3'-UTR.

The MEME software suite was subsequently, used to identify conserved motifs (1-10 motifs) in the OsAKR proteins to elucidate their structural features (Figure 4.7). The motif arrangement in OsAKR

proteins is in line with their classification into different families. Each group has its characteristic motif arrangement. Most *OsAKR* proteins of family-1 have a 9-4-8-2 motif arrangement in the N-terminal and a 6-5-7 arrangement in the C-terminal. However, family-2 proteins have a characteristic 8-2-3 motif core conserved among themselves. Further, family-3 is characterized by an 8-2-3-1-6 motif core, whereas a 3-1-6-5-7 motif arrangement is found in family-4 proteins. family-5 proteins have a 9-4-8-2-3-6 motif arrangement, except for *OsAKR21.1* and 26.1 proteins. Motif-10 is specific to *OsAKR8* proteins and is located at the N-terminus of proteins. Similarly, family-3 and 4 have a distinct motif-1, which is absent in other *OsAKR* proteins. *OsAKR21.1* has just two motifs i.e. motif 4 and 7, which is an exception as all other *OsAKRs* have multiple motifs.

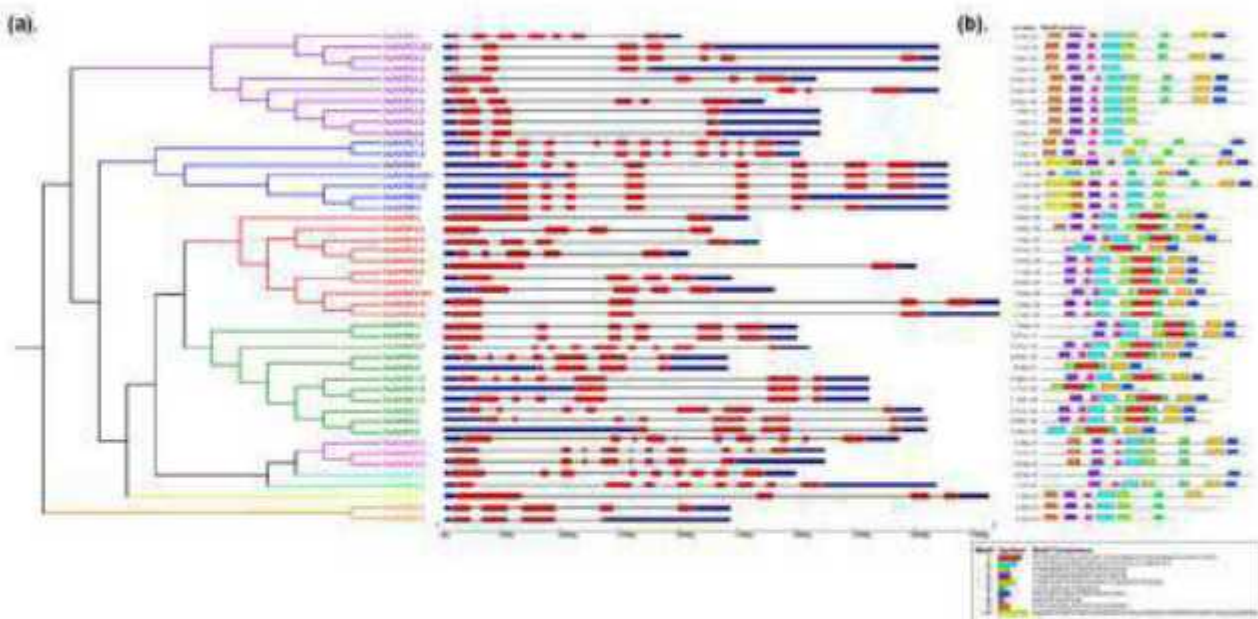


Figure 4.7: Analysis of the gene and protein structure of AKR family members in rice. (a). Gene structure of the 23 *AKR* genes and their isoforms. The UTR regions, exons and introns are represented by blue, red and black colour, respectively. **(b).** Motif distribution in *OsAKR* proteins. The ten different motifs are represented by different colours and their combined *p*-values are also shown on the left side of the figure.

4.2.2.6 Analysis of CREs in the *OsAKR* gene promoters

Promoters (1kb upstream of the transcription start site) of the *OsAKR* genes were scanned for the presence of *cis*-elements and therefore, to understand the mechanisms regulating the expression of *OsAKR* genes (Figure 4.8). For simplification, the various *cis*-elements were divided into 4 broad categories: (a). light-responsive elements, (b). defence and stress-responsive elements, (c). hormone-responsive elements, and (d). growth and development responsive elements (Additional file 1: Table S5). *OsAKR1* was found to be enriched in hormone-responsive elements, like methyl jasmonate, and hence, may be regulated by jasmonic acid. Similarly, *OsAKR7* and 9 are comparatively rich in light-responsive elements, and thus, may be involved in photoregulation.

CpG island prediction in the promoters of *AKR* genes revealed that only a few *AKR* genes, i.e., *AKR2*, *6*, *8*, *12*, *14* and *21*, had only 1 CpG island, while most of the *AKR* gene promoters were devoid of any CpG island. This indicates that epigenetics may have a very limited role in the expression of *AKR* genes.

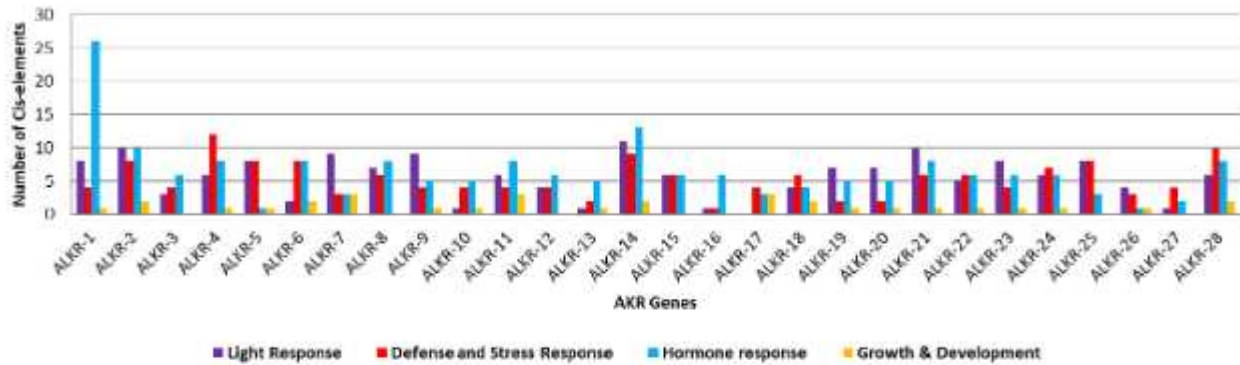


Figure 4.8: Prediction of CREs present in the 1kb promoter region of *OsAKR* genes.

4.2.2.7 Elucidating AKR interactome in rice

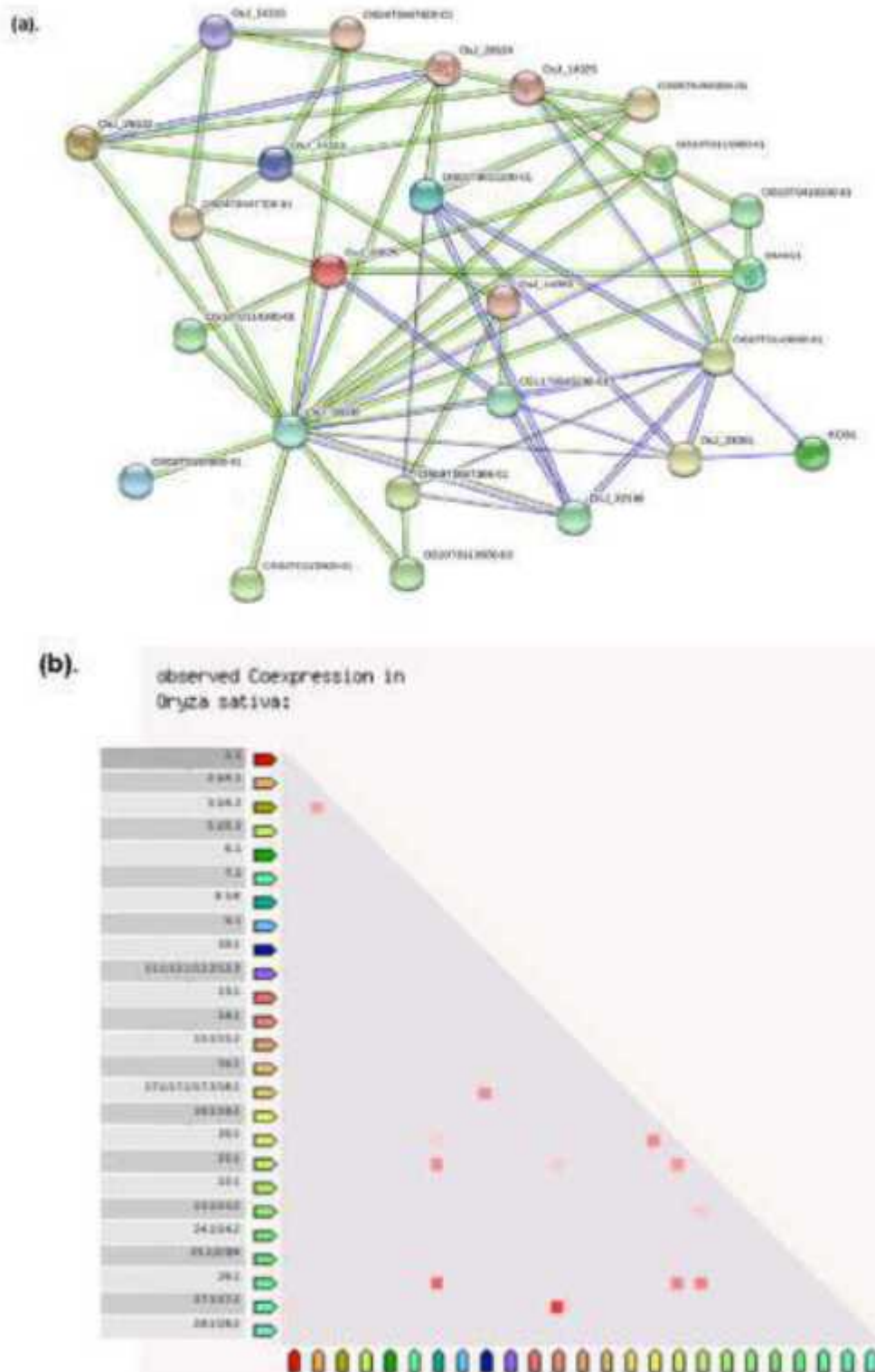
To gain an in-depth understanding of the regulatory mechanisms controlling *OsAKR* proteins, we investigated the interactome made with all the identified *OsAKR* proteins using the STRING database (Figure 4.9a). Each *OsAKR* protein was individually analysed for its interacting members. It was found that most of the members had their unique interactome with a few common and other novel interactors (Additional file: Table S7). For instance, *OsJ_28161* (a putative sorbitol dehydrogenase) reacts with all members of AKR family-3 and 4, except AKR7. Similarly, all members of AKR family 3 & 4, except AKR5 and 7, also interact with phosphoglycerate kinase (*OS10T0442100-00*). Further, NAD(P)H-hydrate epimerase (*OS10T0377800-01*) interacts with all the members of AKR family-1 and 5, besides AKR26 and 27. Surprisingly, few members of AKR family 1 & 4 (AKR3, 4, 9, 14, 15, 16, 23 and 24) were found to interact with *GLYI-8* (*OS05T0295800-01*). Similarly, *GLYI-8* (*OS03T0659300-01*) was found to interact with AKR9, 14, 15 and 23, all belonging to family-3. AKR17 and 18 share all interactors, and hence, it can be assumed that they perform similar functions. Further, there were proteins that had unique interactions, indicating functional diversification. For instance, AKR7 and 22 of family-3 were the only AKRs found to interact with a MATE efflux protein (*OsJ_11501*) and ferredoxin (*OS03T0659200-01*).

The AKR genes are known to have oxidoreductase activity (Mindnich and Penning, 2009). Based on the various interacting patterns, it was found that *OsAKR* enzymes are involved in various processes, including fructose biosynthesis, formation of xylulose-5-phosphate, and glutathione conjugation.

Co-expression analysis using STRINGS database revealed co-expression among *OsAKR* genes (Figure 4.9b). We found a total of 12 incidences of co-expression of *OsAKR* genes. While *OsAKR14*

gene show very strong co-expression with *OsAKR27*, it shows a very weak co-expression with *OsAKR21*. *OsAKR20* is expressed along with other *OsAKR* genes like *OsAKR8*, *19*, *21* and *26*. Similarly, *OsAKR10*, *17* and *18* are also co-expressed.

We also analysed the occurrence of *AKR* genes in other organisms through the STRINGS database (Figure 4.9c). We found that *AKR* genes are found in all organisms from archaea to plants and animals. While there is a very high sequence similarity among the various rice and other plant species, sequence similarity decreases as the evolutionary distance increase.



(c).

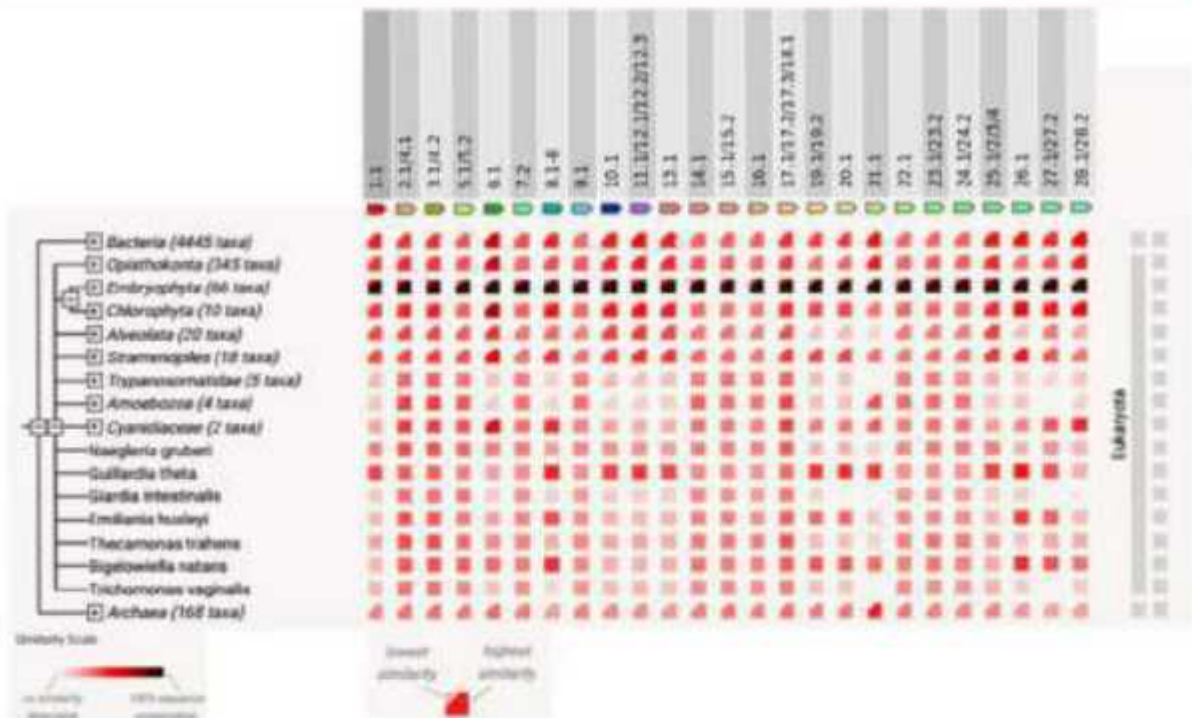


Fig. 4.9: Interactome analysis of OsAKR proteins. (a). OsAKR interacting partners; (b). Coexpression analysis; (c). Gene co-occurrence.

4.2.2.8 Temporal, spatial and stress-based expression profiling of *OsAKR* genes

A publically available Genevestigator database was used to extract the expression profile of rice *AKR* genes. Since AKRs are critical for plant growth and development, we determined the temporal (Figure 4.10a) and spatial expression (Figure 4.10b) profile of these genes. *OsAKR14* data wasn't available in the databases and hence, is not included in our analysis.

During the developmental stages, most of the *OsAKR* genes showed low to moderate expression. While 13 *OsAKR* genes showed no expression, 7 genes showed very high expression during the different developmental stages. During stage-2 and 5 of development, all the *OsAKR* genes showed varying degrees of expression. During stages 1 and 4, several genes showed minimal or no expression. However, *OsAKR1*, 2 and *OsAKR7* genes were highly expressed at developmental stages 6, 4 and 3, respectively. Developmental stage 6 has generally higher expression of *OsAKR* genes compared to other stages.

Analysing expression levels in tissues revealed variations in the transcript profile of AKRs. Most of the tissues showed very low expression of *OsAKR* genes. However, *OsAKR11* showed very high expression in all tissues, except primary cells. While *OsAKR3* and 13 genes have a high expression in callus, *OsAKR21* and 23 showed high expression in shoots. Similarly, *OsAKR5* and 18 have a high expression in inflorescence and primary cells. *OsAKR5* had generally higher expression across all tissues analysed.

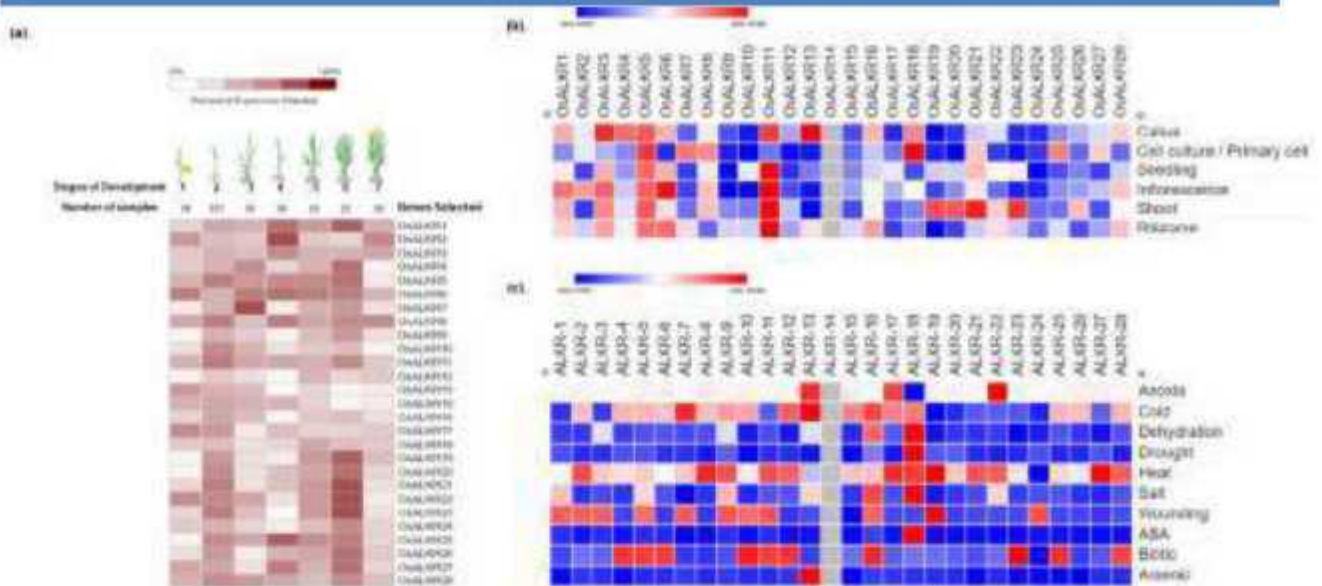


Fig. 4.10: Tissue-specific, developmental and stress-based variations in the expression profile of *OsAKR* genes. Expression was analysed across different (a). developmental stages (b). tissues and, (c). stress conditions. Expression level are based on both the Affymetrix GeneChip and the mRNA-seq datasets ($P < 0.05$) retrieved from Genevestigator database. The transcript level is expressed as Log₂ FC.

4.2.2.9 Effect of *OsAKR* genes under different stress conditions

To investigate the stress-mediated variations in the expression of *OsAKR* genes, we used the publicly available Genevestigator data for extracting the transcript profile of AKR genes. This was followed by validating the expression of a few genes with higher expression by qRT-PCR. Data for *OsAKR14* wasn't available and hence, is excluded from the analysis. Expression profile under different stress conditions, like anoxia, cold, dehydration, drought, heat, salt, wounding, ABA, biotic and arsenic treatments, revealed that not all *OsAKR* genes are stress-inducible (Figure 4.10c). For instance, *OsAKR18* was highly expressed under osmotic stress caused due to dehydration, drought, heat, salt and ABA treatments but wasn't induced in response to anoxia, biotic or arsenic treatment. Cold and heat stress witnessed the expression of many *OsAKR* genes. Few stresses caused the expression of only one or two genes, like *OsAKR13* in cold and *OsAKR18* in drought and under ABA treatment. Similarly, only the *OsAKR13* gene was expressed in arsenic stress, as compared to *OsAKR16* and *18* under dehydration stress. While *OsAKR23* is highly expressed under biotic stress, *OsAKR13* is dominant under anoxia, cold and arsenic stresses.

4.2.2.10 Comparative response of AKR genes to stress in IR64 and Pokkali

Based on the expression profile of *OsAKR* genes under different conditions, six genes (3, 5, 11, 16, 18 & 22) were short-listed for further analysis by qRT-PCR (Table 4.2) under different stress conditions, i.e., salinity, oxidative, heat, drought, cold, heavy metal and biotic stress. We chose two

different rice varieties, salt-sensitive IR64 and salt-tolerant Pokkali. The expression profile was determined at different time points (1, 2, 8, 12 and 24 hrs).

Under salinity stress (Figure 4.11a), in IR64, only the *AKR5* gene showed an initial ~2.5-fold upregulation at 1 hr, whereas the other five *AKR* genes didn't show any drastic change in expression pattern. However, the *AKR3* gene showed ~7.5-fold up-regulation by 24 hrs of salinity stress. In Pokkali, the *AKR* genes didn't show any significant change in expression pattern, except *AKR22* which showed an expression peak (~10-fold induced) at 8 hrs.

Expression of *AKR* genes in IR64 under oxidative stress (Figure 4.11b) showed a slight decrease with time especially at 8 hrs followed by restoration at 24 hrs. Similarly, in Pokkali, there was an initial down-regulation for most of the genes at 1 hr, followed by restoration at 24 hrs. However, *AKR18* and *AKR22* genes peaked with respectively 5- and 10- fold induction at 8 hrs.

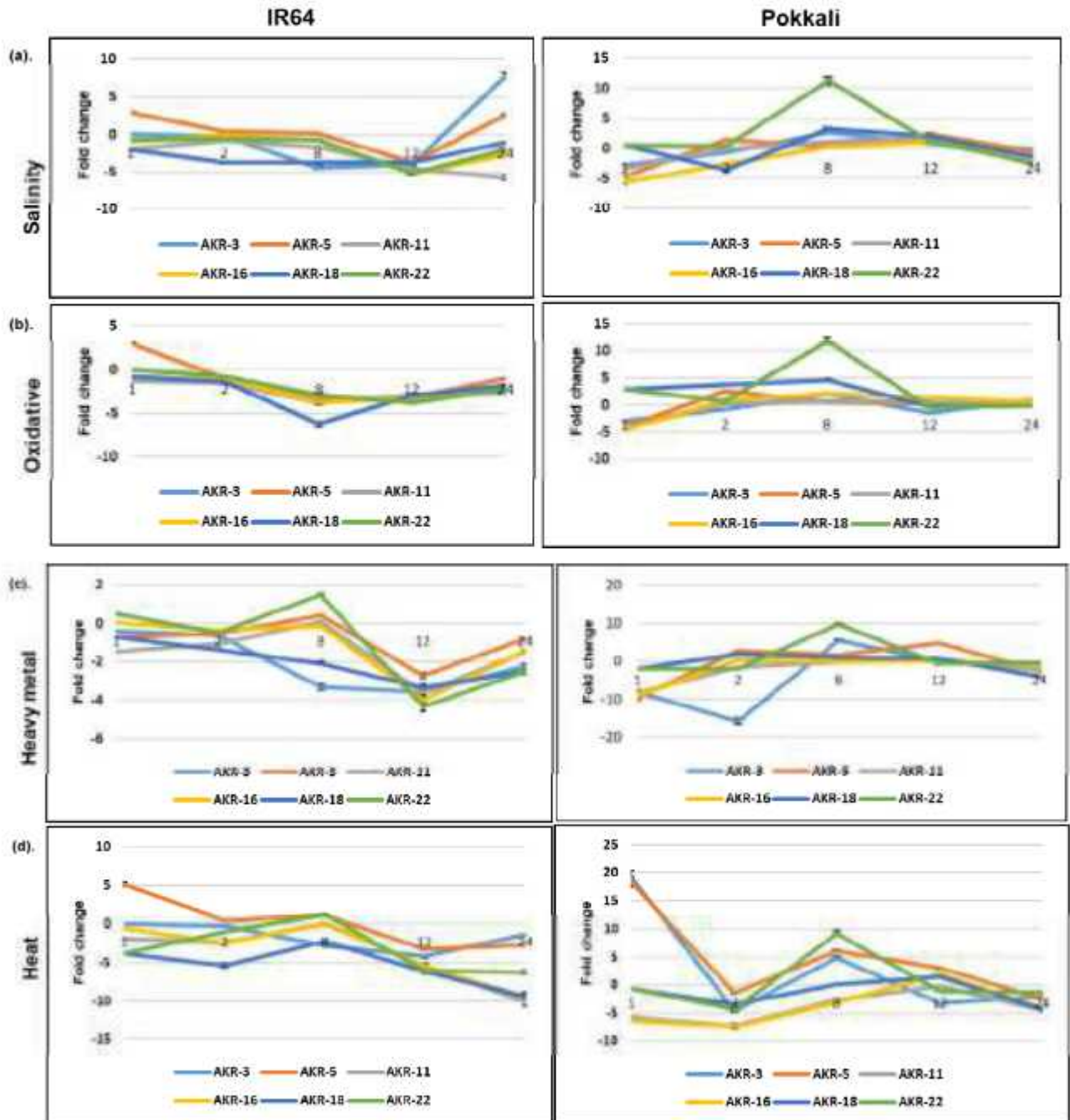
Under heavy metal stress (Figure 4.11c) in IR64, *AKR3* and *AKR18* genes showed gradual down-regulation from 1 hr to 24 hrs. In Pokkali under heavy metal stress, *AKR* genes showed down-regulation at 1 hr followed by stabilization to levels equivalent to untreated control by 24 hrs. Again, *AKR3* and *AKR22* showed an expression peak at 8 hrs, whereas *AKR5* showed an expression peak at 12 hrs. Notably, *AKR22* was also induced in IR64 at 8 hrs.

Under heat stress (Figure 4.11d), *AKR* genes in IR64 showed a general down-regulation. Only the *AKR5* gene was ~5-fold up-regulated (at 1 hr), whereas the other five *AKR* genes were down-regulated. Comparatively, under heat stress, Pokkali generally showed no change in transcript levels of the *AKR* genes at 24 hrs though levels fluctuated at earlier time points. At 1 hr, *AKR3* and *AKR5* were more than 15-fold up-regulated. However, expression levels of *AKR3*, *AKR5* and *AKR22* genes also peaked at 8 hrs of stress in Pokkali.

In IR64, under cold stress (Figure 4.11e), only *AKR5* was ~5-fold up-regulated after 1 hr, whereas other genes were down-regulated. *AKR16* and *22* genes showed slight up-regulation at 8 hrs, whereas *AKR3* was slightly up-regulated by 24 hrs. *AKR* genes in Pokkali under cold stress showed variations in expression patterns. After 8 hrs, *AKR22* expression increased significantly and by 24 hrs, expression of all genes was upregulated.

Under drought stress (Figure 4.11f), *AKR* genes in IR64 showed a general decline in expression over the 24-hr stress, whereas, in Pokkali, there was an increase in expression of a few *AKR* genes, especially at 8 hrs. Specifically, *AKR3* and *22* were up-regulated after 8 hrs of drought stress.

Under simulated biotic stress (Salicylic acid, 100 μ M) (Figure 4.11g), *AKR5*, *11* and *22* genes in IR64 showed upregulation at 8 hrs but showed a decrease after 24 hr period, whereas, in Pokkali, *AKR* genes showed an increase in expression after 24 hrs. In Pokkali, *AKR5* and *AKR16* were significantly up-regulated.



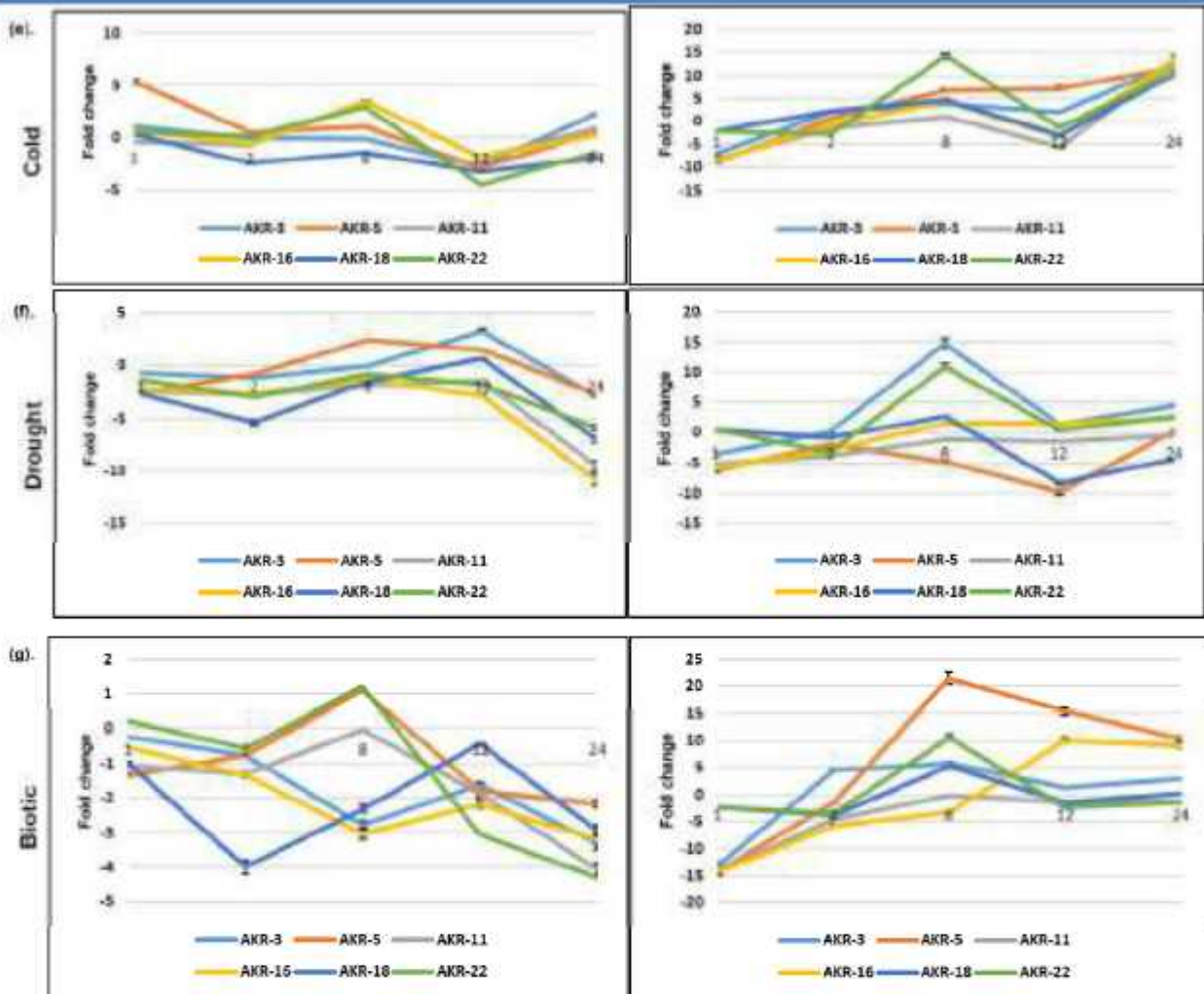


Figure 4.11: Temporal variations in the expression of selected *OsAKR* genes under different stress conditions in two-week old seedlings of IR64 and Pokkali as determined by qRT-PCR. Transcript levels were determined in response to (a). salinity, (b). oxidative, (c). heavy metal, (d). heat, (e). cold, (f). drought, and (g). biotic stress. Untreated samples served as controls. Expression level is expressed as an average of \log_2 of fold change (\log_2 FC) from three replicates ($n=3$).

4.2.3 Screening of drought-tolerant rice varieties using MG as a biomarker

In plants, MG can serve as a possible biomarker of the stress response (Kaur et al. 2014). Assay of MG levels can be utilized to screen for plants with the potential for enhanced stress tolerance. To prove our hypothesis, we screened a population of mutant rice plants generated via gamma radiation. The screened plants were further analysed for drought tolerance. The detailed methodology has been described in the Materials and Methods section (Chapter 3). IR64 rice seeds of M3 mutant lines were procured from Professor Ashwani Pareek's lab at JNU, New Delhi. Mutant rice lines with the lowest MG readings as well as healthy phenotype were selected for further studies (Table 2; Figure 3).

4.2.3.1 Screening of gamma-mutant rice plants

MG levels were measured in a total of 100 mutant rice lines (Figure 4.12). Non-irradiated IR64 (as a sensitive check) and N22 (as a tolerant check) were also analysed. The MG levels in the mutant lines

varied from as low as 1.86 μ M to as high as 29.85 μ M per gram fresh weight. Many mutants had MG levels even lower than the salt-tolerant N22, whereas many lines had MG levels even higher than the salt-sensitive IR64. Based on MG levels and healthy phenotype, a total of nine mutant lines were shortlisted for further analysis (Table 4.2).

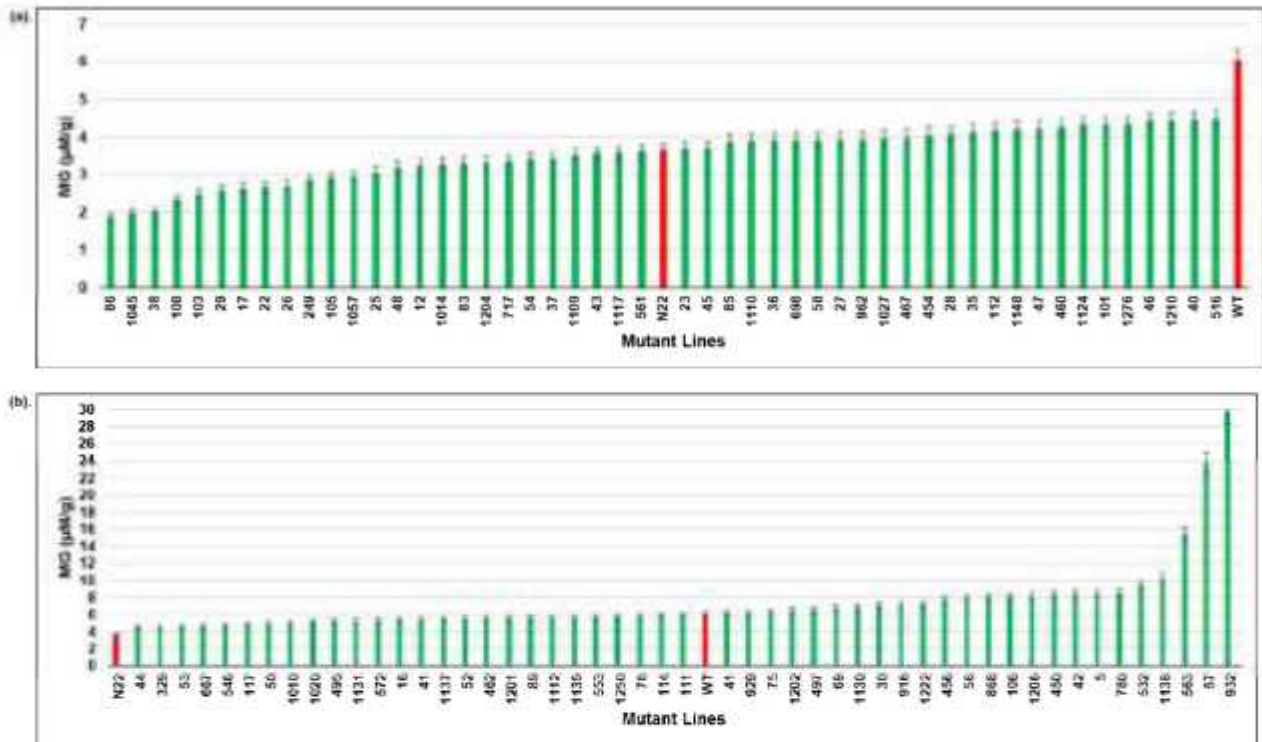


Figure 4.12: Estimation of methylglyoxal in the selected M3 mutant rice lines. All the readings are mean of three biological and two technical replicates.

Table 4.2: Mutant rice lines selected for in-depth analysis for the drought stress imposed at the reproductive stage.

Yield Data of selected Mutants		
S.No.	Selected Mutants	MG (μ M)
1	86	5.41
2	38	6.05
3	108	7.37
4	22	8.58
5	1109	12.05
6	1204	15.94
7	12	16.49
8	48	32.44
9	37	35.3

4.2.3.2 Evaluation of the stress-tolerance ability of gamma-mutant rice

The shortlisted mutant rice lines were subjected to drought stress, with IR64 and N22 as controls. The tissue was harvested just before recovery from drought stress. As expected, the control samples had lower MG levels compared to the drought-stressed samples (Figure 4.13). In some mutant lines, the change in MG levels was negligible like lines 108 and 86, whereas, in other lines, the change was significant. This analysis helped us to identify the mutant lines which are much better adapted for drought stress. Further, it also indicates that the lines with negligible change in MG levels due to drought stress, i.e., 108 and 86, might have a much more efficient MG detoxification mechanism compared to other mutant lines.

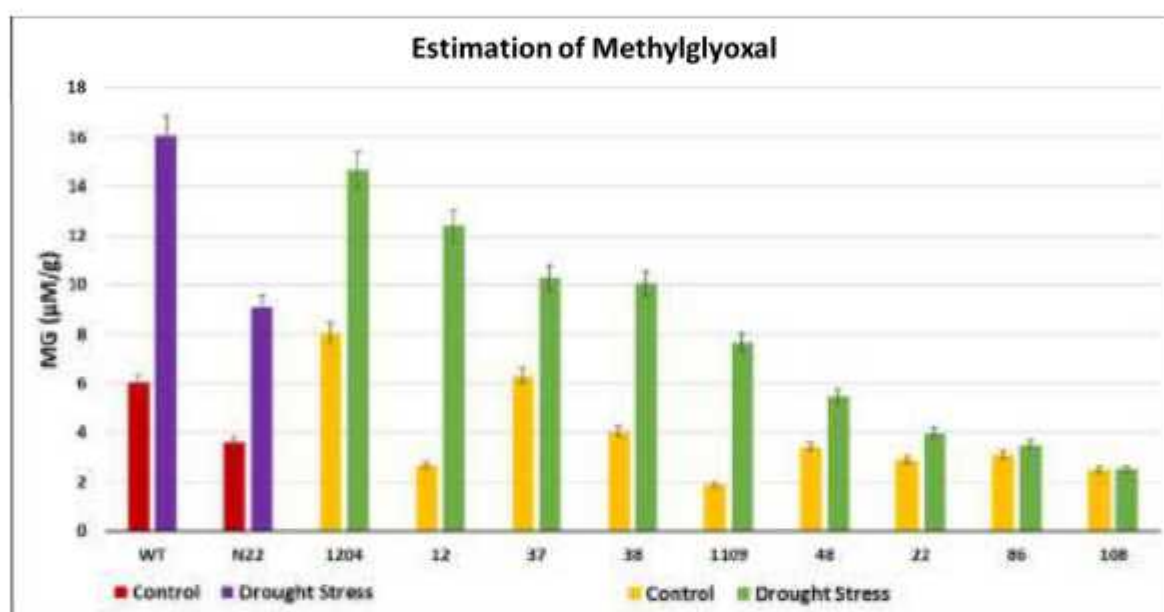


Figure 4.13: Estimation of methylglyoxal in the selected M3 mutant rice lines. All the readings were mean of three biological and two technical replicates.

4.4 Discussion

4.4.1. Spatio-temporal profiling of MG and *Gly* gene expression

Cumulative analysis of MG levels, *Gly* gene expression and their enzyme activities is of significance to understand the MG generation and detoxification process in plants. Through this study, we found leaves generally have higher MG levels in plants in comparison to other tissues, as seen in IR64 and Pokkali genotypes, probably due to the photosynthetic activity of leaves. In particular, flag leaves accumulate maximum MG compared to other leaves in both IR64 and Pokkali. In agreement, leaves displayed a higher accumulation of *Gly* transcripts except for *GlyII-1*, which is shown previously to be root-specific. Likewise, GLY activity is also higher in leaves. Contrary to the general observation of higher MG levels in IR64 than Pokkali across all developmental stages, we instead found Pokkali

to have higher MG, specifically in shoots of the flowering stage plant. In agreement with this observation, *Gly* gene expression and activities were also higher during this stage in Pokkali.

Among the different *Gly* gene isoforms, *GlyI-11.1* and *GlyII-2* were the highly expressed *Gly* genes across different stages of development, both in IR64 and Pokkali. However, *GlyIII* genes showed developmental stage-specific expression. Both IR64 and Pokkali had higher expression of *GlyIII-2* and *GlyIII-5* at the germination stage. At the tillering stage, while IR64 had an increased abundance of *GlyIII-2* transcripts, Pokkali expressed *GlyIII-5* at higher levels among other genes but lower levels when compared to IR64. At the flowering stage, IR64 expressed more of *GlyIII-1* and *GlyIII-2*, while Pokkali showed an abundance of *GlyIII-2*, *GlyIII-4* and *GlyIII-5* compared to other genes. At the maturation stage, IR64 expressed more of *GlyIII-2* and *GlyIII-5*, whereas *GlyIII-4* was the most highly expressed *GlyIII* gene in Pokkali tissues.

Usually higher GLY enzyme activity was observed in tissues with high MG levels and high *Gly* gene expression. Further, IR64 had higher GLYI activity than Pokkali at all developmental stages, except in the flowering stage. The difference between GLYI activity of IR64 and Pokkali tissues is higher when compared to GLYII and GLYIII. In the maturation stage, the GLYIII activity of IR64 and Pokkali seems to be almost similar.

4.4.2. Genome-wide analysis of *AKR* family in rice and expression under different stress conditions

AKR family proteins are found in all organisms, from the archaea to plants and animals, as revealed by gene co-occurrence. However, over the course of evolution, these *AKR* proteins haven't diversified much in terms of acquisition of other domains (Additional file 1: Table S2). Further, they are varied in size and variations are also observed in their physio-chemical properties.

Considering the diverse role of *AKR* proteins in various metabolic reactions, their localization in different cellular compartments is completely justified. The number of *AKR* proteins predicted in the nucleus was much higher compared to mitochondria. This is surprising because mitochondria is the central hub of metabolic reactions occurring in the cell (Aon and Camara, 2015). However, a large number of *AKR* proteins in the nucleus may be necessary to support the redox-dependent activities of the nucleus (Go and Jones, 2010). Some *AKR*s were also present in chloroplast and cytoplasm. Further, extracellular localization of two *AKR*s may be necessary to support the large number of metabolic activities taking place in the apoplastic space (Guerra-Guimarães *et al.*, 2016).

There were just two instances of WGD of *AKR* genes compared to numerous single-gene duplications. This indicates that the expansion of the *AKR* gene family in rice occurred mainly due to single-gene duplications, which are known to play an instrumental role in imparting key environmental adaptations (Freeling, 2009; Panchy *et al.*, 2016; Wang *et al.*, 2012). Hence, it can be

assumed that the AKR family in rice expanded for better adaptation to the evolving environmental conditions.

Gene and protein structure analysis justifies the classification of the OsAKR proteins into nine distinct families. Each family has its characteristic motif arrangement. The presence of certain motifs in proteins of a particular family identify unique characteristics of that protein and indicates functional distinctness. For instance, OsAKR8 protein has a unique motif-10 at the N-terminal which is absent in other OsAKR proteins. Similarly, motif-1 is present only in the proteins of family-3 and 4. The presence of unique motifs in certain OsAKR proteins thus indicates their structural and functional uniqueness.

CREs present in promoters may give important hints regarding the regulatory mechanism behind gene expression (Bilal *et al.*, 2016). However, extra layers of regulation, such as enhancers, silencers and epigenetic modifications, can make identification of such regulatory mechanisms quite difficult through promoter studies. When the CRE profile of *OsAKR* promoters was compared with the expression data, we couldn't identify any concrete relationship. Since there were very few CpG islands in the promoters, it's quite possible that DNA methylation may not be playing a major role in the expression of these genes. Hence, one possible explanation is that there may be CREs beyond the analysed 1kb region, which may be playing a role in the expression of *OsAKR* genes. Further, other epigenetic modifications like histone modifications may be involved in the expression of these genes (Pikaard and Scheid, 2014).

The interactome analysis of OsAKR proteins helped to identify genes with similar interacting partners and which may indicate their functional relatedness. For example, OsAKR17 and 18 have the same set of interactors. Similarly, co-expression analysis correlated with the identified gene duplication events. For instance, *OsAKR19 & 20* which are tandem duplicates, are co-expressed. Similarly, the dispersed duplicates *OsAKR17 & 18* are also co-expressed. Hence, we can conclude that *OsAKR17* and *18* are duplicates that are co-expressed and may be performing the same function.

Careful analysis of publically available data can help to deduce important information of many genes. For instance, *OsAKR19-20* and *OsAKR23-24* are tandemly duplicated genes, and may have the same expression. However, only *OsAKR19-20* have the same expression in tissues, as well as developmental stages but showed different responses under stress conditions. On the other hand, *OsAKR23-24* didn't show any similarity in expression pattern, except developmental stages. Similarly, *OsAKR2* was expressed in inflorescence tissue as well as during heat stress. Since heat stress mainly affects inflorescence causing sterility in pollens, *OsAKR2* can be a probable candidate for heat stress tolerance. Similarly, *OsAKR5, 10, 11* and *16* were expressed under wounding as well as biotic stress. Hence, these can be ideal candidates for tolerance against insects feeding on rice plants. Further, most of the *OsAKR* genes, like 1, 21, 22 and 23, were found to exhibit high expression

during the grain filling stage. At this stage, plants are involved in filling the grains and hence, there is a high fructose metabolism, which is one of the metabolic pathways controlled by AKRs. Callus, inflorescence and shoots in rice are often metabolically very active. While callus is actively involved in utilizing the medium, the inflorescence is metabolically active during the grain filling stage. Similarly, shoots are also involved in the transport of metabolites as well as photosynthesis in rice. Hence, it is quite understandable to have a high expression of *OsAKR* genes.

Expression analysis of selected AKR genes in IR64 and Pokkali gave useful insights regarding their function during stress conditions. In IR64, the *AKR* genes showed significant changes in gene expression up to 12 hrs. On the contrary, in Pokkali, a few *AKR* genes reached their expression peaks by 8 hrs and then subsided. The expression peak at 8 hrs of a few *AKR* genes in Pokkali under most of the stress conditions indicates a regulatory mechanism different from IR64. *OsAKR22* was highly up-regulated in Pokkali and IR64 under different stresses, and hence, can be an important candidate for further investigation. Similarly, *OsAKR5* was upregulated specifically under biotic stress in both varieties. Under drought conditions, *OsAKR3* was particularly up-regulated and hence, may have a drought-specific role.

4.4.3. MG as a biomarker for stress

Identification of stress-tolerant lines from a large population based on morphological parameters is a tedious process. Evaluation at the genomic or proteomic level is a good strategy, but it gives a partial picture. Since the plant system is complex, a minor change can initiate a cascade of changes, and hence, evaluating just one gene or protein to derive a conclusion is not sufficient. MG is a direct consequence of stress in plants, and hence, it gives a much better estimation of the stress-tolerance capacity of plants. Our experiments to use MG estimation assay as a biomarker for stress tolerance helped in the identification of mutant rice lines which are unaffected by drought stress. Hence, this MG estimation assay can be applied as an effective strategy for screening stress-tolerant lines.

4.4 Conclusions

The *Gly* pathway is the major MG detoxification mechanism operative in plants. However, the large number of *Gly* gene isoforms adds to the complexity of the detoxification mechanism. Spatial-temporal profiling of *Gly* genes in IR64 and Pokkali, coupled with MG estimation and GLY enzyme activity, gave useful insights regarding the MG detoxification process. The different isoforms of *GlyI* & *II* genes had different expression patterns, but a few isoforms were expressed at all developmental stages. On the contrary, *GlyIII* isoforms showed different expression under different developmental stages. Further, photosynthetically active leaves were found to have higher MG levels, which may be

the reason for higher GLY activity observed in those leaves. Comparative analysis between IR64 and Pokkali, two contrasting genotypes for stress tolerance revealed IR64 to have higher MG levels compared to Pokkali during all the developmental stages, except flowering. To cope with higher MG levels, IR64 has a higher expression of *Gly* genes in all the tissues and stages, except flowering. In the flowering stage, Pokkali has a higher expression of *Gly* genes, as well as GLY enzyme activity. In a few cases, high MG levels were observed in a few tissues, but the corresponding expression of *Gly* genes and GLY activity was not enough. This was an indication that other detoxification pathways may also be playing a role in MG detoxification. Being functionally diverse in terms of substrate specificity, AKRs are crucial in almost every metabolic pathway, including MG detoxification. Genome-wide identification of AKR genes in rice, followed by expression analysis helped to identify of *OsAKR3*, 5 and 22 as potential AKR genes that can be explored further for stress tolerance. However, the excessive focus on the detoxification of MG has ignored the physiological significance of MG. The use of MG as a biomarker for stress tolerance in plants can assist in the development of newer screening strategies during breeding programs.

Systemic variations in the amino acid profile of rice due to MG using metabolomics

5.1 Background

Systemic view of the rice plants in response to MG is important to understand its full impact on plant physiology. GC-MS is one of the most powerful techniques used for the study of plant metabolomics (Fiehn *et al.*, 2001). It is considered a “gold standard” in metabolomics. It has been used for the analysis of primary metabolites (Schauer *et al.*, 2005; Lisec *et al.*, 2006; Williams *et al.*, 2007; Lytovchenko *et al.*, 2009; Shuman *et al.*, 2011) as well as secondary metabolites (Field and Osbourn, 2008; Klahre *et al.*, 1998).

Plants have evolved different strategies to overcome stress conditions. One effect of these stresses is the disturbance of cellular osmotic potential (Boudsocq and Laurière, 2005), which in turn affects photosynthesis and ultimately, plant growth (Da Silva and Arrabaça, 2004). Accumulation of different osmolytes, such as carbohydrates, betaine, proline and other amino acids, helps the plant to maintain turgor pressure (Da Silva and Arrabaça, 2004). Amino acid metabolism is one of the strategies devised by plants to overcome the adverse effects of different stresses. Amino acids, besides being important building blocks of the plant, are a critical component of the plant metabolic machinery. Their concentrations are known to vary under different stress conditions.

The first impact of different abiotic stresses (salinity, drought and cold) is on the cellular turgor pressure due to disturbance in the plant-water homeostasis (Boudsocq and Lauriere, 2005). This in turn affects the different physiological processes, such as photosynthesis, and reduces plant growth (Da Silva and Arrabaca, 2004). As an adaptive mechanism to overcome these adverse effects, plants accumulate different osmolytes to maintain turgor pressure, such as carbohydrates, proline, betaine and amino acids (Da Silva and Arrabaca, 2004). The amino acids play an important role in maintenance of turgor pressure, including opening and closing of stomata (Kamran, 2009). They are known to increase amination in plants that leads to increase in crop yields without the use of fertilizers (Shekari and Javanmardi, 2017). Amination is known to increase the total content of proteins, fats and carbohydrates in plants, leading to increase in cellulose yield and plant's dry weight (Osuji *et al.*, 2011). Amino acids also play an important role in C₄ metabolism, ammonium fixation and biosynthesis of different metabolites, like flavonoids, catechins, lignins and many others (Fraser and Chapple, 2011).

Many studies have successfully established a direct or indirect correlation between amino acids and stress response. For instance, drought in maize led to a 4-fold increase in proline, a 4.5-fold increase in asparagine, a 2.2-fold increase in serine and glycine, and a 1.5-fold increase in aspartic acid (Mohammadkhani and Heidari, 2008; Shekari and Javanmardi, 2017). Amino acid-derived

metabolites help in stress tolerance through different mechanisms, like providing mechanical strength to plant cells, pollen viability, and protection against UV (Nair and Harris, 2004).

In this study, we have, therefore, determined the metabolite profile of rice seedlings to assess the effect of MG on metabolic functions. Seeds of IR64 rice were grown in petri-plates under different concentrations of MG (0, 2, 4, 6, 8 and 10 mM) for five days (Figure 5.1). The seedlings were harvested after 5 days for further processing. The roots and shoots were harvested separately and processed for GC-MS as described in the Materials and Methods section (Chapter 3).

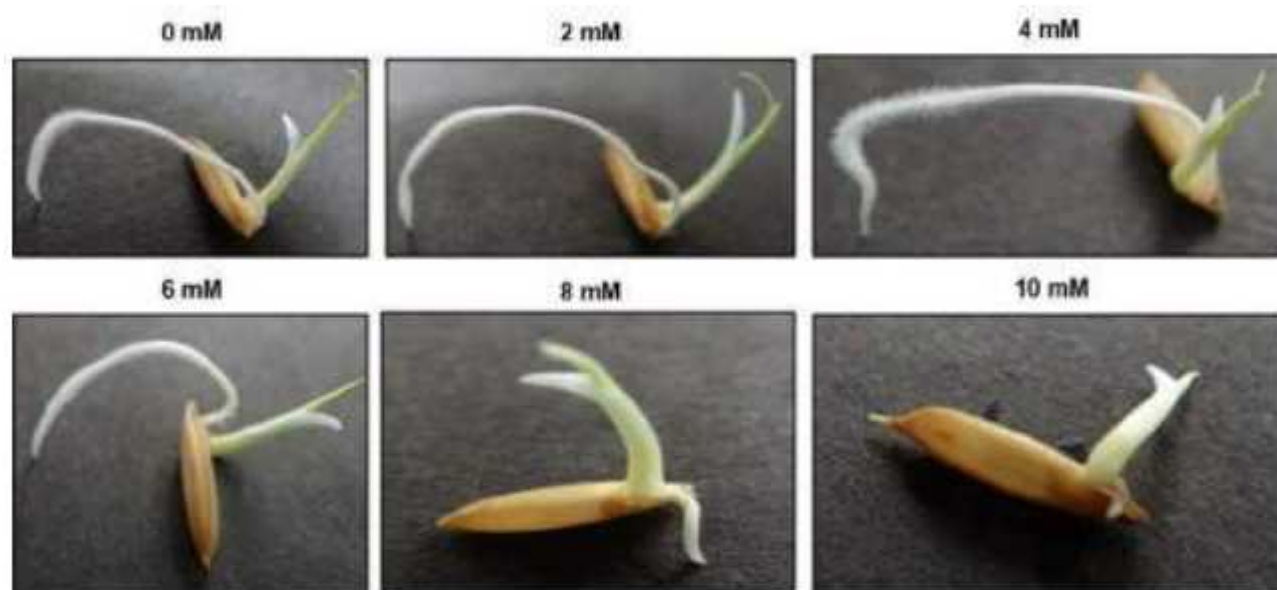


Figure 5.1: Growth of rice seedlings in the presence of a different concentration of MG.

5.2 Results

When grown in the presence of MG, rice seedlings showed stunted growth as the concentration of MG increases. While the shoots show a decrease in length, the roots are drastically affected with almost no growth at 8- and 10-mM MG stress (Also discussed in section 6.2.1). Repeated experiments established 8- and 10-mM MG to be lethal for rice. Hence, GC-MS analysis of both root and shoot metabolites was done separately to assess the effect of MG in these tissues, especially the amino acids. The experiment was repeated twice with two biological and two technical duplicates. A total of 96 chromatograms were obtained in this study. Representative chromatograms are shown in Figure 5.2.

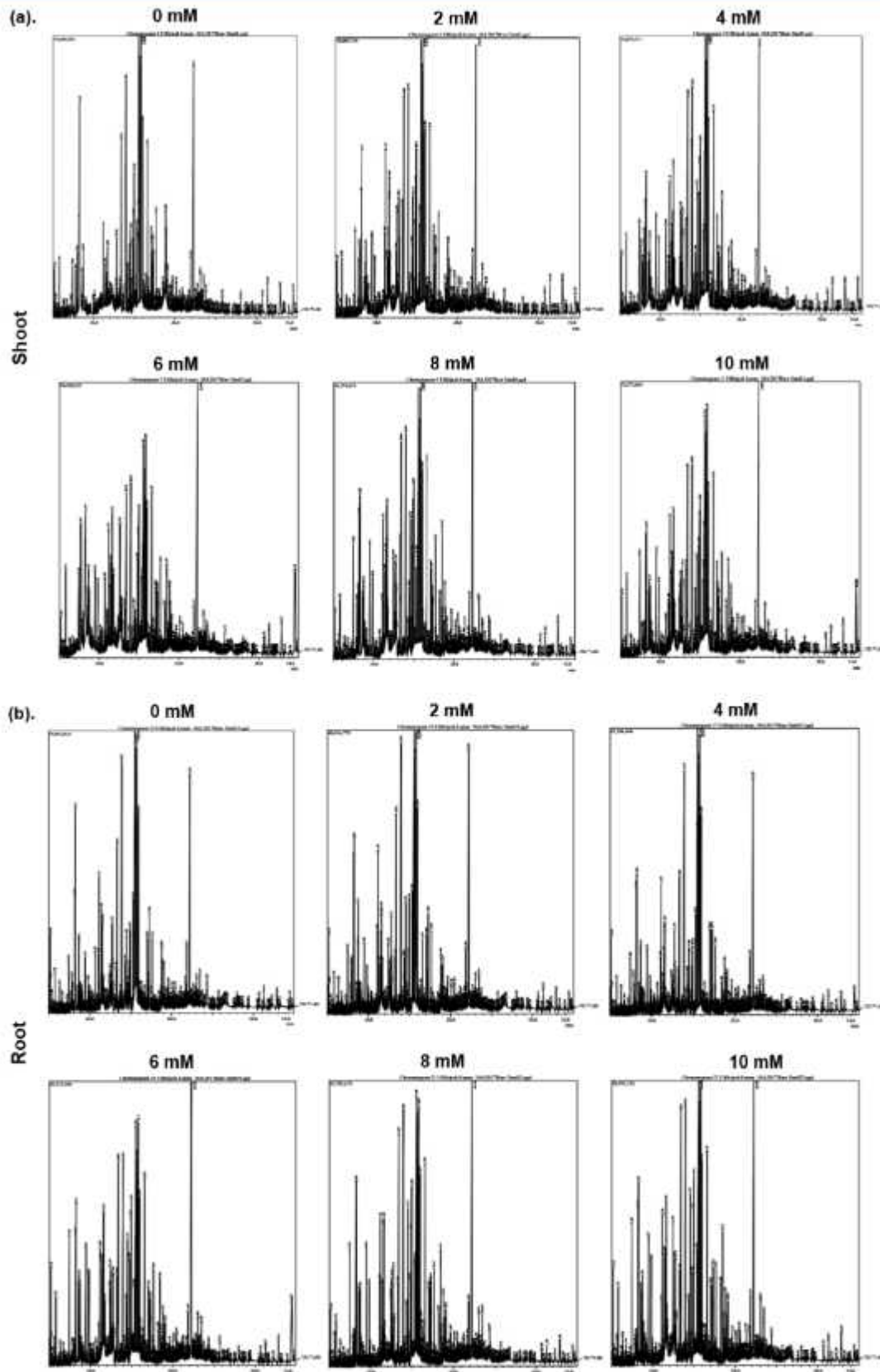
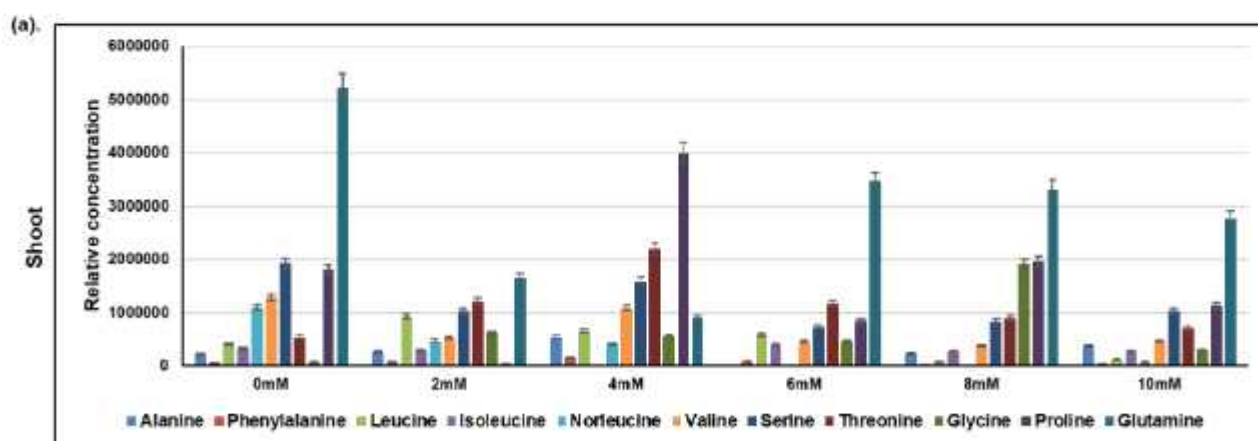


Figure 5.2: Representative GC-MS chromatograms depicting elution profile of metabolites in response to different concentrations of MG in (a). Shoot, and (b). Root.

Metabolites identified through GC-MS can be classified into different chemical groups, and include both primary as well as secondary metabolites. However, we focussed on the metabolites which showed variations in response to MG. Amino acids are one such category of metabolites showing variations with increasing doses of MG (Figure 5.3).

We could not detect alanine in roots of seedlings grown in the presence of MG whereas, in shoots, alanine is present with the highest levels at 4mM MG. Similarly, phenylalanine levels don't increase significantly in shoots in response to different MG concentrations, with the highest levels being observed at 4mM MG concentration. In roots, however, phenylalanine levels are elevated at higher (8-10 mM) MG concentrations. Further, leucine levels are highest at 2mM MG in shoots and 6mM MG in roots, with no specific pattern of fluctuations. On the other hand, isoleucine levels don't seem to vary in response to increasing the dose of MG in shoots. At 4mM MG we could not detect isoleucine in rice seedlings and at 8 and 10mM MG stress, no isoleucine was detected in roots. On the other hand, norleucine levels decrease both in shoots and roots as MG concentration increases and could not be detected in shoots after 4mM MG while roots maintain a relatively higher abundance of norleucine levels. Further, valine levels decrease in shoots with increasing MG concentration but increase at higher MG concentration in roots. Likewise, serine follows a pattern similar to valine in roots. Notably, threonine abundance is highest at 4mM MG in shoots whereas, in roots, it increases with an increase in MG concentration. In contrast, glycine levels decrease in roots with increasing MG concentration. Like threonine, proline levels peak in shoots at 4mM MG and then decline. On the contrary, in roots, proline levels increase to significant levels only at higher MG concentrations of 8 and 10mM. Similarly, glutamine was detected in roots only at 8 and 10mM MG. Whereas in shoots, its levels declined slightly with increasing MG concentration.



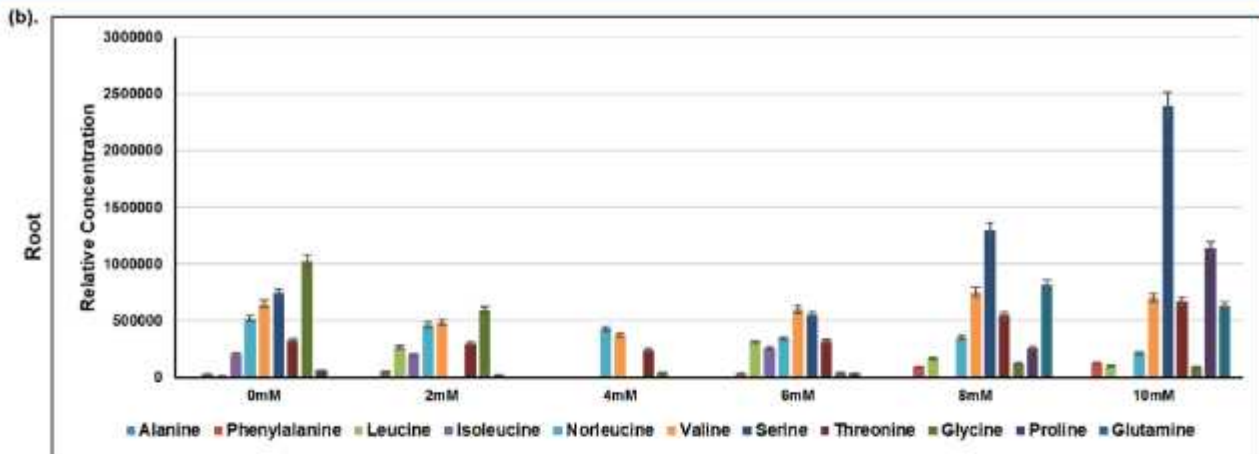


Figure 5.3: Variations in the amino acid profile of rice seedling under different concentrations of MG in (a). Shoots, and (b). Roots.

5.3 Discussion

Amino acids play an important role in plant physiology. Their accumulation as osmolyte is a common response against abiotic stresses in plants (Planchet *et al.*, 2011; Singh *et al.*, 1972). Besides being useful for recovery from stress, they act as a source of carbon and nitrogen for the synthesis of specific enzymes and precursors of various secondary metabolites, such as lignins and flavonoids. They are also involved in pH regulation and ROS detoxification (Rizwan *et al.*, 2016). They even act as regulatory and signalling molecules (Munns, 2002). Studies have established the importance of amino acid accumulation for stress acclimation. However, in some cases, they also act as a sink for excess nitrogen, indicating cellular damage (Munns, 2002).

Proline and glutamate are two well-studied amino acids that are known to play a role in stress tolerance in plants. Proline is involved in the induction of stress tolerance in plants. Its accumulation helps in the maintenance of cellular osmotic potential (Haffani *et al.*, 2017). Besides, it also helps to maintain ion homeostasis under ionic stress (Kamyab *et al.*, 2016). It even upregulates the levels of various antioxidant enzymes (Butt *et al.*, 2016; Nayyar and Walia, 2003). Further, it scavenges ROS and acts as a potential antioxidant itself (Kaul *et al.*, 2008). Glutamate is known to play multiple roles in plant metabolism under normal as well as stress conditions. It is the precursor for many other essential amino acids, such as lysine, arginine and ornithine, and hence, indirectly regulates many metabolic activities (Slocum, 2005). It is involved in the glutamate-glutamine cycle, and thus, regulates glutamine metabolism. It is responsible for cellular osmotic adjustment, especially in guard cells, and thus, regulates stomatal movement. Importantly, it is also involved in the biosynthesis of glutathione which is a vital component of the antioxidant system, and hence, again indirectly controls

the antioxidant defence. It is also involved in chlorophyll and vitamin B9 synthesis, and thus, affects photosynthesis (Hanson and Gregory, 2011).

Since MG is also considered a stressor for plants, we determined changes in amino acid profile in response to MG. For many amino acids, it was observed that their levels increase till 4 or 6mM MG stress, and then decrease. In other cases, the opposite was true, i.e., amino acid levels decrease till 4 or 6mM MG stress, and then increase at higher concentrations of MG stress. This observation correlates with the alteration in plant growth in response to an increasing dose of MG. Up to 6mM MG stress, the plant can tolerate MG and grow well, although growth is a bit stunted compared to control plants. However, at higher MG concentrations of 8 and 10mM MG stress, the plants show retarded shoot growth with almost no roots. The sub-lethal MG stress up to 6mM MG stress may prime the seeds for better stress tolerance and hence, the plants can survive. However, at the lethal dose of MG (8-10mM), the plants get into the defensive mode and try to negate the lethal effects of MG by rerouting their metabolism via alteration of various amino acids. Proline and glutamate are two important amino acids that are known to play important role in stress tolerance in plants. While in shoots, proline peaks at 4mM MG and then declines, the glutamine levels decline till 4mM MG stress, and then again increase. However, in roots, both proline and glutamine have very low levels till 6mM MG and then increase at higher MG concentration. Since roots are more affected than shoots due to MG at higher concentrations, the increase in proline and glutamine indicates that the plants are trying to counter the negative effects of MG through these amino acids. The results concur with the earlier reports that implicate the role of amino acids in stress tolerance. Hence, it can be said that plants try to tolerate MG stress by manipulating the level of different amino acids, especially proline and glutamine.

5.4 Conclusion

Rice seedlings, under different concentrations of MG stress, modulate their amino acids levels depending on the dosage of MG stress. Under sub-lethal stress, most of the amino acids, like phenylalanine, leucine, threonine and proline, peak at either 4 or 6mM MG stress, and then decrease at higher concentrations. However, in roots, amino acids usually increase with increasing MG stress, like in the case of phenylalanine, serine, threonine, proline and glutamine. In the case of norleucine, their levels decrease with increasing concentration of MG. At the lethal dose of MG, amino acids like proline and glutamine increase in roots. These results indicate that rice plants accumulate amino acids to overcome MG stress.

Effect of MG at the cellular and molecular level

6.1 Background

MG is a cellular toxin that causes cell death at higher MG concentrations. However, nothing is known regarding the effect of MG at the cellular and molecular levels. Hence, we decided to explore this aspect of MG in plants.

In this study, we studied the effect of varying concentrations of MG on rice seedlings at the cellular and molecular level using microscopic techniques. To study the effect of MG on cell death, we used the TUNEL (Terminal deoxynucleotidyl transferase-mediated dUTP Nick-End Labeling) assay. Further, we used TEM (Transmission Electron Microscopy) to analyse the effect of MG on the ultrastructure of the cells and even studied the effect of MG on mitochondrial function using fluorescent dyes.

6.2 Results

6.2.1 TEM analysis indicates disruption of cellular organelles at higher MG concentrations

Rice seeds were soaked in water overnight, followed by germination in Petri-plates under different MG concentrations (0, 2, 4, 6, 8, 10 mM). After 5 days of germination, the rice seedlings were harvested for further analysis. The samples were processed for TEM analysis as described in the Materials and Methods section (Chapter 3). However, since the root growth was almost negligible in rice seedlings under 8 and 10mM MG, we couldn't process the same for TEM analysis.

Root and shoot lengths were measured before harvesting the samples (Figure 6.1). It was observed that the growth of rice seedlings became stunted as the MG concentration increased (Figure 6.1a). Both shoots and roots showed a gradual decrease in length as MG concentrations increased (Figure 6.1b). At 6mM MG concentration, the difference between shoot and root length was maximum. After 6mM MG concentration, the effect on overall growth was severe, especially, the roots which were almost absent at higher MG concentrations.

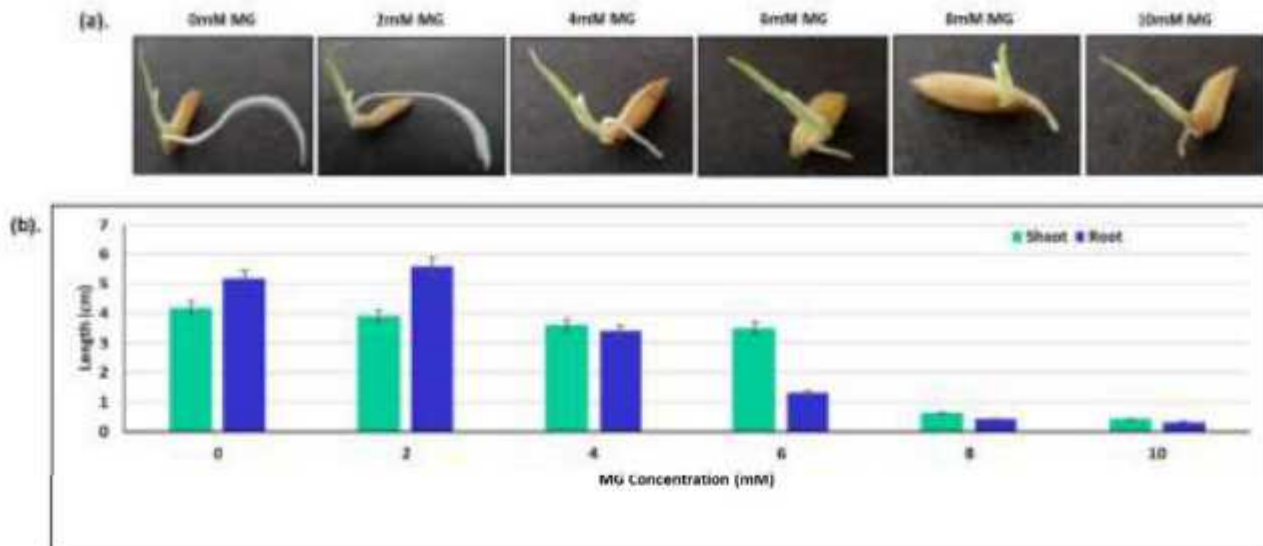


Figure 6.1: Growth of rice seedlings germinated in the presence of MG. (a). Seedlings after 5 days of germination in the presence of different concentrations (0-10 mM) of MG. **(b).** Shoot and root length of rice seedlings after 5 days of growth at different MG concentrations. The results shown are mean of five measurements.

The TEM analysis of these samples gave useful insights regarding the structural integrity of the cell. Roots and shoots were analysed separately. In shoot samples (Figure 6.2), we could observe mitochondria, Golgi bodies, plasma membrane and chloroplast. The mitochondria were structurally intact. Their shape and size appeared to be normal. Similarly, the Golgi bodies and membranes also appear to be normal. However, starch accumulation could be seen in 6mM MG treated chloroplasts, though its morphology was unchanged.

In root samples (Figure 6.3), we could observe mitochondria, Golgi bodies, plasma membrane, vacuoles and cell wall. The mitochondria appeared to be intact at 2mM MG. However, some degradation was visible at 4mM and 6mM MG concentrations. The Golgi bodies also could be seen disintegrating at 6mM MG concentration. Notably, the plasma membrane became curvy at 2mM MG and vacuoles showed signs of disintegration at 4mM MG onwards. However, the cell wall appeared healthy and intact till 4mM MG concentration but appeared disintegrated at 6mM MG.

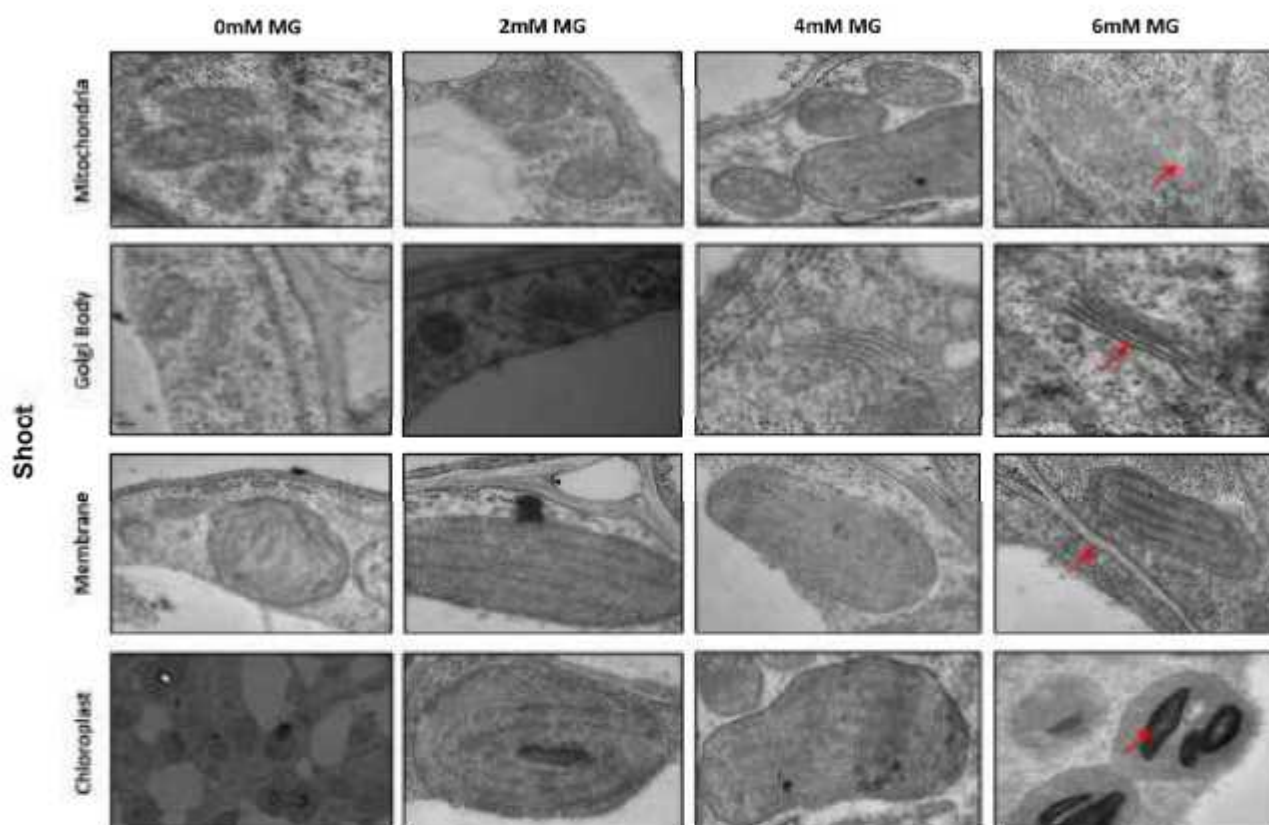


Figure 6.2: Examination of cellular structures in shoots of rice seedlings grown in the presence of MG through TEM. Red arrows show the damage in cellular structures.

6.2.2 MG causes cell death in a dose-dependent manner

In situ TUNEL (Terminal deoxynucleotidyl transferase (TdT) dUTP Nick-End Labelling) assay is used for the detection of *in situ* nuclear fragmentation which is observed during cell death. This assay is based on the incorporation of labelled dUTP at the free 3'-termini of damaged DNA caused due to breaks in inter-nucleosomal DNA. This assay can also be used to detect Programmed Cell Death (PCD).

Rice seeds were soaked for 1 day in the water, and transferred to petri-plates containing different concentrations of MG (0, 2, 4, 6, 8, 10 mM). The petri-plates were covered with aluminium foil and kept in a growth chamber. After 2 days of germination in dark, aluminium foil was removed and the seedlings were allowed to grow in light in the growth chamber. The samples were harvested for analysis after 5 days of growth in light. The samples were processed for TUNEL assay as described in the Materials and Methods section (Chapter 3). Propidium iodide (PI) was used to stain dead cells. Dnase-treated samples were used as a positive control. The slides were prepared and visualized under a confocal microscope (Nikon A1R Microscope).

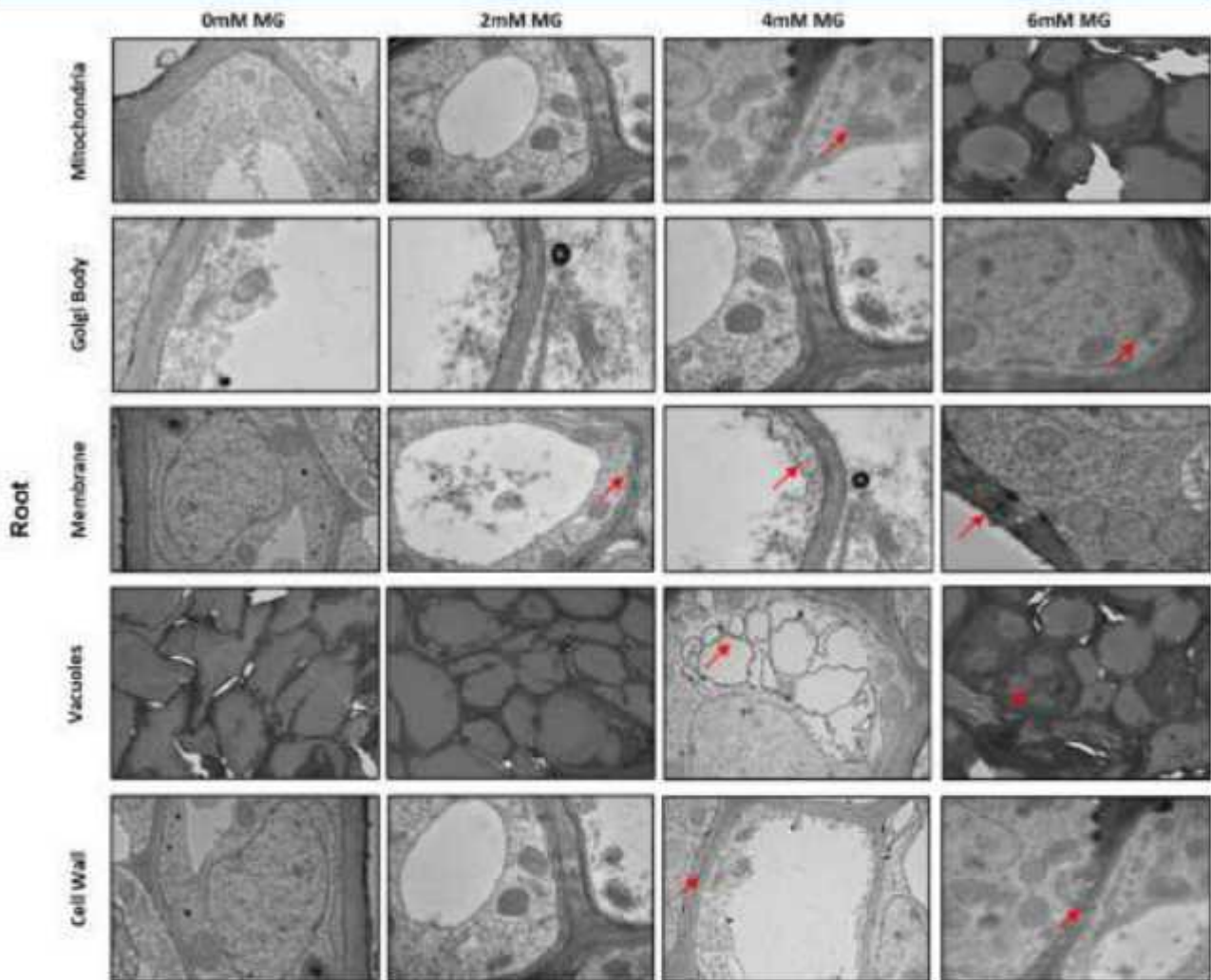


Figure 6.3: Examination of cellular structures in roots of rice seedlings grown in the presence of MG through TEM. Red arrows show the damage in cellular structures.

The positive control sample, i.e., Dnase-treated samples showed a higher number of fluorescein-labelled cells (green), indicating a successful TUNEL reaction (Figure 6.4). In control samples (0 mM MG), almost negligible fluorescence was observed for fluorescein. As the dosage of MG increased from 0 to 10 mM, a corresponding increase in the green fluorescence of fluorescein-labelled cells was observed. Significant fragmentation was visible at 6mM MG and thereafter, as evident from the increasing fluorescence of fluorescein in cells. Similarly, the fluorescence of propidium iodide (PI) which stains dead cells also increased with increasing MG concentration. Figure 6.5 shows DNA damage occurring in seedlings upon exposure to MG. DAPI was also used to stain cells.

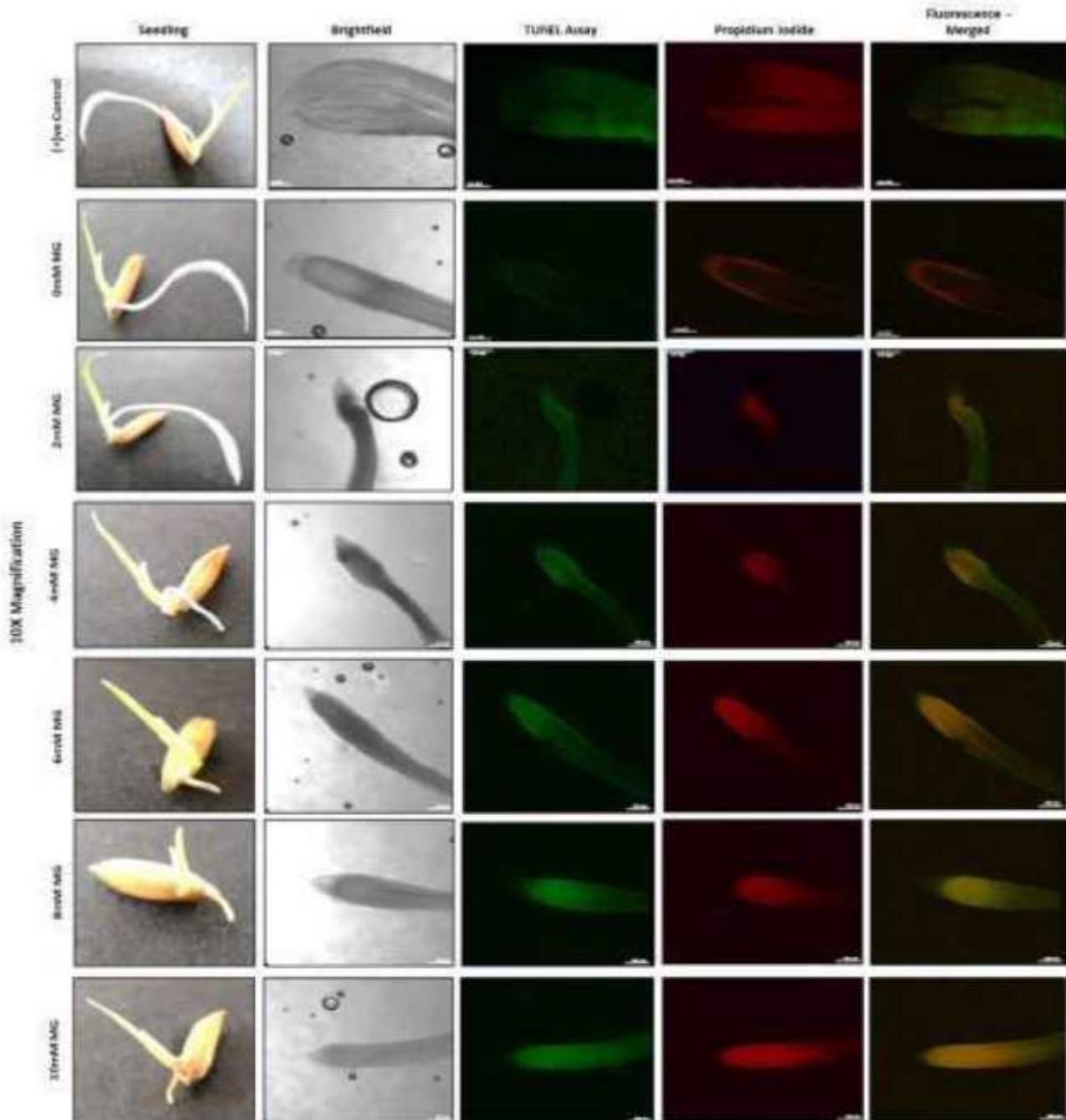


Figure 6.4: Terminal deoxynucleotidyl transferase dUTP Nick-End Labelling (TUNEL) assay to detect *in situ* nuclear fragmentation in rice seedlings grown in different concentrations of MG. Rice seedlings were grown in different concentrations of MG (0, 2, 4, 6, 8, 10 mM) for 5 days in a growth chamber. The samples were fixed with 4% paraformaldehyde (vacuum filtration) followed by permeabilization with 0.1% Triton X-100. Approximately 1 cm of the fixed root tip was cut and used for the TUNEL assay. The reaction was stopped by washing the samples in water and mounted on slides, followed by visualization under a confocal microscope. The green colour represents fluorescein (TUNEL assay), whereas red is for propidium iodide (PI). The image is at 10X magnification.

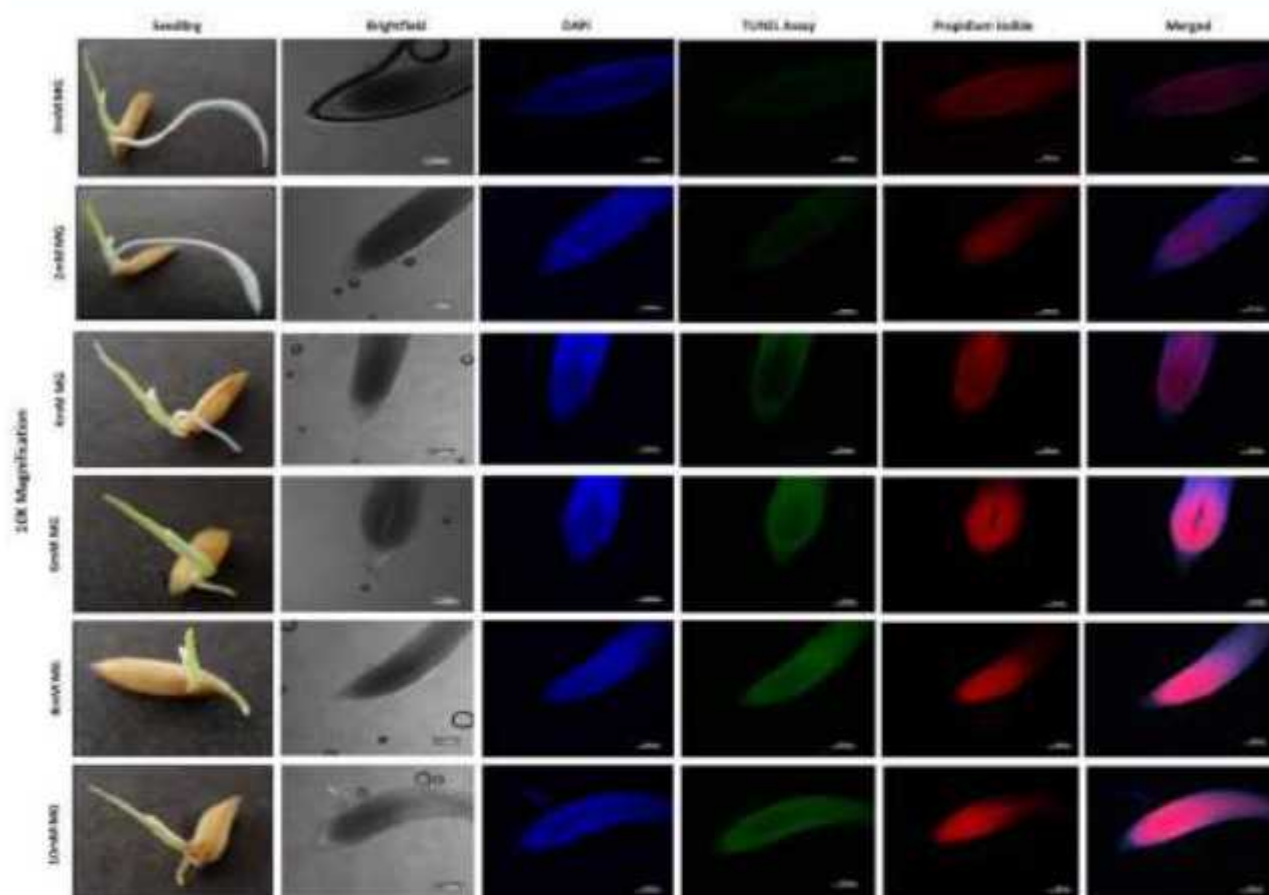


Figure 6.5: Terminal deoxynucleotidyl transferase dUTP Nick-End Labelling (TUNEL) assay to detect *in situ* nuclear fragmentation in rice seedlings grown in the presence of different concentrations of MG at 10X magnification. Rice seedlings were grown in different concentrations of MG (0, 2, 4, 6, 8, 10 mM) for 5 days in a growth chamber. The samples were fixed with 4% paraformaldehyde (vacuum filtration) followed by permeabilization with 0.1% Triton X-100. Approximately 1 cm of the fixed root tip was cut and used for the TUNEL assay. The reaction was stopped by washing the samples in water and mounted on slides, followed by visualization under a confocal microscope. The green colour represents fluorescein (TUNEL assay), whereas red is for propidium iodide (PI) and blue indicates DAPI.

6.2.3 Mitochondrial potential is impaired at higher MG concentrations

TEM analysis indicated that mitochondrial structure is affected due to higher concentrations of MG. Hence, to evaluate the mitochondrial function, rice seedlings grown for 5 days under different MG conditions (0, 2, 4, 6, 8 and 10 mM) were used. Since the roots were most severely affected, we analysed roots using MitoTracker Orange CM-TMROS that stains the cells orange depending on the mitochondrial potential. The non-fluorescent form of the dye is oxidized by molecular oxygen in actively respiring cells, and hence, it can be used in the determination of the oxidative activity of mitochondria. As the membrane potential decreases, mitochondrial oxidative activity decreases and hence, the fluorescence decreases.

The roots of the different samples were stained using the protocol described in the Materials and Methods section (Chapter 3). Figure 6.6 shows the roots stained with MitoTracker Orange. It can be seen that as the MG concentration is increased, fluorescence decreased. This indicates that mitochondrial potential decreases at higher MG concentrations and as a result, the oxidative activity of mitochondria is decreased, affecting its function.

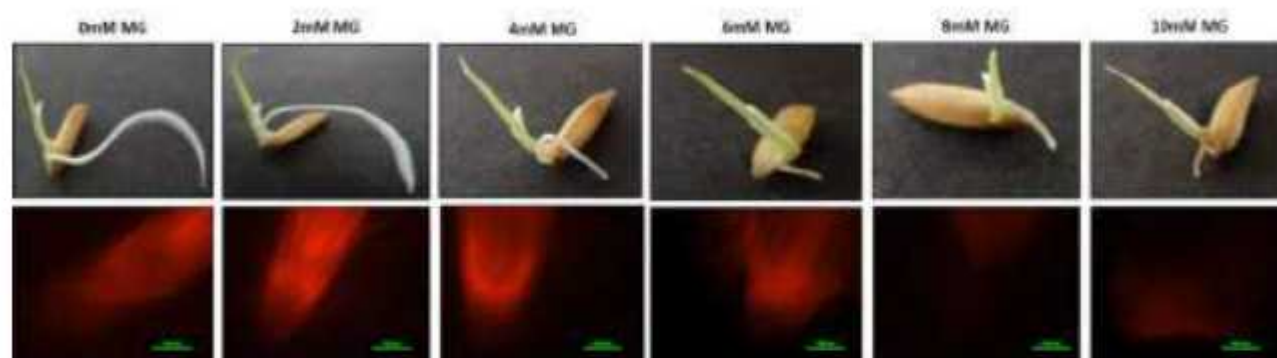


Figure 6.6: MitoTracker Orange staining to analyse mitochondrial activity in roots of rice seeds grown under different concentrations of MG. The MitoTracker Orange dye fluorescence (red) decreases as the MG concentration increases. The image is at 10X magnification.

6.3 Discussion

6.3.1 Cellular organelles show degradation at higher MG concentrations

MG is known to affect the growth of the seedlings. While shoot length showed marginal decline till 6mM MG, root growth was more sensitive to MG decreasing drastically at 6mM MG concentration. At 8 and 10mM MG, both shoot and root growth showed severe retardation. Further, to study how cellular morphology is affected in response to MG, ultrastructures of seedlings grown at different MG concentrations were analysed using TEM. In shoots, most of the organelles, like mitochondria, chloroplast and Golgi bodies appeared to be intact, in terms of shape and size. However, the chloroplasts appeared to be accumulating starch at 6mM MG. Contrary to shoots, cellular organelles in roots appear to be more sensitive to MG.

Mitochondria are the energy centre of the cell. Since exposure to higher MG levels causes growth retardation, energy generation is likely to be affected. Although mitochondria appeared to be intact in shoots, their activity must have been affected in response to MG. In roots, where the effect of MG stress was more profound, the mitochondria did appear to disintegrate. The Golgi bodies also appeared to have lost their normal stacked structure at 6mM MG in roots. Similarly, the plasma membrane also appeared wavy at 2mM MG and onwards. MG is known to react with proteins and lipids, and hence, distortion of membranes of cell organelles is quite possible. The disintegration of

vacuoles into many smaller vacuoles in roots may be an adaptive mechanism to survive the MG stress. MG is known to react with macromolecules, like carbohydrates. The cell wall is composed of cellulose and its interaction with higher concentrations of MG may be responsible for its disintegration.

6.3.2 MG causes cell death in a dose-dependent manner

MG is known to cause cell death at higher concentrations. The reduction in the growth of rice seedlings at higher MG concentrations may be a result of cell death. TUNEL assay confirmed that the DNA was damaged at higher MG concentrations. Besides, PI staining also confirmed cell death at higher MG concentrations.

6.3.3 Mitochondrial activity is affected at higher MG concentrations

Impairment of mitochondrial function exerts a negative impact on plant growth. We thus, hypothesise that the retardation in the growth of rice seedlings at higher MG concentrations may also be due to impairment of mitochondrial function.. Staining with MitoTracker Orange confirmed that there was a decrease in the potential of mitochondria at higher MG concentrations. This probably impaired the functioning of the mitochondrial and caused decrease in the energy level of the plants. Hence, the observed retardation in growth.

6.4 Conclusion

In this chapter, we have studied the effect of MG at the cellular and molecular levels. We found that at higher MG concentrations, the growth of rice seedlings is retarded. The ultrastructure analysis of cell organelles revealed their disintegration, most probably due to the interaction of MG with cellular macromolecules. This impacted the structure of cellular membranes, causing disintegration of organelles, including mitochondria. TUNEL assay and PI staining further confirmed DNA damage and probably, cell death at higher MG concentrations. Staining with MitoTracker Orange dye confirmed that the mitochondrial activity was impacted at higher MG concentrations.

Role of MG as a signal molecule

7.1 Background

MG is known to have adverse effects on the growth and development of all living organisms, including plants. However, some lower organisms like *E. coli* can produce MG via catalytic reactions. MG synthase is one such enzyme that converts DHAP to MG and inorganic phosphate (Ferguson *et al.*, 1998). But again, this enzyme has only been reported in lower organisms so far and any evidence of its presence in higher organisms is still lacking (Kaur *et al.*, 2014). Further, MG is maintained at certain threshold levels in all organisms even under normal conditions (Singla-Pareek *et al.*, 2003; Yadav *et al.*, 2005; Hossain *et al.*, 2009). This raises an important question regarding the physiological relevance of MG in organisms. Several studies have reported the involvement of MG in stress signalling in plants and yeast but this area warrants detailed investigations to elucidate the precise role of MG and its signalling pathway in plants. Therefore, here we have investigated various aspects of MG signalling. We analysed the promoters of *Gly* genes to identify their broad putative role in plant physiology. Further, we identified putative MGREs from the promoters of *GLY* genes as well as promoters of genes expressed under MG treatment.

7.2 Results

7.2.1 *Gly* genes are differentially regulated as evident from the presence of Cis Response Elements (CREs) responsive to diverse factors

Promoter sequences (1000 bp upstream) of *Gly* genes in rice were extracted from the RGAP database (Kawahara *et al.*, 2013) and the CREs present in them were identified using the PlantCare database (Rombauts *et al.*, 1999; Lescot *et al.*, 2002) as described in the Materials and Methods section (Chapter 3).

The CREs identified in the promoter sequences of all the *Gly* genes were classified into four broad functional categories: (a). light response (A-box, Box 4, GATA-motif, G-box, GC-motif, GT1-motif, Sp1, TCCC-motif, ACE, and I-box); (b). defense and stress response (as-1, LTR, STRE, DRE Core, WRE3, ABRE3a, ABRE4, ATCT-motif, W-box, and GARE-motif); (c). hormone response (ABRE, ARE, CGTCA-motif, MYC, TCA-element, TGACG-motif, and ERE); and, (d). growth & development (CAT-box, O2-site, and AC-I) (Figure 7.1). Among the *GlyI* genes, *OsGlyI-3* has a higher number of CREs compared to other *GlyI* genes. *OsGlyI-3* has the highest number of CREs related to hormone response. In agreement, we found that only *OsGlyI-3* is expressed under ABA hormone treatment (Figure 7.1a). Further, only five *OsGlyI* (1, 8, 9, 10, 11) genes have CREs related

to growth & development. Among *OsGlyII* genes, *OsGlyII-2* has a lesser number of CREs compared to other *GlyII* genes. Also, only *OsGlyII-1* has CREs related to growth and development. *OsGlyII-1* has maximum expression under stress conditions as well as under ABA treatment, which corresponds well with the expression data (Figure 7.1b). Among *OsGlyIII* genes, *OsGlyIII-5* has the maximum number of defence & stress and hormone-related CREs (Figure 7.1c). Only *OsGlyIII-4* & 5 have a few CREs responsible for growth & development. Overall observation indicates that the general pattern of CREs in all the *OsGly* genes is similar.

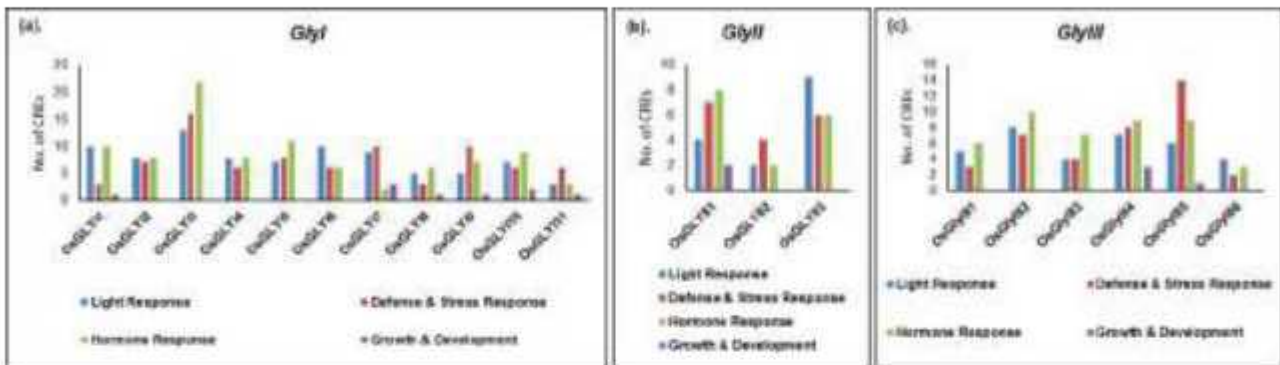


Figure 7.1: The different CREs identified in the promoter sequences of the Gly genes (a). *GlyI*; (b). *GlyII*; and (c). *GlyIII*.

7.2.2 Promoter analysis of *Gly* genes to predict possible MG-response elements (MGREs)

The putative promoter sequences of rice Gly genes were analysed for the prediction of probable MGREs. For this, 1 kb upstream sequences were analysed using the MEME software suite as described in the Materials and Methods section (Chapter 3).

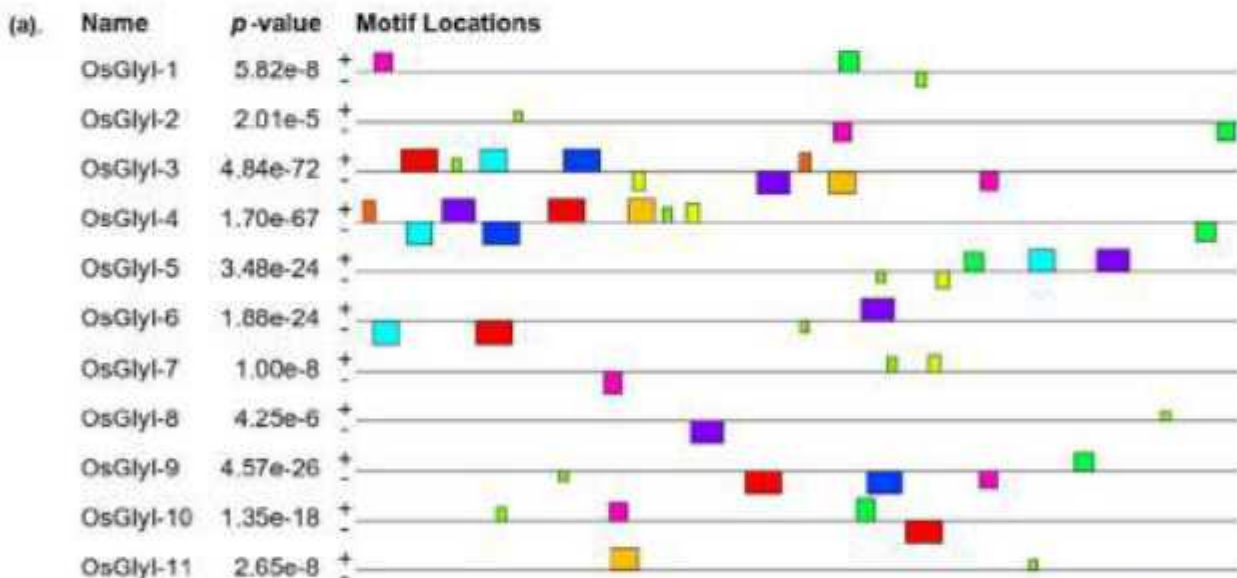
The ten predicted MGREs for all the *OsGly* genes were depicted on the promoters and the length & frequency of the predicted MGREs is given with the expected *p*-value (Figure 7.2). Among the MGREs for *OsGlyI* genes, motifs – 1, 3 and 10 are significant in length and frequency. Similarly, *OsGlyII* has motifs – 2, 3 and 10, while *OsGlyIII* has motifs - 1, 2, 4, 5 & 7.

Similar to MGREs predicted using *Gly* promoters, MGREs were previously predicted using the promoter of genes expressed in response to MG. Genes that were either up-or down-regulated 10-fold or more under MG stress were selected and searched for these predicted MGREs in them (Figure 7.3) (Kaur *et al.*, 2015).

To experimentally validate the existence of MGREs identified in our study, the promoters of genes having the predicted MGREs (Figure 7.4) were PCR amplified from IR64 genomic DNA and cloned in the promoterless vector pCAMBIA1391Z (Figure 7.5). The pCAMBIA-promoter constructs were then transformed in *Agrobacterium* strain LBA4404 and colonies so obtained were screened for

positives through PCR using promoter-specific primers, as well as *vir* gene primers. The band was observed with *vir* gene primers in colonies of all genes screened (Figure 7.6a). Likewise, the band corresponding to the respective amplicon size was obtained when screened using promoter-specific primers (Figure 7.6b). Following confirmation of positive clones, *Agrobacterium* cells harbouring these constructs were used for the transformation of rice calli as per the method standardized in our lab (Sahoo *et al.*, 2011). Calli derived from mature rice seeds were initially agro-infected, followed by three rounds of hygromycin antibiotic selection. Transformed calli were then regenerated to yield rice plantlets (Figure 7.7). The plants were further screened for transgene insertion by PCR using hygromycin and promoter-specific primers (Figure 7.8). However, the transgenic plants yielded no seeds may be due to some unknown issues and hence, no further studies could be carried out for the *in vivo* validation of MGREs.

Using *in silico* approaches, we tried to determine if any specific transcription factor binds to our predicted MGREs. Using the PlantRegMap website (Jin *et al.*, 2015; Jin *et al.*, 2017; Tian *et al.*, 2020), we analysed the predicted MGREs for probable TFs which bind to it. We got only one TF, i.e., ERF which binds to two of our predicted MGREs (Figure 7.9). ERF proteins are DNA binding proteins that have important functions in the transcriptional regulation of a variety of biological processes, including stress tolerance.



Motif	Symbol	Motif Consensus
1.	[Red]	GRKCVGTSCASBTCHNSGCSCTCCRHNTCCCYTCTKCCG
2.	[Cyan]	CBRCCTCTCCCTCTCCCTCANYHTCCCG
3.	[Green]	CCCMACCCGC
4.	[Purple]	GGASRVGABRANGRSTGGCCGCKMGGCTGTMAC
5.	[Orange]	STCCNSNYCKBCWCZGMRGRCRANRGGCC
6.	[Yellow]	AAATCSANYRCTGANACAGAT
7.	[Blue]	CTCCCCZSYGRRVMSGGPCDCYRFBSSGGYCRGSCASACRS
8.	[Pink]	AAKRNATTAARCHATEKCCAC
9.	[Brown]	CCCKYGCOCGGC
10.	[Light Green]	GTGATCCCTCCG

S. No.	P-value	Length	Repeats
1	2.70E+00	5	41
2	1.10E+00	4	29
3	1.40E+00	11	10
4	2.70E+00	5	37
5	8.40E+00	3	30
6	1.90E+04	4	21
7	5.00E+04	3	41
8	4.10E+04	6	21
9	1.70E+00	2	13
10	2.40E+00	4	14

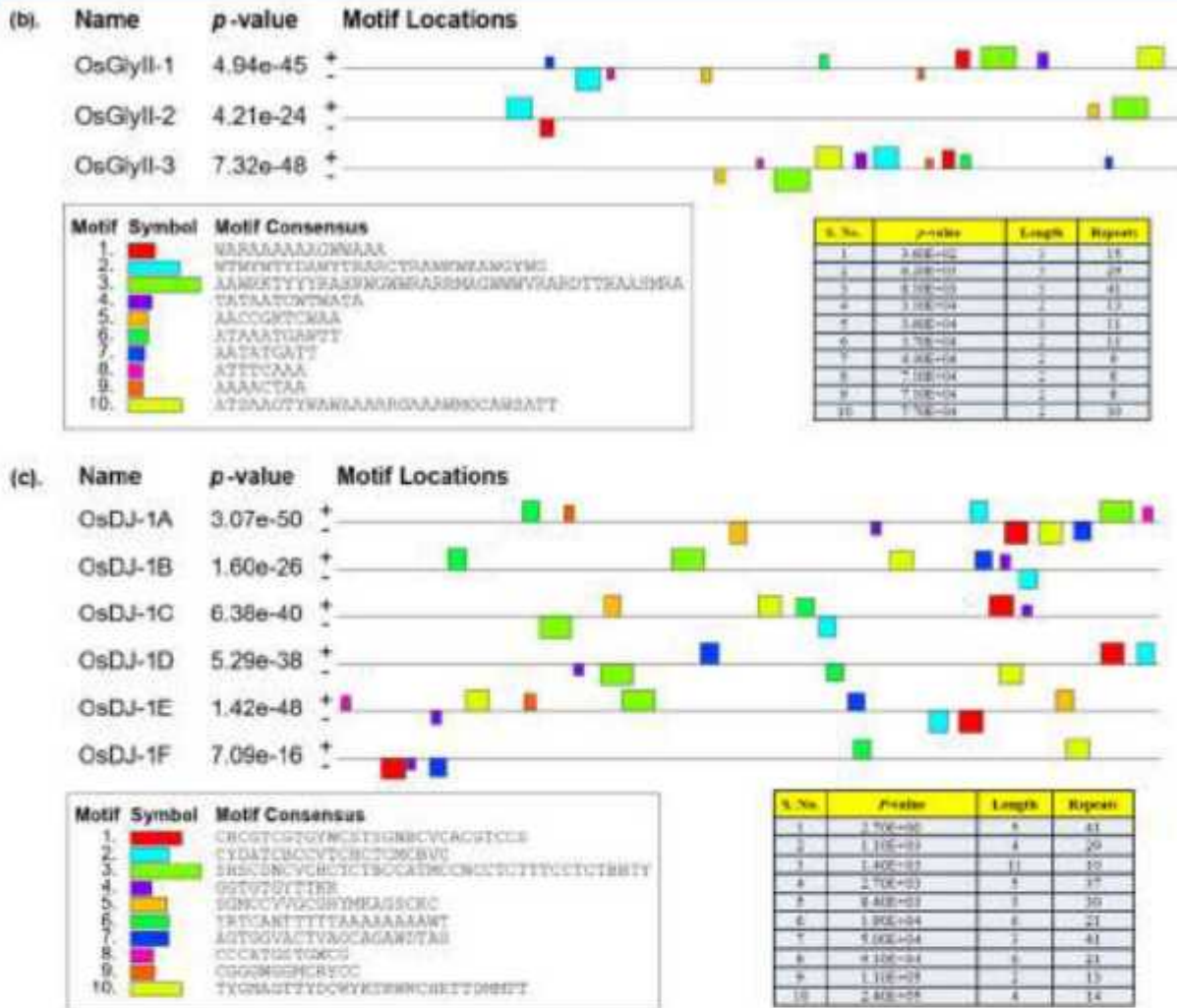


Figure 7.2: Prediction of putative MG-responsive elements (MGREs) by detecting conserved motifs in the promoter region of (a). *GlyI*; (b). *GlyII*; (c). *GlyIII* genes from rice. The 1 kb upstream region of MG-responsive genes was used for the identification of motifs using MEME software.

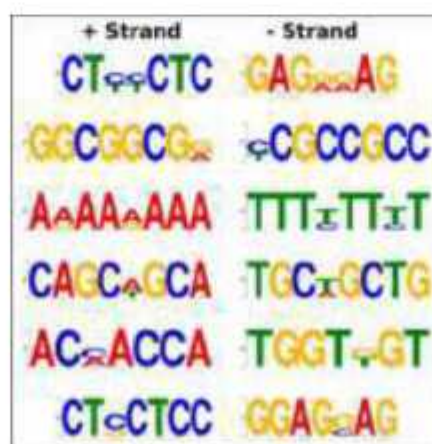


Figure 7.3 Putative MG-responsive elements (MGREs) predicted in the promoter region of MG-responsive genes by (Kaur *et al.*, 2015). The genes with more than 10-fold alteration in expression upon MG treatment were selected for analysis. The 1 kb upstream region of MG-responsive genes was used for the identification of motifs using MEME software.

>Histone like (LOC_Os01g70890)
 CAAGGCACTCTCTATGTCAAAGAAAAAATCAAAATTTATATAGGAATCTAGACATAAATATGTAGAAAAGTIGATTATTTTGGGAT
 GGACAAAGTAAATAAAAACATATACGGTGCATTTTGTAGCGACATATATACCTCGGTTTGTGTATCAATCAACATACAAGCATATTA
 TGTAGTTAAATTAATCTGA

>GATA (LOC_Os03g05160)
 ACATACAAATATTTATGGGAGGGAGTTGGTTGTGTTGGTTTCATTATTGCATCTATTAGGAATCTAAGGAATAAATAGTTATACAA
 TTCGTTAGAAAATTAAGCGGATATTTGTAGTTTTTAAGTTACGGGTAAAGGAGTAATATTGTTTGT

>MYB (LOC_Os03g20090)
 CCAAGCCAGCCACTGCAACAAAAACACAGAAAAAAGGGAGAAGGATCGAGTGCAGGAGATCACAAATTAGCGCGGAGAAATGG
 GCCGGGGATATTGTGTACTAGTGTACCACTCGGCTCGACACTGGAGCATGCAAAGTGTAGT

>B3 (LOC_Os04g49230)
 TCTTTGTACGTCCTTAAGAATTTATGGGCGACGACGACCTCAGTTCCTACTTCTACTGTTGCTGTAGGAGTAGGAACAGTGTGTTG
 CGTAGATGTACACATCTCTTGCAGAAAGTAGTCTTGGTGTCTATTCTGTTCTTGTGCTTTTGTGTACGAATCAGTAAAGAGAGAGCAT
 GTGTTGTGCGTTTA

>Hairpin (LOC_Os09g09460)
 TCTCAGAACTCATCTCAGGGTTAGAAAAAAGGGTTAATTTGATCTATGTCATTGCAAAATGTGCTTATTCAGACAAATATTATGCAAT
 TCGTCTATTTCGTAAACCGTGTCACTCAAATTTTACAAAATTGGAAACCATAATCATCGCCGTCACGTTTTCATCTTCCCTCCATCTGCAT
 GTGGGACCCATCCGTCATATTCATCTCTCCAACTACAATGTCACCACCTCAGCTACGCGCATGACGGCGGCTAGCTGTCTGTCGCTT
 CTTG

Figure 7.4: Upstream sequences of genes highly induced in response to MG and possessing the predicted MGREs. Nucleotides highlighted in yellow are the predicted MGREs and were included in the amplified promoter construct.

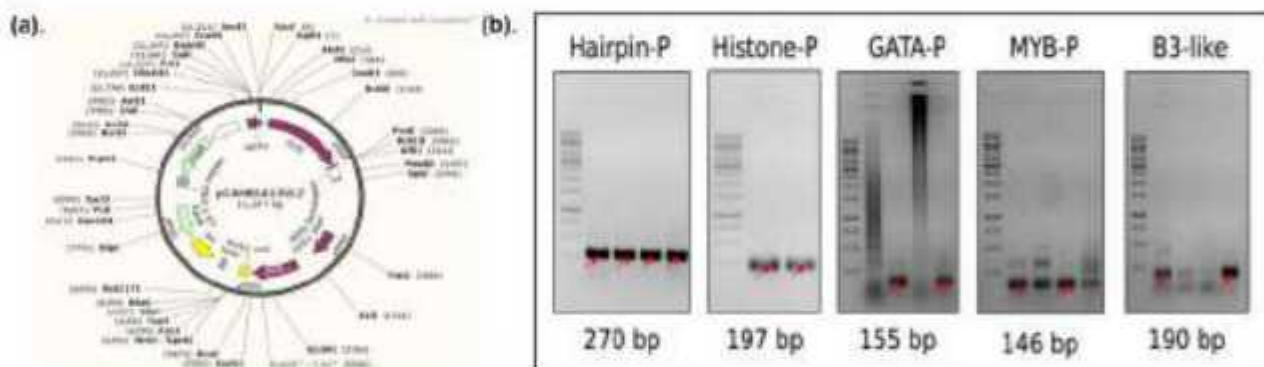


Figure 7.5: Cloning of promoter fragments of MG-responsive genes in pCAMBIA1391Z. (a). Vector map of pCAMBIA1391Z; (b). Colony PCR of promoter constructs cloned in pCAM1391Z vector using promoter-specific primers.

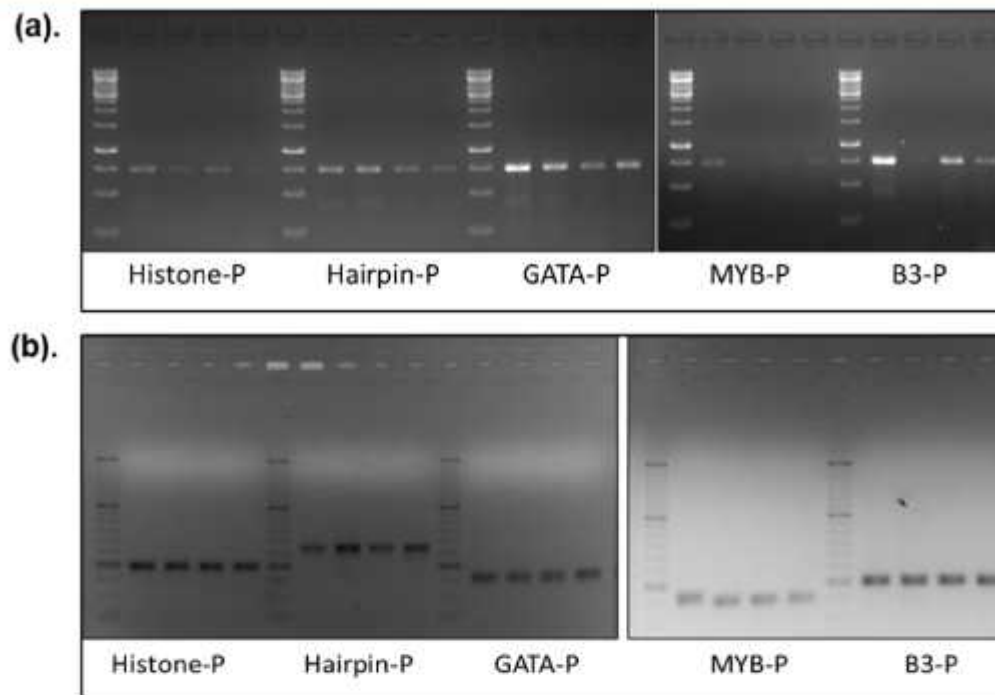


Figure 7.6: Screening of promoter_pCAMBIA1391Z constructs transformed in *Agrobacterium* LBA4404 cells by colony PCR using (a). *vir* gene primers, and (b). promoter-specific primers.

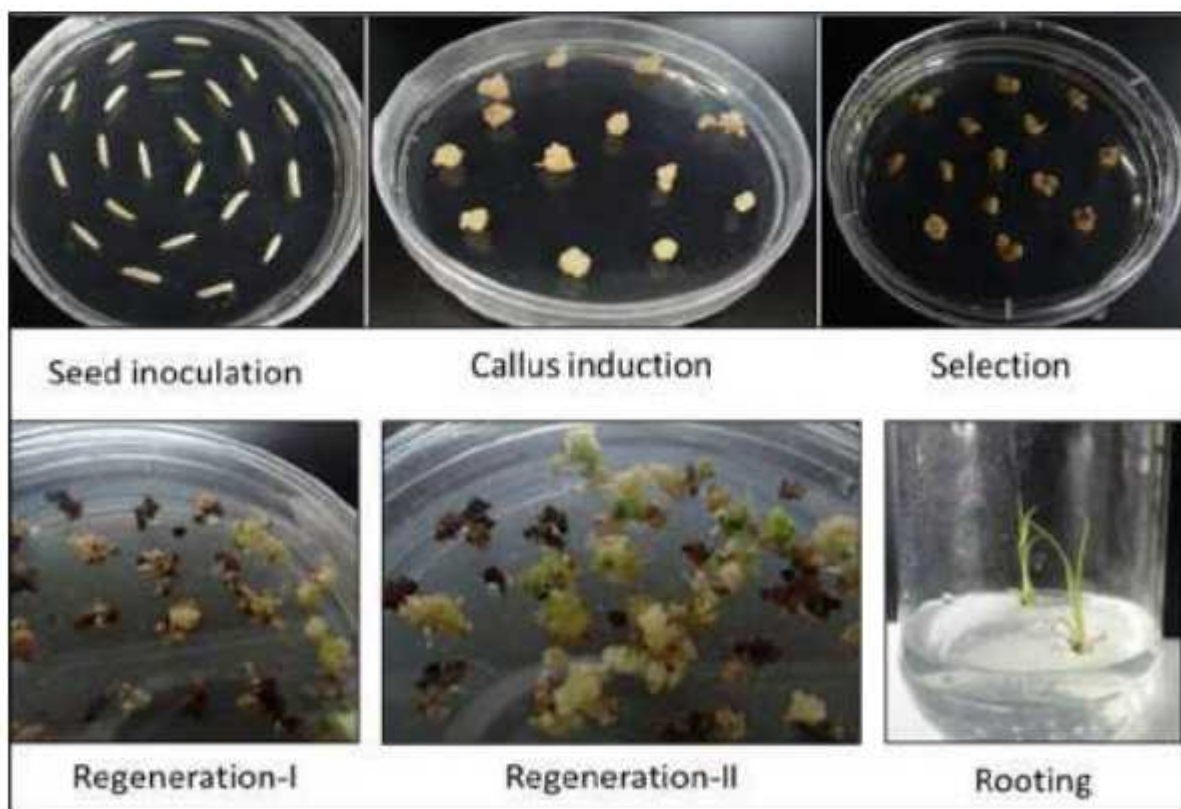


Figure 7.7: Depiction of different stages in tissue culture to raise transgenic rice (PB1) plants containing promoter_pCAMBIA1391Z constructs.

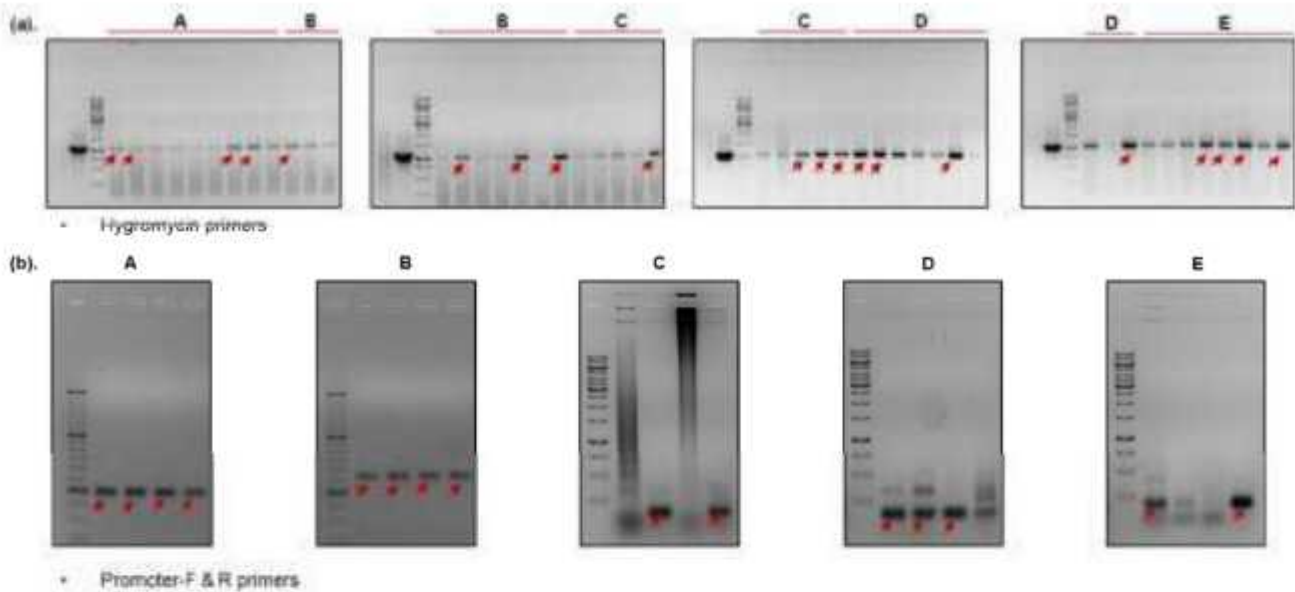


Figure 7.8: PCR for confirmation of transgenic plants using (a). hygromycin-specific, and (b). promoter-specific primers for (A). Hairpin; (B). Histone; (C). GATA; (D). MYB; (E). B3-like genes. Transgenics with strong hygromycin expression (red arrows) were further screened using promoter-specific primers, where positive bands are labelled with red arrows.

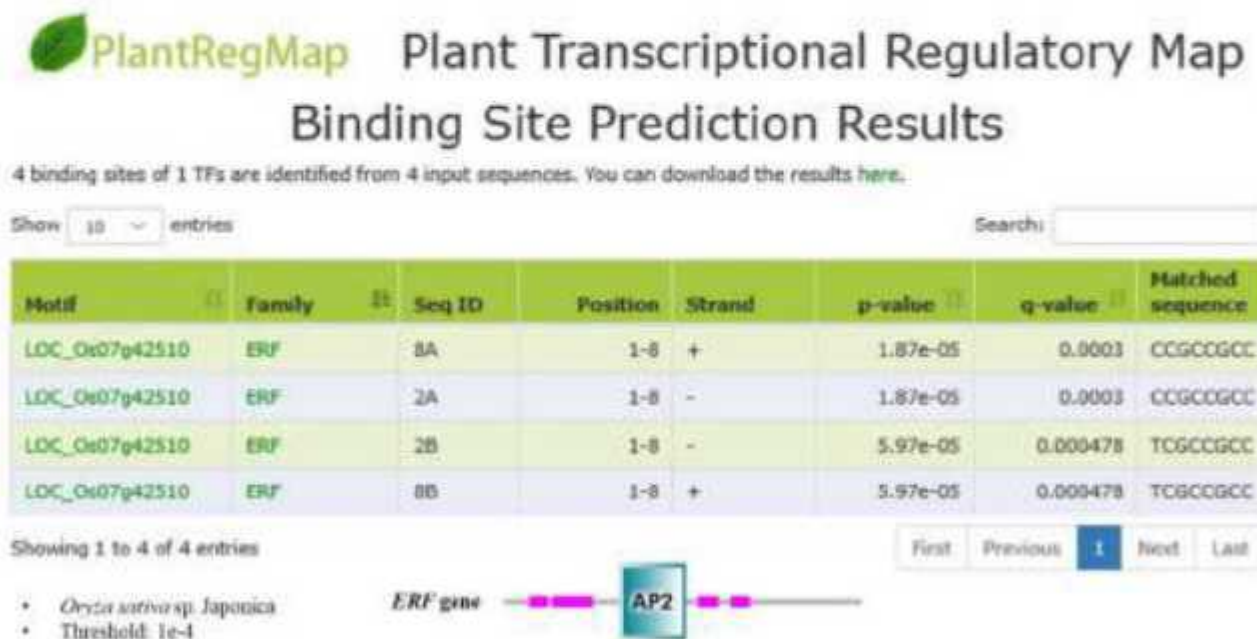


Figure 7.9: Prediction of TFs binding to the predicted MGREs. The predicted MGREs were fed as input to the PlantRegMap website to find the probable TFs binding to our predicted MGREs. *ERF* gene was the only hit binding to two of our input MGREs. This ERF protein contains an AP2 domain.

7.2.3 Identification of genes involved in MG signalling using inhibitors

In plants, signalling is primarily accomplished through many mechanisms, among which MAP kinases and calcium ions are very important. Earlier reports have implicated MAP kinases to be involved in MG signalling (Aguilera *et al.*, 2005; Inoue *et al.*, 1998; Nomura *et al.*, 2017). Hence, we investigated the role of MAP kinases and Ca^{2+} in MG signalling. We used the inhibitor of MAP2K, namely PD98059 (Salam *et al.*, 2013). Similarly, we also used the calcium chelator EGTA (Knight *et al.*, 1997).

In this experiment, two weeks old rice seedlings were exposed to 10 mM MG for 24 hrs, in the presence or absence of Ca^{2+} chelator and MAP2K inhibitor (Table 7.1). Samples were collected after 24 hrs and stored at -80°C for further processing. RNA was isolated from the samples, followed by cDNA preparation. Expression of *Gly* genes was analysed in these samples using primers as described in the Materials and Methods section (Chapter 3).

Table 7.1: Experimental conditions for the study of the role of Ca^{2+} and MAPK in MG signalling.

Experiment	Conditions
1	Control
2	10mM MG (24hrs)
3	250uM EGTA
4	100uM PD98059
5	10mM MG (2hrs), followed by 250uM EGTA (24hrs)
6	10mM MG (2hrs), followed by 100uM PD98059 (24hrs)
7	250uM EGTA (2hrs), followed by 10mM MG (24hrs)
8	100uM PD98059 (2hrs), followed by 10mM MG (24hrs)

Monitoring the growth of plants during the stress experiment gave important hints to the role of Ca^{2+} and MAPK in MG signalling (Figure 7.10). There was no visible difference between untreated control (1) and MG treated (2) plants after 24 hrs. Similarly, treatment of rice seedling with Ca^{2+} chelator (3) and MAP2K inhibitor (4) for 24 hrs didn't show any visible effect. In the next set, seedlings were exposed to 10mM MG for 2 hrs and then treated with Ca^{2+} chelator (5) and MAP2K inhibitor (6) for 24 hrs. Still, the plants didn't show any drastic visible change. However, when the seedlings were first treated with Ca^{2+} chelator (7) and MAP2K inhibitor (8) for 2 hrs and then subjected to 24 hrs of MG treatment, the plants showed wilting.

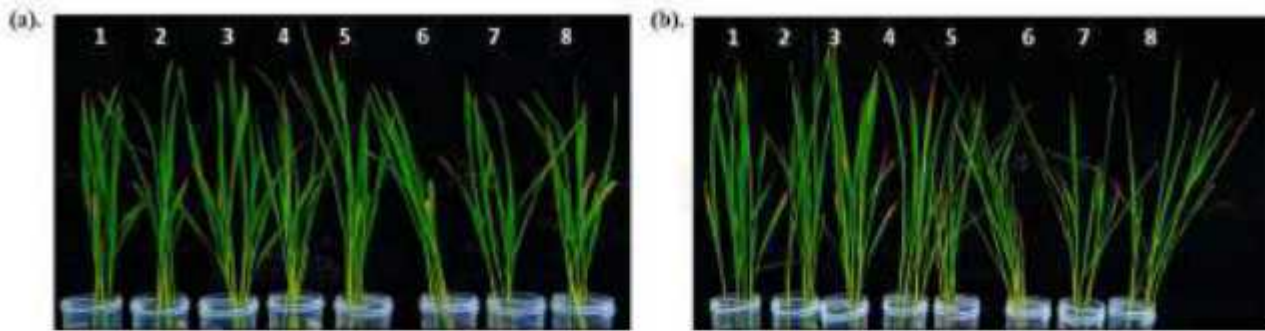


Figure 7.10: Growth of rice seedlings subjected to 10mM MG in combination with Ca^{2+} or MAPK signalling pathway inhibitors. 20-day old rice seedlings were exposed to eight different sets of experimental conditions as mentioned in Table 7.2. (a). Plants before being subjected to MG and/or inhibitors of Ca^{2+} / MAPK signalling pathway (b). Plants after treatment with MG and/or signalling pathway inhibitors. Numbers on the top indicate different treatments as listed in Table 7.1.

Expression analysis of *OsGly* and *OsMAPK* genes in response to MG and signalling pathway inhibitors

Expression of *OsGly* (Figure 7.11) and *OsMAPK* (Figure 7.12) genes were subsequently analysed in the seedlings treated with MG and Ca^{2+} / MAPK signalling pathway inhibitors. In the case of *OsGly* genes, expression of *GlyI-11.1*, *GlyII-2* and *GlyIII-3* was assessed, as representative of *GlyI*, *GlyII* and *GlyIII* families, respectively. For MAPK signalling pathway analysis, we determined the transcript levels of various *OsMPK* and *OsMCK* genes, which have been used in a previous study (Rao *et al.*, 2011). The qRT-PCR was done as described in the Material and Methods section (Chapter 3), and the samples without any stress were used as reference.

Upon 10mM MG treatment (2), the expression of *OsGlyI-11.1* and *OsGlyIII-3* genes almost doubled compared to the untreated sample (1). Ca^{2+} chelator (3) and MAP2K inhibitor (4) did not significantly affect the expression of both *OsGlyI-11.1* and *OsGlyII-2* genes but *OsGlyIII-3* expression was downregulated compared to control upon treatment with MAP2K inhibitor. Further, MG followed by Ca^{2+} chelator treatment (5) led to *OsGly* gene response similar to what was observed with only MG treatment (2). Relative expression of *OsGlyI-11.1* and *OsGlyIII-3* was marginally higher than that observed with only MG treatment (2). However, upon MAP2K inhibition (6), *OsGlyI-11.1* was down-regulated, but *OsGlyIII-3* was up-regulated. When the seedlings were first treated with Ca^{2+} chelator followed by 10mM MG (7), *OsGlyI-11.1* expression was similar to control (1), but *OsGlyIII-3* expression levels were similar to those observed under MG (2). Exposure to MAP2K inhibitor before MG treatment (8) led to down-regulation of both *OsGlyI-11.1* and *OsGlyIII-3* compared to that observed with only MG treatment (2).

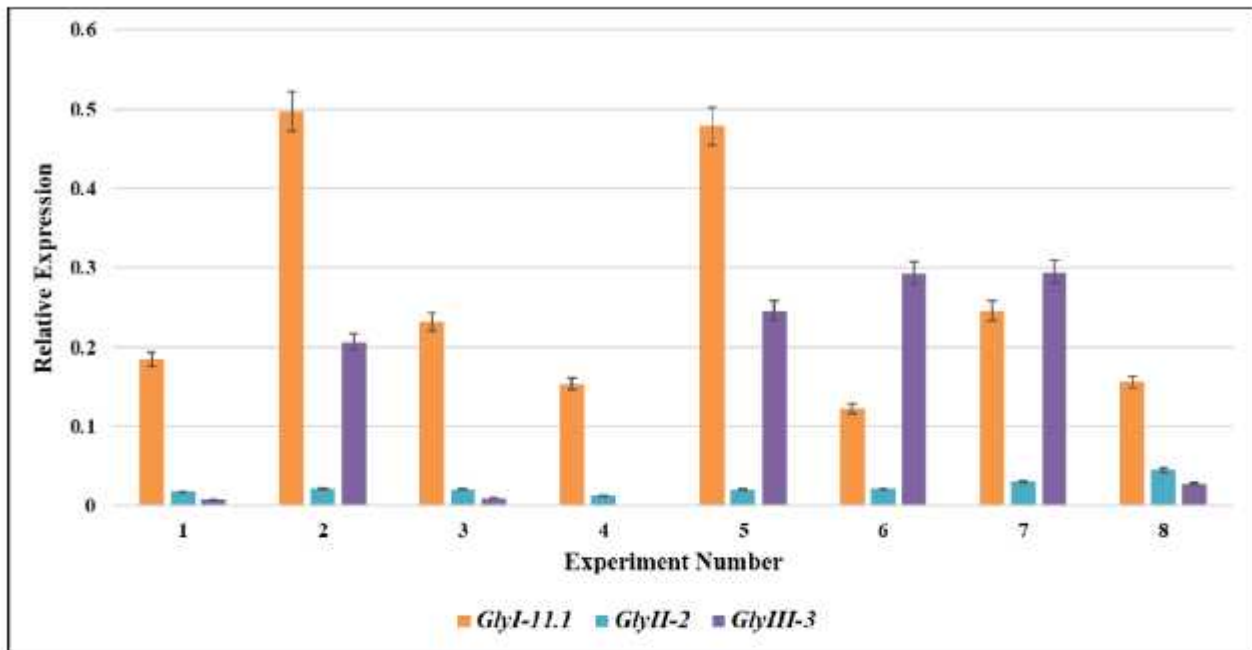


Figure 7.11: Expression profile of *OsGly* genes subjected to MG treatment in combination with inhibitors of Ca^{2+} and MAPK pathway. Rice seedlings were exposed to eight different experimental conditions mentioned in Table 7.2. Relative expression of representative *OsGly* genes was analysed through qRT-PCR. Two biological and two technical replicates were used for the study.

Analysis of the expression profile of MAPK pathway genes upon MG exposure led to some important insights (Figure 7.12). In response to 10mM MG (2), few *MAPK* genes were found to be upregulated in comparison to control. For instance, *OsMPK3*, 4, 16-2, 20-1/2/4, 21-1/2 (Figure 7.12a) and *MKK4* were highly expressed under MG stress (Figure 7.12b). Control plants treated with Ca^{2+} chelator (3) induced only a few *MAPK* genes, like *MPK14*, *MKK1*, 3 and 4. Similarly, plants treated with MAP2K inhibitor (4) had most of the *MAPK* genes down-regulated, except *MKK6*. Further, when plants were exposed to MG for 2hrs, followed by treatment with Ca^{2+} chelator (5), *MPK3*, 6, 20-1/4 and *MKK4* were upregulated to some extent. When MG treatment was followed by MAP2K inhibitor treatment (6), most of the *MAPK* genes were down-regulated to levels similar to control plants, except *MPK3* and *MPK4* which showed upregulation. Likewise, when plants treated with Ca^{2+} chelator were exposed to MG (7), the expression of most of the *MAPK* genes was not affected except a few which were downregulated. Following a similar trend, plants treated with MAP2K inhibitor before MG treatment (8) had most of the *MAPK* genes down-regulated.

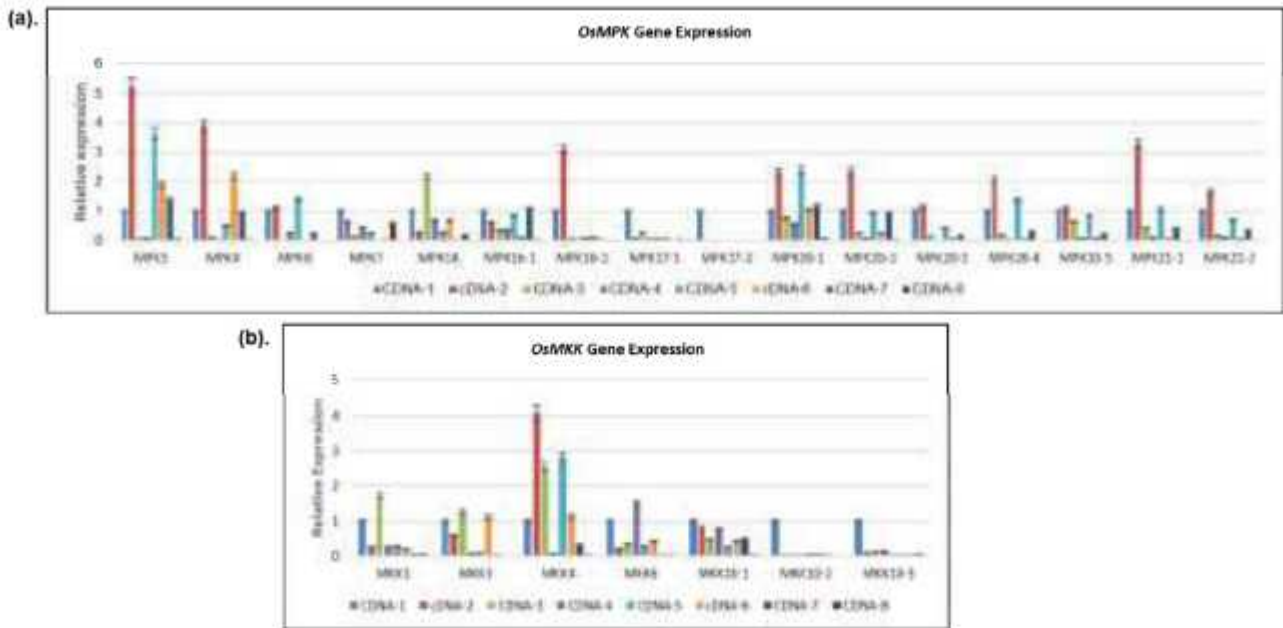


Figure 7.12: Expression analysis of MAPK genes in response to MG treatment given in combination with Ca^{2+} and MAPK signalling inhibitors. Rice seedlings were exposed to eight different experimental conditions as mentioned in Table 7.2. Relative expression of *OsMPK* and *OsMKK* genes was analysed by qRT-PCR. Two biological and two technical replicates were used for the study. The untreated seedling (sample 1) served as the control sample.

7.2.4 Possible interactors of GLY proteins were identified using the STRING database

To understand the interactome and get clues regarding the regulation of GLY proteins, we studied the interacting partners of all the GLY proteins using the *in-silico* STRING database (Szklarczyk *et al.*, 2019). All the different GLY proteins, i.e., 11 GLYI, 3 GLYII and 6 GlyIII, were fed as input and possible interacting partners were obtained from the STRING database with medium stringency and confidence (Table 7.2).

Many of the GLYI proteins had common interacting partners, which indicates that they may be performing similar functions via similar mechanisms. Most of the GLYI proteins interacted/co-expressed with other GLYI proteins (Lactoylglutathione lyase). For instance, GLYI-1, 5, and 10 co-expressed/interacted with GLYI-11. Similarly, some GLYI proteins also interacted with GLYII proteins. For example, GLYI-2, 3, 7, 8 and 11 proteins interact with GLYII (Os09g0516600). However, other GLYI proteins have some interesting interactors, such as GLYI-1 and 6 which interact with MTP7 (Metal tolerance protein). GLYI-2, 7 and 11 interacts with tryptophan synthase (OsJ_25962).

GLYII proteins also have many potential interacting proteins. However, most of them are uncharacterised. GLYII-2 interacts with the peroxiredoxin-2E (PRXIIIE-1) protein which protects cells from oxidative stress by detoxifying peroxides. GLYII-3 reacts with DHAR (Dehydroascorbate

reductase) and MDAR (Monodehydroascorbate reductase), both of which help in oxidative stress tolerance. Further, GLYIII-1, 2, 5 (DJ-1 A, B, E) interact with SUMO E3 ligase which is known to be involved in phosphate starvation response.

Table 7.2: Potential interacting partners of GLY proteins as identified using STRINGS database.

GLYI Protein			
Names	Total	Interactors	Description
OsGLYI_1, 3, 4, 5, 6, 9, 10	1	OS06T0158700-01	Lactoylglutathione lyase-like; Os06g0158700
OsGLYI_1, 3, 5, 6, 9, 10	1	OS03T0277500-01	Os03g0277500 protein
OsGLYI_2, 3, 7, 8, 11	1	OS09T0516600-01	Glyoxalase II; Os09g0516600
OsGLYI_2, 4, 7, 8, 11	1	OS01T0667200-01	Os01g0667200 protein; Putative glyoxalase II
OsGLYI_3, 4, 5, 9, 10	1	OsJ_23167	Os07g0160400 protein
OsGLYI_1, 5, 10, 11	1	GLYI-11	Lactoylglutathione lyase
OsGLYI_2, 5, 7, 10	1	OS05T0230900-01	Lactoylglutathione lyase
OsGLYI_3, 5, 8, 10	1	OS05T0295800-01	Lactoylglutathione lyase
OsGLYI_3, 5, 9, 10	1	OsJ_25424	Glyoxalase family-like protein; Os07g0657100 protein
OsGLYI_1, 4, 9	1	OS03T0659300-01	Glyoxalase family protein expressed; Os03g0659300 protein
OsGLYI_2, 7, 11	1	OsJ_25962	Tryptophan synthase
OsGLYI_1, 3	1	OS01T0173600-00	protein; Putative uncharacterized protein OSJNBa0089K24.11
OsGLYI_1, 6	1	MTP7	Metal tolerance protein 7; Involved in sequestration of excess metal in the cytoplasm into vacuoles to maintain metal homeostasis
OsGLYI_3, 6	1	OsJ_17296	4530.OS05T0171900-01; Os05g0171900 protein; Putative uncharacterized protein P0685E10.15
GLYII Proteins			
Names	Total	Interactors	Description
OsGLYII_1, 2, 3	3	OS05T0295800-01, OS05T0230900-01, GLYI-11	Lactoylglutathione lyase
OsGLYII_2, 3	6	OsJ_10718, OS11T0591550-00, OS09T0516600-01, OS07T0163000-01, OsJ_16581, OS02T0280500-01	Hydroxyacylglutathione hydrolase; Intracellular protease; Glyoxalase II;

RESULTS AND DISCUSSION (PART-IV)

OsGLYII_1	8	OsMCP, OS12T0608600-01, OS01T0667200-01, CPX, OsJ_19113, APR1, OS02T0167100-01, OS12T0442800-01	Moco containing protein; Histone acetyltransferase; Vesicle transport protein-like; Auxiliary protein in the Cpx two-component envelope stress response system;
OsGLYII_2	2	PRXIII-1, OsJ_25825	Peroxiredoxin-2E, chloroplastic precursor; ribosomal protein L13
OsGLYII_3	2	DHAR, MDAR4	PTS-dependent dihydroxyacetone kinase operon regulatory protein; Monodehydroascorbate reductase 4

GLYIII Proteins

Names	Total	Interactors	Description
OsDJ-1_A, B, C, D, E	1	P0483D07.2	Os05g0519600 protein; Putative uncharacterized protein P0599F04.12; cDNA clone:J023077J08
OsDJ-1_A, B, E	1	SIZ2	E3 SUMO-protein ligase SIZ2; Probable SUMO E3 ligase that may regulate Pi starvation responses
OsDJ-1_C, D, F	1	OsJ_16581	OSJNBb0004A17.4 protein

7.2.5 Identification and validation of interacting proteins using Yeast-2-Hybrid

To identify the interacting partners of GLY proteins using *in vivo* techniques, we used a technique based on Bio-Layer Interferometry (BLI) known as Octet. In this technique, the Octet sensor probe is coated with our protein of interest and is dipped in crude protein extract. The interacting protein forms a complex with the protein on the sensor and is extracted. This bound protein is then extracted and analysed using a mass spectrometer. The complete procedure has been described in the Materials and Methods section (chapter 3).

In our experiments, we identified four different proteins (ABC, CGT3, G-20-Oxidase, and ADH) which interacted with GLYIII-3 protein (Table 7.3). We cloned the four proteins (Figure 7.13) and tried validation using yeast-2-hybrid as described in the Materials and Methods section (chapter 3). However, all the four types of yeast cells, transformed with our four different constructs, failed to grow on the 4-drop out media. Hence, it was concluded that the identified interacting partners were false positives.

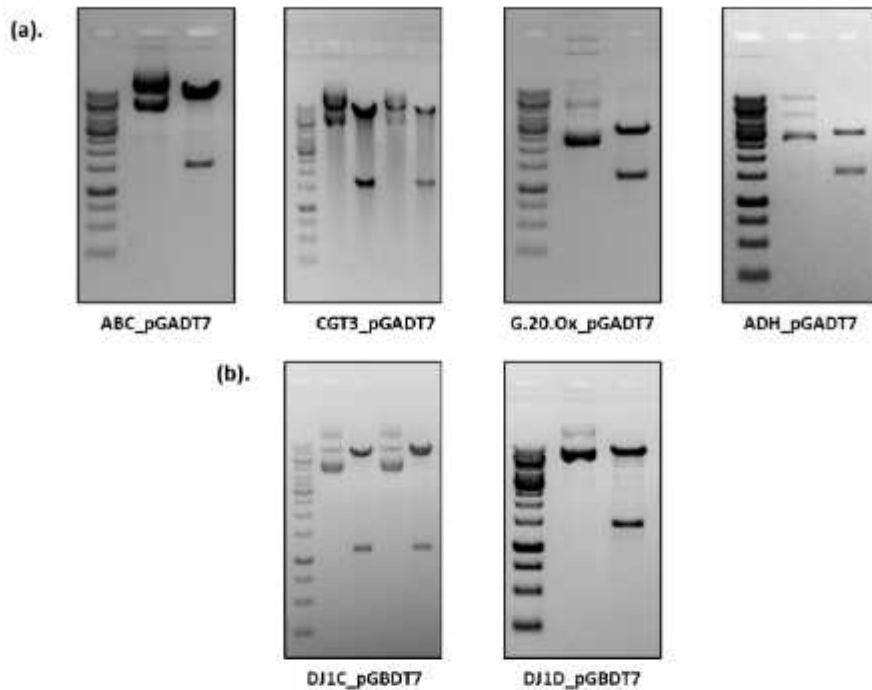


Figure 7.13: Restriction digestion of the four cloned genes identified through octet for validation in yeast vectors. (a). pGADT7 constructs for the four genes. **(b).** pGBDT7 constructs of two *GlyIII* genes (AtDJ1-C and -D).

7.3 Discussion

7.3.1 *OsGly* genes harbour CREs for stress response

Promoter analysis of *Gly* genes led to the identification of a large number of CREs most of which are involved in stress and hormone response. Apart from these, there were many CREs involved in light response, indicating that the expression of these *Gly* genes may be light-regulated. This hypothesis is supported by the fact that MG is produced more in metabolically active tissues. It is in accordance with our finding of chapter-1, where more MG was found in metabolically active leaves compared to other tissues. Chloroplasts in the leaves produce MG, both during the light and dark reactions of photosynthesis, and thus MG production is light regulated. Further, leaves had a higher expression of *Gly* genes. This suggests that leaves that are photosynthetically more active, produce more MG. This photosynthetic process is possible only during the daytime, and hence, it is logical that more *Gly* genes are expressed during light conditions and less during the dark. Thus, *Gly* genes must be light-regulated and so, a greater number of light responsive CREs were found in their upstream regions. GLY pathway is essentially involved in stress response. Stress regulation is a complex process involving other players as well such as secondary messengers and hormones. Therefore, the presence of CREs for hormonal regulation in promoter regions of *Gly* genes is justified. Stress & defence response CREs are important for the *Gly* genes to be active during the different stressful conditions.

The presence of a low number of CREs related to growth and development is in accordance with earlier reports, where *Gly* genes had a low expression under normal conditions (Mustafiz *et al.*, 2011; Ghosh *et al.*, 2016). Similarly, our finding of this chapter (see section 7.3.3) also reports a low level of *Gly* gene expression under normal conditions, compared to the stress conditions.

7.3.2 ERF proteins may bind to the predicted MGREs and regulate stress gene regulation

We believe that there are MGREs present in the promoters of various genes, just like other CREs, to which proteins bind, to regulate respective gene expression in response to MG. MGREs were predicted by screening conserved motifs in the rice *Gly* gene promoters as they are generally MG-inducible. Besides, a previous transcriptomic and promoter analysis study also identified putative MGREs by scanning the promoter sequences of genes whose expression was altered in response to MG (Kaur *et al.*, 2015). The promoter sequences of genes affected by MG were analysed and five sequences containing the predicted MGREs were cloned in promoter-less pCAMBIA1391Z vector. A total of five types of rice transgenics containing these five promoter constructs were developed through tissue culture for analysis. These plants were confirmed using hygromycin and promoter-specific primers. But unfortunately, the plants didn't bear any filled seeds, and hence, the experiment had to be discarded. The reason for the failure of the transgenic plants to bear any seeds couldn't be ascertained.

The *in-silico* approach to identify proteins binding to the predicted MGREs using the PlantRegMap (Tian *et al.*, 2020; Jin *et al.*, 2017) led to the identification of an ERF protein. This protein contains an AP2 domain that binds to the DNA and is involved in various physiological processes, including stress response.

7.3.3 MG signalling in rice plants is mediated via Ca²⁺ and MAPK

MG signalling experiments with Ca²⁺ chelator and MAP2K inhibitor helped to derive some important conclusions regarding the MG signal transduction in rice plants. Even after 24 hrs of MG stress, the plants didn't show any drastic visible change, indicating that the MG detoxification mechanisms are efficient enough to protect the plant from the harmful effects of MG. Treatment with only EGTA and MAP2K inhibitor didn't affect the plant visibly, indicating that the plant may have redundancies to overcome their toxic effects. Even when plants are first exposed to MG and then treated with the chelator and inhibitor, nothing happened to their growth as plants showed no visible changes. This indicates that the MG detoxification mechanisms get activated within the first 2 hrs itself, and hence, the MG signal blocking via Ca²⁺ chelator and MAP2K inhibitor has no effect. On the contrary, when we block the Ca²⁺ and MAP2K signal transduction pathway first and then provide MG, the plants

show wilting. This indicates that plants are not able to sense MG and hence, do not start the MG detoxification mechanism. Hence, we can speculate that Ca^{2+} and MAPK are both involved in the MG signal transduction mechanism.

OsGly gene expression analysis in seedlings exposed to Ca^{2+} and MAP2K signalling inhibitors helps to understand the signalling process of MG in plants. 10mM MG treatment (1) leads to upregulation of *OsGlyI-11.1* as well as *OsGlyIII-3*. Treatment with Ca^{2+} and MAP2K signalling inhibitors after exposure to MG stress doesn't affect the MG detoxification process, and hence, the plants showed no visible effects. However, *OsGlyI-11.1* expression is decreased to a much larger extent due to MAP2K inhibitor, compared to EGTA inhibitor. To compensate for the greater decrease in *OsGlyI-11.1* expression, *OsGlyIII-3* expression is higher in sample-6 compared to sample-5. From sample-4, 5 & 6, it seems that MAP2K inhibitors down-regulate *OsGlyI-11.1* expression and this down-regulation is more pronounced when used before MG stress imposition. On the other hand, Ca^{2+} inhibitor decreases *OsGlyI-11.1* expression in sample-7, but not in sample-5, indicating that Ca^{2+} is important for inducing *OsGlyI-11.1* expression. MAP2K inhibitors decrease the expression of *OsGlyI-11.1* irrespective of whether it is applied before or after MG stress induction. On the contrary, *OsGlyIII-3* expression decreases only when MAP2K inhibitor is applied before MG stress imposition. These observations indicate that Ca^{2+} ions are required for the induction of the *OsGlyI-11.1* gene, but has no role to play in *OsGlyIII-3* gene expression. MAPK is required for the induction as well as maintenance of *OsGlyI-11.1* expression, while it is only required for the induction of *OsGlyIII-3* expression.

Analysis of the role of MAPK in MG signal transduction in rice plants using inhibitors helped to identify some important candidate genes. Some *MAPK* genes, like *MPK3*, *4*, *16-2*, *20-1/2/4*, *21-1/2* and *MKK4* were upregulated under 10mM MG stress. These *MAPK* genes are MG-responsive and are likely to play an important part in MG stress signal transduction. Control plants treated with Ca^{2+} chelator showed upregulation of MAPK genes which were different from the MG-responsive MAPK genes. On the contrary, MAP2K inhibitor down-regulated the expression of most of the MAPK genes to below control plant levels, except *MKK6*. Plants exposed initially to MG stress, upon treatment with Ca^{2+} chelator led to the down-regulation of a few MG-responsive MAPK genes. For instance, while *MPK3* and *MKK4* were down-regulated a little, others like *MPK4* and *21-1* are down-regulated to the level of control plants. While *MPK16-2* expression is completely abolished upon due to the Ca^{2+} chelator, *MPK20-1* expression is not at all affected. Similarly, when plants exposed to MG stress are treated with MAP2K inhibitor, few MG-responsive genes, like *MPK3* and *MPK4* are slightly down-regulated, while others like *MPK20-1* and *MKK4* are down-regulated to control levels. Further, when initially treated with Ca^{2+} chelator, rice plants upon exposure to MG stress, most of the MG-responsive MAPK genes are down-regulated to basal levels. When the same experiment is performed

with MAP2K inhibitor, the MG-responsive MAPK genes expression becomes negligible. These observations, along with the expression analysis of *Gly* genes, confirms that Ca^{2+} ions are required for the expression of a few MG-responsive MAPK genes, like MPK3, 4, 20-1/2 and MKK4, whereas it is essential for the expression of others like MPK16-2, 20-4 and 21-1/2. Further, treatment with MAP2K inhibitor pre-and-post MG stress treatment helps to confirm that the identified genes are truly MG-responsive.

7.3.4 *In silico* analysis indicates possible GLY interactors

The STRING database derives data from a variety of databases, including transcriptome data (Szkarczyk *et al.*, 2019; von Mering *et al.*, 2003). The OsGLYI proteins are predicted to interact with many other GLYI and GLYII proteins. Since both the proteins are part of the same detoxification pathway, it is quite probable that these interactors are co-expressed genes. Other GLYI interactors such as MTP7 (Metal tolerance protein) indicates that GLYI proteins may be involved in heavy metal tolerance as as been shown earlier (Singla-Pareek *et al.*, 2006). Similarly, the interaction of GLYI protein with tryptophan synthase indicates that GLYI protein may be influencing amino acid metabolism. This in fact, aligns with our GC-MS results where MG stress was found to manipulate amino acid metabolism, especially in roots. Interaction of OsGLYII proteins with DHAR (Dehydroascorbate reductase), MDAR (Monodehydroascorbate reductase) and PRXIII (peroxiredoxin) proteins further support the already known fact that the GLY pathway is involved in oxidative stress tolerance. Similarly, the interaction of OsGLYIII proteins with E3 SUMO ligase, known to be involved in phosphate starvation tolerance, reconfirms the role of GLY proteins in various stress tolerance mechanisms.

7.4 Conclusion

MG has been hypothesized to act as a signal molecule. We tried to study different aspects of MG signalling. In the first section, we identified the various CREs present in the promoters of *Gly* genes and found them to be mostly stress & defence related. However, the presence of a large number of light-responsive CREs also indicates their regulation by light/dark conditions.

In the second section, we predicted MGREs using promoters of *Gly* genes and genes responsive to MG. We tried to develop five types of rice transgenics using promoter constructs containing the predicted MGREs. However, the transgenics failed to develop any mature seeds probab. Our *in-silico* approach to predict proteins binding to the predicted MGREs led to the identification of ERF genes that are involved in stress response.

In the third section, we tried to identify if Ca^{2+} and MAPK are involved in MG signal transduction. Our observations indicate that Ca^{2+} and MAPK both are involved in MG signal transduction. While Ca^{2+} is required for inducing *Gly* gene expression, MAPK is required for the induction as well as maintenance of *Gly* gene expression. Further, we also identified MG-responsive MAPK genes, such as MPK3, MPK4, MPK16-2, MPK20-1/2/4, MPK21-1/2 and MKK4.

In the last section, we tried to identify the various interactors of OsGLY proteins using both *in silico* as well *in vitro* techniques. Using the STRING database, we identified OsGLYI proteins interacting with genes involved in stress tolerance, like tryptophan synthase and metal tolerance protein (MTP7). Similarly, OsGLYII was found to interact with proteins involved in antioxidant stress tolerance, like DHAR, MDAR and PRXIII. OsGLYIII proteins were also found to interact with a protein involved in phosphate starvation tolerance. These results confirm the role of GLY proteins in stress tolerance. Using the Octet technique, we identified four proteins as probable interacting partners. However, validation using yeast-2-hybrid confirmed them to be false positives.

SUMMARY AND CONCLUSIONS

8.1 Summary & Conclusions

Plants are unique in terms of their ability to survive under diverse environmental conditions. The various physiological mechanisms operative in plants at the cellular and molecular level enable them to respond appropriately to different kinds of biotic and abiotic stress. The complexity of these physiological processes poses a great challenge in designing stress tolerant crops plants to ensure global food security. Hence, it is imperative to precisely decipher various physiological mechanisms operational at the cellular and molecular levels.

Methylglyoxal is a common metabolite that is ubiquitous throughout the tree of life. Its high reactivity makes it a potent cytotoxin. Importantly, plants have developed various detoxification mechanisms for its metabolism such as the glyoxalase pathway and the aldo-keto reductase enzymes. Despite its cytotoxic but ubiquitous nature, the role of MG in plant physiology is little studied.

Our experiments into the various aspects of MG in plants led us to derive some important conclusions which are summarized below:

1. MG estimation in different tissues revealed that metabolically more active tissues, such as flag leaves, tend to produce more MG. Further, the salt-sensitive IR64 in general, accumulates more MG particularly at germination and tillering stage, as compared to the salt-tolerant Pokkali rice variety.
2. The spatial-temporal profiling of *Gly* genes in the two contrasting rice varieties, i.e., salt-sensitive IR64 and salt-tolerant Pokkali, revealed alterations in expression of *Gly* genes and helped to better understand the role of *Gly* genes in MG detoxification in rice. *OsGlyI-11.1* and *OsGlyII-2* were found to be the most expressed *Gly* genes. The expression of different *OsGlyIII* genes were found to be specific to different developmental stages.
3. In few tissues, the OsGLY enzyme activity did not correlate exactly with the MG levels during the spatio-temporal profiling. Low levels of MG were observed despite low GLY activity. This indicates that other detoxification systems may also be involved in maintaining MG homeostasis in the plant system.
4. Genome-wide analysis of *AKR* genes in rice helped to identify 28 *AKR* genes coding for 46 proteins. Expression analysis using publically available expression data followed by qPCR analysis under different stress conditions helped to identify *OsAKR3*, *AKR5* and *AKR22* as potential candidate genes for stress tolerance.

5. Our results indicate that plants screened for drought tolerance based on low MG levels as a selection criterion indeed showed better drought tolerance. Hence, MG can be used as a biomarker for stress tolerance in plants.
6. Systemic analysis of rice seeds grown under MG stress using GC-MS analysis revealed modulation of amino acids, especially in roots. Accumulation of amino acids like proline, at higher MG concentrations, indicates a strategy by the plants to overcome MG stress.
7. Ultrastructure analysis of rice roots using Transmission Electron Microscopy (TEM) revealed the disintegration of cellular organelles at higher MG concentrations, most probably due to the interaction of MG with various structural macromolecules. In agreement, analysis of mitochondrial activity using MitoTracker Orange dye confirmed that there is a decrease in mitochondrial potential and hence, activity as MG concentration is increased. Further, TUNEL assay and PI staining confirmed DNA damage that probably led to cell death at higher MG concentrations.
8. Promoters of *OsGly* genes were analysed for Cis Regulatory Elements (CREs) to reveal that *OsGly* genes are mostly involved in stress and defence response. Besides, they may be photo-regulated and play a very limited role in growth and development. In fact, the predicted MGREs obtained from analysis of promoter sequences of *OsGly* and other genes affected due to MG stress were shown to bind ERF protein, which is involved in stress response.
9. The role of Ca^{2+} and *MAPK* genes in MG signal transduction was confirmed. MG-responsive *MAPK* genes, such as *MPK3*, *MPK4*, *MPK16-2*, *MPK20-1/2/4*, *MPK21-1/2* and *MKK4*, were identified.
10. A metal tolerance protein was found to interact with OsGLYI proteins. Besides, the interaction of OsGLYI proteins with tryptophan synthase supports our metabolomics results that MG stress leads to modulation of amino acid levels. OsGLYII and OsGLYII were found to interact with proteins involved in oxidative stress and phosphate starvation tolerance, respectively.

REFERENCES

1. **Aguilera, J. and Prieto, J.A.** (2001) The *Saccharomyces cerevisiae* aldose reductase are implied in the metabolism of methylglyoxal in response to stress conditions. *Curr. Genet.*, 39, 273–283.
2. **Aguilera, J., Rodriguez-Vargas, S. and Prieto, J.A.** (2005) The HOG MAP kinase pathway is required for the induction of methylglyoxal-responsive genes and determines methylglyoxal resistance in *Saccharomyces cerevisiae*. *Mol. Microbiol.*, 56(1), 228-39.
3. **Ahmed, M.U., Brinkmann Frye, E., Degenhardt, T.P., Thorpe, S.R. and Baynes, J.W.** (1997) N ϵ -(Carboxyethyl)lysine, a product of the chemical modification of proteins by methylglyoxal, increases with age in human lens proteins. *Biochem. J.*, 324(2), 565-70.
4. **Ahmed, M.U., Thorpe, S.R. and Baynes, J.W.** (1986) Identification of N(ϵ)-carboxymethyllysine as a degradation product of fructoselysine in glycated protein. *J. Biol. Chem.*, 261, 4886-4894.
5. **Aon, M.A. and Camara, A.K.S.** (2015) Mitochondria: hubs of cellular signaling, energetics and redox balance. A rich, vibrant, and diverse landscape of mitochondrial research. *Front. Physiol.*, 6, 94.
6. **Auiyawong, B., Narawongsanont, R. and Tantitadapitak, C.** (2017) Characterization of AKR4C15, a Novel Member of Aldo–Keto Reductase, in Comparison with Other Rice AKR(s). *Protein J.*, 36, 257–269.
7. **Bartels, D. and Sunkar, R.** (2005) Drought and salt tolerance in plants. *Crit Rev Plant Sci.*, 24, 23–58.
8. **Bechtold, U., Rabbani, N., Mullineaux, P.M. and Thornalley, P.J.** (2009) Quantitative measurement of specific biomarkers for protein oxidation, nitration and glycation in *Arabidopsis* leaves. *Plant J.*, 59(4), 661-671.
9. **Bilas, R., Szafran, K., Hnatuszko-Konka, K. and Kononowicz, A.K.** (2016) Cis-regulatory elements used to control gene expression in plants. *Plant Cell. Tissue Organ Cult.*, 127, 269–287.

10. **Bilova, T., Greifenhagen, U., Paudel, G., et al.** (2016) Glycation of Plant Proteins under Environmental Stress — Methodological Approaches, Potential Mechanisms and Biological Role. In *Abiotic and Biotic Stress in Plants - Recent Advances and Future Perspectives*. InTech.
11. **Booth, I.R., Ferguson, G.P., Miller, S., Li, C., Gunasekera, B. and Kinghorn, S.** (2003) Bacterial production of methylglyoxal: A survival strategy or death by misadventure? In *Biochemical Society Transactions*, 31(6), 1406-8.
12. **Borysiuk, K., Ostaszewska-Bugajska, M., Vaultier, M.N., Hasenfratz-Sauder, M.P. and Szal, B.** (2018) Enhanced formation of methylglyoxal-derived advanced glycation end products in Arabidopsis under ammonium nutrition. *Front. Plant Sci.*, 9, 667.
13. **Boudsocq, M. and Laurière, C.** (2005) Osmotic signaling in plants. Multiple pathways mediated by emerging kinase families. *Plant Physiology*, 138(3), 1185–1194.
14. **Boudsocq, M. and Laurière, C.** (2005) Osmotic signaling in plants. Multiple pathways mediated by emerging kinase families. *Plant Physiol.*, 138(3), 1185-94.
15. **Brocker, C., Vasiliou, M., Carpenter, S., Carpenter, C., Zhang, Y., Wang, X., Kotchoni, S. O., Wood, A. J., Kirch, H., Kopecny, D. and Nebert, D. W.** (2012) Aldehyde dehydrogenase (ALDH) superfamily in plants: gene nomenclature and comparative genomics. *Planta*, 237(1), 189-210.
16. **Butt, M., Ayyub, C.M., Amjad, M. and Ahmad, R.** (2016) Proline application enhances growth of chilli by improving physiological and biochemical attributes under salt stress. *Pakistan J. Agric. Sci.*, 53(01), 43-49.
17. **Cameron, A.D., Olin, B., Ridderström, M., Mannervik, B. and Jones, T.A.** (1997) Crystal structure of human glyoxalase I - evidence for gene duplication and 3D domain swapping. *EMBO J.*, 16, 3386-3395.
18. **Campbell, A.K., Naseem, R., Holland, I.B., Matthews, S.B. and Wann, K.T.** (2007) Methylglyoxal and other carbohydrate metabolites induce lanthanum-sensitive Ca²⁺ transients and inhibit growth in *E. coli*. *Arch. Biochem. Biophys.*

19. **Campos-Bermudez, V.A., Morán-Barrio, J., Costa-Filho, A.J. and Vila, A.J.** (2010) Metal-dependent inhibition of glyoxalase II: A possible mechanism to regulate the enzyme activity. *J. Inorg. Biochem.*, 468(1), 107-13.
20. **Casazza, J.P., Felver, M.E. and Veech, R.L.** (1984) The metabolism of acetone in rat. *J. Biol. Chem.*, 29(2), 541-9.
21. **Cooper, R.A.** (1974) Methylglyoxal Formation during Glucose Catabolism by *Pseudomonas saccharophila* Identification of Methylglyoxal Synthase. *Eur. J. Biochem.*, 44(1), 81-6.
22. **Cooper, R.A. and Anderson, A.** (1970) The formation and catabolism of methylglyoxal during glycolysis in *Escherichia coli*. *FEBS Lett.*, 11, 273-276.
23. **Da Silva, J.M. and Arrabaça, M.C.** (2004) Photosynthesis in the water-stressed C4 grass *Setaria sphacelata* is mainly limited by stomata with both rapidly and slowly imposed water deficits. *Physiologia Plantarum*, 121(3), 409-420.
24. **Degenhardt, T.P., Thorpe, S.R. and Baynes, J.W.** (1998) Chemical modification of proteins by methylglyoxal. *Cell. Mol. Biol. (Noisy-le-grand)*, 44, 1139-1145.
25. **Deswal, R., Chakaravarty, T.N. and Sopory, S.K.** (1993) The glyoxalase system in higher plants: Regulation in growth and differentiation. *Biochem. Soc. Trans.*, 21, 527-530.
26. **Esterbauer, H., Cheeseman, K.H., Dianzani, M.U., Poli, G. and Slater, T.F.** (1982) Separation and characterization of the aldehydic products of lipid peroxidation stimulated by ADP-Fe²⁺ in rat liver microsomes. *Biochem. J.*, 208, 129-140.
27. **Fareleira, P., Legall, J., Xavier, A. V. and Santos, H.** (1997) Pathways for utilization of carbon reserves in *Desulfovibrio gigas* under fermentative and respiratory conditions. *J. Bacteriol.*, 179(12), 3972-3980.
28. **Ferguson, G.P., Töttemeyer, S., MacLean, M.J. and Booth, I.R.** (1998) Methylglyoxal production in bacteria: Suicide or survival? *Arch. Microbiol.*, 170, 209-218.
29. **Ferguson, G.P., Töttemeyer, S., MacLean, M.J. and Booth, I.R.** (1998) Methylglyoxal production in bacteria: Suicide or survival? *Arch. Microbiol.*, 170, 209-218.

30. **Fiehn, O., Kloska, S. and Altmann, T.** (2001) Integrated studies on plant biology using multiparallel techniques. *Curr. Opin. Biotechnol.*, 12(1), 82-86.
31. **Field, B. and Osbourn, A.E.** (2008) Metabolic diversification - Independent assembly of operon-like gene clusters in different plants. *Science*, 320(5875), 543-547.
32. **Foerster, A. and Henle, T.** (2003) Glycation in food and metabolic transit of dietary AGEs (advanced glycation end-products): Studies on the urinary excretion of pyrroline. *Biochemical Society Transactions*, 31(Pt 6), 1383-1385.
33. **Fortpied, J., Gemayel, R., Stroobant, V. and Schaftingen, E. Van** (2005) Plant ribulosamine/erythrulosamine 3-kinase, a putative protein-repair enzyme. *Biochem. J.*, 388(Pt 3), 795-802.
34. **Fraser, C.M. and Chapple, C.** (2011) The phenylpropanoid pathway in Arabidopsis. *The Arabidopsis Book*, 9, e0152.
35. **Freeling, M.** (2009) Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.*, 60, 433-453.
36. **Fujita, M., Fujita, Y., Maruyama, K., Seki, M., Hiratsu, K., Ohme-Takagi, M., Tran, L.S.P., Yamaguchi-Shinozaki, K. and Shinozaki, K.** (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J.*, 39(6), 863-876.
37. **Ghosh, A., Kushwaha, H.R., Hasan, M.R., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L.** (2016) Presence of unique glyoxalase III proteins in plants indicates the existence of shorter route for methylglyoxal detoxification. *Sci. Rep.*, 6, 18358.
38. **Ghosh, A., Kushwaha, H.R., Hasan, M.R., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L.** (2016) Presence of unique glyoxalase III proteins in plants indicates the existence of shorter route for methylglyoxal detoxification. *Sci. Rep.*, 6, 18358.
39. **Ghosh, A., Kushwaha, H.R., Hasan, M.R., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L.** (2016) Presence of unique glyoxalase III proteins in plants indicates the existence of shorter route for methylglyoxal detoxification. *Sci. Rep.*, 6, 18358.
40. **Go, Y.M. and Jones, D.P.** (2010) Redox control systems in the nucleus: Mechanisms and functions. *Antioxidants Redox Signal.*, 13, 489-509.

41. **Gruber, P. and Hofmann, T.** (2005) Chemoselective synthesis of peptides containing major advanced glycation end-products of lysine and arginine. *J. Pept. Res.*, 66(3), 111-124.
42. **Guerra-Guimarães, L., Pinheiro, C., Chaves, I., Barros, D.R. and Ricardo, C.P.** (2016) Protein dynamics in the plant extracellular space. *Proteomes*, 4, 22.
43. **Gupta, B.K., Sahoo, K.K., Ghosh, A., et al.** (2018) Manipulation of glyoxalase pathway confers tolerance to multiple stresses in rice. *Plant Cell Environ.*, 41, 1186–1200.
44. **Haffani, S., Mezni, M., Nasri, M. Ben and Chaibi, W.** (2017) Comparative leaf water relations and anatomical responses of three vetch species (*Vicia narbonensis* L., *V. sativa* L. and *V. villosa* Roth.) to cope with water stress. *Crop Pasture Sci.*, 68(7), 691-702.
45. **Hanson, A.D. and Gregory, J.F.** (2011) Folate biosynthesis, turnover, and transport in plants. *Annu. Rev. Plant Biol.*, 62, 105-125.
46. **Hasim, S., Hussin, N.A., Alomar, F., Bidasee, K.R., Nickerson, K.W. and Wilson, M.A.** (2014) A glutathione-independent glyoxalase of the DJ-1 superfamily plays an important role in managing metabolically generated methylglyoxal in *Candida albicans*. *J. Biol. Chem.*, 289, 1662–1674.
47. **He, M.M., Clugston, S.L., Honek, J.F. and Matthews, B.W.** (2000) Determination of the structure of *Escherichia coli* glyoxalase I suggests a structural basis for differential metal activation. *Biochemistry*, 39, 8719–8727.
48. **Henle, T., Walter, A.W., Haebner, R. and Klostermeyer, H.** (1994) Detection and identification of a protein-bound imidazolone resulting from the reaction of arginine residues and methylglyoxal. *Z. Lebensm. Unters. Forsch.*, 199, 55–58.
49. **Himo, F. and Siegbahn, P.E.M.** (2001) Catalytic mechanism of glyoxalase I: A theoretical study. *J. Am. Chem. Soc.*, 123(42), 10280–10289.
50. **Hopper, D.J. and Cooper, R.A.** (1972) The purification and properties of *Escherichia coli* methylglyoxal synthase. *Biochem. J.*, 128(2), 321–329.
51. **Hossain, M.Z.A., Hossain, M.Z.A. and Fujita, M.** (2009) Stress-induced changes of methylglyoxal level and glyoxalase I activity in pumpkin seedlings and cDNA cloning of glyoxalase I gene. *Aust. J. Crop Sci.*, 3, 53–64.

52. **Hu, B., Jin, J., Guo, A.Y., Zhang, H., Luo, J. and Gao, G.** (2015) GSDDS 2.0: An upgraded gene feature visualization server. *Bioinformatics*, 31, 1296–1297.
53. **Huang, K.X., Rudolph, F.B. and Bennett, G.N.** (1999) Characterization of methylglyoxal synthase from *Clostridium acetobutylicum* ATCC 824 and its use in the formation of 1,2-propanediol. *Appl. Environ. Microbiol.*, 65(7), 3244-3247.
54. **Inoue, Y., Tsujimoto, Y. and Kimura, A.** (1998) Expression of the glyoxalase I gene of *Saccharomyces cerevisiae* is regulated by high osmolarity glycerol mitogen-activated protein kinase pathway in osmotic stress response. *J. Biol. Chem.*, 273(5), 2977-83.
55. **Inoue, Y., Tsujimoto, Y. and Kimura, A.** (1998) Expression of the glyoxalase I gene of *Saccharomyces cerevisiae* is regulated by high osmolarity glycerol mitogen-activated protein kinase pathway in osmotic stress response. *J. Biol. Chem.*, 273, 2977–2983.
56. **Jagt, D.L. Vander** (1993) Glyoxalase II: Molecular characteristics, kinetics and mechanism. *Biochemical Society Transactions*, 21(2), 522-7.
57. **Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J. and Gao, G.** (2015) An arabidopsis transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. *Mol. Biol. Evol.*, 32, 1767–1773.
58. **Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J. and Gao, G.** (2017) PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.*, 45, D1040–D1045.
59. **Kalapos, M.P.** (1999) On the promine/retine theory of cell division: Now and then. *Biochim. Biophys. Acta*, 1426(1), 1-16.
60. **Kalapos, M.P. and Biology, T.** (1999) Methylglyoxal in living organisms - Chemistry, biochemistry, toxicology and biological implications. *Toxicol. Lett.*, 110, 145–175.
61. **Kamran, M., Shahbaz, M., Ashraf, M. and Akram, N.A.** (2009) Alleviation of drought-induced adverse effects in spring wheat (*Triticum aestivum* L.) using proline as a pre-sowing seed treatment. *Pakistan Journal of Botany*, 41(2), 621–632.
62. **Kamyab, M., Kafi, M., Shahsavand, H., Goldani, M. and Shokouhifar, F.** (2016) Exploring ion homeostasis and mechanism of salinity tolerance in primary tritipyrum

- lines (Wheat× *Thinopyrum bessarabicum*) in the presence of salinity. *Aust. J. Crop Sci.*, 10, 911.
63. **Kaul, S., Sharma, S.S. and Mehta, I.K.** (2008) Free radical scavenging potential of L-proline: Evidence from in vitro assays. *Amino Acids*. 34(2), 315-20.
64. **Kaur, C., Ghosh, A., Pareek, A., Sopory, S.K.S.K. and Singla-Pareek, S.L.** (2014) Glyoxalases and stress tolerance in plants. *Biochem. Soc. Trans.*, 42, 485–490.
65. **Kaur, C., Kushwaha, H.R., Mustafiz, A., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L.** (2015) Analysis of global gene expression profile of rice in response to methylglyoxal indicates its possible role as a stress signal molecule. *Front. Plant Sci.*, 6, 682.
66. **Kaur, C., Kushwaha, H.R., Mustafiz, A., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L.** (2015) Analysis of global gene expression profile of rice in response to methylglyoxal indicates its possible role as a stress signal molecule. *Front. Plant Sci.*, 6, 682.
67. **Kaur, C., Sharma, S., Hasan, M.R., Pareek, A., Singla-Pareek, S.L. and Sopory, S.K.** (2017) Characteristic variations and similarities in biochemical, molecular, and functional properties of glyoxalases across prokaryotes and eukaryotes. *Int. J. Mol. Sci.*, 18, 250.
68. **Kawahara, Y., la Bastide, M. de, Hamilton, J.P., et al.** (2013) Improvement of the *oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice*, 6, 4.
69. **Kawahara, Y., la Bastide, M. de, Hamilton, J.P., et al.** (2013) Improvement of the *oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice*, 6, 4.
70. **Klahre, U., Noguchi, T., Fujioka, S., Takatsuto, S., Yokota, T., Nomura, T., Yoshida, S. and Chua, N.H.** (1998) The arabidopsis DIMINUTO/DWARF1 gene encodes a protein involved in steroid synthesis. *Plant Cell*, 10(10), 1677-90.
71. **Knight, H., Trewavas, A.J. and Knight, M.R.** (1997) Calcium signalling in *Arabidopsis thaliana* responding to drought and salinity. *Plant J.*, 12, 1067–1078.

72. **Koop, D.R. and Casazza, J.P.** (1985) Identification of ethanol-inducible P-450 isozyme 3a as the acetone and acetol monooxygenase of rabbit microsomes. *J. Biol. Chem.*, 260(25), 13607-12.
73. **Kosmachevskaya, O. V., Shumaev, K.B. and Topunov, A.F.** (2015) Carbonyl Stress in Bacteria: Causes and Consequences. *Biochem.*, 80, 1655–1671.
74. **Krajčovičová-Kudláčková, M., Šebeková, K., Schinzel, R. and Klvanová, J.** (2002) Advanced glycation end products and nutrition. *Physiol. Res.*, 51(3), 313-6.
75. **Kwon, K., Choi, D., Hyun, J.K., Jung, H.S., Baek, K. and Park, C.** (2013) Novel glyoxalases from *Arabidopsis thaliana*. *FEBS J.*, 280, 3328–3339.
76. **la Maza, M.P. De, Bravo, A., Leiva, L., et al.** (2007) Fluorescent serum and urinary advanced glycoxidation end-products in non-diabetic subjects. *Biol. Res.*, 40(2), 203-212.
77. **Lee, J., Song, J., Kwon, K., Jang, S., Kim, C., Baek, K., Kim, J. and Park, C.** (2012) Human DJ-1 and its homologs are novel glyoxalases. *Hum. Mol. Genet.*, 21, 3215–3225.
78. **Lescot, M., Déhais, P., Thijs, G., et al.** (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.*, 30, 325–327.
79. **Li, Y., Cohenford, M.A., Dutta, U. and Dain, J.A.** (2008) In vitro nonenzymatic glycation of guanosine 5'-triphosphate by dihydroxyacetone phosphate. *Anal. Bioanal. Chem.*, 392(6), 1189-96.
80. **Lin, F., Xu, J., Shi, J., Li, H. and Li, B.** (2010) Molecular cloning and characterization of a novel glyoxalase I gene TaGly I in wheat (*Triticum aestivum* L.). *Mol. Biol. Rep.*, 37, 729–735.
81. **Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. and Fernie, A.R.** (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.*, 1(1), 387-96.
82. **Lyles, G.A. and Chalmers, J.** (1992) The metabolism of aminoacetone to methylglyoxal by semicarbazide-sensitive amine oxidase in human umbilical artery. *Biochem. Pharmacol.*, 7(1), 1409-1414.

83. **Lytovchenko, A., Beleggia, R., Schauer, N., Isaacson, T., Leuendorf, J.E., Hellmann, H., Rose, J.K. and Fernie, A.R.** (2009) Application of GC-MS for the detection of lipophilic compounds in diverse plant tissues. *Plant Methods*, 5, 4.
84. **Maeta, K., Izawa, S. and Inoue, Y.** (2005) Methylglyoxal. A metabolite derived from glycolysis, functions as a signal initiator of the high osmolarity glycerol-mitogen-activated protein kinase cascade and calcineurin/crz1-mediated pathway in *Saccharomyces cerevisiae*. 280, 253–260.
85. **Mering, C. von, Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B.** (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.*, 31, 258–261.
86. **Misra, K., Banerjee, A.B., Ray, S. and Ray, M.** (1995) Glyoxalase III from *Escherichia coli*: a single novel enzyme for the conversion of methylglyoxal into D-lactate without reduced glutathione. *Biochem. J.*, 305, 999–1003.
87. **Mohammadkhani, N. and Heidari, R.** (2008) Drought-induced Accumulation of Soluble Sugars and Proline in Two Maize Varieties. *World Appl. Sci. J.*, 3(3), 448–453.
88. **Mosa, K.A., Ismail, A. and Helmy, M.** (2017) Introduction to Plant Stresses. In Springer, Cham, pp. 1–19.
89. **Munns, R.** (2002) Comparative physiology of salt and water stress. *Plant, Cell Environ.*, 25, 239–250.
90. **Murata, K., Fukuda, Y., SIMOSAKA, M., WATANABE, K., SAIKUSA, T. and KIMURA, A.** (1985) Metabolism of 2-oxoaldehyde in yeasts: Purification and characterization of NADPH-dependent methylglyoxal-reducing enzyme from *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, 151, 631–636.
91. **Murata, K., Fukuda, Y., Watanabe, K., Saikusa, T., Shimosaka, M. and Kimura, A.** (1985) Characterization of methylglyoxal synthase in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, 131(1), 190–8.
92. **Murata, k., Saikusa, T., Fukuda, Y., Watanabe, K., Inoue, Y., Shimosaka, M. and Kimura, A.** (1986) Metabolism of 2-oxoaldehydes in yeasts: Possible role of glycolytic bypath as a detoxification system in l-threonine catabolism by *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, 157, 297–301.

93. **Mustafiz, A., Singh, A.K., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L.** (2011) Genome-wide analysis of rice and Arabidopsis identifies two glyoxalase genes that are highly expressed in abiotic stresses. *Funct. Integr. Genomics*, 11, 293–305.
94. **Mustafiz, A., Singh, A.K., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L.** (2011) Genome-wide analysis of rice and Arabidopsis identifies two glyoxalase genes that are highly expressed in abiotic stresses. *Funct. Integr. Genomics*, 11, 293–305.
95. **Narayana, N.K., Vemanna, R.S., Kodekallu Chandrashekar, B., Rao, H., Vennapusa, A.R., Narasimaha, A., Makarla, U. and Basavaiah, M.R.** (2017) Aldo-ketoreductase 1 (AKR1) improves seed longevity in tobacco and rice by detoxifying reactive cytotoxic compounds generated during ageing. *Rice*, 10, 11.
96. **Nayyar, H. and Walia, D.P.** (2003) Water stress induced proline accumulation in contrasting wheat genotypes as affected by calcium and abscisic acid. *Biol. Plant*, 46, 275–279.
97. **Nemet, I., Varga-Defterdarović, L. and Turk, Z.** (2006) Methylglyoxal in food and living organisms. *Mol. Nutr. Food Res.*, 50(12), 1105-17.
98. **Nettleton, J.A., Steffen, L.M., Schulze, M.B., Jenny, N.S., Barr, R.G., Bertoni, A.G. and Jacobs, D.R.** (2007) Associations between markers of subclinical atherosclerosis and dietary patterns derived by principal components analysis and reduced rank regression in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am. J. Clin. Nutr.*, 85(6), 1615-25.
99. **Nomura, W., Maeta, K. and Inoue, Y.** (2017) Phosphatidylinositol 3,5-bisphosphate is involved in methylglyoxal-induced activation of the Mpk1 mitogen-activated protein kinase cascade in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, 292, 15039–15048.
100. **Okado-Matsumoto, A. and Fridovich, I.** (2000) The role of α,β -dicarbonyl compounds in the toxicity of short chain sugars. *J. Biol. Chem.*, 275(45), 34853-7.
101. **Osuji, G.O., Brown, T.K., South, S.M., Duncan, J.C. and Johnson, D.** (2011) Doubling of crop yield through permutation of metabolic pathways. *Advances in Bioscience and Biotechnology*, 2, 364–379.

102. **Oya, T., Hattori, N., Mizuno, Y., Miyata, S., Maeda, S., Osawa, T. and Uchida, K.** (1999) Methylglyoxal modification of protein. Chemical and immunochemical characterization of methylglyoxal-arginine adducts. *J. Biol. Chem.*, 274(26), 18492-18502.
103. **Panchy, N., Lehti-Shiu, M. and Shiu, S.H.** (2016) Evolution of gene duplication in plants. *Plant Physiol.*, 171, 2294–2316.
104. **Philips, S.A. and Thornalley, P.J.** (1993) The formation of methylglyoxal from triose phosphates: Investigation using a specific assay for methylglyoxal. *Eur. J. Biochem.*, 212, 101–105.
105. **Pikaard, C.S. and Scheid, O.M.** (2014) Epigenetic regulation in plants. *Cold Spring Harb. Perspect. Biol.*, 6, a019315.
106. **Planchet, E., Rannou, O., Ricoult, C., Boutet-Mercey, S., Maia-Grondard, A. and Limami, A.M.** (2011) Nitrogen metabolism responses to water deficit act through both abscisic acid (ABA)-dependent and independent pathways in *Medicago truncatula* during post-germination. *J. Exp. Bot.*, 62(2), 605-15.
107. **Racker, E.** (1951) The mechanism of action of glyoxalase. *J. Biol. Chem.*, 190, 685–696.
108. **Ramasamy, R., Yan, S.F. and Schmidt, A.M.** (2006) Methylglyoxal comes of AGE. *Cell*, 124, 258–260.
109. **Rao, K.P., Vani, G., Kumar, K., Wankhede, D.P., Misra, M., Gupta, M. and Sinha, A.K.** (2011) Arsenic stress activates MAP kinase in rice roots and leaves. *Arch. Biochem. Biophys.*, 506, 73–82.
110. **Ray, M. and Ray, S.** (1984) Purification and partial characterization of a methylglyoxal reductase from goat liver. *BBA - Gen. Subj.*, 802, 119–127.
111. **Ray, S. and Ray, M.** (1981) Isolation of methylglyoxal synthase from goat liver. *J. Biol. Chem.*, 256, 6230–6233.
112. **Rhee, H.I., Watanabe, K., Murata, K. and Kimura, A.** (1987) Metabolism of 2-Ketoaldehydes in Bacteria: Oxidative Conversion of Methylglyoxal to Pyruvate by an Enzyme from *Pseudomonas putida*. *Agric. Biol. Chem.*, 51, 1059–1066.
113. **Richard, J.P.** (1984) Acid-Base Catalysis of the Elimination and Isomerization Reactions of Triose Phosphates. *J. Am. Chem. Soc.*, 106, 4926–4936.

114. **Rizwan, M., Ali, S., Adrees, M., Rizvi, H., Zia-ur-Rehman, M., Hannan, F., Qayyum, M.F., Hafeez, F. and Ok, Y.S.** (2016) Cadmium stress in rice: toxic effects, tolerance mechanisms, and management: a critical review. *Environ. Sci. Pollut. Res.*, 23(18), 17859-79.
115. **Rombauts, S., Déhais, P., Montagu, M. Van and Rouzé, P.** (1999) PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Res.*, 30, 325–327.
116. **Rout, G.R. and Das, A.B.** (2013) *Molecular stress physiology of plants*. Springer.
117. **Sahoo, K.K., Tripathi, A.K., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L.** (2011) An improved protocol for efficient transformation and regeneration of diverse indica rice cultivars. *Plant Methods*, 7, 49.
118. **Salam, M.A., Jammes, F., Hossain, M.A., Ye, W., Nakamura, Y., Mori, I.C., Kwak, J.M. and Murata, Y.** (2013) Two guard cell-preferential MAPKs, MPK9 and MPK12, regulate YEL signalling in Arabidopsis guard cells. *Plant Biol.*, 15, 436–442.
119. **Sankaranarayanan, S., Jamshed, M. and Samuel, M.A.** (2015) Degradation of glyoxalase I in Brassica napus stigma leads to self-incompatibility response. *Nat. Plants*, 1, 15185.
120. **Sankaranarayanan, S., Jamshed, M. and Samuel, M.A.** (2015) Degradation of glyoxalase I in Brassica napus stigma leads to self-incompatibility response. *Nat. Plants*, 1, 15185.
121. **Schauer, N., Zamir, D. and Fernie, A.R.** (2005) Metabolic profiling of leaves and fruit of wild species tomato: A survey of the Solanum lycopersicum complex. In *Journal of Experimental Botany*, 56(410), 297-307.
122. **Schwarzenbolz, U., Henle, T., Haebßner, R. and Klostermeyer, H.** (1997) On the reaction of glyoxal with proteins. *Eur. Food Res. Technol.*, 205, 121-124.
123. **Šebeková, K., Krajčovičová-Kudláčková, M., Schinzel, R., Faist, V., Klvanová, J. and Heidland, A.** (2001) Plasma levels of advanced glycation end products in healthy, long-term vegetarians and subjects on a western mixed diet. *Eur. J. Nutr.*, 40(6), 275-81.

124. **Sengupta, D., Naik, D. and Reddy, A. R.** (2015) Plant aldo-keto reductases (AKRs) as multi-tasking soldiers involved in diverse plant metabolic processes and stress defense: A structure-function update. *J Plant Phys*, 179, 40–55.
125. **Shekari, G. and Javanmardi, J.** (2017) Effects of foliar application pure amino acid and amino acid containing fertilizer on broccoli (*Brassica oleracea* L. var. italica) transplants. *Advances in Crop Science and Technology*, 5(3), 1–4.
126. **Shekari, G. and Javanmardi, J.** (2017) Effects of Foliar Application Pure Amino Acid and Amino Acid Containing Fertilizer on Broccoli (*Brassica oleracea* L. var. italica) Transplants. *Adv. Crop Sci. Technol.*, 5(3), 280.
127. **Shipanova, I.N., Glomb, M.A. and Nagaraj, R.H.** (1997) Protein modification by methylglyoxal: Chemical nature and synthetic mechanism of a major fluorescent adduct. *Arch. Biochem. Biophys.*, 344(1), 29-36.
128. **Shuman, J.L., Cortes, D.F., Armenta, J.M., Pokrzywa, R.M., Mendes, P. and Shulaev, V.** (2011) Plant metabolomics by GC-MS and differential analysis. *Methods Mol. Biol.*, 678, 229-46.
129. **Silva, J.M. Da and Arrabaça, M.C.** (2004) Photosynthesis in the water-stressed C4 grass *Setaria sphacelata* is mainly limited by stomata with both rapidly and slowly imposed water deficits. *Physiol. Plant.*, 121(3), 409-420.
130. **Silva, M.S., Barata, L., Ferreira, A.E.N., Romão, S., Tomás, A.M., Freire, A.P. and Cordeiro, C.** (2008) Catalysis and structural properties of *Leishmania infantum* glyoxalase II: Trypanothione specificity and phylogeny. *Biochemistry*, 47(1), 195-204.
131. **Singh, T.N., Aspinall, D. and Paleg, L.G.** (1972) Proline accumulation and varietal adaptability to drought in barley: A potential metabolic measure of drought resistance. *Nat. New Biol.*, 236, 188-190.
132. **Singla-Pareek, S.L., Reddy, M.K. and Sopory, S.K.** (2003) Genetic engineering of the glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *Proc. Natl. Acad. Sci. U. S. A.*, 100, 14672–14677.
133. **Singla-Pareek, S.L., Reddy, M.K. and Sopory, S.K.** (2003) Genetic engineering of the glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *Proc. Natl. Acad. Sci. U. S. A.*, 100, 14672–14677.

134. **Singla-Pareek, S.L., Yadav, S.K., Mustafiz, A. and Sopory, S.K.** (2008) Detoxification of Methylglyoxal by Genetic Manipulation of Glyoxalase Pathway Leads to Salt Stress Tolerance in Crop Plants. *Plant Cell Physiol. Suppl.*, 49, S0026–S0026.
135. **Singla-Pareek, S.L., Yadav, S.K., Mustafiz, A. and Sopory, S.K.** (2008) Detoxification of Methylglyoxal by Genetic Manipulation of Glyoxalase Pathway Leads to Salt Stress Tolerance in Crop Plants. *Plant Cell Physiol.*, 49, S0026–S0026.
136. **Singla-Pareek, S.L., Yadav, S.K., Pareek, A., Reddy, M.K. and Sopory, S.K.** (2006) Transgenic Tobacco Overexpressing Glyoxalase Pathway Enzymes Grow and Set Viable Seeds in Zinc-Spiked Soils. *Plant Physiol.*, 140, 613–623.
137. **Slocum, R.D.** (2005) Genes, enzymes and regulation of arginine biosynthesis in plants. *Plant Physiol. Biochem.*, 43(8), 729-745.
138. **Sousa Silva, M., Gomes, Ricardo A, Ferreira, A.E.N., Freire, A.P. and Cordeiro, C.** (2013) The glyoxalase pathway: the first hundred years . . . and beyond. *Biochem. J.*, 453, 1–15.
139. **Speer, O., Morkunaite-Haimi, S., Liobikas, J., Franck, M., Hensbo, L., Linder, M.D., Kinnunen, P.K.J., Wallimann, T. and Eriksson, O.** (2003) Rapid suppression of mitochondrial permeability transition by methylglyoxal: Role of reversible arginine modification. *J. Biol. Chem.*, 278(37), 34757-63.
140. **Subedi, K.P., Choi, D., Kim, I., Min, B. and Park, C.** (2011) Hsp31 of Escherichia coli K-12 is glyoxalase III. *Mol. Microbiol.*, 81, 926–936.
141. **Sukdeo, N. and Honek, J.F.** (2007) Pseudomonas aeruginosa contains multiple glyoxalase I-encoding genes from both metal activation classes. *Biochim. Biophys. Acta - Proteins Proteomics*, 1774, 756–763.
142. **Szklarczyk, D., Gable, A.L., Lyon, D., et al.** (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, 47, D607–D613.
143. **Thornalley, P J** (1996) Pharmacology of methylglyoxal: formation, modification of proteins and nucleic acids, and enzymatic detoxification--a role in pathogenesis and antiproliferative chemotherapy. *Gen. Pharmacol.*, 27, 565–573.

144. **Thornalley, P.J.** (1990) The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. *Biochem. J.*, 269, 1–11.
145. **Thornalley, P.J.** (2008) Protein and nucleotide damage by glyoxal and methylglyoxal in physiological systems--role in ageing and disease. *Drug Metabol. Drug Interact.*, 23, 125–50.
146. **Thornalley, P.J., Langborg, A. and Minhas, H.S.** (1999) Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.*, 344 Pt 1, 109–16.
147. **Tian, F., Yang, D.C., Meng, Y.Q., Jin, J. and Gao, G.** (2020) PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Res.*, 48, D1104–D1113.
148. **Touchette, B.W., Smith, G.A., Rhodes, K.L. and Poole, M.** (2009) Tolerance and avoidance: Two contrasting physiological responses to salt stress in mature marsh halophytes *Juncus roemerianus* Scheele and *Spartina alterniflora* Loisel. *J. Exp. Mar. Bio. Ecol.*, 380, 106–112.
149. **Urata, G. and Granick, S.** (1963) Biosynthesis of alpha-aminoketones and the metabolism of aminoacetone. *J. Biol. Chem.*, 238, 811-820.
150. **Uribarri, J., Cai, W., Sandu, O., Peppas, M., Goldberg, T. and Vlassara, H.** (2005) Diet-derived advanced glycation end products are major contributors to the body's AGE pool and induce inflammation in healthy subjects. *Annals of the New York Academy of Sciences*, 1043, 461-6.
151. **Usui, T., Shimohira, K., Watanabe, H. and Hayase, F.** (2007) Detection and determination of glyceraldehyde-derived pyridinium-type advanced glycation end product in streptozotocin-induced diabetic rats. *Biosci. Biotechnol. Biochem.*, 71(2), 442-8.
152. **Veena, Reddy, V.S. and Sopory, S.K.** (1999) Glyoxalase I from *Brassica juncea*: molecular cloning, regulation and its over-expression confer tolerance in transgenic tobacco under stress. *Plant J.*, 17, 385–95.
153. **Vemanna, R.S., Babitha, K.C., Solanki, J.K., Amarnatha Reddy, V., Sarangi, S.K. and Udayakumar, M.** (2017) Aldo-keto reductase-1 (AKR1) protect cellular

- enzymes from salt stress by detoxifying reactive cytotoxic compounds. *Plant Physiol. Biochem.*, 113, 177–186.
154. **Wang, Y., Wang, X. and Paterson, A.H.** (2012) Genome and gene duplications and gene expression divergence: A view from plants. *Ann. N. Y. Acad. Sci.*, 1256, 1–14.
155. **Wells-Knecht, K.J., Zyzak, D. V., Litchfield, J.E., Thorpe, S.R. and Baynes, J.W.** (1995) Mechanism of Autoxidative Glycosylation: Identification of Glyoxal and Arabinose as Intermediates in the Autoxidative Modification of Proteins by Glucose. *Biochemistry*, 34(11), 3702-9.
156. **Wermuth, B.** (1985). Aldo-keto reductases, In T. G. Flynn and H. Weiner (ed.), *Enzymology of carbonyl metabolism 2: aldehyde dehydrogenase, aldo-keto reductase, and alcohol dehydrogenase*. Alan R. Liss, Inc., New York, p. 209–230.
157. **Williams, B.J., Cameron, C.J., Workman, R., Broeckling, C.D., Sumner, L.W. and Smith, J.T.** (2007) Amino acid profiling in plant cell cultures: An inter-laboratory comparison of CE-MS and GC-MS. *Electrophoresis*, 28(9), 1371-9.
158. **Wolff, S.P. and Dean, R.T.** (1987) Glucose autoxidation and protein modification. The potential role of “autoxidative glycosylation” in diabetes. *Biochem. J.*, 245(1), 243-250.
159. **Yadav, S.K., Singla-Pareek, S.L. and Sopory, S.K.** (2008) An overview on the role of methylglyoxal and glyoxalases in plants. *Drug Metabol. Drug Interact.*, **23**, 51–68.
160. **Yadav, S.K., Singla-Pareek, S.L., Ray, M., Reddy, M.K. and Sopory, S.K.** (2005) Methylglyoxal levels in plants under salinity stress are dependent on glyoxalase I and glutathione. *Biochem. Biophys. Res. Commun.*, 337, 61–67.
161. **Yadav, S.K., Singla-Pareek, S.L., Ray, M., Reddy, M.K. and Sopory, S.K.** (2005) Methylglyoxal levels in plants under salinity stress are dependent on glyoxalase I and glutathione. *Biochem. Biophys. Res. Commun.*, 337, 61–67.
162. **Yadav, S.K., Singla-Pareek, S.L., Reddy, M.K. and Sopory, S.K.** (2005) Methylglyoxal detoxification by glyoxalase system: A survival strategy during environmental stresses. *Physiol Mol Biol Plants*, 11, 11.

163. **Yamauchi, Y., Ejiri, Y., Toyoda, Y. and Tanaka, K.** (2003) Identification and biochemical characterization of plant acylamino acid-releasing enzyme. *J. Biochem.*, 134(2), 251-257.
164. **You, X., Zhang, W., Hu, J., et al.** (2019) FLOURY ENDOSPERM15 encodes a glyoxalase I involved in compound granule formation and starch synthesis in rice endosperm. *Plant Cell Rep.*, 38, 345–359.
165. **Zang, T.M., Hollman, D.A., Crawford, P.A., Crowder, M.W. and Makaroff, C.A.** (2001) Arabidopsis Glyoxalase II Contains a Zinc/Iron Binuclear Metal Center That Is Essential for Substrate Binding and Catalysis. *J. Biol. Chem.*, 276(7), 4788-95.
166. **Zhang, H., Zhao, Y. and Zhu, J.K.** (2020) Thriving under Stress: How Plants Balance Growth and the Stress Response. *Dev. Cell.*, 55(5), 529-543.

PUBLICATIONS

Research paper

- Kumar, B.; Kaur, C.; Pareek, A.; Sopory, S. K.; Singla-Pareek, S. L. Tracing the Evolution of Plant Glyoxalase III Enzymes for Structural and Functional Divergence. *Antioxidants*, **2021**, 10(648), 648. <https://doi.org/10.3390/antiox10050648>

Review

- Singla-Pareek, S. L.; Kaur, C.; Kumar, B.; Pareek, A.; Sopory, S. K. Reassessing Plant Glyoxalases: Large Family and Expanding Functions. *New Phytol.* **2020**, 227 (3), 714–721. <https://doi.org/10.1111/nph.16576>.

Poster Presentations

- Kumar, B.; Hasan, R.; Pareek, A.; Singla-Pareek, S. L. “Expressway for cellular detox under stress conditions”: In-house Symposium, ICGER (Best Poster Award) – 18th July, 2017
- Kumar, B.; Hasan, R.; Pareek, A.; Singla-Pareek, S. L. “Cellular detoxification of methylglyoxal under stress conditions”: International Conference on “Next Gen Crops for Sustainable Agriculture”, Chandigarh - 19-20th July 2018
- Kumar, B.; Hasan, R.; Pareek, A.; Singla-Pareek, S. L. “An alternate route for methylglyoxal detoxification in rice”: “Plant Response to Light and Stress” Workshop, ICGER - 10-13th October 2018

ADDITIONAL FILES

Table S1: List of the *OsAKR* genes and proteins in rice.

RGAP ID	AKR gene	Isoforms
LOC_Os01g43090	<i>OsAKR1</i>	1
LOC_Os01g62860	<i>OsAKR2</i>	1
LOC_Os01g62870	<i>OsAKR3</i>	2
LOC_Os01g62880	<i>OsAKR4</i>	2
LOC_Os02g03100	<i>OsAKR5</i>	2
LOC_Os02g57240	<i>OsAKR6</i>	1
LOC_Os03g13390#	<i>OsAKR7</i>	1
LOC_Os03g41510	<i>OsAKR8</i>	5*
LOC_Os04g08550	<i>OsAKR9</i>	1
LOC_Os04g26870	<i>OsAKR10</i>	1
LOC_Os04g26910	<i>OsAKR11</i>	1
LOC_Os04g26920	<i>OsAKR12</i>	3
LOC_Os04g27060	<i>OsAKR13</i>	1
LOC_Os04g37470	<i>OsAKR14</i>	1
LOC_Os04g37480	<i>OsAKR15</i>	2
LOC_Os04g37490	<i>OsAKR16</i>	1
LOC_Os05g38230	<i>OsAKR17</i>	3
LOC_Os05g39690	<i>OsAKR18</i>	1
LOC_Os07g04990	<i>OsAKR19</i>	2
LOC_Os07g05000	<i>OsAKR20</i>	1
LOC_Os09g39390	<i>OsAKR21</i>	1
LOC_Os10g02380	<i>OsAKR22</i>	1
LOC_Os10g02480	<i>OsAKR23</i>	2
LOC_Os10g02490	<i>OsAKR24</i>	1*
LOC_Os10g28320	<i>OsAKR25</i>	3*
LOC_Os10g37330	<i>OsAKR26</i>	1
LOC_Os11g42540	<i>OsAKR27</i>	2
LOC_Os12g29760	<i>OsAKR28</i>	2

Table S2: A catalogue of OsAKR proteins with their HMM profiles.

Gene Name	Family		Description	Domain		Length	Bit Score	Domain E-values	
	ID	Accession		Start	End			Individual	Conditional
<i>OsAKR1.1</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	24	318	295	256.7	2.50E-76	1.40E-80
<i>OsAKR2.1</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	15	284	270	164.37	3.30E-48	1.90E-52
<i>OsAKR3.1</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	15	284	270	161.74	2.10E-47	1.20E-51
<i>OsAKR3.2</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	5	210	206	113.35	1.20E-32	6.50E-37
<i>OsAKR4.1</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	15	277	263	131.14	8.40E-46	2.50E-42
<i>OsAKR4.2</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	1	185	185	76.63	9.10E-28	9.90E-26
<i>OsAKR5.1</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	79	365	287	171.64	2.00E-50	1.10E-54
<i>OsAKR5.2</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	79	358	280	172.04	1.50E-50	8.60E-55
<i>OsAKR6.1</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	15	318	304	230.08	3.20E-68	1.80E-72
<i>OsAKR7.2</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	12	290	279	163.45	6.40E-48	3.60E-52
<i>OsAKR8.1/2</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	63	397	335	231.85	9.30E-69	5.20E-73
<i>OsAKR8.3</i>	Aldo_ket_red	PF00248.21	Aldo/keto reductase family	59	392	334	222.42	6.90E-66	3.90E-70

ADDITIONAL FILES

<i>OsAKR8.4</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	63	250	188	134.08	5.60E-39	3.10E-43
<i>OsAKR8.5/6/7</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	1	277	277	189.81	5.90E-56	3.30E-60
<i>OsAKR8.8</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	63	266	204	144.56	3.60E-42	2.00E-46
<i>OsAKR9.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	33	308	276	148.65	2.10E-43	1.10E-47
<i>OsAKR10.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	27	320	294	237.64	1.60E-70	8.90E-75
<i>OsAKR11.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	28	319	292	250	2.70E-74	1.50E-78
<i>OsAKR12.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	28	149	122	86.34	2.00E-24	1.10E-28
<i>OsAKR12.2</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	28	189	162	137.38	5.60E-40	3.10E-44
<i>OsAKR12.3</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	28	200	173	144.36	4.20E-42	2.30E-46
<i>OsAKR13.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	30	323	294	240.52	2.10E-71	1.20E-75
<i>OsAKR14.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	30	304	275	181.22	2.50E-53	1.40E-57
<i>OsAKR15.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	20	293	274	162.77	1.00E-47	5.70E-52
<i>OsAKR15.2</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	33	254	222	125.16	2.90E-36	1.60E-40
<i>OsAKR16.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	53	327	275	177.06	4.50E-52	2.50E-56
<i>OsAKR17.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	52	307	256	163.33	6.90E-48	3.80E-52
<i>OsAKR17.2</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	29	299	271	169.47	9.30E-50	5.20E-54

ADDITIONAL FILES

<i>OsAKR17.3</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	10	170	161	91.61	4.90E-26	2.70E-30
<i>OsAKR18.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	24	288	265	166.65	6.80E-49	3.80E-53
<i>OsAKR19.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	62	364	303	199.35	7.40E-59	4.10E-63
<i>OsAKR19.2</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	62	298	237	153.37	7.50E-45	4.20E-49
<i>OsAKR20.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	62	364	303	208.85	9.40E-62	5.30E-66
<i>OsAKR21.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	62	367	306	206.1	6.50E-61	3.60E-65
<i>OsAKR22.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	18	293	276	171.62	2.10E-50	1.20E-54
<i>OsAKR23.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	15	293	279	174.82	2.20E-51	1.20E-55
<i>OsAKR23.2</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	15	249	235	140.44	6.50E-41	3.60E-45
<i>OsAKR24.1/2</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	29	308	280	158.64	1.90E-46	1.00E-50
<i>OsAKR25.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	20	314	295	263.42	2.20E-78	1.20E-82
<i>OsAKR25.2/3</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	20	245	226	199.93	4.90E-59	2.70E-63
<i>OsAKR25.4</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	20	145	126	108.32	4.00E-31	2.20E-35
<i>OsAKR26.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	59	360	302	221.32	1.50E-65	8.30E-70
<i>OsAKR27.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	16	351	336	252.85	3.70E-75	2.10E-79
<i>OsAKR27.2</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	16	332	317	203.5	4.00E-60	2.20E-64

<i>OsAKR28.1</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	15	292	278	180.28	4.70E-53	2.60E-57
<i>OsAKR28.2</i>	Aldo ket red	PF00248.21	Aldo/keto reductase family	15	234	220	145.8	1.50E-42	8.40E-47

Table S3: Physical and chemical characterization of OsAKR proteins.

Locus ID	OsAKR Name	Number of amino acids	Molecular weight Mw/Da	Theoretical pI	Number of negatively amino acid (Asp + Glu)	Number of positively amino acid (Arg + Lys)	Instability index	Stability	Aliphatic index	Average of hydropathicity
LOC_Os01g43090.1	OsAKR1.1	344	37915.24	6.27	45	42	35.88	Stable	89.33	-0.299
LOC_Os01g62860.1	OsAKR2.1	311	34640.6	5.89	39	32	42.75	Unstable	92.09	-0.229
LOC_Os01g62870.1	OsAKR3.1	311	34634.66	6.56	36	34	43.21	Unstable	90.93	-0.269
LOC_Os01g62870.2	OsAKR3.2	237	26665.56	6.24	28	25	49.29	Unstable	89.28	-0.317
LOC_Os01g62880.1	OsAKR4.1	334	33435.37	5.85	38	33	41.96	Unstable	90.46	-0.147
LOC_Os01g62880.2	OsAKR4.2	212	23660.14	5.76	27	23	47.47	Unstable	88.3	-0.245
LOC_Os02g03100.1	OsAKR5.1	388	43394.74	8.15	42	44	44.78	Unstable	88.99	-0.146
LOC_Os02g03100.2	OsAKR5.2	375	41802.8	8.16	42	44	42.88	Unstable	89.23	-0.147
LOC_Os02g57240.1	OsAKR6.1	328	36449.44	6.92	37	37	34.14	Stable	88.26	-0.27
LOC_Os03g13390.2	OsAKR7.1	318	35421.58	7.04	40	40	40.1	Unstable	82.86	-0.247
LOC_Os03g41510.1/2	OsAKR8.1/2	405	45663.95	8.13	47	49	37.6	Stable	83.28	-0.391
LOC_Os03g41510.3	OsAKR8.3	400	45192.4	8.13	47	49	38.37	Stable	82.62	-0.412
LOC_Os03g41510.4	OsAKR8.4	293	33261.1	8.78	31	35	45.04	Unstable	86.83	-0.317
LOC_Os03g41510.5/6/7	OsAKR8.5/6/7	285	32144.82	8.21	34	36	28.47	Stable	90.28	-0.287
LOC_Os03g41510.8	OsAKR8.8	267	30424.48	8.33	31	33	43.14	Unstable	78.46	-0.481

ADDITIONAL FILES

LOC_Os04g08550.1	OsAKR9.1	337	35765.81	6.1	35	29	35.92	Unstable	86.91	0.071
LOC_Os04g26870.1	OsAKR10.1	350	38118.8	6.52	41	39	33.04	Stable	84.49	-0.149
LOC_Os04g26910.1	OsAKR11.1	351	38233.54	6.03	44	39	37.58	Stable	86.5	-0.261
LOC_Os04g26920.1	OsAKR12.1	158	17011.53	7.76	16	17	36.99	Stable	82.15	-0.104
LOC_Os04g26920.2	OsAKR12.2	189	20179.35	7.76	21	22	31.37	Stable	94.5	0.019
LOC_Os04g26920.3	OsAKR12.3	201	21633.91	6.38	25	23	34.66	Stable	94.18	-0.043
LOC_Os04g27060.1	OsAKR13.1	355	38273.46	5.66	43	35	34.28	Stable	84.65	-0.121
LOC_Os04g37470.1	OsAKR14.1	333	36502.17	7.62	36	37	47.98	Unstable	89.82	-0.1
LOC_Os04g37480.1	OsAKR15.1	323	36123.47	5.8	43	38	44.73	Unstable	91.18	-0.184
LOC_Os04g37480.2	OsAKR15.2	284	31864.75	6.02	36	33	45.25	Unstable	92.64	-0.158
LOC_Os04g37490.1	OsAKR16.1	357	39541.57	8.6	42	45	44.01	Unstable	88.29	-0.202
LOC_Os05g38230.1	OsAKR17.1	334	37674.56	8.34	40	43	52.66	Unstable	93.68	-0.172
LOC_Os05g38230.2	OsAKR17.2	326	36029.45	6.42	37	34	44.01	Unstable	90.31	-0.131
LOC_Os05g38230.3	OsAKR17.3	197	21663.93	7.06	23	23	41.7	Unstable	88.58	-0.235
LOC_Os05g39690.1	OsAKR18.1	318	35589.69	6.32	43	39	28.47	Stable	83.43	-0.409
LOC_Os07g04990.1	OsAKR19.1	376	40553.12	6.68	43	43	33.72	Stable	85.45	-0.244
LOC_Os07g04990.2	OsAKR19.2	306	32964.55	7.6	34	35	32.62	Stable	88.4	-0.153
LOC_Os07g05000.1	OsAKR20.1	377	40601.25	8.51	40	43	26.99	Stable	89.6	-0.178
LOC_Os09g39390.1	OsAKR21.1	385	42217.35	7.64	43	44	41.01	Unstable	89.43	-0.191
LOC_Os10g02380.1	OsAKR22.1	320	35577.88	5.33	41	34	42.63	Unstable	86.25	-0.135
LOC_Os10g02480.1	OsAKR23.1	330	37016.64	6.27	42	39	33.29	Stable	81.27	-0.248
LOC_Os10g02480.2	OsAKR23.2	273	30552.58	7.56	29	30	39.94	Stable	84.36	-0.092
LOC_Os10g02490.1/2	OsAKR24.1/2	342	38024.61	5.64	43	37	41.7	Unstable	85.91	-0.188
LOC_Os10g28320.1	OsAKR25.1	343	37643.18	5.81	44	39	33.1	Stable	93.21	-0.136
LOC_Os10g28320.2/3	OsAKR25.2/3	249	27659.81	5.59	32	26	32.06	Stable	96.67	-0.019
LOC_Os10g28320.4	OsAKR25.4	153	16769.14	5.93	20	18	27.61	Stable	85.95	-0.165
LOC_Os10g37330.1	OsAKR26.1	375	41141.77	9.42	29	40	46.02	Unstable	90.29	-0.099
LOC_Os11g42540.1	OsAKR27.1	358	39976.64	6.27	41	37	58.25	Unstable	84.25	-0.249
LOC_Os11g42540.2	OsAKR27.2	339	37565.88	6.22	38	34	59.53	Unstable	82.36	-0.229

ADDITIONAL FILES

LOC_Os12g29760.1	OsAKR28.1	316	34230.23	5.33	40	31	37.21	Stable	99.34	0.008
LOC_Os12g29760.2	OsAKR28.2	236	25493.16	5.73	28	24	37.5	Stable	99.92	0.054

Table S4: Subcellular localization of *OsAKR* genes.

Gene Name	Localization	Gene Name	Localization
<i>OsAKR1.1</i>	nucleus	<i>OsAKR25.2</i>	cytoplasm
<i>OsAKR10.1</i>	chloroplast	<i>OsAKR25.3</i>	cytoplasm
<i>OsAKR11.1</i>	chloroplast	<i>OsAKR25.4</i>	extracellular space
<i>OsAKR12.1</i>	chloroplast outer membrane	<i>OsAKR26.1</i>	mitochondrion
<i>OsAKR12.2</i>	chloroplast outer membrane	<i>OsAKR27.1</i>	nucleus
<i>OsAKR12.3</i>	chloroplast outer membrane	<i>OsAKR27.2</i>	nucleus
<i>OsAKR13.1</i>	mitochondrion	<i>OsAKR28.1</i>	cytoplasm
<i>OsAKR14.1</i>	chloroplast	<i>OsAKR28.2</i>	nucleus
<i>OsAKR15.1</i>	cytoplasm	<i>OsAKR3.1</i>	cytoplasm
<i>OsAKR15.2</i>	cytoplasm	<i>OsAKR3.2</i>	nucleus
<i>OsAKR16.1</i>	chloroplast	<i>OsAKR4.1</i>	cytoplasm
<i>OsAKR17.1</i>	chloroplast	<i>OsAKR4.2</i>	nucleus
<i>OsAKR17.2</i>	extracellular space	<i>OsAKR5.1</i>	chloroplast
<i>OsAKR17.3</i>	cytoplasm	<i>OsAKR5.2</i>	chloroplast
<i>OsAKR18.1</i>	cytoplasm	<i>OsAKR6.1</i>	chloroplast
<i>OsAKR19.1</i>	mitochondrion	<i>OsAKR7.2</i>	nucleus
<i>OsAKR19.2</i>	mitochondrion	<i>OsAKR8.1</i>	chloroplast
<i>OsAKR2.1</i>	cytoplasm	<i>OsAKR8.2</i>	chloroplast
<i>OsAKR20.1</i>	mitochondrion	<i>OsAKR8.3</i>	chloroplast
<i>OsAKR21.1</i>	chloroplast outer membrane	<i>OsAKR8.4</i>	chloroplast
<i>OsAKR22.1</i>	cytoplasm	<i>OsAKR8.5</i>	nucleus
<i>OsAKR23.1</i>	cytoplasm	<i>OsAKR8.6</i>	nucleus
<i>OsAKR23.2</i>	cytoplasm	<i>OsAKR8.7</i>	nucleus
<i>OsAKR24.1</i>	chloroplast	<i>OsAKR8.8</i>	chloroplast
<i>OsAKR24.2</i>	chloroplast	<i>OsAKR9.1</i>	chloroplast
<i>OsAKR25.1</i>	cytoplasm		

Table S5: Promoter analysis of 1kb upstream promoters of *OsAKR* genes.

Functional Classification	Motif Name
Light Response	A-box
	Box 4
	GATA-motif
	G-box
	GC-motif
	GT1-motif
	Sp1
	TCCC-motif

ACE
I-box

Defense and Stress Response

as-1
LTR
STRE
DRE core
WRE3
ABRE3a
ABRE4
ATCT-motif
W box
GARE-motif

Hormone response

ABRE
CGTCA-motif
MYC
TCA-element
TGACG-motif
ERE

Growth & Development

CAT-box
O ₂ -site
AC-I

Table S6: Gene duplications of *OsAKR* genes.

Dispersed			Transposed		
Genes		E-value	Genes		E-value
<i>OsAKR1</i>	<i>OsAKR25</i>	3.00E-173	<i>OsAKR2</i>	<i>OsAKR5</i>	3.00E-60
<i>OsAKR2</i>	<i>OsAKR4</i>	1.00E-148	<i>OsAKR2</i>	<i>OsAKR18</i>	1.00E-76
<i>OsAKR3</i>	<i>OsAKR17</i>	6.00E-151	<i>OsAKR6</i>	<i>OsAKR8</i>	1.00E-33
<i>OsAKR4</i>	<i>OsAKR17</i>	2.00E-132	<i>OsAKR6</i>	<i>OsAKR27</i>	3.00E-33
<i>OsAKR5</i>	<i>OsAKR3</i>	3.00E-64	<i>OsAKR6</i>	<i>OsAKR28</i>	1.00E-21
<i>OsAKR6</i>	<i>OsAKR25</i>	5.00E-32	<i>OsAKR7</i>	<i>OsAKR9</i>	2.00E-92
<i>OsAKR7</i>	<i>OsAKR24</i>	2.00E-140	<i>OsAKR21</i>	<i>OsAKR26</i>	2.00E-13
<i>OsAKR8</i>	<i>OsAKR27</i>	1.00E-98			
<i>OsAKR9</i>	<i>OsAKR22</i>	4.00E-91			
<i>OsAKR10</i>	<i>OsAKR11</i>	0			
<i>OsAKR11</i>	<i>OsAKR13</i>	0			
<i>OsAKR12</i>	<i>OsAKR10</i>	2.00E-114			

Proximal		
Genes		E-value
<i>OsAKR12</i>	<i>OsAKR13</i>	3.00E-103

<i>OsAKR13</i>	<i>OsAKR25</i>	2.00E-128
<i>OsAKR14</i>	<i>OsAKR16</i>	4.00E-130
<i>OsAKR15</i>	<i>OsAKR7</i>	2.00E-122
<i>OsAKR16</i>	<i>OsAKR22</i>	1.00E-119
<i>OsAKR17</i>	<i>OsAKR18</i>	7.00E-82
<i>OsAKR19</i>	<i>OsAKR26</i>	5.00E-78
<i>OsAKR20</i>	<i>OsAKR26</i>	7.00E-76
<i>OsAKR21</i>	<i>OsAKR19</i>	2.00E-18
<i>OsAKR22</i>	<i>OsAKR23</i>	5.00E-125
<i>OsAKR23</i>	<i>OsAKR15</i>	7.00E-115
<i>OsAKR24</i>	<i>OsAKR16</i>	5.00E-113
<i>OsAKR25</i>	<i>OsAKR26</i>	1.00E-25
<i>OsAKR27</i>	<i>OsAKR25</i>	2.00E-27
<i>OsAKR28</i>	<i>OsAKR25</i>	5.00E-19

Whole Genome		
Genes		E-value
<i>OsAKR2</i>	<i>OsAKR17</i>	2.00E-145
<i>OsAKR7</i>	<i>OsAKR22</i>	9.00E-143

Tandem		
Genes		E-value
<i>OsAKR2</i>	<i>OsAKR3</i>	0
<i>OsAKR3</i>	<i>OsAKR4</i>	1.00E-149
<i>OsAKR11</i>	<i>OsAKR12</i>	2.00E-127
<i>OsAKR14</i>	<i>OsAKR15</i>	2.00E-127
<i>OsAKR15</i>	<i>OsAKR16</i>	0.00E+00
<i>OsAKR19</i>	<i>OsAKR20</i>	0.00E+00
<i>OsAKR23</i>	<i>OsAKR24</i>	1.00E-178

Table S7: Interactome analysis of OsAKR proteins.

S. No.	Description	Common	Genes
1	Os01g0183800 protein	OS01T0183800-01	1 & 25
2	Os01g0559600 protein; Putative C13 endopeptidase NP1	OS01T0559600-01	1 & 25
3	NAD(P)H-hydrate epimerase	OS10T0377800-01	11,25,10, 27, 1, 13, 19, 20, 26
4	Histidinol dehydrogenase, chloroplatic; Catalyzes the sequential NAD-dependent oxidations of L- histidinol to L-histidinaldehyde and then to L-histidine	HDH	2,3,4,17,18
5	Putative aldose reductase	OsJ_26534	2,4
6	3,4-dihydroxy-2-butanone kinase; Putative DAK2 domain containing protein	OsJ_12371	14,15,16,2,4,9,23,24
7	Glucose and ribitol dehydrogenase homolog; May act as a short alcohol-polyol-sugar dehydrogenase possibly related to carbohydrate metabolism and the acquisition of desiccation tolerance. May also be involved in signal transduction (By similarity)	OS05T0140800-01	2,3,4,17,18
8	Putative uncharacterized protein	B1036C05.9	2,4,9,14,15,16,22,23,24
9	4-alpha-glucanotransferase DPE2; Cytosolic alpha-glucanotransferase essential for the cytosolic metabolism of maltose, an intermediate on the pathway by which starch is converted to sucrose in leaves at night	DPE2	2,3,4,5
10	Os08g0545200 protein; Putative sorbitol dehydrogenase	OsJ_28161	16,17,18,2,4,5,14,22,23,24,3,9,15
11	Os10g0403700 protein; Putative oxidoreductase	OS10T0403700-00	2,4,9,14,15,16,23,24
12	Phosphoglycerate kinase ; Belongs to the phosphoglycerate kinase family	OS10T0442100-00	2,3,4,9,14,15,16,17,22,23,24
13	GSI-like protein; HAD-superfamily hydrolase, subfamily IA, variant 3 containing protein, expressed; Os10g0464400 protein; Putative glutamine synthetase	GLN	2,4,17,18
14	Lactoylgutathione lyase; Catalyzes the conversion of hemimercaptal, formed from methylglyoxal and glutathione, to S-lactoylgutathione (GlyI-7)	OS05T0230900-01	3,15
15	Lactoylgutathione lyase; Catalyzes the conversion of hemimercaptal, formed from methylglyoxal and glutathione, to S-lactoylgutathione (GlyI-8)	OS05T0295800-01	3,9,14,15,16,23,24
16	annotation not available	OS08T0148267-00	2,4,9,14,15,16,23,24
17	MATE efflux family protein, expressed ; Belongs to the multi antimicrobial extrusion (MATE) (TC 2.A.66.1) family	OsJ_11501	22,7

18	Ferredoxin; Ferredoxins are iron-sulfur proteins that transfer electrons in a wide variety of metabolic reactions	OS03T0659200-01	22,7
19	Os01g0174300 protein; Putative cytochrome b5 reductase; Uncharacterized protein ; Belongs to the flavoprotein pyridine nucleotide cytochrome reductase family	OsJ_00570	10,17,18
20	annotation not available	OS01T0627933-00	10,11
21	Os05g0488900 protein; Putative NADH-cytochrome b5 reductase	OS05T0488900-01	10,17,18
22	Probable aldo-keto reductase 3 ; Belongs to the aldo/keto reductase family	OsJ_14323	13,22
23	Probable NAD(P)H-dependent oxidoreductase 1; May play a role in auxin-induced cell growth by generating hydroxyl radicals, which tends to increase cell wall loosening	OS10T0113000-01	13,22
24	Glyoxalase family protein, expressed; Os03g0659300 protein (GlyI-4)	OS03T0659300-01	9,14,15,23
25	Os01g0174300 protein; Putative cytochrome b5 reductase; Uncharacterized protein ; Belongs to the flavoprotein pyridine nucleotide cytochrome reductase family	OsJ_00570	10,17,18
26	Os03g0850000 protein; Oxidoreductase, short chain dehydrogenase/reductase family protein	OS03T0850000-01	17,18
27	cDNA clone:001-207-F04	OS05T0456300-01	17,18
28	Os05g0488900 protein; Putative NADH-cytochrome b5 reductase	OS05T0488900-01	10,17,18
29	Putative P-glycoprotein 1	OS08T0564300-04	17,18
30	Aspartate kinase-homoserine dehydrogenase; Os09g0294000 protein	OS09T0294000-01	17,18
31	Thioredoxin-like protein CDSP32, chloroplastic; Thiol-disulfide oxidoreductase that may participate in various redox reactions. May act as electron donor to the BAS1 peroxiredoxin. Possesses low insulin disulfide bonds reducing activity	CDSP32	19,20
32	Os10g0474800 protein	OS10T0474100-00	19,20
33	annotation not available	OS10T0474800-00	19,20
34	GMC oxidoreductase family protein, expressed; Putative oxidase	OS10T0475100-00	19,20
35	Putative transaldolase ToTAL2	OsJ_26073	14,15,23

*Tansley insight*

Reassessing plant glyoxalases: large family and expanding functions

Author for correspondence:

Sneh L. Singla-Pareek

Tel: +91 11 26741358

Email: sneh@icgeb.res.in

Received: 27 December 2019

Accepted: 7 March 2020

Sneh L. Singla-Pareek¹ , Charanpreet Kaur² , Brijesh Kumar¹ , Ashwani Pareek² and Sudhir K. Sopory¹¹Plant Stress Biology, International Centre for Genetic Engineering and Biotechnology, New Delhi 110067, India; ²Stress Physiology and Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

Contents

Summary			
I. Introduction	714	VI. Discovering new functions of plant glyoxalases: taking leads from other organisms	718
II. Role of glyoxalases: classical function in stress adaptation	715	VII. Conclusions and future perspectives	719
III. Multiple isoforms of glyoxalases in plants offer scope for novel functions	716	Acknowledgements	719
IV. Evolutionary aspects related to emergence of glyoxalases as a multi-membered family in plant systems	717	References	719
V. Different shades of methylglyoxal: cytotoxin or a signal molecule	718		

New Phytologist (2020) 227: 714–721

doi: 10.1111/nph.16576

Key words: aging, glycation, glyoxalase, methylglyoxal, plants, stress response.

Summary

Methylglyoxal (MG), a reactive carbonyl compound, is generated during metabolism in living systems. However, under stress, its levels increase rapidly leading to cellular toxicity. Although the generation of MG is spontaneous in a cell, its detoxification is essentially catalyzed by the glyoxalase enzymes. In plants, modulation of MG content via glyoxalases influences diverse physiological functions ranging from regulating growth and development to conferring stress tolerance. Interestingly, there has been a preferred expansion in the number of isoforms of these enzymes in plants, giving them high plasticity in their actions for accomplishing diverse roles. Future studies need to focus on unraveling the interplay of these multiple isoforms of glyoxalases possibly contributing towards the unique adaptability of plants to diverse environments.

I. Introduction

Cellular metabolic rewiring is an important mechanism regulating cell proliferation and growth. This tuning occurs not only under stress but also during various physiological functions and triggers the formation of a reactive dicarbonyl compound, methylglyoxal (MG), which spontaneously glycates nucleic acids, proteins and lipids, forming advanced glycation end products (AGEs) (Thornalley, 2008). MG, which is generated from triosephosphate sugars (Box 1), is an unavoidable outcome of primary metabolism contributing to aging and disease in humans (Richard, 1993).

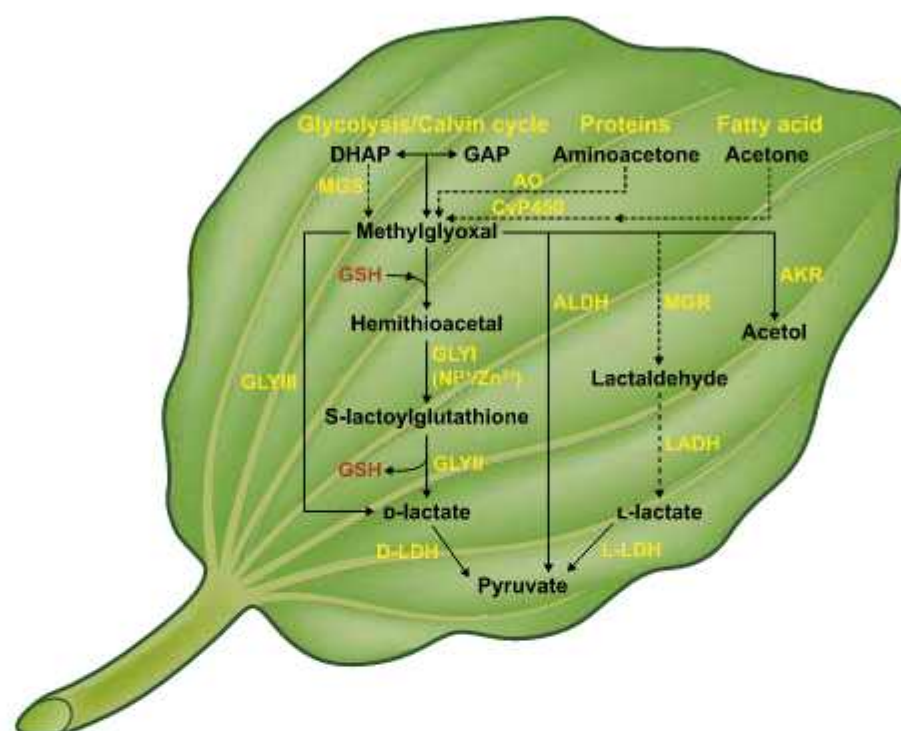
Several conditions such as diabetes, cardiovascular disease, cancer and neurodegenerative disorders are associated with an increase in MG content (Nigro *et al.*, 2019). Likewise, MG is also a universal consequence of stress in plants (Yadav *et al.*, 2005a), affecting their growth and development (Box 1).

Because MG toxicity severely affects cellular functions, nature has evolved a highly efficient detoxification system for its metabolism via the glyoxalase pathway (Thornalley, 1993). This two-step enzyme system comprises glyoxalase I (GLYI) that catalyzes conversion of hemithioacetal, which is spontaneously generated by adduct formation between reduced glutathione

Box 1

Generation of MG

In plants, MG is generated primarily from triosephosphate sugars produced during glycolysis and the Calvin cycle through nonenzymatic reactions. In addition, enzymatic pathways have also been proposed to generate MG but have not yet been validated in plants. These include MG synthase (MGS; involved in glycolytic bypass in prokaryotes), cytochrome P450 (CYP450; acetone metabolism) and amine oxidase (AO; amino acid breakdown).



Detoxification of MG

Glyoxalase pathway: The major route of MG breakdown in plants is the glyoxalase pathway comprising the enzymes glyoxalase I (GLYI; lactoylglutathione lyase, EC 4.4.1.5) and glyoxalase II (GLYII; hydroxyacylglutathione hydrolase, EC 3.1.2.6). MG spontaneously reacts with glutathione (GSH) to form a hemithioacetal adduct which is metabolized by GLYI to form S-lactoylglutathione (SLG). GLYI is a metalloenzyme requiring divalent metal ions, either Ni²⁺ or Zn²⁺, for activation. Subsequently, GLYII converts SLG to D-lactate, recycling GSH in the process. The D-lactate so formed is finally converted to pyruvate through D-lactate dehydrogenase (D-LDH), which then enters the Krebs cycle.

Glyoxalase III (GLYIII): GLYIII enzymes (EC 4.2.1.130) can directly convert MG to D-lactate in one step without requiring GSH and divalent metal ions and no SLG is formed in the process. Besides GLYIII activity, these enzymes also possess other catalytic functions such as chaperone and protease activities.

Other enzymes: Aldoketo reductases (AKRs), aldehyde dehydrogenases (ALDHs) and methylglyoxal reductases (MGRs) can also metabolize MG. While AKRs form acetol from MG, ALDHs form pyruvate and MGRs convert MG to lactaldehyde which is subsequently converted to L-lactate via lactaldehyde dehydrogenases (LADH). L-Lactate is finally converted to pyruvate through the enzyme L-lactate dehydrogenase (L-LDH). The pyruvate thus formed can enter the Krebs cycle. Unlike the glyoxalase pathway, these enzymes have broad substrate specificity and can catalyze reduction of various ketones and aldehydes.

The pathways shown by dotted lines have not yet been established in plant systems.

(GSH) and MG, to S-lactoyl glutathione (SLG) involving divalent metal ions as cofactors. Glyoxalase II (GLYII) further converts SLG to D-lactate with the release of GSH (Box 1). In addition, a shorter route for detoxification of MG also exists, which involves GLYIII proteins directly catalyzing the conversion of MG to D-lactate in one step without involving GSH or metal ions (Misra *et al.*, 1995). Furthermore, aldo-keto

reductases and aldehyde dehydrogenases can also metabolize MG but, unlike GLYI-GLYII, they have broad substrate specificity (reviewed by Chakraborty *et al.*, 2014). By facilitating MG-mediated damage control, glyoxalases and other MG-metabolizing enzymes can be said to be components of the metabolite damage repair machinery of cells (Hanson *et al.*, 2016).

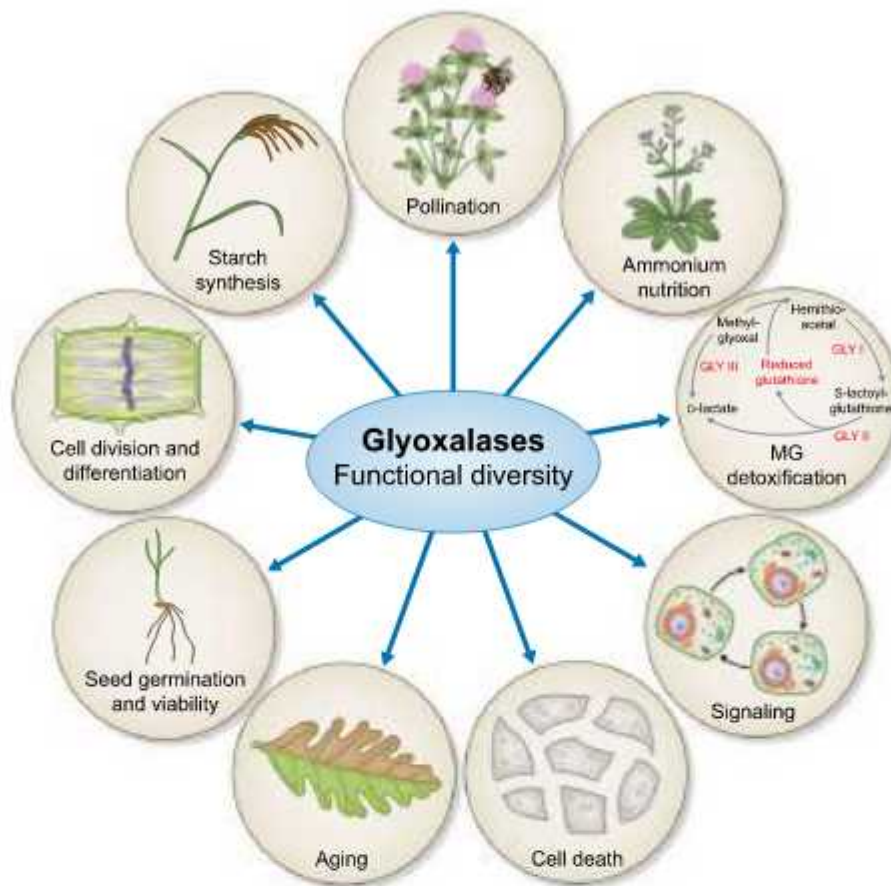


Fig. 1 Glyoxalase (GLY) enzymes show functional diversity in plants. MG, methylglyoxal.

The role of the MG–glyoxalase module in plant stress adaptation has been unambiguously established (Singla-Pareek *et al.*, 2003; Gupta *et al.*, 2018). Herein, we highlight the expansion of glyoxalases as multiple isoforms, a feature possibly linked to sessile organisms such as plants, in mediating a robust response to diverse environmental stresses. We also discuss newer roles of plant glyoxalases, including some recent experimental evidence (Sankaranarayanan *et al.*, 2015; You *et al.*, 2019), to appreciate the need for multiplicity in executing cellular and physiological functions. Altogether, we believe that glyoxalases have a far more impactful role in living systems given the direct and critical association of their substrate MG with glycolysis, an essential component of the metabolic machinery.

II. Role of glyoxalases: classical function in stress adaptation

Discovered in 1913 as a single enzyme, glyoxalases were initially considered as a mainstream enzyme of glycolysis converting MG to lactate. However, with elucidation of the Embden–Meyerhof scheme of glucose breakdown and the fact that MG metabolism by glyoxalase yielded D-lactate and not L-lactate, which is the end product of glycolysis, studies on this enzyme–substrate module decreased. In 1965 Szent-Georgyi revived interest in MG–glyoxalase research with the proposition of the promine–retine theory, identifying the role of glyoxalases in cell division (Szent-Georgyi, 1965).

Glyoxalase activity was first demonstrated in plants in 1981 in Douglas fir needles, and this initiated plant glyoxalase research. In line with Szent-Georgyi's promine–retine theory regarding the MG–glyoxalase module, correlations between GLYI activity and cell proliferation were established in several plant species (reviewed by Kaur *et al.*, 2014a) but due to a lack of experimental evidence on the relationship underlying enzyme activity and cell division, no functional implications could be drawn.

An initial report from our laboratory led to elucidation of the most significant and widely accepted role of this pathway in stress adaptation response in plants (Singla-Pareek *et al.*, 2003). We reported that MG levels increase following exposure to salinity stress (Yadav *et al.*, 2005a), which could be resisted by overexpression of glyoxalase enzymes in transgenic systems (Yadav *et al.*, 2005b). Because MG levels are generally induced under stress due to alterations in metabolic flux, the glyoxalase pathway serves an important detoxification role by converting this cytotoxin to D-lactate, which can subsequently enter the Krebs cycle as pyruvate via D-lactate dehydrogenases. Overexpression of glyoxalase genes, either individually (GLYI/GLYII) or as a complete pathway (GLYI + GLYII), in plant species such as rice, tobacco, tomato and Carrizo citrange rootstock has resulted in plants with improved tolerance to salinity, drought, heavy metals and oxidative stresses (reviewed by Kaur *et al.*, 2014a). In addition, enhancing MG detoxification in plants via glyoxalase overexpression also improves tolerance to biotic stresses (Gupta *et al.*, 2018).

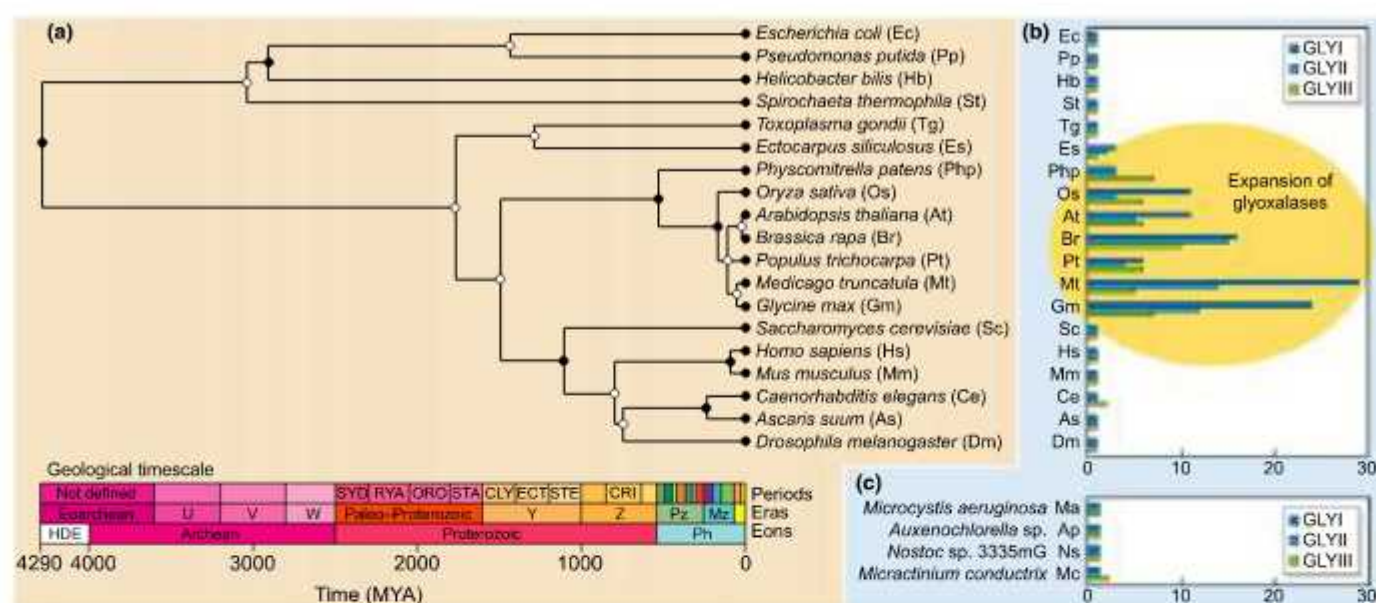


Fig. 2 Glyoxalase family beyond plant systems. (a) Time tree of species used in the present study. Geological timescale is in million years ago (MYA). The tree has been created using TIME TREE (Kumar *et al.*, 2017). Periods: SYD, Syderian; RYA, Ryacian; ORO, Orosirian; STA, Statherian; CLY, Calymmian; ECT, Ectasian; STE, Stenian; CRI, Cryogenian. Eras: U, Paleoproterozoic; V, Mesoarchean; W, Neoarchean; Y, Meso-Proterozoic; Z, Neo-Proterozoic; Pz, Paleozoic; Mz, Mesozoic. Eons: HDE, Hadean; Ph, Phanerozoic. (b) Graph indicating the number of glyoxalase genes (GLYI, GLYII, GLYIII) in various representative species from prokaryotes to eukaryotes. (c) Graph indicating the number of glyoxalase genes (GLYI, GLYII, GLYIII) in algal species.

III. Multiple isoforms of glyoxalases in plants offer scope for novel functions

Genome-wide analysis has identified multiple isoforms of GLYI, GLYII and GLYIII in rice (Mustafiz *et al.*, 2011; Ghosh *et al.*, 2016). Evolutionary studies have also established their presence as ubiquitous enzymes across all the kingdoms of life and have also highlighted the essential acquisition of both metal-dependent forms of GLYI in plants (Kaur *et al.*, 2013). Interestingly, along with these different metal-dependent forms (Ni^{2+} and Zn^{2+}), plants also possess GLYI-like proteins, which have probably diverged in their functional roles during evolution (Kaur *et al.*, 2014b; Schmitz *et al.*, 2018). One such case is the OsGLYII-1 protein, a member of the rice GLYII family that lacks GLYII activity and instead encodes a stress-responsive putative sulphur dioxygenase function (Kaur *et al.*, 2014b). Similarly, AtGLX2-3 from Arabidopsis, the OsGLYII-1 ortholog, also encodes sulphur dioxygenase activity, being involved in amino acid catabolism during carbohydrate starvation and embryo development in Arabidopsis (Krübel *et al.*, 2014). Like GLYI/GLYII genes, there has been an expansion in the GLYIII family of plants as well, a family which exhibits multiple functions apart from MG detoxification such as chaperone and protease activities (Chen *et al.*, 2010; Ghosh *et al.*, 2016; Lewandowska *et al.*, 2019). A more important role of these enzymes is speculated in conditions where the GLYI/GLYII system is less effective due to reduced levels of GSH, such as chronic oxidative stress, sulfur limitation and stationary growth phase (Hasim *et al.*, 2014). However, much remains to be discovered about these proteins and discussion of them warrants a separate review.

The ubiquitous nature of MG and of glyoxalases as multiple members in plants indicates that the functionally active glyoxalases have more important physiological roles where, by manipulating MG levels, they indirectly dictate outcome of a process. In this regard, novel roles beyond stress response but involving metabolic tuning are being reported. For instance, the requirement of GLYI activity for regulating pollen–pistil interactions has been shown in *Brassica napus* (Sankaranarayanan *et al.*, 2015). Increased glycolysis to generate energy for growing pollen tube leads to a surge in MG levels in stigma, which in turn correlates with changes in glyoxalase activity. Consequently, increased GLYI activity is needed for pollen germination on stigma. Moreover, compound starch granule formation and starch synthesis in the rice endosperm also involves GLYI (You *et al.*, 2019). The involvement of glyoxalases in nutritional responses has also been reported, as indicated by enhanced MG generation during NH_4^+ nutrition in plants (Borysiuk *et al.*, 2018) (Fig. 1).

IV. Evolutionary aspects related to emergence of glyoxalases as a multi-membered family in plant systems

The search for glyoxalase proteins from prokaryotes to eukaryotes has indicated that the glyoxalase family expanded exclusively in plant systems (Fig. 2a,b). This expansion probably resulted from several gene/genome duplication events which have together shaped plant development and the modular lifestyle of plants and, consequently, have driven trait evolution (Schrantz *et al.*, 2012). The emergence of photosynthesis and loss of motility are two such plant-specific traits, although not the only ones, but may have contributed significantly towards driving such expansion.

Photosynthesis is one of the major sources of MG generation in plants and appears as a more likely cause behind expansion of glyoxalases. However, analysis of GLY genes in blue-green algal genomes, which are also photosynthetic organisms, reveals no support for this hypothesis (Fig. 2c). These algal species possess a single copy of GLY genes and thus clearly do not support a link for the cause of evolutionary expansion of multiple glyoxalases with photosynthesis. However, to some extent, expansion of glyoxalases as multigene families appears to be linked to loss of motility in plants, as motile organisms such as animals and humans harbor only a single copy of glyoxalases. A sessile nature predisposes plants to adverse environmental conditions from which they cannot escape. The presence of multiple gene copies may therefore act as a buffering mechanism providing redundancy in gene functions in order to boost plant defenses and other vital functions. In a similar context, GLYIII genes have also undergone expansion in their numbers in plants (Fig. 2a,b). Notably, the moss *Physcomitrella patens*, where stomata first evolved, shows the highest number of GLYIII genes (Fig. 2b).

V. Different shades of methylglyoxal: cytotoxin or a signal molecule

The toxicity of MG is well known to be due to its irreversible effects on protein structure and function (reviewed by Thornalley, 2008). MG-mediated glycation reactions directed to arginine and lysine residues have critical implications, as arginine is more commonly located in the functional sites of proteins and may lose its ligand-binding ability upon reacting with MG (Ahmed *et al.*, 2005). The formation of hydroimidazolones (MG-Hs), as in human serum albumin and collagen, causes structural distortion, loss of side chain charge and functional impairment (Ahmed *et al.*, 2005; Dobler *et al.*, 2006). Generally, hydroimidazolones have short chemical half-lives (2–6 wk) under physiological conditions, and their concentration depends on the balance between their rates of formation and decomposition.

In terms of morphological changes, we have shown that plants with high endogenous MG levels under stress conditions exhibit growth and development defects (Singla-Pareek *et al.*, 2003, 2006, 2008; Kaur *et al.*, 2017; Gupta *et al.*, 2018). Furthermore, MG at 1 mM inhibits seed germination and root elongation and induces chlorosis in *Arabidopsis* (Hoque *et al.*, 2012). In view of MG-mediated glycation, researchers have identified 'dicarbonyl proteomes' in animals and plants, comprising proteins susceptible to MG-mediated modification. In humans, these include hemoglobin, transcription factors and mitochondrial proteins, which may be associated with mitochondrial dysfunction in disease, aging, stress and apoptosis (Rabbani & Thornalley, 2012). In plants, studies of glycation hotspots in *Arabidopsis*, *B. napus* and recently in legume nodules indicate heavy glycation of plant proteomes in comparison to animals (Shumilina *et al.*, 2019) and that stress conditions such as drought, heat and light primarily target glycations to the chloroplast, the site of production of higher levels of sugars (Chaplin *et al.*, 2019).

However, evidence indicates that MG, similar to reactive oxygen species (ROS), can serve dual roles in organisms in a dose-dependent manner. At low doses, MG priming has been shown to

induce salt and cadmium tolerance in wheat seedlings (Li *et al.*, 2017, 2018a). The signaling role of MG has, in fact, been clearly demonstrated in yeast where it acts via the HOG-MAPK pathway, facilitating an osmotic stress response. In plants, initial indication towards a signaling role of MG has been reported from our laboratory based on assessment of the rice transcriptome in response to MG. By triggering large-scale perturbations in signaling-related genes, MG is proposed to act as a novel stress signal molecule (Kaur *et al.*, 2015). It is now known that MG-mediated stress adaptation in plants may act via a complex cascade of various signaling molecules such as Ca^{2+} , ROS, K^+ and ABA (reviewed by Kaur *et al.*, 2014a; Mostofa *et al.*, 2018). In addition, hydrogen sulphide can also mediate MG priming-induced heat tolerance in maize (Li *et al.*, 2018b).

Likewise, protein carbonylation mediated by MG and other dicarbonyl compounds occurs not only as a consequence of external oxidative stress but also as a part of plant development. A specific set of embryo proteins have been shown to be carbonylated during release of seed dormancy in sunflower (Oracz *et al.*, 2007). Furthermore, protein carbonylation occurs even during germination of fully viable *Arabidopsis* seeds (Job *et al.*, 2005). Depending on dose and duration, glycation can thus either enhance or inhibit protein degradation and thereby facilitate reprogramming of cellular functions. Altogether, MG and MG-mediated glycations do not have just a lethal effect on plant functions but also serve important physiological roles as well.

VI. Discovering new functions of plant glyoxalases: taking leads from other organisms

Plant diabetes

In humans, dicarbonyl stress is considered to be one of the drivers of pathogenesis in obesity and diabetes, particularly in associated vascular complications. MG levels increase under both conditions due to hyperglycemia in diabetes and increased triosephosphate flux in glyceroneogenesis, associated with adipocyte expansion. These conditions also lead to downregulation of GLYI expression, which augments accumulation of MG-derived AGEs (Rabbani & Thornalley, 2018). In this context, GLYI has been found to cooperate with fatty acid synthesis to protect animals against sugar toxicity, as evident from the fact that lipogenesis-deficient animals are highly sensitive to dietary sugar due to accumulation of MG-derived AGEs (Garrido *et al.*, 2015). A similar correlation can be reasonably assumed to be true for plant glyoxalases as well. In fact, Takagi *et al.* (2014) suggest that plants are likely to suffer from diabetes-like conditions in environments where photosynthesis is enhanced, as in high light or high CO_2 concentrations. This is because, under conditions which promote photosynthesis, higher levels of MG are produced in chloroplasts due to increased flux of the Calvin cycle. In the day, MG is preferentially reduced by photosystem I (PSI) without evolving O_2 , but reducing it to O_2^- (Saito *et al.*, 2011). The oxidative stress that results has been referred to as 'Plant diabetes' by the authors as it originates from a common metabolite of the primary pathways of sugar metabolism in all organisms.

Cell death

MG is a well-known inducer of apoptosis in animals, which acts by triggering p38 mitogen-activated protein kinase and involves oxidative stress (Fukunaga *et al.*, 2005). In human leukocytes, this phenomenon is accompanied by morphological changes and alterations in expression of antioxidant, apoptotic and glycation-responsive genes (de Souza Prestes *et al.*, 2019). In this regard, the GLYII gene has been shown to be a pro-survival factor of the p53 family, the classic tumor suppressor gene (Xu & Chen, 2006). Furthermore, the apoptotic process in animals and programmed cell death (PCD) pathways in plants commonly involve cytochrome C (CYTc). This offers another possible link between MG and apoptosis because D-lactate dehydrogenase, which specifically metabolizes D-lactate (the product of GLYII) requires CYTc as an electron acceptor (Welchen *et al.*, 2016). In addition, hexokinase activity (glucose metabolism) is also known to control CYTc release, linking MG with PCD (Kim *et al.*, 2006). In support of this, we have shown that plants overexpressing the glyoxalase pathway exhibit a 'stay-green' phenotype under stress conditions that would otherwise induce senescence (Gupta *et al.*, 2018). Hence, glyoxalases can possibly regulate various developmental processes that involve PCD in plants such as xylem formation, seed maturation and leaf senescence (Fig. 1).

Aging

An age-related decline in GLYI activity has been observed in arterial tissue, human lens and brain (Xue *et al.*, 2011). A decline in GLYI activity leads to increased formation of AGEs in *Caenorhabditis elegans*, resulting in mitochondrial ROS production which thereby restricts lifespan (Morcos *et al.*, 2008). Furthermore, MG-mediated loss of proteasomal activity or impaired proteostasis due to the formation of proteolysis-resistant AGEs can also contribute to aging (Nigro *et al.*, 2019). On similar lines, age-related changes have also been observed in the Arabidopsis glycated proteome (Bilova *et al.*, 2017) and also during natural leaf senescence in winter wheat (Havé *et al.*, 2015), indicating the existence of age-related glycation hotspots. Together, glycation may reasonably be termed as a nonenzymatic post-translational modification marker guiding proteins to degradation and organs to senescence. Plant glyoxalases, by affecting glycations, may therefore serve as critical components regulating the aging process (Fig. 1).

VII. Conclusions and future perspectives

Energy adjustments occurring during diverse cellular processes such as stress adaptation and development, generate MG in the process. By facilitating its detoxification, plant glyoxalases can be conveniently assigned diverse physiological roles, being involved in the detox pathway central to sugar metabolism and as crucial components of metabolite damage repair systems (Fig. 1). Here, although we have ascribed various functional roles of these plant glyoxalases by drawing parallels from other kingdoms, the same

need to be validated in future to determine the actual significance of these enzymes in plant physiology.

Future research on glyoxalases may focus on providing evidence to confirm the proposed role of the MG–glyoxalase module in processes such as apoptosis and aging. Furthermore, although the stress-mitigating role of glyoxalases has been well established, little is presently known regarding the interplay of different isoforms and the signaling mechanisms that contribute to these stress adaptation responses. In addition, GLYIII proteins, which are novel but less well-studied enzymes of MG detoxification, also merit detailed investigations, especially in context of their multiple roles. We believe that a comprehensive assessment of the two pathways, GLYI/GLYII and GLYIII, with respect to their regulation will delineate the factors and conditions that govern which pathway for MG detoxification is activated.

Acknowledgements

We apologize to all those colleagues whose work has not been cited because of space constraints but have immensely contributed to our understanding of glyoxalases in plants. SLS-P acknowledges funds received from ICGEB. CK and BK acknowledge DST-INSPIRE Faculty Award (IFA-14/LSPA-24) from DST and Senior Research Fellowship from DBT, Government of India, respectively. SKS is grateful to the Science and Engineering Research Board (SERB) for the award of a Distinguished fellowship. AP and SLS-P would like to acknowledge the funds received from IAEA CRP D23031.

Author contributions

SLS-P and AP conceived the study, CK and BK wrote the manuscript, and SLS-P, SKS and AP finalized the manuscript.

ORCID

Charanpreet Kaur  <https://orcid.org/0000-0003-1542-0131>
Brijesh Kumar  <https://orcid.org/0000-0001-7461-4660>
Ashwani Pareek  <https://orcid.org/0000-0002-2923-0681>
Sneh L. Singla-Parcek  <https://orcid.org/0000-0002-0521-2622>

References

- Ahmed N, Dobler D, Dean M, Thornalley PJ. 2005. Peptide mapping identifies hotspot site of modification in human serum albumin by methylglyoxal involved in ligand binding and esterase activity. *Journal of Biological Chemistry* **280**: 5724–5732.
- Bilova T, Paudel G, Shilyaev N, Schmidt R, Brauch D, Tarakhovskaya E, Milrud S, Smolikova G, Tissier A, Vogt T *et al.* 2017. Global proteomic analysis of advanced glycation end products in the Arabidopsis proteome provides evidence for age-related glycation hot spots. *Journal of Biological Chemistry* **292**: 15758–15776.
- Borysiuk K, Ostaszewska-Bugajska M, Vaultier MN, Hasenfratz-Sauder MP, Szal B. 2018. Enhanced formation of methylglyoxal-derived advanced glycation end products in *Arabidopsis* under ammonium nutrition. *Frontiers in Plant Science* **9**: 667.

- Chakraborty S, Karmakar K, Chakravorty D. 2014. Cells producing their own nemesis: Understanding methylglyoxal metabolism. *TUBMB Life* 66: 667–678.
- Chaplin AK, Chernukhin I, Bechtold U. 2019. Profiling of advanced glycation end products uncovers abiotic stress-specific target proteins in *Arabidopsis*. *Journal of Experimental Botany* 70: 653–670.
- Chen J, Li L, Chin LS. 2010. Parkinson disease protein DJ-1 converts from a zymogen to a protease by carboxyl-terminal cleavage. *Human Molecular Genetics* 19: 2395–2408.
- de Souza Prestes A, dos Santos MM, Ecker A, de Macedo GT, Fachinetto R, Bressan GN, da Rocha JB, Barbosa NV. 2019. Methylglyoxal disturbs the expression of antioxidant, apoptotic and glycation responsive genes and triggers programmed cell death in human leukocytes. *Toxicology in Vitro* 55: 33–42.
- Dobler D, Ahmed N, Song L, Eboighodin KE, Thornalley PJ. 2006. Increased dicarbonyl metabolism in endothelial cells in hyperglycemia induces anoikis and impairs angiogenesis by RGD and GFOGER motif modification. *Diabetes* 55: 1961–1969.
- Fukunaga M, Miyata S, Higo S, Hamada Y, Ueyama S, Kasuga M. 2005. Methylglyoxal induces apoptosis through oxidative stress-mediated activation of p38 mitogen-activated protein kinase in rat Schwann cells. *Annals of the New York Academy of Sciences* 1043: 151–157.
- Garrido D, Rubin T, Poidevin M, Maroni B, Le Rouzic A, Parvy JP, Montagne J. 2015. Fatty acid synthase cooperates with glyoxalase 1 to protect against sugar toxicity. *PLoS Genetics* 11: e1004995.
- Ghosh A, Kushwaha HR, Hasan MR, Pareek A, Sopory SK, Singla-Pareek SL. 2016. Presence of unique glyoxalase III proteins in plants indicates the existence of shorter route for methylglyoxal detoxification. *Scientific Reports* 6: 18358.
- Gupta BK, Sahoo KK, Ghosh A, Tripathi AK, Anwar K, Das P, Singh AK, Pareek A, Sopory SK, Singla-Pareek SL. 2018. Manipulation of glyoxalase pathway confers tolerance to multiple stresses in rice. *Plant, Cell & Environment* 41: 1186–1200.
- Hanson AD, Henry CS, Fiehl O, de Crécy-Lagard V. 2016. Metabolite damage and metabolite damage control in plants. *Annual Review of Plant Biology* 67: 131–152.
- Hasim S, Hussin NA, Alomar F, Bidasec KR, Nickerson KW, Wilson MA. 2014. A glutathione-independent glyoxalase of the DJ-1 superfamily plays an important role in managing metabolically generated methylglyoxal in *Candida albicans*. *Journal of Biological Chemistry* 289: 1662–1674.
- Havé M, Leitao L, Bagard M, Castell JF, Repellin A. 2015. Protein carbonylation during natural leaf senescence in winter wheat, as probed by fluorescein-5-thiosemicarbazide. *Plant Biology* 17: 973–979.
- Hoque TS, Urabi M, Tuya A, Nakamura Y, Murata Y. 2012. Methylglyoxal inhibits seed germination and root elongation and up-regulates transcription of stress-responsive genes in ABA-dependent pathway in *Arabidopsis*. *Plant Biology* 14: 854–858.
- Job C, Rajjou L, Lovigny Y, Belghazi M, Job D. 2005. Patterns of protein oxidation in *Arabidopsis* seeds and during germination. *Plant Physiology* 138: 790–802.
- Kaur C, Kushwaha HR, Mustafiz A, Pareek A, Sopory SK, Singla-Pareek SL. 2015. Analysis of global gene expression profile of rice in response to methylglyoxal indicates its possible role as a stress signal molecule. *Frontiers in Plant Science* 6: 682.
- Kaur C, Singla-Pareek SL, Sopory SK. 2014a. Glyoxalase and methylglyoxal as biomarkers for plant stress tolerance. *Critical Reviews in Plant Sciences* 33: 429–456.
- Kaur C, Mustafiz A, Sarkar AK, Ariyadasa TU, Singla-Pareek SL, Sopory SK. 2014b. Expression of abiotic stress inducible ETHE1-like protein from rice is higher in roots and is regulated by calcium. *Physiologia Plantarum* 152: 1–16.
- Kaur C, Tripathi AK, Nutan KK, Sharma S, Ghosh A, Tripathi JK, Pareek A, Singla-Pareek SL, Sopory SK. 2017. A nuclear-localized rice glyoxalase I enzyme, OsGLYI-8, functions in the detoxification of methylglyoxal in the nucleus. *The Plant Journal* 89: 565–576.
- Kaur C, Vishnoi A, Ariyadasa TU, Bhattacharya A, Singla-Pareek SL, Sopory SK. 2013. Episodes of horizontal gene-transfer and gene-fusion led to co-existence of different metal-ion specific glyoxalase I. *Scientific Reports* 3: 3076.
- Kim M, Lim JH, Ahn CS, Park K, Kim GT, Kim WT, Pai HS. 2006. Mitochondria-associated hexokinases play a role in the control of programmed cell death in *Nicotiana benthamiana*. *Plant Cell* 18: 2341–2355.
- Krübel L, Junemann J, Wirtz M, Birke H, Thornton JD, Browning LW, Poschet G, Hell R, Balk J, Braun HP *et al.* 2014. The mitochondrial sulfur dioxygenase ETHYLMALONIC ENCEPHALOPATHY PROTEIN1 is required for amino acid catabolism during carbohydrate starvation and embryo development in *Arabidopsis*. *Plant Physiology* 165: 92–104.
- Kumar S, Stecher G, Suleski M, Heddes SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution* 34: 1812–1819.
- Lewandowska A, Vo T, Nguyen TD, Wahni K, Vertommen D, Van Breusegem F, Young D, Messens J. 2019. Bifunctional chloroplastic DJ-1B from *Arabidopsis thaliana* is an oxidation-robust holdase and a glyoxalase sensitive to H₂O₂. *Antioxidants* 8: 8.
- Li ZG, Duan XQ, Min X, Zhou ZH. 2017. Methylglyoxal as a novel signal molecule induces the salt tolerance of wheat by regulating the glyoxalase system, the antioxidant system, and osmolytes. *Protoplasma* 254: 1995–2006.
- Li ZG, Long WB, Yang SZ, Wang YC, Tang JH. 2018b. Signaling molecule methylglyoxal-induced thermotolerance is partly mediated by hydrogen sulfide in maize (*Zea mays* L.) seedlings. *Acta Physiologiae Plantarum* 40: 76.
- Li ZG, Nie Q, Yang CL, Wang Y, Zhou ZH. 2018a. Signaling molecule methylglyoxal ameliorates cadmium injury in wheat (*Triticum aestivum* L.) by a coordinated induction of glutathione pool and glyoxalase system. *Ecotoxicology and Environmental Safety* 149: 101–107.
- Misra K, Banerjee AB, Ray S, Ray M. 1995. Glyoxalase III from *Escherichia coli*: a single novel enzyme for the conversion of methylglyoxal into D-lactate without reduced glutathione. *Biochemical Journal* 305: 999–1003.
- Morcos M, Du X, Pfisterer F, Hutter H, Sayed AA, Thornalley P, Ahmed N, Baynes J, Thorpe S, Kukudov G *et al.* 2008. Glyoxalase-1 prevents mitochondrial protein modification and enhances lifespan in *Caenorhabditis elegans*. *Aging Cell* 7: 260–269.
- Mostafa MG, Ghosh A, Li ZG, Siddiqui MN, Fujita M, Tran LS. 2018. Methylglyoxal – a signaling molecule in plant abiotic stress responses. *Free Radical Biology & Medicine* 122: 96–109.
- Mustafiz A, Singh AK, Pareek A, Sopory SK, Singla-Pareek SL. 2011. Genome-wide identification of glyoxalase genes and their expression profiling during development and in response to abiotic stresses in *Arabidopsis* and rice. *Functional & Integrative Genomics* 11: 293–305.
- Nigro C, Leone A, Fiory F, Prevezano I, Nicolò A, Mirra P, Beguinot F, Miele C. 2019. Dicarbonyl stress at the crossroads of healthy and unhealthy aging. *Cells* 8: 749.
- Oracz K, Bouteau HE, Farrant JM, Cooper K, Belghazi M, Job C, Job D, Corbineau F, Bailly C. 2007. ROS production and protein oxidation as a novel mechanism for seed dormancy alleviation. *The Plant Journal* 50: 452–465.
- Rabbani N, Thornalley PJ. 2012. Methylglyoxal, glyoxalase I and the dicarbonyl proteome. *Amino Acids* 42: 1133–1142.
- Rabbani N, Thornalley PJ. 2018. Glyoxalase I modulation in obesity and diabetes. *Antioxidants & Redox Signaling* 30: 354–374.
- Richard JP. 1993. Mechanism for the formation of methylglyoxal from triosephosphates. *Biochemical Society Transactions* 21: 549–553.
- Saito R, Yamamoto H, Makino A, Sugimoto T, Miyake C. 2011. Methylglyoxal functions as Hill oxidant and stimulates the photoreduction of O₂ at photosystem I: A symptom of plant diabetes. *Plant, Cell & Environment* 34: 1454–1464.
- Sankaranarayanan S, Jamshed M, Samuel MA. 2015. Degradation of glyoxalase I in *Brassica napus* stigma leads to self-incompatibility response. *Nature Plants* 1: 15185.
- Schmitz J, Rossoni AW, Maurino VG. 2018. Dissecting the physiological function of plant glyoxalase I and glyoxalase I-like proteins. *Frontiers in Plant Science* 9: 1618.
- Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: The WGD Radiation Lag-Time Model. *Current Opinion in Plant Biology* 15: 147–153.
- Shumilina J, Kusnetsova A, Tsarev A, Janse van Rensburg HC, Medvedev S, Demidchik V, Van den Ende W, Frolov A. 2019. Glycation of plant proteins: regulatory roles and interplay with sugar signalling? *International Journal of Molecular Sciences* 20: 2366.
- Singla-Pareek SL, Reddy MK, Sopory SK. 2003. Genetic engineering of glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *Proceedings of the National Academy of Sciences, USA* 100: 14672–14677.

- Singla-Pareek SL, Yadav SK, Pareek A, Reddy MK, Sopory SK. 2006. Transgenic tobacco overexpressing glyoxalase pathway enzymes grow and set viable seeds in zinc spiked soils. *Plant Physiology* 140: 613–623.
- Singla-Pareek SL, Yadav SK, Pareek A, Reddy MK, Sopory SK. 2008. Enhancing salt tolerance in a crop plant by overexpression of glyoxalase II. *Transgenic Research* 17: 171–180.
- Szent-Georgyi A. 1965. Cell division and cancer. *Science* 149: 34–37.
- Takagi D, Inoue H, Odawara M, Shimakawa G, Miyake C. 2014. The Calvin cycle inevitably produces sugar-derived reactive carbonyl methylglyoxal during photosynthesis: a potential cause of plant diabetes. *Plant and Cell Physiology* 55: 333–340.
- Thornalley PJ. 1993. The glyoxalase system in health and disease. *Molecular Aspects of Medicine* 14: 287–371.
- Thornalley PJ. 2008. Protein and nucleotide damage by glyoxal and methylglyoxal in physiological systems—role in ageing and disease. *Drug Metabolism and Drug Interactions* 23: 125–50.
- Welchen E, Schmitz J, Fuchs P, Garcia L, Wagner S, Wienstroer J, Schertl P, Braun HP, Schwarzländer M, Gonzalez DH *et al.* 2016. D-Lactate dehydrogenase links methylglyoxal degradation and electron transport through cytochrome C. *Plant Physiology* 172: 901–912.
- Xu Y, Chen X. 2006. Glyoxalase II, a detoxifying enzyme of glycolysis byproduct methylglyoxal and a target of p63 and p73, is a pro-survival factor of the p53 family. *Journal of Biological Chemistry* 281: 26702–26713.
- Xue M, Rabbani N, Thornalley PJ. 2011. Glyoxalase in ageing. *Seminars in Cell & Developmental Biology* 22: 293–301.
- Yadav SK, Singla-Pareek SL, Ray M, Reddy MK, Sopory SK. 2005a. Methylglyoxal levels in plants under salinity stress are dependent on glyoxalase I and glutathione. *Biochemical Biophysical Research Communications* 337: 61–67.
- Yadav SK, Singla-Pareek SL, Reddy MK, Sopory SK. 2005b. Transgenic tobacco plants overexpressing glyoxalase enzymes resist an increase in methylglyoxal and maintain higher reduced glutathione levels under salinity stress. *FEBS Letters* 579: 6265–6271.
- You X, Zhang W, Hu J, Jing R, Cai Y, Feng Z, Kong F, Zhang J, Yan H, Chen W *et al.* 2019. FLOURY ENDOSPERM15 encodes a glyoxalase I involved in compound granule formation and starch synthesis in rice endosperm. *Plant Cell Reports* 16: 1–5.






About New Phytologist

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit www.newphytologist.com to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit www.newphytologist.com

Article

Tracing the Evolution of Plant Glyoxalase III Enzymes for Structural and Functional Divergence

Brijesh Kumar ^{1,†}, Charanpreet Kaur ^{1,2,†}, Ashwani Pareek ², Sudhir K. Sopory ¹
and Sneh L. Singla-Pareek ^{1,*}

¹ Plant Stress Biology, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110067, India; brijeshkumar@icgeb.res.in (B.K.); charanpreet06@gmail.com (C.K.); sopory@icgeb.res.in (S.K.S.)

² Stress Physiology and Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India; ashwanip@mail.jnu.ac.in

* Correspondence: sneh@icgeb.res.in; Tel.: +91-11-26741358 (ext. 363); Fax: +91-11-26742316

† Authors with equal contribution.

Abstract: Glyoxalase pathway is the primary route for metabolism of methylglyoxal (MG), a toxic ubiquitous metabolite that affects redox homeostasis. It neutralizes MG using Glyoxalase I and Glyoxalase II (GLYI and GLYII) enzymes in the presence of reduced glutathione. In addition, there also exists a shorter route for the MG detoxification in the form of Glyoxalase III (GLYIII) enzymes, which can convert MG into D-lactate in a single-step without involving glutathione. GLYIII proteins in different systems demonstrate diverse functional capacities and play a vital role in oxidative stress response. To gain insight into their evolutionary patterns, here we studied the evolution of GLYIII enzymes across prokaryotes and eukaryotes, with special emphasis on plants. GLYIII proteins are characterized by the presence of DJ-1_PfpI domains thereby, belonging to the DJ-1_PfpI protein superfamily. Our analysis delineated evolution of double DJ-1_PfpI domains in plant GLYIII. Based on sequence and structural characteristics, plant GLYIII enzymes could be categorized into three different clusters, which followed different evolutionary trajectories. Importantly, GLYIII proteins from monocots and dicots group separately in each cluster and the each of the two domains of these proteins also cluster differentially. Overall, our findings suggested that GLYIII proteins have undergone significant evolutionary changes in plants, which is likely to confer diversity and flexibility in their functions.

Keywords: glyoxalase III; DJ-1_PfpI domains; evolution; horizontal gene transfer; methylglyoxal



Citation: Kumar, B.; Kaur, C.; Pareek, A.; Sopory, S.K.; Singla-Pareek, S.L. Tracing the Evolution of Plant Glyoxalase III Enzymes for Structural and Functional Divergence. *Antioxidants* **2021**, *10*, 648. <https://doi.org/10.3390/antiox10050648>

Academic Editor: Cinzia Antognelli

Received: 1 January 2021

Accepted: 8 March 2021

Published: 23 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Glyoxalase (GLY) pathway is well-known for its role in plant stress adaptation as well as in disease conditions in animals. The two enzymes of this pathway, glyoxalase I (GLYI) and glyoxalase II (GLYII) act by metabolizing methylglyoxal (MG), an α -ketoaldehyde produced primarily by spontaneous dephosphorylation of triose phosphates [1,2], to a relatively non-toxic compound, D-lactate. The toxicity of MG is attributed to its highly reactive nature towards proteins, lipids, and nucleic acids. Being a carbonyl electrophile, MG can readily modify arginine, lysine, and cysteine amino acids of proteins as well as adenine, cytosine, and guanine residues of nucleic acids, to form advanced glycation end products (AGEs) in turn, altering the functional properties of these biomolecules [3]. Besides triose sugars, other minor sources of MG formation also exist and include catabolism of acetone and threonine [3,4]. At the same time, prokaryotes demonstrate the enzyme-catalyzed formation of MG as well, which takes place via the dephosphorylation of dihydroxyacetone phosphate (DHAP) [1,2]. These biomolecular modifications due to reactivity of MG leads to changes in the redox status of the cellular milieu. Since the generation of MG in the cell

is inevitable, almost all organisms possess glyoxalase enzymes that restrict increase in MG levels in their cellular milieu and hence, contribute to redox homeostasis [5,6].

Based on the glutathione (GSH) requirement, glyoxalases were categorized into two types, i.e., the GSH-dependent and the GSH-independent form [7–9]. The GSH-dependent system is a two-enzyme pathway involving GLYI and GLYII enzymes [10], whereas the GSH-independent system involves a single glyoxalase III (GLYIII) enzyme [7–9]. The metalloenzyme GLYI (requiring either Ni^{2+} or Zn^{2+}) converts hemithioacetal formed from the spontaneous combination of MG and GSH to R-(S)-lactoylglutathione [11,12], which is further hydrolyzed by GLYII to give D-lactate and GSH [13]. On the other hand, GLYIII enzymes catalyze the direct conversion of MG to D-lactate without requiring GSH [7–9] (Figure 1).

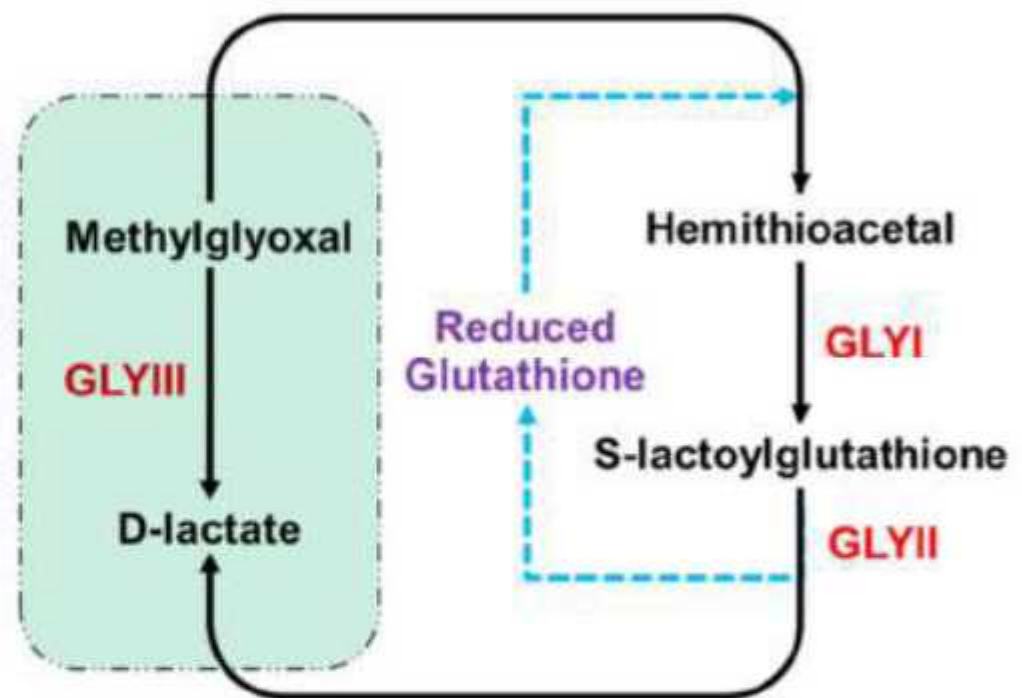


Figure 1. The glyoxalase pathway for metabolism of methylglyoxal. The classical two-step glyoxalase pathway converts methylglyoxal (MG) to D-lactate with the help of Glyoxalase I (GLYI) and Glyoxalase II (GLYII) enzymes and uses reduced glutathione (GSH) as cofactor. Glyoxalase III (GLYIII) constitutes the shorter pathway of MG detoxification which directly converts MG to D-lactate in a single step, without involving GSH.

While the GSH-dependent glyoxalase pathway is highly specific for the detoxification of MG and other dicarbonyls, the GSH-independent GLYIII proteins exhibit varied catalytic functions. These proteins possess the DJ-1_PfpI domains and belong to the DJ-1 enzyme superfamily, the members of which in addition to the GLYIII activity, exhibit various other physiological functions, such as the protease [14], chaperone [15], deglycase [16,17], esterase [18], anti-oxidant stimulant [19], mitochondrial regulation, and transcriptional regulator [20] functions. On similar lines, the human DJ-1 protein was reported to show protease and chaperone functions along with the GLYIII activity [20]. Besides, it also acts as a sensor of oxidative stress and is a major determinant for the early onset of Parkinson's disease, with its loss leading to neurodegeneration [21]. In plants, the role of GLYIII proteins largely remains unexplored, with a few studies demonstrating their role in response towards environmental stresses [22] and chloroplast development [23]. Overexpression of DJ-1 (Hsp31) protein from *E. coli* was shown to confer tolerance against biotic as well as abiotic stresses in tobacco [24]. Similarly, ectopic expression of DJ-1_PfpI domain containing GLYIII from *Erianthus arundinaceus*, a wild-relative of sugarcane confers drought tolerance in commercially cultivated sugarcane hybrid [25]. With respect to their

glyoxalase activity, it was reported that the GLYIII proteins have lower catalytic efficiencies as compared to the conventional GSH-dependent GLY pathway [9] but are particularly important under conditions of diminished GLYI/GLYII activity as a result of low GSH levels, such as in chronic oxidative stress, stationary phase, or sulfur limitation [26]. In addition to possessing various catalytic functions, the varied sub-cellular localization of GLYIII proteins makes them even more versatile and functionally diverse [9]. Hence, GLYIII proteins appear to be important for the normal functioning of living systems.

The availability of completely sequenced genomes of many organisms has allowed comparative analysis of various gene families across species, giving an insight into how the evolution and diversification of a protein family can help organisms evolve, as conditions change. One of our recent studies highlighted the expansion of glyoxalase family, including GLYIII, exclusively in plants and established the lack of motility as the major reason for the multiplicity of the glyoxalase gene family in plants [27]. The sessility of plants makes them more vulnerable to the different kind of stresses, and hence, these multigenic glyoxalase families provide redundancy in gene function and act as a contingency under extreme conditions [27]. Another study unraveled interesting information on the evolutionary trajectory of GLYI proteins, tracing the origin of the two metal-dependent forms to different prokaryotic ancestors along with elucidating the components of the plant GLYI family [6]. Since the discovery of GLYIII pathway raises many questions regarding the defense strategy of plants against MG stress, it will be worth investigating the distribution and evolution of these proteins across the different species. While the prokaryotes have just one catalytic (DJ-1_PfpI) domain [28], plant genomes possess GLYIII proteins with one or two catalytic domains [28]. How and when did the two-domain GLYIII proteins come into existence? How can the different domain architecture of these proteins be correlated to their diverse functions? These are the few questions that the present study aims to address via an investigation of the evolutionary trajectory of GLYIII proteins along with deciphering their structural and functional diversity in the plant kingdom.

2. Materials and Methods

2.1. Data Retrieval and Sequence Analysis

To extract amino acid sequences of GLYIII proteins from different organisms, either the conserved DJ-1_PfpI domain, PF01965 [29] or the sequence of HsDJ1-1 (BAG34938.1) and ScDJ-1 (EWG86848.1) proteins, was used as query. The NCBI database [30] was used to retrieve protein sequences for archaea, bacteria, protists, fungi, and animals using HsDJ1-1 and ScDJ-1 proteins as a query, while the Phytozome database [31] was used to retrieve plant DJ-1 proteins using Hidden Markov Model (HMM) profile search for the conserved DJ-1_PfpI domain. BLAST was done using the default settings. All the retrieved sequences were screened using SMART [32,33] and Pfam [29] for the presence of DJ-1_PfpI and other domains.

A total of 69 species were selected to cover the entire tree of life (representing five kingdoms, covering sub-divisions as well), and subsequently, GLYIII proteins from those species were used for further analysis, similar to that done for GLYI proteins [6]. The amino acid sequences of GLYIII used in the study are given in File S1 in Supplementary Materials. Sequences were aligned using ClustalW in MEGA 7.0 [34] with default settings.

2.2. Phylogenetic and Subcellular Localization Analysis

Phylogenetic trees were constructed using the amino acid sequence of the full-length GLYIII proteins or the corresponding domains. The sequences were aligned using MUSCLE [35], and subsequently, Maximum Likelihood tree was constructed using MEGA 7.0 [34] with default settings and 1000 bootstrap replicates. The tree was visualized and modified in iTol [36]. Subcellular localization of the proteins was predicted using online servers, CELLO [37,38] and WolfpSORT [39].

2.3. Motif and Three-Dimensional Structural Analysis of GLYIII Proteins

The plant GLYIII protein sequences were analyzed for the presence of conserved motifs using the MEME software suite [40]. Secondary structure predictions were made using the ESPript 3.0 program [41]. For homology modelling studies, SWISS-MODEL [42] tool of the ExPasy server was used and then visualized in Pymol software [43].

3. Results

3.1. GLYIII Proteins Contain DJ-1_PfpI Domain in Combination with Various Additional Domains

GLYIII protein sequences were extracted from 69 species representing different phyla to cover all the five kingdoms (Table S1). All GLYIII proteins possessed the hallmark DJ-1_PfpI domain, belonging to the respective superfamily (Figure 2). Besides, other domains are present in these proteins that may confer diverse functional properties (Figure 2). Of the five archaeal species used in the study, *Natronococcus occultus* was found to possess a catalase (PF00199) and a catalase-rel (catalase-related, PF06628) domain in addition to the DJ-1_PfpI domain (Table S2 in Supplementary Materials). These domains are present in catalase enzymes that protect the cells from toxic effects of hydrogen peroxide [44]. The catalase-rel domain additionally carries the immune-responsive amphipathic octapeptide [45]. Interestingly, this catalase/catalase-rel domain was not observed in the GLYIII proteins from other kingdoms used in this study. Likewise, domain analysis of the bacterial GLYIII proteins revealed the presence of additional HTH_18 (PF12833) domain in about 9 of the 45 analyzed bacterial GLYIII proteins (Table S2 in Supplementary Materials). Of the 9 GLYIII proteins having this additional HTH_18 domain, six belonged to *Pseudomonas* sp. GM30 strain, which had a total of 11 GLYIII proteins, the other three belonged to two different bacterial species viz. *Rhizobium* sp. CF080 and *Desulfovibrio alkalitolerans* (Table S2 in Supplementary Materials). The helix-turn-helix (HTH) domain is a major structural motif capable of binding DNA, indicating the ability of such proteins to interact with nucleic acids as well. In contrast, GLYIII proteins from protists do not possess additional catalytic domains besides the DJ-1_PfpI domain. Although few protist GLYIII proteins have an N-terminal transmembrane domain required for their association with membranes, but may not have any additional specific catalytic functions (Figure 2). Importantly, the first instance of the presence of two DJ-1_PfpI domains could be seen in this kingdom as observed for *Dictyostelium discoideum* where in, it possessed two domains with unusually shorter (83 and 97 aa) lengths (Table S2 in Supplementary Materials). On the other hand, fungal GLYIII proteins which were thought to be evolutionary more advanced than the protists, had only single DJ-1_PfpI domain (Figure 2). However, they had a signal peptide sequence at the N-terminus, in some of the proteins.

In plants, GLYIII proteins majorly existed as two DJ-1_PfpI domain-containing proteins, while some proteins showed only single domain. In addition, some GLYIII proteins (both single as well as double DJ-1_PfpI domain-containing proteins) had a signal peptide or transmembrane domain at the N-terminus similar to that found in protists and fungi. In fact, a DUF111 (PF01969) domain was also found to coexist with DJ-1_PfpI domains as observed in PpGLYIII6 (Table S2 in Supplementary Materials). The presence of single DJ-1_PfpI domain-containing GLYIII proteins was, however, limited to some species, and that the majority of instances being of the dicots, the only exception in monocots being *T. aestivum* (Table S2 in Supplementary Materials). The green algae, *Chlamydomonas reinhardtii* and *Chlorella variabilis*, had only single DJ-1_PfpI domain-containing GLYIII proteins, whereas the moss *Physcomitrella patens*, which is the ancestor of land plants, had both single and double DJ-1_PfpI domain-containing GLYIII proteins. Further, like plants, animals also possessed both single and double DJ-1_PfpI domain-containing GLYIII proteins; however, with a predominance of single-domain proteins (Table S2 in Supplementary Materials). The two domains in case of double DJ-1_PfpI domain-containing proteins as in *Echinops telfairi* (EtGLYIII2) and *Tupaia chinensis* (TchGLYIII3), had domains that are either overlapping or conjoint. Some animal GLYIII proteins harbored SNO (PF01174) or Ribosomal (Ribosomal_L18_c, PF14204 or Ribosomal_L16, PF00252) domains, the latter having implications

in rRNA binding (Figure 2). Notably, some species in bacteria, animals, and even plants possessed GATase3 domain, representing a CobB/CobQ-like glutamine amidotransferase domain, enclosed within the DJ-1_PfpI domain (Figure 2). The GATase3 domain possesses similar catalytic sites as found in DJ-1 proteins, which is the reason for the observed overlap of the two domains.

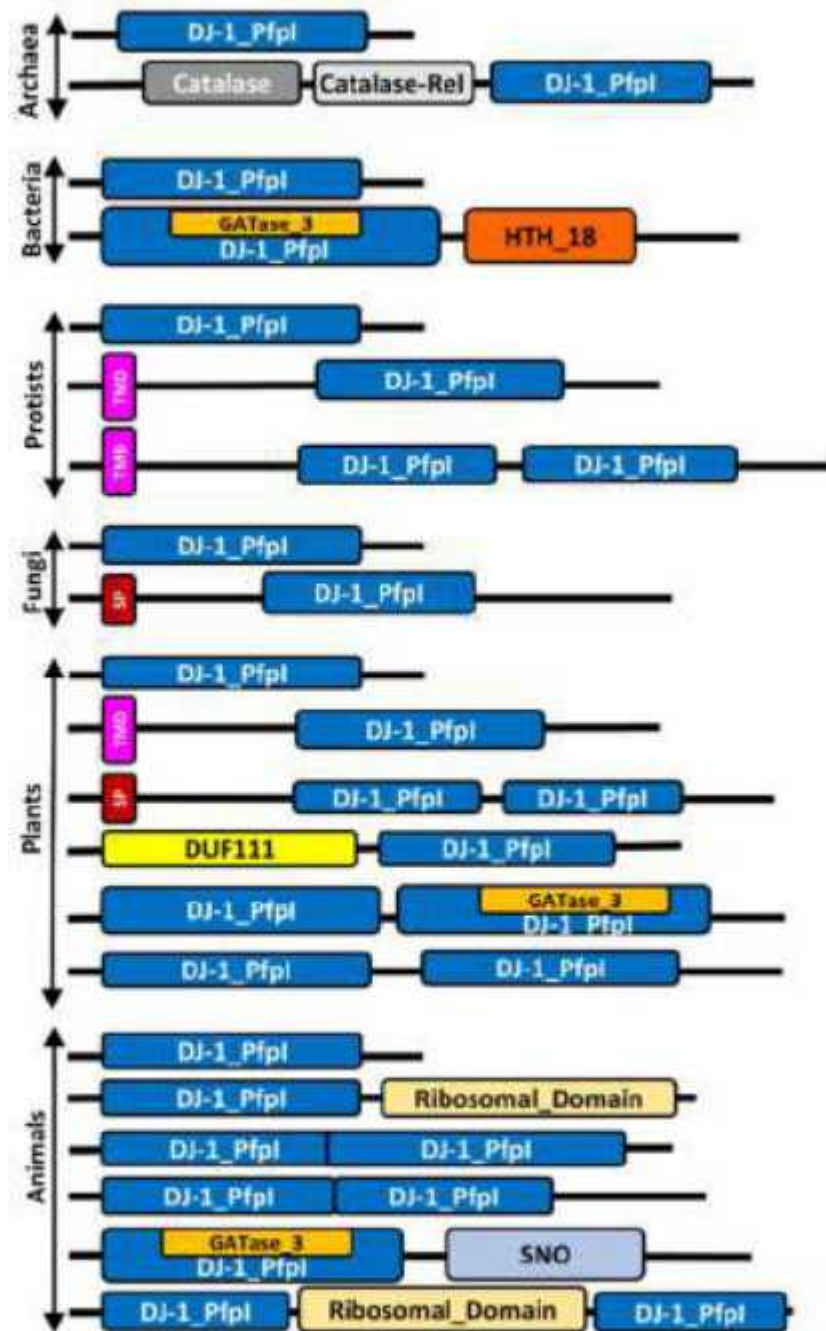


Figure 2. Schematic depiction of domain architecture observed in GLYIII proteins from different kingdoms. GLYIII proteins essentially contain DJ-1_PfpI domains often in combination with other domains such as catalase, HTH_18 (helix-turn-helix), DUF (Domain of unknown function), SNO (glutamine amidotransferase), etc. Two DJ-1_PfpI domains are common to plant GLYIII and some instance of its presence is also seen in animals. Image shown is not to scale. TMD—transmembrane domain, SP—signal peptide.

Further, we found the length of the DJ-1_PfpI domain in GLYIII proteins ranged from 52 to 282 amino acids (Table S1). The shortest domain (52 aa) was observed in *T. aestivum* TaGLYIII4 and the longest (282 aa) in *E. coli* EcGLYIII1 (Table S1). However, the average domain size was about ~162 aa. Though variable in length, the GLYIII domain in plants, based on the domain lengths were majorly of two types, smaller ones, which were approximately 165 aa in length and the larger ones which were about 184 aa long. However, there was no particular pattern of inheritance of these domains in the plant GLYIII family, but the domains of similar size usually co-existed in the two DJ-1_PfpI domain-containing GLYIII proteins (Table S1). In contrast, plant GLYIII proteins having single DJ-1_PfpI domains, generally had very short domain lengths (55–146 aa). Similarly, animal GLYIII proteins having two DJ-1_PfpI domains also had both the domains being very short in length (52–90 aa). But considering the general trend, even animal GLYIII proteins like plants had shorter (165 aa) as well as longer (184 aa) type of domains.

3.2. Plants Have Three Major Types of GLYIII Proteins

To understand the evolutionary relationship of GLYIII proteins from different organisms, a maximum likelihood tree was constructed using ClustalW in MEGA7.0 (Figure 3). Of these, only some proteins have been biochemically characterized and same is indicated in the tree (Figure 3). Through this protein-based phylogenetic analysis, we acquired initial indications on GLYIII protein relationship among different organisms, simultaneously correlating them with their biochemical and molecular features. Further, we determined whether any prokaryotic or eukaryotic GLYIII had the same composition of domains as plant GLYIII, in turn revealing conservation among these proteins. We found, besides plants, several other organisms also possessed multiple GLYIII proteins. The tree clearly showed three major types of plant GLYIII proteins, one closely grouped with the prokaryotic homologs (named as cluster-A), the other (cluster-B) having a closer relationship with the animal and protist GLYIII proteins, and the third (cluster-C) forming a distinct group, not clustering with proteins from any other kingdom (Figure 3). All the three clusters contained single as well as double DJ-1_PfpI domain-containing GLYIII proteins, with proteins from monocots and dicots forming separate sub-clusters within each cluster. However, no significant difference was found between monocot and dicot GLYIII proteins.

Further, GLYIII members of these three clusters were generally longer in length than those from the other organisms due to the presence of two DJ-1_PfpI domains in them as against the rest which possesses single DJ-1_PfpI domain. Moreover, cluster A, B, and C proteins also possessed an additional stretch of amino acids known as the linker region, which connect the two domains (Figure 3; Table S1). Interestingly, the two DJ-1_PfpI domain-containing GLYIII proteins in cluster-A had shorter linker regions (9–20 aa), except for AtGLYIII3, OsGLYIII3, and OsGLYIII6 proteins, while the cluster-B and cluster-C proteins had longer linker regions of approx. 37–39 aa long, as shown in Figure 3. Further, the proteins with longer linker regions had relatively shorter domain lengths (~165 aa for both the domains), while the proteins with shorter linker regions had longer domain lengths (~185 aa for both domains) (Figure 3; Table S1), indicating a tight correlation between the length of the linker region and the domain size.

Another important observation pertains to the presence of GLYIII proteins from green algae in cluster-C of the tree (Figure 3). The results clearly indicated that proteins of the cluster-C represent a plant-specific lineage derived from their green algae ancestors. Further, cluster-B proteins were observed to be a part of the larger cluster comprising of proteins from protists, brown algae, diatoms, and animals, which suggests that members of this cluster have undertaken a vertical path during the course of evolution. On the other hand, cluster-A proteins belonged to a larger cluster comprising majorly of prokaryotic GLYIII, especially grouping with archaeal and bacterial GLYIII proteins with no closer eukaryotic homolog being identified. It is possible that horizontal gene transfer events have shaped their evolution or cluster-A type of proteins were lost or diverged in animals.

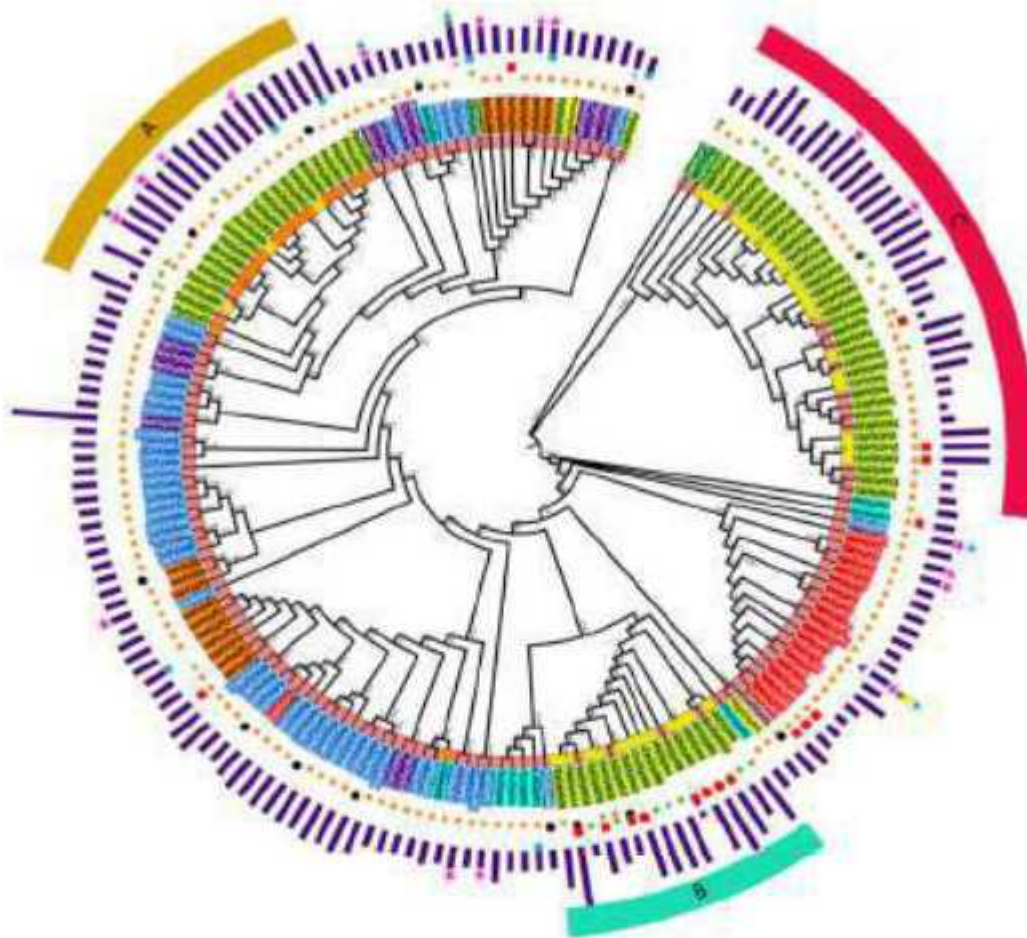


Figure 3. Phylogenetic analysis reveals three types of GLYIII proteins exist in plants. In all, 183 sequences from 69 species were used. The tree indicates the presence of three types of plant GLYIII proteins represented as cluster-A (mustard strip), B (pale-green strip), and C (red strip). Innermost strip indicates the length of linker region separating the two domains. Yellow-green color represents longer linker length (37–39 aa), orange indicates small linker length (9–20 aa), and light coral indicates no linker region. Purple bars on the outer side of the tree indicate protein length. Symbols next to protein names indicate different organellar localization patterns. Orange star indicates cytoplasmic localization, green star—chloroplast, purple star—periplasm, black circle—membrane (inner and outer membrane for prokaryotes, and plasma membrane for eukaryotes), blue circle—extracellular, red square—mitochondria, and purple triangle—nuclear localization. GLYIII proteins belonging to the different phylogenetic domains of life are depicted using specific colors as, archaea (purple), bacteria (blue), protists (cyan), fungi (orange), diatoms (yellow), brown algae (light brown), green algae (dark green), plants (light green), and animals (red). The different colored symbols above the bars (representing protein size) indicate different biochemical activities, i.e., GLYIII (pink star), chaperone (green square), protease (yellow circle), and deglycase (blue triangle). The maximum likelihood tree has been constructed with 1000 bootstrap replicates (indicated over branches) using MEGA-7.0 and visualized using iTOL.

3.3. Cluster-C GLYIII Proteins Share Similarity with Their Ancestral Homologs in Localization Patterns

After gaining initial insights into the evolution of plant GLYIII proteins from their prokaryotic ancestors, we next investigated their subcellular localization patterns to further unravel the evolutionary path of these proteins. The predictions for subcellular localization were made using CELLO [37,38] and WolfpSORT [39]. In prokaryotes, GLYIII proteins were predicted to be generally cytoplasmic, except for very few species, which were predicted to be either extracellular, periplasmic, or membrane-localized (Figure 3). Likewise, our predictions indicated that GLYIII proteins from the kingdom Protists and fungi also have few extracellular, membrane-localized, or even mitochondria-localized isoforms. Further, the majority of the animal GLYIII proteins were predicted to be in the cytoplasm with the exceptions of few mitochondrial candidates.

In comparison to the other kingdoms, GLYIII proteins from plants exhibited greater variations in their localization properties, attributed primarily to the differential predicted localization patterns of the three plant-specific GLYIII clusters. For instance, most of the cluster-A proteins are possibly cytoplasmic except for OsGLYIII3, OsGLYIII6, and PtGLYIII1, which were predicted to be exclusively localized to the chloroplast. In fact, OsGLYIII3 and OsGLYIII6 formed exceptions even in terms of the linker region length, which was not similar to other members of this group, whereas PtGLYIII1 was a single DJ-1_PfpI domain-containing protein, indicating their inherent outlier nature (Figure 3). On the other hand, most of the cluster-B members were predicted to be chloroplast and/or mitochondria localized. Further, GLYIII proteins from green algal ancestors of plants, which were the members of cluster-C, were predicted to be localized to the cytoplasm and/or chloroplast and consequently, most of the members of this group were either cytoplasmic or chloroplastic, with very few (TaGLYIII9, ZmGLYIII1, and ZmGLYIII5) being predicted to co-localize in mitochondria as well. In agreement, GLYIII proteins from the moss, *P. patens*, which is an evolutionary intermediate between the green algae and the land plants, also exhibited putative cytoplasmic and chloroplastic localization of two of its members belonging to this cluster. In fact, it could be generalized that almost all plant species under investigation had at least two members in this cluster, one predicted to encode a cytoplasmic and another a chloroplastic isoform. Overall, assessment of sub-cellular localization suggested that plant GLYIII proteins of the three clusters had different evolutionary trajectories. In particular, proteins of cluster-C were more primitive, sharing their possible localization patterns with ancestral homologs from green algae and the moss *P. patens*.

3.4. Assessment of Domain Relationships of GLYIII Proteins Indicate Primitive Evolution of N-Terminal Domain of Plant GLYIII

Since plant GLYIII possessed two DJ-1_PfpI domains, we analyzed the evolutionary trajectory of each domain. For this, phylogenetic relationships among the DJ-1_PfpI domain sequences from different species were investigated (Figure 4). Both the domains of plant GLYIII proteins formed two separate clusters, with first domain (or N-terminal domain, indicated by suffix 'A') of all proteins grouping together and the second or C-terminal domain (indicated by suffix 'B') forming a separate cluster. Importantly, proteins belonging to cluster-A had both the sub-clusters representing the two domains in the same major cluster-A, with both sub-clusters grouping to the same node (Figure 4). Whereas, the members of cluster-B and cluster-C collectively formed two clusters, representing the two domains, with each domain within cluster-B and C grouping to different nodes (Figure 4).

Each of the two groups representing the two domains in cluster-A could be further sub-grouped into two smaller clusters (A1 and A2), indicating further divergence in domain sequences of cluster-A proteins (Figure 4). Of the two sub-clusters A1 and A2, the bigger sub-cluster A1 comprised of AtGLYIII3 and OsGLYIII3 domains from Arabidopsis and rice, respectively, full-length proteins of which were previously shown to possess high catalytic efficiencies as GLYIII [9,46]. The other sub-cluster A2 comprised of AtGLYIII5, AtGLYIII2, OsGLYIII6, BrGLYIII4, and BrGLYIII8 proteins. Interestingly, GLYIII proteins carrying single DJ-1_PfpI domain from wheat, TaGLYIII21 (clustering with N-terminal domain) and TaGLYIII23 (clustering with C-terminal domain) also grouped in the sub-cluster A2. In fact, analyzing the domain sequence similarity of these plant GLYIII proteins with those of *E. coli* (where GLYIII proteins are well-characterized), revealed greater similarity of cluster-A domains with EcGLYIII3 and of cluster-B/C domains with EcGLYIII2 protein (Table S3 in Supplementary Materials). Few exceptions included AtGLYIII3, BrGLYIII6, and PtGLYIII6 proteins, which have one domain more similar to EcGLYIII2 and the other being closer to EcGLYIII3 in terms of sequence similarity. Further, multiple sequence alignments of plant GLYIII proteins also uncovered the presence of an extra stretch of 4–18 aa in the domains of cluster-A proteins, the same being observed in both the domains of proteins belonging to this cluster (Figure S1 in Supplementary Materials). Another interesting observation pertained to the PpGLYIII1 protein, whose two domains grouped in different

sub-clusters of the bigger cluster-A. Its N-terminal domain (or PpGLYIII_A) grouped with N-terminal GLYIII domain sequences of cluster-A2, whereas the C-terminal domain (PpGLYIII_B) grouped within the C-terminal GLYIII domains of cluster-A1, indicating differential origins of the two domains of cluster-A proteins.

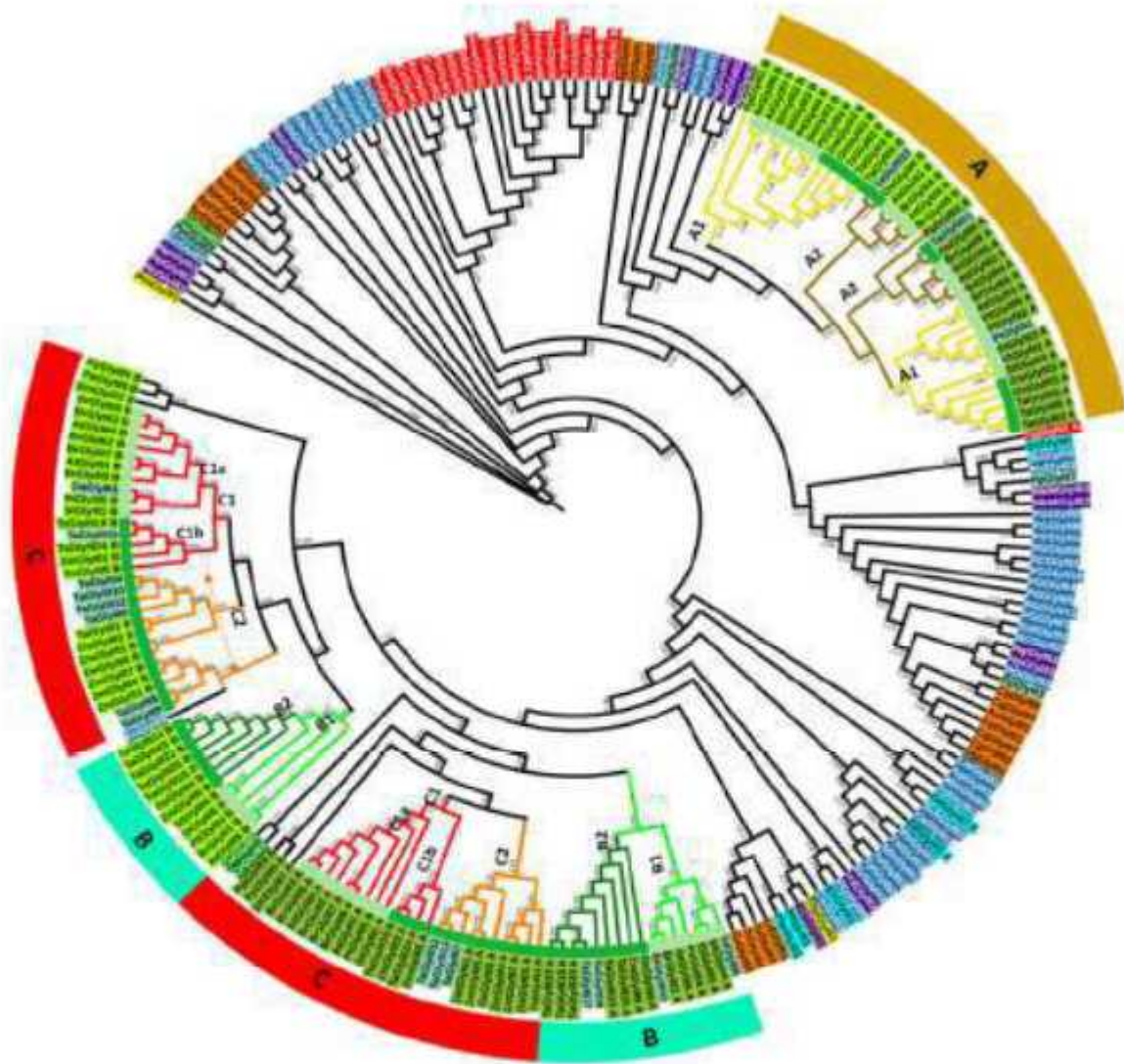


Figure 4. Domain relationship between GLYIII proteins unravel differential evolution of two DJ-I_PfpI domains in plants. The maximum likelihood tree was constructed with 1000 bootstrap replicates (indicated over branches) in MEGA 7.0 using 238 sequences from 183 GLYIII proteins belonging to 69 species. Different sub-clusters (A1&2, B1&2, C1&2) are indicated in the tree and nodes are colored differently to enable ease in understanding. The two domains of plant GLYIII proteins group separately in each cluster and can be identified from the suffix A (for first or N-terminal domain) and B (for second or C-terminal domain). Outermost strip indicates the three types of plant GLYIII proteins represented as cluster-A (mustard strip), B (pale-green strip), and C (red strip). Innermost strip indicates monocot (dark green) and dicot (pale green) species. GLYIII proteins belonging to the different phylogenetic domains of life are depicted using specific colors as, archaea (purple), bacteria (blue), protists (cyan), fungi (orange), diatoms (yellow), brown algae (dark orchid), green algae (dark green), and plants (A-domain: light green, B-domain: dark olive green, plants with single domain: aquamarine) and animals (red).

Similar to cluster-A, the N- and C- terminal domains of GLYIII proteins in cluster-B and cluster-C, also formed separate sub-clusters. Importantly, GLYIII proteins from green algae which possessed single GLYIII domains were more similar to N-terminal domain of GLYIII proteins of cluster-C. Notably, both N and C terminal domains of GLYIII proteins belonging to cluster-B and cluster-C can be further subdivided into smaller clusters, B1-

B2 and C1-C2, respectively, much like that observed for cluster-A proteins (Figure 4). The sub-clusters B1 and B2 of the main cluster-B consisted of GLYIII proteins from dicots and monocots, respectively. Similar grouping can also be observed in cluster-C where cluster C1a and C1b of the sub-cluster-C1 comprised of dicot and monocot GLYIII proteins, respectively. In addition, we observed another cluster in the major cluster-C, which is named as C2, comprising of proteins from only monocots (Figure 4). Importantly, this differential clustering of GLYIII from monocots and dicots was also observed in cluster-A proteins, with both the sub-clusters A1 and A2 being sub-divided into groups containing proteins from either monocots or dicots. Notably, GLYIII domains from the green algae grouped with the N-terminal domain sequences of GLYIII proteins, which suggested that the N-terminal domain of the two-domain GLYIII may have evolved first, which may have subsequently given rise to the C-terminal domain through gene duplication events.

3.5. Extensive Motif Rearrangements Have Shaped the Evolution of Plant GLYIII Proteins

The plant GLYIII proteins are found to be of three different types. Hence, to further understand the evolutionary differences between them, we analyzed the presence of conserved motifs using the MEME suite. In all, ten conserved motifs were predicted upon analysis of 73 GLYIII sequences from different plants and green algae (Figure 5a). It was evident from the assessment that proteins in different clusters varied with respect to both, presence and arrangement of motifs (Figure 5b–d). Assessment of catalytic triad (Cys-His-Asp/Glu) in these proteins also suggested differences in residue conservation among these clusters (Figure S2). Of the three clusters (A, B and C), cluster-A proteins comprising of about seven motifs (3-9-2-1-5-4-6), showed least variations in comparison to the other two cluster proteins (Figure 5b). In agreement, the catalytic triad was highly conserved in cluster-A proteins, with both the domains of most of the cluster-A proteins showing complete conservation of residues (Figure S2). Importantly, proteins belonging to this cluster had 'His' residue essentially conserved in the first domain, with majority conserving 'His' even in second DJ-1_PfpI domain. As against cluster-B and -C, proteins in cluster-A lacked motifs 7, 8, and 10 and instead possessed motifs 5 and 9, which were unique to its members, not detected in group B and C proteins (Figure 5b–d). Closer examination of motif arrangements revealed that the two sub-clusters of the cluster-A also exhibited differences in motifs. The dicot members of the cluster A1 possessed 7 motifs majorly in the order 3-9-2-1-5-4-6, whereas monocot members contained these domains in the order 3-9-1-5-4-2-6 (Figure 5b). The PpGLYIII1 protein is an interesting case, wherein it contained all seven motifs; however, in a different order, with N- and C-terminal motifs being swapped in position (5-4-6-3-9-2-1). That was the reason we observed grouping of both domains of PpGLYIII1 in a different than usual manner (Figure 5). On the other hand, the smaller cluster A2 possessed motifs largely in the order, 3-4-2-1-5-9, with motif 6 being flexible in its location in these proteins (Figure 5b). Further, in comparison to the cluster-A, cluster B and C proteins were disorganized with no predominant pattern being observed (Figure 5c,d). In particular, most of the variations were due to the monocot members of these clusters. Further, as some of the *T. aestivum* GLYIII proteins were single-domain proteins, it was a further source of observed variations in the motif organization pattern. Assessment of triad residues revealed cluster-B and -C proteins to lack conserved 'His' from the triad in both the domains and 'Cys' being absent from the second domain of cluster-B proteins (Figure S2). However, it is to be noted here that these predictions were based on mere sequence alignment and that modelling of proteins may give more accurate prediction of presence/absence of active site residues.

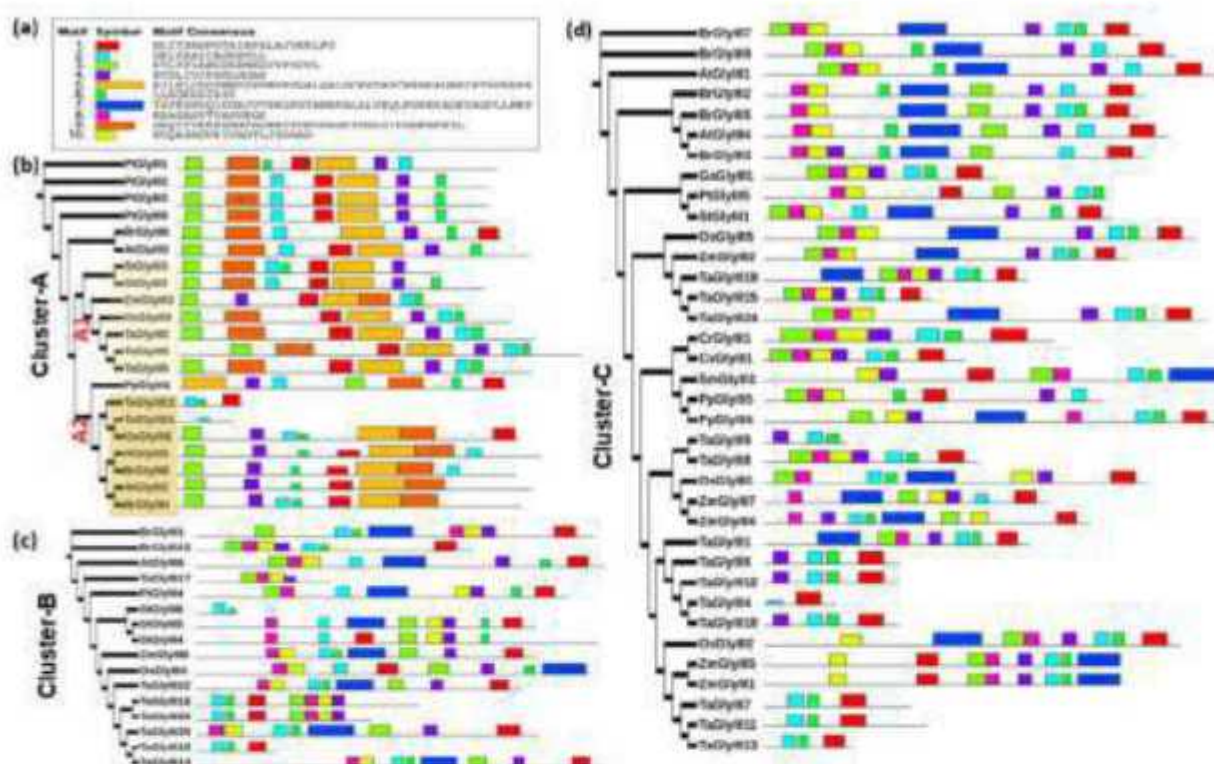


Figure 5. Extensive motif rearrangements have occurred during evolution of plant GLYIII proteins. (a) List of conserved motifs predicted in plant GLYIII proteins. Block diagram depicting motif organization in (b) Cluster-A, (c) Cluster-B, and (d) Cluster-C proteins. Conserved motifs in 73 GLYIII protein sequences from different plant species and green algae were predicted using MEME suite.

3.6. Assessment of Motif Organization in DJ-1_PfpI Domains Reveals Both Vertical and Horizontal Evolutionary Patterns of Plant GLYIII Domains

To understand the origin of individual DJ-1_PfpI domains, we analyzed the motif organization in domains of all the 183 GLYIII proteins (with either one or two domains) under study (Figure 6; Table S4 in Supplementary Materials). In all, 238 domain sequences were used to derive conserved patterns, which gave clear indications regarding domain evolution and also conveyed differences among the three plant GLYIII-specific clusters. Importantly, we analyzed ten motifs in all (which are different from those obtained while analyzing complete protein sequences) (Figure 6a). GLYIII proteins in the cluster-A possessed 7 motifs (3-7-8-4-10-2-1), which were also conserved in archaeal species like *Natronococcus occultus* (NoGLYIII2), *Natrinema pellirubrum* (NpGLYIII3), and *Natronobacterium gregoryi* (NgGLYIII1), and even in bacteria as *Acetobacter pasteurianus* and *Ralstonia solanacearum* (Figure 6b). Both the domains of proteins belonging to cluster-A possessed this seven-motif pattern with few exceptions, especially in cluster-A2. Herein, the first domain of AtGLYIII2/5 and BrGLYIII4 and both the domains of BrGLYIII8 showed a different pattern of motif organization viz. 3-7-8-4-5. Other exceptions pertained to the absence of one or more motifs, the loss of motif 2 being the most common exception in cluster-A2 proteins. Further, members of cluster B (B1 and B2) and C shared the same motif composition and organization, in the form of seven motifs arranged in the order, 3-6-4-5-2-9-1. The larger cluster-C was, however, more consistent in the presence of this motif pattern, while the cluster-B proteins mostly showed absence of either of the motifs from the 4-5-2-9 motif combination. Compared to cluster-A, in cluster-B/C, a single motif 6 replaced the two cluster-A motifs, 7 and 8. Moreover, motifs 5 and 9 were unique to cluster -B and -C members, with the exception of few cluster-A2 members, which also possessed motif 5, as discussed above. Usually, members of cluster-B1 possessed motif 5, whereas those of cluster-B2 possessed motif 9. The general motif pattern of cluster-B/C proteins was also

common to GLYIII proteins from some bacteria (as *Pseudomonas* sp. GM30, PGGLYIII9 and *E. coli*, EcGLYIII2), diatom (like *Thalassiosira pseudonana* CCMP1335, ThpGLYIII1), brown algae (*Ectocarpus siliculosus*, EcsGLYIII1), and protists (*Entamoeba histolytica*, EhGLYIII1 and *Leishmania donovani*, LdGLYIII1), which also inherited these motifs (Figure 6b). However, animal GLYIII despite sharing close sequence similarity with the cluster-B proteins, possessed certain differences with respect to the presence of motifs 2 and 5. As against plant GLYIII, which possessed seven motifs, animal GLYIII has six motifs, though in essentially the same order as of cluster-B proteins, but either motif 2 or motif 5 was absent. Overall, the results indicated that cluster-B and -C members have been probably inherited from prokaryotes through vertical descent as we observed the similar motif architectures in several prokaryotes, plants, and animals. In contrast, cluster-A type of motif organization was restricted to archaeal and few bacterial species along with plants, indicating cluster-A2 proteins have different ancestry than that of cluster B/C proteins and probably have been acquired through horizontal gene transfer events in plants.

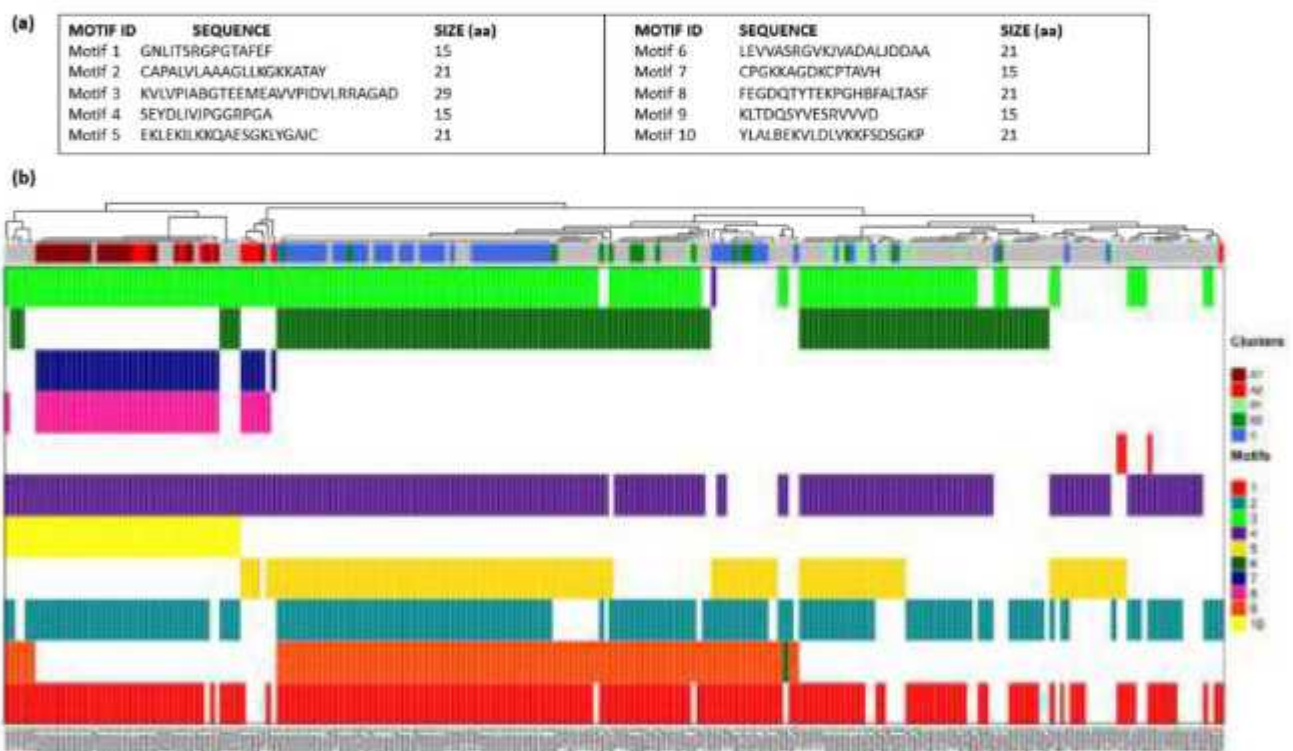


Figure 6. Motif organization pattern of domain sequences of GLYIII proteins from different organisms reveal distinct evolutionary trajectory of plant GLYIII clusters. (a) List of conserved motifs predicted in the domain sequences of GLYIII proteins from 69 species. (b) Heatmap showing distribution of motifs in domain sequences of all organisms under study. A total of ten motifs were identified in different combinations as depicted by different colors. Color strip on the top indicates clusters to which respective proteins belong. Gray indicates proteins not being clustered into either of the three Clusters- A (A1 and A2), B (B1 and B2), or C. MEME suite was used for prediction of motifs.

3.7. Variations in the Three-Dimensional Predicted Structure of Plant GLYIII Proteins from Different Clusters Are Indicative of Their Functional Flexibility

Following the assessment of domain and motif organization, we examined the secondary and three-dimensional (3D) structures of the plant GLYIII proteins. For this, GLYIII proteins from rice viz. OsGLYIII3, OsGLYIII4, and OsGLYIII1 were selected, which served as representatives of clusters A, B, and C, respectively. These were initially analyzed for their secondary structure using ESPrpt 3.0 program (Figure 7a). Subsequently, 3D structures were modelled based on homology using SWISS-MODEL (Figure 7b–f). For both secondary structure prediction and 3D structure modelling, AtDJ-1D protein (PDB

ID: 3uk7.1.A) from *Arabidopsis* was used as the template. The predicted structures were obtained as homo-trimers (Figure 7b) in line with the similar oligomeric nature of the AtDJ-1D protein. An overall high structural similarity could be observed between the three GLYIII proteins (Figure 7b).

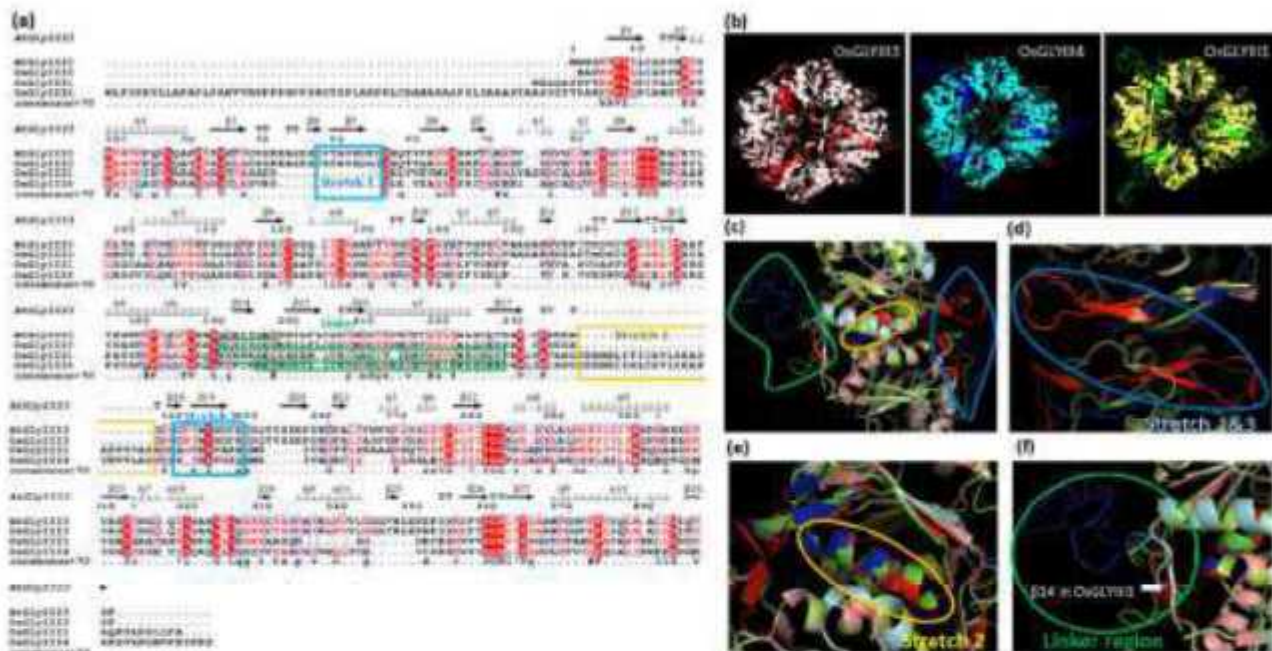


Figure 7. Structural differences in rice GLYIII proteins belonging to three clusters are indicative of their functional flexibility. (a) Sequence alignment and secondary structure information of rice OsGLYIII3, OsGLYIII4, and OsGLYIII1 proteins using ESript3.0. (b) Homotrimeric structure of OsGLYIII3, 4 and 1 as modelled using SWISS-MODELLER. (c) Enlarged view showing regions of structural dissimilarity between the three proteins. Enlarged view showing region corresponding to (d) stretch 1 and 3, (e) stretch 2 and (f) linker region, in the aligned monomers of the three proteins. AtGLYIII3 was used as the template for both secondary and tertiary structure predictions. Structures were visualized in PYMOL.

Secondary structure alignment revealed the presence of several α -helices (12), β -strands (28), and 3_{10} -helices (9) in the GLYIII proteins with a predominance of β -strands. In addition, strict β -turns (TT) were also predicted in the polypeptide chains (Figure 7a). There were certain regions of difference between OsGLYIII3 (belonging to cluster-A) and cluster-B (OsGLYIII4) and -C (OsGLYIII1) members. For instance, the region between residues 40 and 50 corresponding to $\beta 4$ and $\beta 5$ strands (marked by a blue box; stretch 1) was specific to OsGLYIII3 and not present in the other two members. However, a similar motif in the C-terminal domain (represented by $\beta 18$ and $\beta 19$ strands and enclosed in a blue box; stretch 3), was found to be present in all the three proteins. The other significant difference was in the region marked by a yellow box (stretch 2), which was not found in OsGLYIII3, but was present in the other two proteins (Figure 7a,c).

Inspection of these regions of difference in the 3D structure, however, revealed several discrepancies (Figure 7c). While the stretch 1 was predicted not to be present in OsGLYIII1/4, no overlay corresponding to OsGLYIII1/4 was observed even in stretch 3 as against the predicted secondary structure alignment. The regions enclosed in a blue circle (Figure 7c) represents the β -strand encoded by the stretch 1 and 3 residues in OsGLYIII3, enlarged view of which can be observed as red color β -strands (Figure 7d). Further, the stretch 2, which was predicted to be absent in OsGLYIII3 (Figure 7a) was, however, found to be aligned as an α -helix structure, with the other two proteins (Figure 7c,e). Interestingly, linker regions marked in green boxes (Figure 7a) did not align in the overlay of OsGLYIII1 and OsGLYIII4 (Figure 7f). The linker region was generally, shorter in the cluster-A proteins, to which OsGLYIII3 belonged, and in fact, existed as a β -strand ($\beta 14$) in

OsGLYIII3 (Figure 7f). Altogether, these three proteins possessed higher overall similarity in structures, but regions of distinctiveness in their structures may offer scope for functional diversity in GLYIII proteins.

4. Discussion

GLYIII enzymes belong to the DJ-1_PfpI superfamily, with its members mediating diverse functions such as proteases [14] and chaperones [15]. The human DJ-1 protein, one of the most extensively studied members of this superfamily, was reported as the genetic cause of early-onset of Parkinson's disease and found to exhibit weak protease activity, chaperone functions, and even glyoxalase activity [14–16]. Importantly, members of this superfamily possess a catalytic triad, Cys-His-Asp/Glu, known to be required for its catalytic activity [47]. The cysteine residue is functionally essential, being invariantly conserved in all the members of this superfamily and can participate in different functions owing to the presence of a reactive thiol group. The requirement of cysteine was essentially realized in the reaction chemistry of different functions viz. GLYIII, chaperone and protease activities [14–16] and may even act as a sensor of cellular redox homeostasis [48].

Assessing the distribution of DJ-1_PfpI (or GLYIII or DJ-1) proteins across different kingdoms revealed their ubiquitous presence across the tree of life, much like the GLYI and GLYII proteins [6,28]. Interestingly, multiple DJ-1 isoforms are found to be present in all kingdoms and not specifically restricted to plants, as otherwise is the case for GLYI/II genes, which have specifically expanded in plants. Nonetheless, our results suggest that plants do possess more copies of DJ-1 genes than the other species. The presence of multiple DJ-1 in most phyla is possible because, proteins belonging to the DJ-1 superfamily are functionally diverse, performing various essential functions encoded by different members, contrary to the role of GLYI/II proteins, which have a highly specific catalytic function in MG detoxification. Another similar feature of plant GLYIII (or DJ-1) with that of the plant GLYI (Ni-GLYI) is the evolution of two catalytic domain-containing DJ-1 proteins, although single DJ-1 domain-containing proteins are also detected in some plants, especially monocots. In agreement, single domain GLYI also exists in plants that are the Zn/Mn dependent enzymes [6].

The first instance of the presence of two DJ-1 domains in a single polypeptide is seen in the protist *Dictyostelium discoideum* (DdGLYIII2); however, these domains are different from the typical domains, being smaller in length ranging from about 83–97 aa. The first domain has the motif composition 3-6-4, and the second or C-terminal domain contains only two motifs 2-1. This is in sharp contrast to the seven-motif composition of plant GLYIII domains (3-6-4-5-2-9-1). Presence of such DJ-1 protein appears more as an exception, as all the analyzed algal species possess only single DJ-1 domain-containing proteins. Further, the moss *P. patens*, a non-vascular ancestor of land plants, is the first species to harbor both single and two-domain proteins and thereafter, the two DJ-1_PfpI domain-containing proteins can be observed in the other plant lineages. In addition to the DJ-1_PfpI domains, members of this superfamily generally possess other catalytic domains, as well such as a catalase domain in archaea *Natronococcus occultus* or the HTH_18 domain in the bacteria *Pseudomonas* sp. GM30, which may add functional diversity or facilitate the regulation of these proteins. Even the human DJ-1 protein was shown to bind multiple RNA targets in an oxidation-dependent manner [49].

Analysis of phylogenetic relationship among DJ-1 proteins from different kingdoms uncovers an interesting evolutionary path, much different from that observed for the GLYI proteins [6]. Our studies suggested that plant DJ-1 proteins have evolved into three different types of proteins, which we categorized into clusters A, B, and C, with all the clusters being further sub-grouped into smaller clusters. Since these proteins perform multiple functions, it is possible that divergence in sequence between the different clusters is related to the selectivity in functions of these proteins. Cluster-A was more closely related to prokaryotic DJ-1 proteins, being enclosed within the larger clade composed of prokaryotic and yeast DJ-1 proteins. The members of cluster-A had the typical domain motif patterns comprising

of motifs 3-7-8-4-10-2-1. Of the four *E. coli* GLYIII (or DJ-1) proteins, two isoforms viz. EcGLYIII1 (or Hsp31), and EcGLYIII3 (or YhbO), along with yeast DJ-1 proteins SpGLYIII2-SpGLYIII6 (or Hsp3101-Hsp3105) from *S. pombe* and ScGLYIII1 (or Hsp31) from *S. cerevisiae* belong to this clade. However, except for a few archaeal GLYIII (NgGLYIII1, NoGLYIII2, and NpGLYIII3) and a bacterial GLYIII (RsGLYIII1), no other protein from this parent group was found to contain the cluster-A type of motif structure. On the other hand, cluster-B was positioned close to the animal DJ-1 proteins and possessed more similarity to the eukaryotic DJ-1. In fact, protists, brown algae, and diatoms also co-existed within the same larger cluster to which cluster-B belongs. Among the prokaryotes, the GLYIII protein YajL (EcGLYIII2) from *E. coli* shared close similarity to this cluster. Analyzing the motif organization of proteins from plants and the above-mentioned organisms revealed that the members possessed the motif pattern 3-6-4-5-2-9-1, with sub-groups B1 and B2 differing in the presence of motifs 9 and 5, respectively. Animal GLYIII members, however, possess either of the two motifs, 2 or 5, with the rest being the same. The absence of motif 2 in humans, was the main point of distinction between the plant and animal GLYIII sequences. In contrast, cluster-C was comprised exclusively of plant GLYIII proteins, not containing any members from other organisms and possessed the same motif organization pattern as of cluster-B proteins.

Assessment of functional properties of plant GLYIII proteins from the three clusters revealed differential catalytic features of the respective proteins. The OsGLYIII3 (or OsDJ-1c) from rice [9] and AtGLYIII3 (or AtDJ-1d) from Arabidopsis [46] belonging to cluster-A1, exhibited higher levels of GLYIII activity in comparison to the rest of the members of their family, which indicated that cluster-A1 comprised of catalytically more efficient GLYIII proteins. Likewise, EcGLYIII1 and EcGLYIII3 proteins from *E. coli* [50], ScGLYIII1 from *S. cerevisiae* [51], and SpGLYIII2-SpGLYIII6 from *S. pombe* [8] were also demonstrated to possess GLYIII activity. Similar to AtGLYIII3 (AtDJ-1d), all these members can use both MG and glyoxal (GO) as substrates and in fact, show higher activity with GO than MG. While the smaller cluster A2, comprising of proteins like AtGLYIII2 (AtDJ-1e) and AtGLYIII5 (AtDJ-1f), exhibited lower levels of MG-specific GLYIII activity and did not use GO as substrate [46]. Further, proteins in the cluster-C such as AtGLYIII1 (AtDJ-1b) and AtGLYIII4 (AtDJ-1a), were also active as GLYIII enzymes but possessed low levels of activity in comparison to the cluster-A1 representatives from Arabidopsis. Again, the activity of AtDJ-1a and AtDJ-1b was higher when GO but not MG, was used as the substrate [46], much like EcGLYIII3, ScGLYIII1, and SpGLYIII2-6 [8]. The AtDJ-1a protein is involved in the process of ageing, as the loss of its function resulting in accelerated cell death in aging plants [22]. It confers stress protection through cytosolic SOD1 (superoxide dismutase) activation. While AtDJ-1b possesses both holdase chaperone function and GLYIII activity, which is sensitive to H₂O₂ [15]. On the similar lines, EcGLYIII2 and EcGLYIII4 possess GO-specific GLYIII activity and are not able to utilize MG as the substrate [50]. On the other hand, AtGLYIII6 (or AtDJ-1c) in cluster B (specifically cluster-B1, when looking at the domain-specific phylogeny) did not exhibit GLYIII activity and was involved in chloroplast maturation [23], indicating functional divergence in such proteins [46].

Another important observation pertains to the observed high sequence similarity between the two domains of the two DJ-1 domain-containing proteins of cluster A1. The first or N-terminal domain shared more than 50% sequence identity to the second or the C-terminal domain, which was highly indicative of domain duplication followed by domain fusion events behind the emergence of the two-domain DJ-1 proteins in this cluster, whereas proteins in cluster A2 possessed about 30% sequence identity between the two domains. Likewise, less than 50% similarity was also seen between the two domains of each member of the clusters B and C, except for SmGLYIII1, PpGLYIII4, and PpGLYIII5 proteins, which shared 55–59% sequence identity between their domains. In fact, proteins of cluster A and B/C had some characteristic features specific to their groups. For instance, domains of cluster-A proteins were generally longer in length (~184 aa) than the cluster-B/C proteins (~165 aa), but had shorter linker regions (~9–20 aa) than these proteins

(37–39 aa). Importantly, motif analysis of GLYIII proteins revealed proteins of cluster A and B differing with respect to two motifs, being exclusively present in the respective clusters. Indeed, proteins of cluster A had an extra stretch of amino acids, which were also evident from the 3D structures of these proteins and mainly represented loop regions in cluster-B proteins, but present as a beta-strand in cluster-A proteins. This region of variation may be responsible for the difference in substrate specificity of the enzymes.

Further, these plant GLYIII sequences were also very close in similarity to CvGLYIII1 and CrGLYIII1 proteins from green algae, sharing 70–75% sequence similarity to CrGLYIII1. In fact, first or N-terminal domains were more similar to CrGLYIII1 than the second or C-terminal domains. CrGLYIII1 shared 76% sequence similarity (at 91% query coverage; e-val: 5.61E-74) to DJ-1 family protein from *Chloroflexi* bacterium (HGV28139.1), whereas no ortholog or paralog could be obtained for CrGLYIII2, which showed the highest similarity to type 1 glutamine amidotransferase domain-containing protein from *Hymenobacter gelipurpurascens*, a flavobacterium composed of environmental bacteria (49.42% similar at query coverage of 83%, eval: 4.18E-34) and thus, resembled bacterial GLYIII more than the plant ones. Importantly, unlike *C. reinhardtii*, *C. variabilis* another green alga, has only one DJ-1 protein. In such case, it was possible that CrGLYIII2 had been acquired through horizontal gene transfer and hence, constitutes a case specific to *C. reinhardtii* as also, analysis of DJ-1 proteins in other green algal species revealed mostly single gene/protein being present in them [27].

Not only the cluster-A, B, and C proteins differed in their enzyme properties, differences in their subcellular localization patterns also existed. Most of the proteins of cluster-A were predicted to be cytoplasmic in nature with few exceptions mainly in monocots, being chloroplastic proteins and include CrGLYIII2 as well. Since DJ-1 proteins have a role in oxidative stress response, it is believed that their cytoplasmic localization may directly buffer cytosolic redox changes [21]. On similar lines, the human DJ-1 protein is known to be cytoplasmic under basal conditions, but its presence in mitochondria and nucleus is also realized [52] probably to allow the protein to shuttle between different functions based on the redox conditions of the cell. In the cytoplasm, human DJ-1 overexpression can protect against oxidative stress-induced cell death through the suppression of cytoplasmic TDP-43 aggregation [53]. Likewise, AtDJ-1a protein is also a cytosolic protein that facilitates protection against stress by activation of cytosolic SOD [22]. Similar to HsDJ-1 protein, Hsp3101 (SpGLYIII2) and SpDJ-1 (SpGLYIII1) from *S. pombe*, are dually localized proteins, being detected in both cytoplasm and nucleus, whereas SpGLYIII3 (or Hsp3102) is exclusively cytoplasmic [8]. Nuclear localization indicates the role of DJ-1 proteins in regulating gene expression as reported for HsDJ-1 protein, which regulates thioredoxin 1 expression via Nrf2-mediated transcriptional induction [54]. Importantly, cluster-C proteins were predicted to be multi-organelle localized, whereas most of the cluster-B proteins were cytoplasmic, with few being predicted in chloroplast and/or mitochondria as well. Mitochondrial localization of DJ-1 proteins is now well-established to be involved in cytoprotective effects on mitochondria. Human DJ-1 protein is in fact, known to translocate to mitochondria under oxidative stress to protect the cell from damage caused by ROS. To some extent, the translocation was shown to be facilitated by cys-106 oxidation in the human DJ-1, but may not be a prerequisite for translocation [52]. In mitochondria, human DJ-1 was detected both on the outer mitochondrial membrane by subcellular fractionation of M17 neuroblastoma cells [55] and also in the intermembrane space and mitochondrial matrix [56]. In *S. cerevisiae*, in the absence of DJ-1 paralogs, cells exhibited a significant delay in cell-cycle progression due to mitochondrial hyperfusion and partial DNA damage [57].

Further, chloroplastic localization of DJ-1 was observed in several plant DJ-1 proteins, from cluster B and C. AtDJ-1c (or AtGLYIII6) from cluster-C is a chloroplastic protein, which is essentially required for chloroplast development as mutants lack thylakoid membranes and granal stacks [23]. Another protein, AtDJ-1b (or AtGLYIII1) belonging to cluster-B, is also chloroplast-localized, which can act as an oxidation-robust holdase or a GLYIII

enzyme. These discrete functions of AtDJ-1b respond differently to H₂O₂ with only GLYIII activity being sensitive to H₂O₂ [15].

5. Conclusions

The DJ-1_PfpI proteins have come a long way in the evolutionary journey. From modest origins in archaea, these proteins have acquired multiple functions to enable organisms to adapt better to their environment. The shift from aquatic to terrestrial habitation may have been the driving force leading to the evolution of special features of these proteins in plants. In fact, their sessile nature is a major contributor to the multigenic expansion of GLYIII family, which enabled them to supplement their defenses and other critical functions. In lower plants, these proteins acquired an additional DJ-1_PfpI domain, either by fusion or duplication, to help the plants better cope with the terrestrial conditions. Also, with the evolution within plants, the proteins started mobilizing to chloroplast as well. As the plants evolved further, these GLYIII proteins followed three distinct evolutionary paths to give rise to three categories of two-domain GLYIII proteins, each of which possesses specific motifs. These additional motifs may impart added physiological advantages to the GLYIII proteins. Future research is needed to dig out the specific functions of the diverse members of the GLYIII family, which may pave the way towards breeding for higher and sustainable stress tolerance in crop plants mediated by the GLYIII superfamily.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/antiox10050648/s1>, File S1: Sequences of GLYIII proteins used in the study; Figure S1: Multiple sequence alignment of domain sequences of GLYIII proteins from different plant species; Figure S2: Multiple sequence alignment of GLYIII proteins from different plant species; Table S1: Details of GLYIII proteins used in the study. Table S2: List of domains present in GLYIII proteins from different organisms; Table S3: Similarity scores of plant GLYIII proteins with those from *E. coli*; Table S4: Motif patterns observed in domain sequences of all GLYIII proteins.

Author Contributions: S.L.S.-P. conceived the idea and designed the study; B.K. and C.K. analyzed data and wrote the manuscript; S.L.S.-P., S.K.S. and A.P. critically read and finalized the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data and material were available in the supplementary files.

Acknowledgments: S.L.S.-P. acknowledges ICGEB for core grant support. C.K. acknowledges DST-INSPIRE Faculty award (IFA-14/LSPA-24) received from Department of Science and Technology (DST), Govt. of India. B.K. is thankful to Department of Biotechnology (DBT) for Senior Research Fellowship.

Conflicts of Interest: Authors declare no competing interests.

Abbreviations

GLY	glyoxalase
GSH	glutathione
MG	methylglyoxal

References

1. Kalapos, M.P.; Biology, T. Methylglyoxal in living organisms—Chemistry, biochemistry, toxicology and biological implications. *Toxicol. Lett.* **1999**, *110*, 145–175. [[CrossRef](#)]
2. Kalapos, M.P. The tandem of free radicals and methylglyoxal. *Chem. Biol. Interact.* **2008**, *171*, 251–271. [[CrossRef](#)]
3. Thornalley, P.J. Pharmacology of methylglyoxal: Formation, modification of proteins and nucleic acids, and enzymatic detoxification—A role in pathogenesis and antiproliferative chemotherapy. *Gen. Pharmacol.* **1996**, *27*, 565–573. [[CrossRef](#)]
4. Thornalley, P.J. Protein and nucleotide damage by glyoxal and methylglyoxal in physiological systems—role in ageing and disease. *Drug Metabol. Drug Interact.* **2008**, *23*, 125–150. [[CrossRef](#)] [[PubMed](#)]

5. Singla-Pareek, S.L.; Reddy, M.K.; Sopory, S.K. Genetic engineering of the glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 14672–14677. [CrossRef]
6. Kaur, C.; Vishnoi, A.; Ariyadasa, T.U.; Bhattacharya, A.; Singla-Pareek, S.L.; Sopory, S.K. Episodes of horizontal gene-transfer and gene-fusion led to co-existence of different metal-ion specific glyoxalase I. *Sci. Rep.* **2013**, *3*, 3076. [CrossRef]
7. Misra, K.; Banerjee, A.B.; Ray, S.; Ray, M. Glyoxalase III from *Escherichia coli*: A single novel enzyme for the conversion of methylglyoxal into D-lactate without reduced glutathione. *Biochem. J.* **1995**, *305*, 999–1003. [CrossRef]
8. Zhao, Q.; Su, Y.; Wang, Z.; Chen, C.; Wu, T.; Huang, Y. Identification of glutathione (GSH)-independent glyoxalase III from *Schizosaccharomyces pombe*. *BMC Evol. Biol.* **2014**, *14*, 86. [CrossRef]
9. Ghosh, A.; Kushwaha, H.R.; Hasan, M.R.; Pareek, A.; Sopory, S.K.; Singla-Pareek, S.L. presence of unique glyoxalase iii proteins in plants indicates the existence of shorter route for methylglyoxal detoxification. *Sci. Rep.* **2016**, *6*, 18358. [CrossRef]
10. Mannervik, B.; Ridderstrom, M. Catalytic and molecular properties of glyoxalase I. *Biochem. Soc. Trans* **1993**, *21*, 515–517. [CrossRef]
11. Inoue, Y.; Maeta, K.; Nomura, W. Glyoxalase system in yeasts: Structure, function, and physiology. *Semin. Cell Dev. Biol.* **2011**, *22*, 278–284. [CrossRef]
12. Suttisansanee, U.; Honek, J.F. Bacterial glyoxalase enzymes. *Semin. Cell Dev. Biol.* **2011**, *22*, 285–292. [CrossRef]
13. Martins, A.M.; Cordeiro, C.; Freire, A.P. Glyoxalase II in *Saccharomyces cerevisiae*: In situ kinetics using the 5,5'-dithiobis(2-nitrobenzoic acid) assay. *Arch. Biochem. Biophys.* **1999**, *366*, 15–20. [CrossRef] [PubMed]
14. Chen, J.; Li, L.; Chin, L.S. Parkinson disease protein DJ-1 converts from a zymogen to a protease by carboxyl-terminal cleavage. *Hum. Mol. Genet.* **2010**, *19*, 2395–2408. [CrossRef]
15. Lewandowska, A.; Vo, T.N.; Nguyen, T.-D.H.D.H.; Wahni, K.; Vertommen, D.; Van Breusegem, F.; Young, D.; Messens, J. Bifunctional chloroplastic DJ-1B from *Arabidopsis thaliana* is an oxidation-robust holdase and a glyoxalase sensitive to H₂O₂. *Antioxidants* **2019**, *8*, 8. [CrossRef] [PubMed]
16. Richarme, G.; Mihoub, M.; Dairou, J.; Chi Bui, L.; Leger, T.; Lamouri, A. Parkinsonism-associated protein DJ-1/park7 is a major protein deglycase that repairs methylglyoxal- and glyoxal-glycated cysteine, arginine, and lysine residues. *J. Biol. Chem.* **2015**, *290*, 1885–1897. [CrossRef]
17. Richarme, G.; Liu, C.; Mihoub, M.; Abdallah, J.; Leger, T.; Joly, N.; Liebart, J.C.; Jurkunas, U.V.; Nadal, M.; Bouloc, P.; et al. Guanine glycation repair by DJ-1/Park7 and its bacterial homologs. *Science* **2017**, *357*, 208–211. [CrossRef] [PubMed]
18. Vázquez-Mayorga, E.; Díaz-Sánchez, Á.G.; Dagda, R.K.R.Y.; Domínguez-Solís, C.A.; Dagda, R.K.R.Y.; Coronado-Ramírez, C.K.; Martínez-Martínez, A. Novel Redox-dependent Esterase activity (EC 3.1.1.2) for DJ-1: Implications for Parkinson's disease. *Int. J. Mol. Sci.* **2016**, *17*, 1346. [CrossRef] [PubMed]
19. Oh, S.E.; Mouradian, M.M. Cytoprotective mechanisms of DJ-1 against oxidative stress through modulating ERK1/2 and ASK1 signal transduction. *Redox Biol.* **2018**, *14*, 211–217. [CrossRef] [PubMed]
20. Ariga, H.; Takahashi-Niki, K.; Kato, I.; Maita, H.; Niki, T.; Iguchi-Ariga, S.M.M.M. Neuroprotective function of DJ-1 in Parkinson's disease. *Oxid. Med. Cell. Longev.* **2013**, *2013*, 683920. [CrossRef] [PubMed]
21. Bonifati, V.; Rizzu, P.; Van Baren, M.J.; Schaap, O.; Breedveld, G.J.; Krieger, E.; Dekker, M.C.J.; Squitieri, F.; Ibanez, P.; Joosse, M.; et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science* **2003**, *299*, 256–259. [CrossRef] [PubMed]
22. Xu, X.M.; Lin, H.; Maple, J.; Björkblom, B.; Alves, G.; Larsen, J.P.; Møller, S.G. The Arabidopsis DJ-1a protein confers stress protection through cytosolic SOD activation. *J. Cell Sci.* **2010**, *123*, 1644–1651. [CrossRef]
23. Lin, J.; Nazarenus, T.J.; Frey, J.L.; Liang, X.; Wilson, M.A.; Stone, J.M. A plant DJ-1 homolog is essential for *Arabidopsis thaliana* chloroplast development. *PLoS ONE* **2011**, *6*, e23731. [CrossRef]
24. Melvin, P.; Bankapalli, K.; D'Silva, P.; Shivaprasad, P.V. Methylglyoxal detoxification by a DJ-1 family protein provides dual abiotic and biotic stress tolerance in transgenic plants. *Plant Mol. Biol.* **2017**, *94*, 381–397. [CrossRef] [PubMed]
25. Mohanan, M.V.; Pushpanathan, A.; Sasikumar, S.P.T.; Selvarajan, D.; Jayanarayanan, A.N.; Kumar, A.; Ramalingam, S.; Karuppasamy, S.N.; Subbiah, R.; Ram, B.; et al. Ectopic expression of DJ-1/PfpI domain containing *Erianthus arundinaceus* Glyoxalase III (*EaGly III*) enhances drought tolerance in sugarcane. *Plant Cell Rep.* **2020**, *39*, 1581–1594. [CrossRef] [PubMed]
26. Hasim, S.; Hussin, N.A.; Alomar, F.; Bidasee, K.R.; Nickerson, K.W.; Wilson, M.A. A glutathione-independent glyoxalase of the DJ-1 superfamily plays an important role in managing metabolically generated methylglyoxal in *Candida albicans*. *J. Biol. Chem.* **2014**, *289*, 1662–1674. [CrossRef] [PubMed]
27. Singla-Pareek, S.L.; Kaur, C.; Kumar, B.; Pareek, A.; Sopory, S.K. Reassessing plant glyoxalases: Large family and expanding functions. *New Phytol.* **2020**, *227*, 714–721. [CrossRef]
28. Kaur, C.; Sharma, S.; Hasan, M.R.; Pareek, A.; Singla-Pareek, S.L.; Sopory, S.K. Characteristic variations and similarities in biochemical, molecular, and functional properties of glyoxalases across prokaryotes and eukaryotes. *Int. J. Mol. Sci.* **2017**, *18*, 250. [CrossRef]
29. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]
30. National Center for Biotechnology Information. Available online: <https://www.ncbi.nlm.nih.gov/> (accessed on 5 May 2020).
31. Phytozome v12.1: Home. Available online: <https://phytozome.jgi.doe.gov/pz/portal.html> (accessed on 5 May 2020).

32. Letunic, I.; Doerks, T.; Bork, P. SMART: Recent updates, new developments and status in 2015. *Nucleic Acids Res.* **2015**, *43*, D257–D260. [[CrossRef](#)] [[PubMed](#)]
33. Letunic, I.; Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **2018**, *46*, D493–D496. [[CrossRef](#)] [[PubMed](#)]
34. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
35. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
36. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [[CrossRef](#)] [[PubMed](#)]
37. Yu, C.S.; Lin, C.J.; Hwang, J.K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n -peptide compositions. *Protein Sci.* **2004**, *13*, 1402–1406. [[CrossRef](#)] [[PubMed](#)]
38. Yu, C.S.; Chen, Y.C.; Lu, C.H.; Hwang, J.K. Prediction of protein subcellular localization. *Proteins Struct. Funct. Genet.* **2006**, *64*, 643–651. [[CrossRef](#)]
39. Bannai, H.; Tamada, Y.; Maruyama, O.; Nakai, K.; Miyano, S. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **2002**, *18*, 298–305. [[CrossRef](#)]
40. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [[CrossRef](#)]
41. Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, W320–W324. [[CrossRef](#)]
42. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M.C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381–3385. [[CrossRef](#)]
43. DeLano, W.L. The PyMOL Molecular Graphics System. 2002. Available online: <http://www.pymol.org> (accessed on 27 March 2020).
44. Bai, J.; Cederbaum, A.I. Mitochondrial catalase and oxidative injury. *Neurosignals* **2001**, *10*, 189–199. [[CrossRef](#)] [[PubMed](#)]
45. Guy, B.; Krell, T.; Sanchez, V.; Kennel, A.; Manin, C.; Sodoyer, R. Do Th1 or Th2 sequence motifs exist in proteins? Identification of amphipatic immunomodulatory domains in *Helicobacter pylori* catalase. *Immunol. Lett.* **2005**, *96*, 261–275. [[CrossRef](#)]
46. Kwon, K.; Choi, D.; Hyun, J.K.; Jung, H.S.; Baek, K.; Park, C. Novel glyoxalases from *Arabidopsis thaliana*. *FEBS J.* **2013**, *280*, 3328–3339. [[CrossRef](#)] [[PubMed](#)]
47. Subedi, K.P.; Choi, D.; Kim, I.; Min, B.; Park, C. Hsp31 of *Escherichia coli* K-12 is glyoxalase III. *Mol. Microbiol.* **2011**, *81*, 926–936. [[CrossRef](#)] [[PubMed](#)]
48. Wilson, M.A. The role of cysteine oxidation in DJ-1 function and dysfunction. *Antioxidants Redox Signal.* **2011**, *15*, 111–122. [[CrossRef](#)]
49. Van Der Brug, M.P.; Blackinton, J.; Chandran, J.; Hao, L.Y.; Lal, A.; Mazan-Mamczarz, K.; Martindale, J.; Xie, C.; Ahmad, R.; Thomas, K.J.; et al. RNA binding activity of the recessive parkinsonism protein DJ-1 supports involvement in multiple cellular pathways. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10244–10249. [[CrossRef](#)]
50. Lee, C.; Lee, J.; Lee, J.; Park, C. Characterization of the *Escherichia coli* YajL, YhbO and ElbB glyoxalases. *FEMS Microbiol. Lett.* **2016**, *363*, fnv239. [[CrossRef](#)]
51. Bankapalli, K.; Saladi, S.D.; Awadia, S.S.; Goswami, A.V.; Samaddar, M.; D’Silva, P. Robust Glyoxalase activity of Hsp31, a Thij/DJ-1/Pfpl Family Member Protein, Is Critical for Oxidative Stress Resistance in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **2015**, *290*, 26491–26507. [[CrossRef](#)]
52. Junn, E.; Jang, W.H.; Zhao, X.; Jeong, B.S.; Mouradian, M.M. Mitochondrial localization of DJ-1 leads to enhanced neuroprotection. *J. Neurosci. Res.* **2009**, *87*, 123–129. [[CrossRef](#)]
53. Lei, Y.; Zhang, Z.F.; Lei, R.X.; Wang, S.; Zhuang, Y.; Liu, A.C.; Wu, Y.; Chen, J.; Tang, J.C.; Pan, M.X.; et al. DJ-1 Suppresses Cytoplasmic TDP-43 Aggregation in Oxidative Stress-Induced Cell Injury. *J. Alzheimer’s Dis.* **2018**, *66*, 1001–1014. [[CrossRef](#)]
54. Im, J.Y.; Lee, K.W.; Woo, J.-M.; Junn, E.; Mouradian, M.M. DJ-1 induces thioredoxin 1 expression through the Nrf2 pathway. *Hum. Mol. Genet.* **2012**, *21*, 3013–3024. [[CrossRef](#)] [[PubMed](#)]
55. Canet-Avilés, R.M.; Wilson, M.A.; Miller, D.W.; Ahmad, R.; McLendon, C.; Bandyopadhyay, S.; Baptista, M.J.; Ringe, D.; Petsko, G.A.; Cookson, M.R. The Parkinson’s disease DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9103–9108. [[CrossRef](#)] [[PubMed](#)]
56. Zhang, L.; Shimoji, M.; Thomas, B.; Moore, D.J.; Yu, S.-W.; Marupudi, N.I.; Torp, R.; Torgner, I.A.; Ottersen, O.P.; Dawson, T.M.; et al. Mitochondrial localization of the Parkinson’s disease related protein DJ-1: Implications for pathogenesis. *Hum. Mol. Genet.* **2005**, *14*, 2063–2073. [[CrossRef](#)]
57. Bankapalli, K.; Vishwanathan, V.; Susarla, G.; Sunayana, N.; Saladi, S.D.; Peethambaram, D.; D’Silva, P. Redox-dependent regulation of mitochondrial dynamics by DJ-1 paralogs in *Saccharomyces cerevisiae*: Regulation of mitochondrial health by DJ-1 homologs. *Redox Biol.* **2020**, *32*, 101451. [[CrossRef](#)] [[PubMed](#)]